

Harald Rose

Geometrical Charged-Particle Optics

2nd Edition



Springer

Springer Series in Optical Sciences

Volume 142

Founded by

H. K. V. Lotsch

Editor-in-Chief:

W. T. Rhodes

Editorial Board:

Ali Adibi, Atlanta

Toshimitsu Asakura, Sapporo

Theodor W. Hänsch, Garching

Takeshi Kamiya, Tokyo

Ferenc Krausz, Garching

Bo A. J. Monemar, Linköping

Herbert Venghaus, Berlin

Horst Weber, Berlin

Harald Weinfurter, München

For further volumes:

<http://www.springer.com/series/624>

Springer Series in Optical Sciences

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief William T. Rhodes, Georgia Institute of Technology, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.

With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also www.springer.com/series/624

Editor-in-Chief

William T. Rhodes
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250, USA
e-mail: bill.rhodes@ece.gatech.edu

Editorial Board

Ali Adibi
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250, USA
e-mail: adibi@ee.gatech.edu

Toshimitsu Asakura
Hokkai-Gakuen University
Faculty of Engineering
1-1, Minami-26, Nishi 11, Chuo-ku
Sapporo, Hokkaido 064-0926, Japan
e-mail: asakura@eli.hokkai-s-u.ac.jp

Theodor W. Hänsch
Max-Planck-Institut für Quantenoptik
Hans-Kopfermann-Straße 1
85748 Garching, Germany
e-mail: t.w.haensch@physik.uni-muenchen.de

Takeshi Kamiya
Ministry of Education, Culture, Sports
Science and Technology
National Institution for Academic Degrees
3-29-1 Otsuka, Bunkyo-ku
Tokyo 112-0012, Japan
e-mail: kamiyatk@niad.ac.jp

Ferenc Krausz
Ludwig-Maximilians-Universität München
Lehrstuhl für Experimentelle Physik
Am Coulombwall 1
85748 Garching, Germany and
Max-Planck-Institut für Quantenoptik
Hans-Kopfermann-Straße 1
85748 Garching, Germany
e-mail: ferenc.krausz@mpq.mpg.de

Bo A. J. Monemar
Department of Physics and Measurement Technology
Materials Science Division
Linköping University
58183 Linköping, Sweden
e-mail: bom@ifm.liu.se

Herbert Venghaus
Fraunhofer Institut für Nachrichtentechnik
Heinrich-Hertz-Institut
Einsteinufer 37
10587 Berlin, Germany
e-mail: venghaus@hhi.de

Horst Weber
Optisches Institut
Technische Universität Berlin
Straße des 17. Juni 135
10623 Berlin, Germany
e-mail: weber@physik.tu-berlin.de

Harald Weinfurter
Sektion Physik
Ludwig-Maximilians-Universität München
Schellingstraße 4/III
80799 München, Germany
e-mail: harald.weinfurter@physik.uni-muenchen.de

Harald Rose

Geometrical Charged-Particle Optics

Second Edition

 Springer

Harald Rose
Institut für Angewandte Physik
TU Darmstadt
Darmstadt
Germany

ISSN 0342-4111

ISBN 978-3-642-32118-4

ISBN 978-3-540-85915-4 (First edition)

DOI 10.1007/978-3-642-32119-1

Springer Heidelberg New York Dordrecht London

ISSN 1556-1534 (electronic)

ISBN 978-3-642-32119-1 (eBook)

Library of Congress Control Number: 2012955484

© Springer-Verlag Berlin Heidelberg 2009, 2012

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface to the Second Edition

Some errors and misprints that were found in the first edition of this work have been corrected. Although the impetus of the work is on the geometrical properties of charged-particle optics, we have included sections on the propagation of electron waves in macroscopic fields and on the Aharonov–Bohm effect in Chap. 2. The incorporation of wave-optical considerations has been necessary for deriving the resolution limit of electron microscopes and for understanding diffraction and interference phenomena utilized in electron holography.

Chapter 3 has been extended by a section on the calculation of static electromagnetic fields by means of the charge-simulation method. To elucidate the imaging properties of electron lenses in more detail, we have added in Chaps. 4 and 8 the imaging properties of the Glaser model field for a magnetic round lens because this field yields analytical expressions for the paraxial rays and the primary aberrations. Moreover, we have incorporated in Chap. 6 a section on the formation and classification of caustics because they are nowadays widely used for determining the state of alignment of aberration-corrected electron microscopes. Owing to its importance for the performance of systems corrected for primary chromatic and geometrical aberrations, we have added in Chap. 8 a section on fifth-order aberrations of multipole systems with straight axis.

We have added extensive new material to Chap. 12 and rewritten Chap. 14 on relativistic electron motion eliminating several inaccuracies. In particular, we have included aspects of the Stern–Gerlach effect and depicted Lorentz transformations within the frame of relativistic electron motion in Minkowski space. The last chapter is entirely new and discusses the effect of velocity and acceleration on the electromagnetic field of a moving charged particle. We treat this difficult problem by introducing the self-action of the particle in a covariant form. An appendix is added containing a list of symbols used frequently.

I am grateful to several readers who drew my attention to errors and misprints in the first edition and to Mrs. Anna Zilch for skilful drawing of many new figures.

Darmstadt
December 2012

Harald Rose

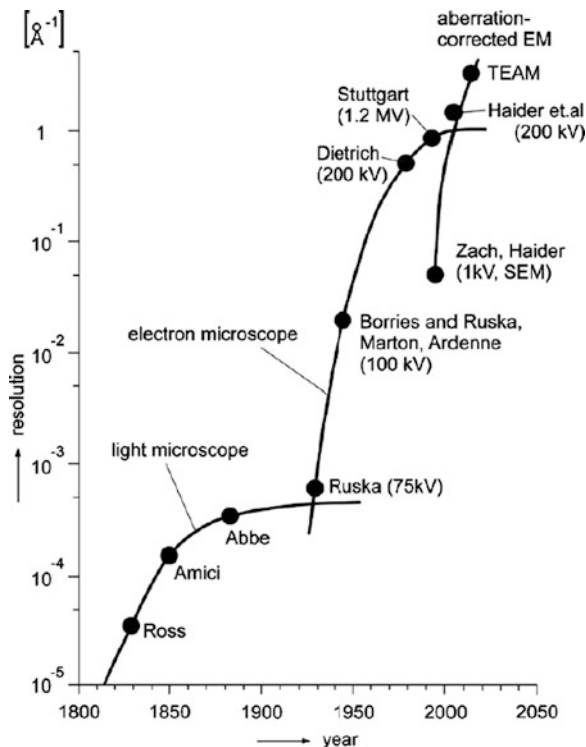
Preface to the First Edition

The resolution of any imaging microscope is ultimately limited by diffraction and can never be significantly smaller than the wavelength λ of the image-forming wave, as realized by Ernst Abbe in 1870. In a visionary statement he argued that there might be some yet unknown radiation with a shorter wavelength than that of light enabling a higher resolution at some time in the future [1]. The discovery of the electron provided such a radiation because its wavelength at accelerating voltages above 1 kV is smaller than the radius of the hydrogen atom. The wave property of the electron was postulated in 1924 by Louis de Broglie [2]. Geometrical electron optics started in 1926 when Busch [3] demonstrated that the magnetic field of a rotationally symmetric coil acts as a converging lens for electrons. The importance of this discovery was subsequently conceived by Ruska and Knoll who had the idea to build an electron microscope by combining a sequence of such lenses [4]. Within a short period of time the resolution of the electron microscope surpassed that of the light microscope, as depicted in Fig. 1. This success resulted primarily from the extremely small wavelength of the electrons rather than from the quality of standard electron lenses which limit the attainable resolution to about 100λ . Therefore, shortening the wavelength by increasing the voltage was the most convenient method for improving the resolution. However, radiation damage by knock-on displacement of atoms limits severely the application of high-voltage electron microscopes. In addition, the so-called delocalization caused by spherical aberration prevents an unambiguous interpretation of images of non-periodic objects such as interfaces and grain boundaries. The correction of the spherical aberration eliminates this deleterious effect.

The successful correction of the spherical aberration can be considered as a quantum step in the development of the electron microscope because it enables one to obtain sub-Å resolution at voltages below the threshold for atom displacement. The threshold voltage depends on the composition of the object and lies in the region between 60 to 300 kV for most materials.

At about the same time as Ruska and Knoll developed the first electron microscope with magnetic lenses, Ernst Brueche at the research department of the AEG in Berlin investigated with his collaborators A. Recknagel and H. Mahl the

Fig. 1 Increase in resolution of transmission microscopy as a function of time



properties of electrostatic round lenses. In order to obtain theoretical assistance, Brueche invited the young Otto Scherzer in 1932 to join his group. Within the short period of two years Otto Scherzer established the theoretical basis of geometrical electron optics. In 1934, he published his results together with Ernst Brueche in the first book on the subject entitled “Geometrische Elektronenoptik” [5]. Scherzer employed for his calculations the so-called trajectory method, which starts from the Newton equation of motion and the Lorentz force [6], whereas Walter Glaser applied the *Hamiltonian formalism* to electron optics to determine the motion of electrons in rotationally symmetric static electromagnetic fields [7]. This method is based on the ideas of Hamilton who showed that the properties of an optical system can be derived from a single characteristic function or eikonal. Because the two calculation procedures differ from each other, they give seemingly different integral expressions for the aberration coefficients. However, the integrals can be transformed in identical forms by partial integrations. Using this method, Scherzer transformed, in 1936, the integral expressions for the coefficients of the spherical and axial chromatic aberrations in such a form that the integrands consist of sums of positive quadratic terms, proving that these coefficients can never change sign [8]. The physical origin for this behavior is due to the fact that the static electromagnetic potentials satisfy the Laplace equation in the domain of the electron trajectories. As a consequence,

the spatial distribution of the index of refraction of electron lenses cannot be formed arbitrarily. Because the potential adopts an extremum at the boundary surfaces, the outer zones of rotationally symmetric electron lenses always focus the rays more strongly than the inner zones, causing unavoidable spherical aberration. Owing to its importance this property has been named “Scherzer Theorem”.

O. Scherzer and W. Glaser are recognized as the founders of theoretical electron optics. The subject up to 1952 was fully summarized in Glaser’s book *Grundlagen der Elektronenoptik*, which served as the standard textbook for several decades [9]. The Hamiltonian approach to electron optics was developed further by P. A. Sturrock [10]. Several other books on the subject appeared in the following years [11, 12]. In particular the treatise *Electron Optics and the Electron Microscope* by Zworykin et al. [13] and P. Grivet’s excellent *Electron Optics* [14] are milestones of the subject. The last approach covering all fields of electron optics was performed by Peter Hawkes and Erwin Kasper with their three volume treatise *Principles of Electron Optics* published in 1996 [15].

The history of electron optics is to a large extent the struggle to overcome the limitations of the resolution of electron microscopes imposed by the unavoidable spherical and chromatic aberrations of round lenses. In 1947, Scherzer demonstrated in another fundamental paper that correction of aberrations is possible by lifting any of the constraints of his theorem, either by abandoning rotational symmetry, or by introducing time-varying fields or space charges [16]. In the following decades intensive experimental efforts to compensate for the resolution-limiting aberrations by means of multipole correctors have been pursued by several groups in Germany [17], England [18] and the USA [19] with disappointing results. The attempts came to an end in the 1980s primarily due to severe problems of precisely aligning the many elements of the correctors during a period of time which is shorter than the overall stability period of the microscope. Moreover, digital processing of through-focus series provided a successful alternative solution for eliminating the spherical aberration of images a posteriori.

As a result, work on electron optics shrunk and was limited to theoretical investigations and to applications in electron lithography and to the design of electron-beam devices for the inspection of wafers [20, 21]. Owing to the advancement in technology and computer-assisted alignment, correction of the resolution-limiting aberrations became very promising again at the beginning of the 90s. In 1992, experimental work started by M. Haider at the EMBL in Heidelberg within the frame of the Volkswagen project, aimed to compensate for the spherical aberration of a transmission electron microscope (TEM) by means of a novel hexapole corrector [22]. One of the main tasks concerned the reduction of the information limit in order that the resolution was limited by the spherical aberration rather than by the incoherent aberrations resulting from instabilities. At about the same time high-performance imaging energy filters became available in commercial electron microscopes leading to a rapid growth of analytical electron microscopy [23]. The successful correction of the spherical aberration in a commercial 200 kV TEM by M. Haider et al. (1997) and by O. Krivanek et al. (1999) in a 100 kV scanning transmission electron microscope (STEM) induced a revival of electron

optics [24, 25]. In the following years, numerous new correctors compensating for chromatic and spherical aberrations were proposed as well as novel high-performance imaging energy filters and monochromators [26, 27]. The revival of electron optics culminated in the TEAM project of the US Department of Energy (DOE) aimed to realize a chromatic and spherically corrected TEM with a resolution limit of $0.5\text{\AA}$.

Geometrical electron optics provides the appropriate tool for designing a large variety of other charged-particle instruments such as electron mirrors, spectrometers, time of flight analyzers, electron guns, accelerators and storage rings. Owing to the large progress in electron optics, electron holography, image formation and design of charged-particle instruments made during the last 15 years, it is impossible to treat all subjects in a single book. Therefore, we confine the content of this book to geometrical electron optics with the impetus on analytical methods for calculating the properties of charged-particle systems and methods for designing optimum electron optical instruments and elements. Diffraction effects resulting from the wave nature of the elementary particles and interactions between electrons within the beam (Boersch effect) will not be covered. Therefore, the content of this book may properly be referred to as a single particle description. Because the effect of the spin on the motion of the electron is very small, it is only treated in the chapter 14 at the end of the book.

The content of this book originated from lectures taught by the author for many years at the Technical University Darmstadt and from courses in charged-particle optics given at the Lawrence Berkeley National Laboratory (BNL) during the period 2003-2005. Therefore, particular attention has been given to the presentation of techniques which would enable the reader not only to “follow the literature” but also to perform electron optical design and calculations on his own. The degree of emphasis which each topic has is a matter of personal judgment. We have not attempted to present an encyclopedia on the subject because it is not possible to include all topics of geometrical electron optics in a single book. For example, model fields providing analytical solutions for the paraxial trajectories of electron lenses have been omitted. They are discussed in great detail in the second volume of *Principles of Electron Optics* by Hawkes and Kasper [15]. Moreover, many computer programs are nowadays available which provide solutions of the paraxial path equations for arbitrary field distributions. Most of the presented material on aberrations, systems with curved axis and aberration correctors is based on research work performed at the University of Darmstadt over a period of several decades. No attempt has been made to provide a complete bibliography. The references have been confined to those which treat specific topics in greater detail. Hence, this selection should not be judged as a ranking and I offer my apologies to the many contributors to the subject whose excellent papers have not been cited. An extensive list of references can be found in Hawkes and Kasper [15].

The book is intended as a textbook for graduate students with good mathematical background and for anyone involved in the design of charged-particle devices ranging from electron lenses to spectrometers. Practical applications of electron optics serve as illustrations of the principles under discussion. Due to the recent

progress in aberration correction, the properties of various corrector types are discussed in detail. The book contains some unpublished material on multipole systems and provides a novel analytical calculation procedure for determining the Gaussian optics and the aberrations of electron guns in the absence of space charge effects. In the last chapter we consider spin precession and radiation effects in the context of relativistic electron motion in electromagnetic fields by employing a novel covariant treatment [28]. By introducing the Lorentz-invariant universal time as independent variable, we extend the Hamilton- Jacobi formalism of classical mechanics from three to four spatial dimensions. This approach allows one to construct a proper four-dimensional covariant Lagrangian, which considers charge, gravitation and spin interactions [28].

I want to thank Dr. Weishi Wan, BNL, for numerical calculations of trajectories and Mrs. Anna Zilch for the skilful making of many drawings. Thanks are due to the members of CEOS (Heidelberg) for helpful discussions and editorial support and to Prof. E. Plies and Dr. Essers for permission to publish drawings.

Darmstadt
August 2008

Harald Rose

Contents

1	Introduction	1
2	General Properties of the Electron	5
2.1	Particle Nature of the Electron	6
2.1.1	Equation of Motion	6
2.1.2	Conservation of Energy	7
2.1.3	Hamilton's Principle	8
2.1.4	Principle of Maupertuis	11
2.1.5	Time of Flight	12
2.2	Wave Properties of the Electron	13
2.2.1	Eikonal and Fermat's Principle	14
2.2.2	Phase, Wavelength, Frequency, Phase and Group Velocity, and Index of Refraction	17
2.3	Aharonov–Bohm Effect	21
2.4	Ray Properties Associated with the Eikonal	26
2.5	Propagation of the Electron Wave in Macroscopic Fields	29
2.5.1	Propagator	30
2.5.2	Wave Propagation in the Infinite Half-Space	33
2.5.3	Fundamentals of Diffraction and Image Formation	38
2.5.4	Point Spread Function	41
3	Multipole Expansion of the Stationary Electromagnetic Field	45
3.1	Scalar Potentials	46
3.1.1	Complex Variables	47
3.1.2	Laplace Equation	49
3.1.3	Planar Fields	50
3.2	Systems with Straight Axis	52
3.2.1	Multipole Expansion of the Scalar Potential	53
3.2.2	Electrostatic Cylinder Lenses	55
3.3	Systems with Curved Axis	58
3.3.1	Recurrence Formula for the Coefficients of the Power Series Expansion	58

3.3.2	Power Series Expansion of the Electric Potential	60
3.3.3	Index of Refraction	62
3.4	Magnetic Vector Potential	63
3.4.1	Rectilinear Systems	66
3.4.2	Magnetic Fields with Special Symmetry	67
3.4.3	Systems with Curved Axis	69
3.5	Integral Representation of the Multipole Components of the Potential	71
3.6	Potentials of Simple Systems	72
3.6.1	Laplace Equation for Oblate Spheroidal Coordinates	73
3.6.2	Solutions with Rotational Symmetry	74
3.6.3	Multipoles	81
3.7	Charge-Simulation Method	84
3.7.1	Systems of Rotational Symmetry	86
4	Gaussian Optics	89
4.1	Paraxial Path Equation	90
4.2	Orthogonal Systems with Midsection Symmetry	98
4.3	Systems with a Straight Optic Axis	102
4.3.1	Systems with an Axis of Rotational Symmetry	102
4.3.2	Wronski Determinant	106
4.3.3	Lagrange–Helmholtz Relation	107
4.3.4	Theorem of Alternating Images	108
4.3.5	Longitudinal Magnification	111
4.3.6	Characteristic Paraxial Rays	112
4.3.7	Thin-Lens Approximation	113
4.4	Magnetic Model Lens	120
4.4.1	Condenser-Objective Lens	124
4.5	Quadrupoles	125
4.5.1	Imaging Properties of a Single Quadrupole	127
4.5.2	Quadrupole Multiplets	130
4.5.3	Strong Focusing	132
4.6	Electrostatic Cylinder Lenses	142
4.6.1	Modified Paraxial Equation	144
4.6.2	Short Cylinder Lenses	146
4.7	General Systems with Straight Axis	149
4.7.1	Inseparable Systems with Straight Axis	150
4.7.2	Generalized Helmholtz–Lagrange Relations	151
4.7.3	Imaging Properties	154
4.7.4	Paraxial Pseudo Rays	159
4.8	Systems with Curved Axis	160
4.8.1	General Systems	161
4.8.2	Systems with Midsection Symmetry	164

4.9	Quadrupole Anastigmat	166
4.9.1	Focal Lengths of the Constituent Quadrupoles of the Anastigmat	168
4.9.2	Cardinal Elements of the Quadrupole Anastigmat	172
4.10	Variable-Axis Lens	176
4.11	Highly Symmetric Telescopic Systems	179
4.11.1	Streak-Free Beam Chopper	187
5	General Principles of Particle Motion	189
5.1	Hamiltonian Formulation	189
5.2	Lagrange Invariants	192
5.3	Liouville's Theorem	195
5.3.1	Paraxial Approximation	197
5.3.2	Abbe Sine Condition	197
5.4	Generalized Symplectic Matrices	199
5.5	Poincaré's Invariant	202
5.6	Eikonals	207
5.6.1	Mixed Eikonal and Sine Condition	210
5.6.2	Perturbation Eikonal	212
5.6.3	Integral Equations of the Path and Momentum Deviations	220
5.7	Poisson Brackets	221
6	Beam Properties	223
6.1	Brightness	224
6.2	Emittance	227
6.2.1	Paraxial Approximation	228
6.2.2	Matching	232
6.3	Propagation of Beam Intensity	233
6.4	Caustics	234
6.4.1	Determination of the Caustic	235
6.4.2	Formation of the Caustic in Free Space	236
6.4.3	Classification of the Fundamental Caustic Forms	241
6.4.4	Third-Order Elementary Caustics	243
6.4.5	Fourth-Order Elementary Caustics	246
7	Path Deviations	251
7.1	Iteration Algorithm	253
7.2	Canonical Representation	256
7.2.1	Recurrence Formula	258
7.2.2	Canonical Representation of the Path Deviations	259
7.3	Expansion Polynomials of the Variational Function	261
7.4	Path Equation Approach	264
7.4.1	Primary Deviations	266

7.5	Second-Rank Path Deviations of Systems with Midsection Symmetry	267
7.5.1	Wien Filter	268
7.5.2	Magnetic Systems	271
7.6	Second-Rank Path Deviations of Systems with Straight Axis	273
7.6.1	Second-Order Path Deviation	275
8	Aberrations	281
8.1	Second-Rank Aberrations	284
8.1.1	Systems with Midsection Symmetry	285
8.1.2	Systems with Straight Optic Axis	289
8.1.3	Axial Chromatic Aberration and Chromatic Distortion	291
8.2	Third-Order Aberrations of Systems with Straight Axis	295
8.2.1	Structure of the Geometrical Eikonal Polynomials	295
8.3	Geometrical Aberrations of Round Lenses	297
8.3.1	Scherzer Theorem	303
8.3.2	Spherical Aberration and Disc of Least Confusion	304
8.3.3	Coma	306
8.3.4	Image Curvature	308
8.3.5	Field Astigmatism	310
8.3.6	Distortion	312
8.4	Primary Aberrations of Glaser's Magnetic Model Lens	313
8.5	Geometrical Aberrations of Quadrupole–Octopole Systems	315
8.5.1	Aperture Aberration of Stigmatic Orthogonal Quadrupole Systems	317
8.5.2	Aberrations Introduced by Octopoles	319
8.5.3	Third-Order Aberrations of Systems with Threefold Symmetry Corrected for Second-Order Aberrations	320
8.5.4	Parasitic Aberrations	322
8.6	Fifth-Order Aberrations of Orthogonal Systems with Straight Axis	327
8.6.1	Fifth-Order Aperture Aberration of Achromatic Third-Order Aplanats	329
9	Correction of Aberrations	333
9.1	Correction of Chromatic Aberrations	336
9.1.1	First-Order Wien Filter	336
9.1.2	Correction of Chromatic Distortions	339
9.1.3	Electrostatic Correction of Chromatic Aberration	342
9.1.4	Chromatic Correction of Systems with Curved Axis	347
9.1.5	Correction of Second-Degree Dispersion	354
9.2	Correction of Geometrical Aberrations	356
9.2.1	Correction of Second-Order Aberrations	356
9.2.2	Correction of Third-Order Spherical Aberration	362

9.2.3	Aplanats	364
9.2.4	Achromatic Aplanats	367
9.2.5	Correction of Third-Order Field Curvature and Field Astigmatism	376
9.2.6	Correction of Coma.....	380
10	Electron Mirrors	387
10.1	Reference Electron	388
10.2	Equation of Motion	389
10.3	Eikonal Approach	391
10.4	Rotationally Symmetric Mirrors	393
10.4.1	Linear Approximation	394
10.4.2	Lateral Fundamental Rays	395
10.4.3	Longitudinal Fundamental Deviations.....	395
10.5	Path Deviations	401
10.6	Electrostatic Mirror	404
10.6.1	Positional Deviations	404
10.6.2	Axial Aberrations	405
11	Optics of Electron Guns	413
11.1	Field Emission Guns	414
11.2	Gaussian Optics	416
11.3	Aberrations	418
11.3.1	Second-Rank Deviations	419
11.3.2	Third-Order Spherical Aberration at the Crossover.....	421
12	Confinement of Charged Particles	425
12.1	Paul Trap.....	425
12.2	Combined Trap	427
13	Monochromators and Imaging Energy Filters	429
13.1	Electrostatic Monochromator	430
13.2	Imaging Energy Filters	432
13.2.1	Types of Imaging Energy Filters	433
13.2.2	MANDOLINE Filter	435
13.2.3	W-Filter	439
14	Relativistic Electron Motion and Spin Precession	443
14.1	Covariant Hamilton Formalism	446
14.2	Propagation in Minkowski Space.....	451
14.2.1	Path Equation	451
14.2.2	Hamiltonian	452
14.2.3	Propagation of Free Particles in Minkowski Space	456
14.3	Four-Dimensional Hamilton–Jacobi Equation	460
14.4	Relativistic Maupertuis Principle	463
14.5	Approximate Relativistic Canonical Momentum and Hamiltonian in the Laboratory System	467

14.6	Spin Precession	469
14.7	Stern–Gerlach Effect Revisited	474
15	Electron Self-Action	477
15.1	Electromagnetic Field of the Moving Electron	478
15.1.1	Field of an Electron Moving with Constant Velocity	481
15.2	Radiation Field Emitted by the Accelerated Electron	482
15.2.1	Dipole Approximation	485
15.2.2	Radiation Field of the Moving Electron	486
	References	489
	Nomenclature	495
	Index	497

Chapter 1

Introduction

Geometrical charged-particle optics describes the motion of charged particles in *macroscopic* electromagnetic fields by employing the well-established notations and concepts of *light* optics. Macroscopic fields are produced by macroscopic elements, such as solenoids, electric, and magnetic multipoles, or by voltages applied to conducting devices, for example, cylinders or apertures. We define the atomic fields within solid or biological objects as *microscopic* fields. The propagation of the particles in these fields will not be considered within the frame of *geometrical charged-particle optics*.

The description of the particle motion from the point of view of light optics is reasonable because the elementary particles have particle and wave properties. The similarity between the propagation of light and particles is documented by the equivalent mathematical treatments [29]. Moreover, the properties of particle-optical instruments and their constituent components are described most appropriately in light-optical terms, which have been established at a time when charged particles were still unknown. The treatment of particle motion by means of optical concepts has been proven extremely useful for the design of beam-guiding systems, the electron microscope in particular. This microscope has developed over the years from an image-forming system to a sophisticated analytical instrument yielding structural and chemical information about the object on an atomic scale.

Within the frame of validity of charged-particle optics, we describe electrons and ions by the same formalism because their propagation in macroscopic fields depends only on their mass and charge, respectively. The effect of the spin on the motion of charged particle is of the same order of magnitude as that resulting from diffraction. The influence of diffraction becomes negligibly small in the limit that the index of refraction does not change appreciably over a distance of several wavelengths λ . The limit $\lambda \rightarrow 0$ represents the domain of *geometrical charged-particle optics*.

For reasons of simplicity, we restrict our further investigations to electrons. Nevertheless, we can use all results for ions as well if we substitute their charge and rest mass for the corresponding quantities of the electron. Geometrical light optics describes the properties of optical elements by means of their effects on the light rays along which the point-like *photons* propagate. The rays form straight lines

in the region outside the lenses. These rays are either refracted at the surfaces of the lenses where the index of refraction changes abruptly or are deflected steadily if the index of refraction changes gradually, as in the case of the atmosphere due to the varying density with respect to the distance from the earth. The so-called *gradient-index* lenses have an index of refraction, which increases quadratic with the distance from their *optic axis*.

In close analogy to light optics, geometrical electron optics conceives the path of an electron as a geometrical line or trajectory, respectively. However, contrary to light optics, all electron optical elements form gradient-index lenses because the electrons must travel in vacuum where the electromagnetic fields produced by the exterior currents and charges vary continuously. The word electron originates from the Greek word *ηλεκτρον* meaning amber. In 1890, *Stoney* introduced this word for denoting the elementary charge because amber charges up by friction.

Electron optics is based on two fundamental discoveries made in 1924 by *Louis de Broglie* [2] and in 1926 by *Hans Busch* [3]. De Broglie postulated on ground of theoretical considerations that one must attribute a wave to each elementary particle. At about the same time, Busch discovered that the magnetic field of a solenoid acts on electrons in exactly the same way as a glass lens on the light rays. It had been these two important discoveries, which lead *Ernst Ruska* to the conclusion that it must be possible to build a microscope, which uses electrons instead of photons. He realized successfully the first electron microscope in 1931 [30]. The development of the electron microscope, oscillographs, and cathode-ray tubes gave rise to the science of electron optics. The guiding of charged particles is also of great importance in accelerators and spectrometers employed in nuclear physics [31, 32]. However, the close analogy between these instruments and the classical electron optical devices was not widely recognized. For the development of the latter instruments, it proved extremely useful to utilize the concepts and notations employed in light optics. Subsequently, one applied and expanded these methods in the context of designing aberration correctors, monochromators, and imaging energy filters composed of non-rotationally symmetric elements such as dipoles and quadrupoles. Unfortunately, the designers of accelerators and spectrometers in nuclear physics did not take notice of these developments established much earlier. As a result, different notations exist for the same device or property. This unfortunate situation causes quite often confusion among the nonexperts. This situation dates back from the early days of charged-particle optics, when each group entering this field of research introduced its own nomenclature. In this book, we use the notation and terminology introduced by Scherzer [6, 33]. Within the frame of this terminology, we distinguish between *planes* and *sections*. Planes are plane surfaces perpendicular to the optic axis regardless of whether it is straight or curved. Sections are surfaces which contain the optic axis. Unlike a plane, a section can be curved, as what happens in systems with curved axis.

The main task of charged-particle optics is the manipulation of ensembles of rays each originating from a common point. Important collective properties of optical elements are, for example, the focusing of homocentric bundles of rays in order to form an image and the guiding of particles in accelerators or storage rings [34],

respectively. We do not consider methods for producing charged particles in the frame of geometrical charged-particle optics. Although this approximation is well suited to describe the action of optical elements, it fails to provide information about the intensity in the region of the caustic formed by the loci of the intersections of rays emanating from the same origin. Because a plane partial wave is associated to each trajectory, strong interference effects arise at the vicinity of the caustic, as it is the case in the image plane of an electron microscope.

Chapter 2

General Properties of the Electron

Abstract The fundamental properties of the electron are outlined, and the dualism of its particle and wave nature is discussed. Starting from the particle nature and the concepts of classical mechanics, the principles of Hamilton and Maupertuis are introduced, and the conservation of energy is derived. The concepts of light optics are employed for defining the index of refraction, wavelength, phase and group velocity, and the eikonal. Electron holography, the Aharonov–Bohm effect, and the propagation of the electron wave in macroscopic fields are chosen for demonstrating effects resulting from the wave nature of charged particles.

Elementary particles exhibit a wave and a particle nature depending on the specific experiment. Owing to its relatively small rest energy $E_0 = m_e c^2 \approx 0.51 \text{ MeV}$, the electron approaches roughly half the speed of light $c \approx 3 \times 10^8 \text{ m/s}$ at an accelerating voltage $U \approx 60 \text{ kV}$. Therefore, it is necessary to consider relativistic effects for accelerating voltages larger than about 100 kV . Despite the fact that we can consider the electron as a *point-like* particle, it has an angular momentum associated with a magnetic moment

$$\mu = \frac{eg\mu_0}{2m_e}s = \frac{e\mu_0\hbar}{2m_e}. \quad (2.1)$$

Here, $e \approx 1.6 \cdot 10^{-19} \text{ A s}$ and $\hbar \approx 6.58 \cdot 10^{-16} \text{ eV s}$ are the charge of the electron and the Planck constant, respectively; μ_0 is the permeability of the vacuum. We use SI units, which now are universally accepted. From the point of view of classical electrodynamics, a magnetic moment originates from a rotating charge of finite extension forming a magnetic dipole. However, the measured ratio of the magnetic moment and the angular momentum or spin $s = |\vec{s}| = \hbar/2$ of the electron is twice as large as predicted by classical electrodynamics. This discrepancy, which requires an empirical *Lande factor* $g = 2$, can only be explained by means of the relativistic electron theory of Dirac [35, 36]. The spin $\vec{s}$ of the electron is comparable with the *polarization* of the light.

2.1 Particle Nature of the Electron

Within the frame of geometrical charged-particle optics, one considers the electron as a point-like charged mass, whose motion is governed by the laws of *classical mechanics* [37]. We do not consider the *precession* of the electron spin because it does not appreciably affect the motion, and we do not consider polarized electron beams. Nevertheless, we can take into account the spin precession sufficiently accurate by means of the so-called BMT equation without the need of quantum-mechanical calculations [38].

2.1.1 Equation of Motion

The Lorentz force [37] determines the motion of a particle with charge q in an external electromagnetic field

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}) \quad (2.2)$$

Here, $\vec{E}$ and $\vec{B}$ are the electric field strength and the magnetic induction, respectively. The magnetic force vanishes if the velocity $\vec{v}$ of the particle is zero. According to Newton's law, the force acting on the particle is equal to the temporal change of its kinetic momentum $\vec{p}_k = m\vec{v}$:

$$\frac{d\vec{p}_k}{dt} = \frac{d(m\vec{v})}{dt} = q(\vec{E} + \vec{v} \times \vec{B}). \quad (2.3)$$

$$m = \gamma m_e, \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \quad (2.4)$$

The mass m of the electron is proportional to the relativistic factor γ , which depends on the relative particle velocity $\vec{\beta} = \vec{v}/c$. Accordingly, the kinetic momentum increases very strongly if the velocity of the particle approaches that of light. For the electron, the charge $q = -e$ is negative. The equation of motion is valid only if the particle propagates in vacuum where it does not collide with other particles. In order to realize approximately this ideal situation, the distance along which the particle travels must be smaller than the mean free path length within the residual gas. Unfortunately, Eq. (2.2) can be solved analytically only in rather trivial cases.

To obtain an insight in the general properties of the particle motion, it is advantageous to solve the equation approximately for specific configurations of the electrodes and magnets, which produce the external fields. The development of such calculation procedures is the main task of geometrical charged-particle optics. However, we face in almost all cases the inverse problem to find the electromagnetic field, which affects the path of the particles in a distinct way. Then it is necessary to

find calculation procedures, which yield information on the required course of the trajectories and the geometry and arrangement of the field-producing electrodes and pole pieces.

2.1.2 Conservation of Energy

The electromagnetic field of most electron optical devices does not depend on time. In this case, we can readily obtain a *first integral* of the second-order differential equation (2.3) by scalar multiplication with the differential path length $d\vec{r} = \vec{v} dt$ and subsequent integration over t , giving

$$\int_{t_0}^t \frac{d(m\vec{v})}{dt} \vec{v} dt = \int_{v_0}^v \vec{v} d(m\vec{v}) = -e \int_{\vec{r}_0}^{\vec{r}} \vec{E} d\vec{r}. \quad (2.5)$$

The magnetic term of the Lorentz force does not contribute because it is perpendicular to the velocity.

In the case of stationary magnetic fields ($\partial \vec{A}/\partial t = 0$), we can readily evaluate the last integral by employing the relation

$$\vec{E} = -\text{grad } \varphi - \dot{\vec{A}} = -\text{grad } \varphi. \quad (2.6)$$

The resulting voltage $U = \varphi - \varphi_0$ is the difference between the electric potential φ at the point of observation $\vec{r}$ and the potential φ_0 at the initial point $\vec{r}_0$. We can also evaluate analytically the second integral in (2.5) by partial integration, yielding

$$\begin{aligned} \int \vec{v} d(m\vec{v}) &= m\vec{v}^2 - \int m\vec{v} d\vec{v} = m\vec{v}^2 - \frac{m_e c^2}{2} \int \frac{d(\beta^2)}{\sqrt{1-\beta^2}} \\ &= m\vec{v}^2 + m_e c^2 \sqrt{1-\beta^2} = \frac{m_e c^2}{\sqrt{1-\beta^2}} = mc^2. \end{aligned} \quad (2.7)$$

By inserting this result into Eq. (2.5) and considering the relation (2.6), we obtain the conservation of energy in the relativistic form

$$E = E_0 + E_k + E_p = m_e c^2 + (m - m_e) c^2 - e\varphi = mc^2 - e\varphi = m_0 c^2 - e\varphi_0 = \text{const.} \quad (2.8)$$

The index 0 indicates the value taken at the initial position $\vec{r}_0$. We should not confuse the symbol E for the energy with the vector symbol $\vec{E}$ for the electric field strength. The potential energy $E_p = -e\varphi$ is not a measurable quantity because

the electric potential is not gauge invariant. The kinetic energy $E_k = (m - m_e)c^2$ approaches the classic expression $m_e \tilde{v}^2/2$ in the nonrelativistic limit $\beta \rightarrow 0$.

In the following, we use the gauge such that $\varphi_0 = 0$ at the surface of the cathode where $\tilde{v}_0 = 0$. Then the potential at the point of observation is identical with the voltage U applied between this point and the cathode. Moreover, the constant on the right-hand side adopts the value $\text{const.} = E_0 = m_e c^2$, which coincides with the rest energy E_0 of the electron. In this case, we derive from (2.8) for the velocity v and the kinetic momentum p_k of the electron the expressions

$$v = c \sqrt{\frac{2eU}{E_0}} \frac{\sqrt{1 + eU/2E_0}}{1 + eU/E_0}, \quad p_k = mv = \frac{eU}{c} \sqrt{1 + \frac{2E_0}{eU}}, \quad \gamma = \frac{m}{m_e} = 1 + \frac{eU}{E_0}. \quad (2.9)$$

At the limit $eU \gg E_0$, the velocity approaches the velocity of light c . Any further acceleration increases only the mass and the kinetic momentum in proportion to U (Fig. 2.1).

2.1.3 Hamilton's Principle

We can also derive the Newtonian path equation (2.3) from *Hamilton's principle* of classical mechanics [39]. Hamilton demonstrated that it is possible to obtain the optical laws from a single characteristic function, which one later called *eikonal*, derived from the Greek word $\varepsilon\iota\kappa\omicron\nu$ (icon) meaning image [40]. Hamilton himself showed that the techniques he had developed for handling optical problems are also applicable in mechanics. This is the reason why it is advantageous to treat many problems of charged-particle optics most effectively by means of the eikonal method. We obtain this function most conveniently by employing Hamilton's principle. It states that the true path $\vec{r} = \vec{r}(t)$ of a particle traveling from the initial point $\vec{r}_0$ at time t_0 to the point $\vec{r}$ makes the *action* an *extremum*:

$$W = W(\vec{r}, t; \vec{r}_0, t_0) = Ex \int_{t_0}^t L(\vec{r}(t'), \dot{\vec{r}}(t'), t') dt'. \quad (2.10)$$

The action is a minimum if the point of observation $\vec{r}$ is located in front of the caustic formed by the loci of the points of intersection of adjacent trajectories starting from the common origin $\vec{r}_0$. However, the action may adopt a maximum if the caustic is located between the origin and the point of observation. The caustic can degenerate to a point, which represents the so-called *conjugate* point with respect to the origin $\vec{r}_0$. If we can achieve this condition for all points of a given object plane, we obtain a perfect image of this plane at the corresponding conjugate image plane.

The *Lagrangian* L , which is a function of the position $\vec{r}(t')$ and the velocity $\vec{v} = \dot{\vec{r}}(t')$ of the particle, must be a *Lorentz-invariant* scalar quantity since we consider

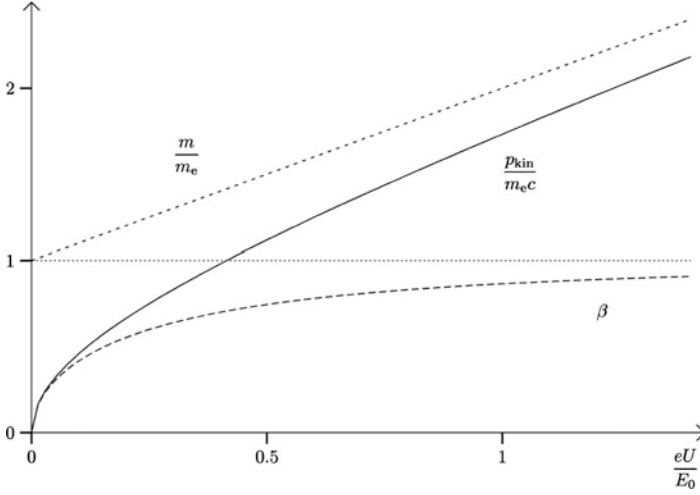


Fig. 2.1 Normalized mass $m/m_e = \gamma$, relative velocity $\beta = v/c$, and normalized kinetic momentum $p_k/m_e c$ as functions of the normalized kinetic energy eU/E_0 , $m_0 = m_e$

relativistic particles. In the classical case, the Lagrangian is the difference between the kinetic energy and the potential energy. To obtain a *covariant* expression for L , we assume the simple case that it is a scalar product of two 4-vectors, one of which is the path length element. In order to avoid the use of metric coefficients, we describe the 4-vectors in Minkowski space. In this case, we have a four-dimensional *pseudo-Euclidian* space where the fourth (time-like) component of any 4-vector is purely imaginary. For example, the four-dimensional position vector has the form $\vec{R} = (x, y, z, ict)$. Using this representation, we obtain for the components of the four-dimensional differential path length element the expressions

$$dx_1 = dx, \quad dx_2 = dy, \quad dx_3 = dz, \quad dx_4 = ic \, dt. \quad (2.11)$$

In order to obtain an action, the other 4-vector must have the dimension of a momentum. The appropriate vector is the *canonical momentum* 4-vector $\vec{P} = (\vec{p}, p_4)$ with the spatial component

$$\vec{p} = m\vec{v} - e\vec{A}, \quad (2.12)$$

where $\vec{A}$ is the magnetic vector potential and the *time-like* imaginary component

$$p_4 = m\dot{x}_4 - eA_4 = i(mc - e\phi/c) = iE/c. \quad (2.13)$$

The comparison of this result with the expression (2.8) shows that the fourth component of the canonical momentum represents the energy up to the imaginary

factor i/c . Scalar multiplication of the canonical momentum 4-vector with the velocity 4-vector yields the Lagrangian in covariant form

$$L = \sum_{\mu=1}^4 p_{\mu} \frac{dx_{\mu}}{dt} = m(\vec{v}^2 - c^2) - e\vec{v}\vec{A} + e\phi = -m_e c^2 \sqrt{1 - \beta^2} + e\phi - e\vec{v}\vec{A}. \quad (2.14)$$

We can readily verify the correctness of this Lagrangian by means of the *Euler–Lagrange* equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_{\mu}} \right) - \frac{\partial L}{\partial x_{\mu}} = 0, \quad \mu = 1, 2, 3. \quad (2.15)$$

We derive these equations from the action function (2.10) by employing the condition $\delta W = 0$ and by keeping the initial and final positions fixed ($\delta \vec{r}_0 = 0, \delta \vec{r} = 0$). By inserting the expression (2.14) for L into Eq. (2.15), we eventually obtain the path equation (2.3). Hence, we can readily determine the action function W if we insert the solutions of this equation in the integrand (2.14) of the integral (2.10) and perform the integration with respect to the independent time variable.

If we vary slightly the coordinates of the point of observation by $\delta \vec{r}$ and the time of observation by δt , we change the path of the particle to a neighboring trajectory starting from the fixed origin $\vec{r}_0$. As a result, the action changes by

$$\delta W = W(\vec{r} + \delta \vec{r}, t + \delta t; \vec{r}_0, t_0) - W(\vec{r}, t; \vec{r}_0, t_0) = \sum_{\mu=1}^4 p_{\mu} \delta x_{\mu}. \quad (2.16)$$

Since we can perform the infinitesimal displacement arbitrarily, we select any one of the four infinitesimal displacements δx_{μ} as nonzero, resulting in

$$\frac{\partial W}{\partial x_{\mu}} = p_{\mu} = m\dot{x}_{\mu} - eA_{\mu} \Rightarrow \frac{\partial W}{\partial x_{\mu}} + eA_{\mu} = m\dot{x}_{\mu}. \quad (2.17)$$

Summation of the squares of the second relation yields the *relativistic Hamilton–Jacobi equation* for the electron

$$\sum_{\mu=1}^4 m^2 \dot{x}_{\mu}^2 = \sum_{\mu=1}^4 (p_{\mu} + eA_{\mu})^2 = \sum_{\mu=1}^4 \left(\frac{\partial W}{\partial x_{\mu}} + eA_{\mu} \right)^2 = m_e^2 \frac{\vec{v}^2 - c^2}{1 - \vec{v}^2/c^2} = -m_e^2 c^2. \quad (2.18)$$

To separate the time-like component from the spatial components, we rewrite the equation in the form

$$\left(\vec{\nabla} W + e\vec{A} \right)^2 - \frac{1}{c^2} \left(\frac{\partial W}{\partial t} - e\phi \right)^2 + m_e^2 c^2 = 0. \quad (2.19)$$

Contrary to the Hamilton–Jacobi equation of classical mechanics, (2.19) is of second degree in the time derivative of the action function W . This behavior results from the condition that the relativistic correct equations must be Lorentz invariant.

A constant action $W = W(\vec{r}, t; \vec{r}_0, t_0) = W_0$ represents a hypersurface in four-dimensional space. We can depict this surface approximately by a discrete set of surfaces $W_n = W(\vec{r}, n\Delta t, \vec{r}_0) = W_0$, $n = 1, 2, \dots$, in the conventional three-dimensional space. If both the magnetic vector potential $\vec{A}$ and the electric potential φ do not depend on the time t , the action function decomposes as

$$W(\vec{r}, t; \vec{r}_0, t_0) = S(\vec{r}, \vec{r}_0, E) + E(t - t_0). \quad (2.20)$$

The *reduced action* or *eikonal* S is a function of the position coordinates and the energy E . By inserting the relation (2.20) into (2.19) and choosing the gauge for φ such that $E = E_0 = m_e c^2$, we obtain the so-called *eikonal equation* of the electron

$$(\vec{\nabla} S + e\vec{A})^2 = m^2 v^2 = 2m_e e \varphi^*. \quad (2.21)$$

Here, $\vec{\nabla} = \text{grad}$ is the *nabla* operator. For reasons of simplicity, we have introduced the *relativistic modified* electric potential

$$\varphi^* = \varphi \left(1 + \frac{e\varphi}{2m_e c^2} \right) \approx \varphi \left(1 + \frac{e\varphi}{1.02 \text{ MeV}} \right), \quad (2.22)$$

The eikonal represents a characteristic function, which governs the imaging properties of the optical system. This function has the properties of an optical potential.

2.1.4 Principle of Maupertuis

The *principle of Maupertuis* or *principle of least action* is the special case of Hamilton’s principle for *conservative* systems. Since the action can also be a maximum, it is more appropriate to use the expression “*principle of stationary action*.” For conservative systems, the total energy $E = -ip_4 c = m c^2$ is constant. As a result, the action (2.10) adopts the form

$$W = E x \sum_{\mu=1}^4 \int_{\vec{R}_0}^{\vec{R}} p_\mu dx_\mu = E x \int_{\vec{r}_0}^{\vec{r}} \vec{p} d\vec{r} + p_4(x_4 - x_{40}) = S - E(t - t_0), \quad (2.23)$$

where $\vec{R} = (\vec{r}, ict)$ denotes the four-dimensional position vector. It readily follows from the relations (2.23) that the reduced action or eikonal

$$S = S(\vec{r}, \vec{r}_0, E) = Ex \int_{\vec{r}_0}^{\vec{r}} \vec{p} \, d\vec{r} \quad (2.24)$$

is also an extremum. This finding is Maupertuis's principle, which we may also write as

$$\delta S = \delta \int_{\vec{r}_0}^{\vec{r}} (m\vec{v} - e\vec{A}) d\vec{r} = 0. \quad (2.25)$$

In order to derive the corresponding Euler–Lagrange equations, we must fix the origin $\vec{r}_0$ and the point of observation $\vec{r}$. If we vary the coordinates of the position vector $\vec{r}$, we readily obtain the relation

$$\vec{\nabla} S + e\vec{A} = m\vec{v}. \quad (2.26)$$

Hence, the direction of the particle velocity $\vec{v}$ is perpendicular to the surfaces of constant reduced action

$$S_\nu(\vec{r}, \vec{r}_0, E) = E(t_\nu - t_0), \quad \nu = 1, 2, \dots, \quad (2.27)$$

only in the absence of a magnetic field ($\vec{A} = 0$), as illustrated in Fig. 2.2. We can interpret the continuous set of wave surfaces (2.27) as a sequence of instant photographs of the propagating discontinuity surface $W = 0$, which are taken at regular intervals of time. The external fields may deform this surface considerably, but they can never tear it into pieces. In the presence of a magnetic field, the actual paths of the particles do *not* coincide with orthogonal trajectories.

By taking the square of the relation (2.26), we readily derive the eikonal equation (2.21). The eikonal (2.24) describes the propagation of an *ensemble* of charged particles, which originate from a common point source.

2.1.5 Time of Flight

We define the time of flight $T = t - t_0$ as the time which the particle needs to travel from its origin $\vec{r}_0$ at time t_0 to the point of observation $\vec{r}$. For reasons of simplicity, we assume stationary electromagnetic fields. In this case, we obtain from the relation (2.8) the expression

$$v = \frac{ds}{dt} = c \sqrt{1 - \frac{E_0^2}{(E + e\phi)^2}}. \quad (2.28)$$

Here, $ds = |d\vec{r}|$ is the differential path length element. The integration of the differential equation (2.28) along the particle trajectory from its origin to its endpoint yields directly the time of flight

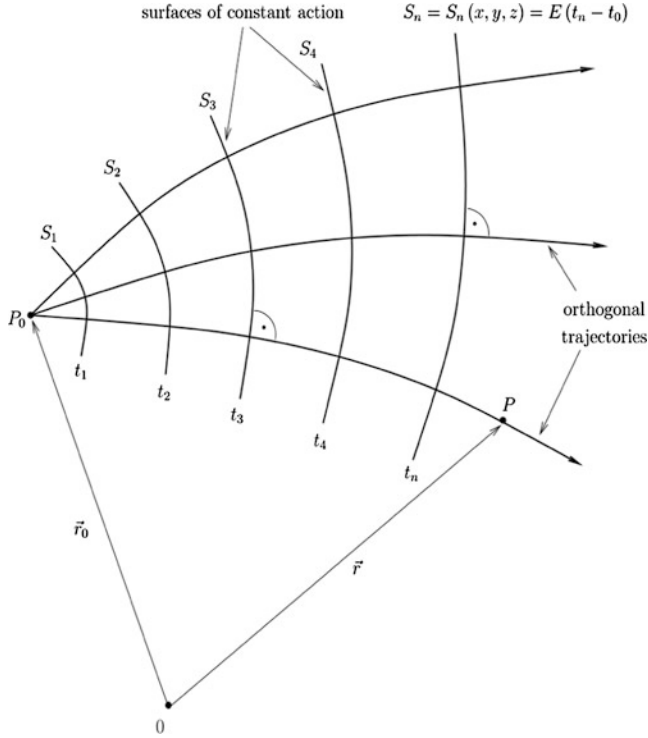


Fig. 2.2 Homocentric paths of electrons in the case $\vec{A} = 0$, $\varphi \neq 0$ representing the orthogonal trajectories of the set of surfaces of constant reduced action $S_v = S_v(\vec{r}, \vec{r}_0; E) = E(t_v - t_0)$

$$T = \frac{1}{c} \int_{\vec{r}_0}^{\vec{r}} \frac{E + e\varphi}{\sqrt{(E + e\varphi)^2 - E_0^2}} ds = \frac{\partial S}{\partial E}. \quad (2.29)$$

By differentiating the expression (2.27) with respect to the total energy E and putting $t_v = t$, we readily obtain the second relation in (2.29).

2.2 Wave Properties of the Electron

Already in 1828, *Hamilton* discovered the close connection existing between the laws of geometrical light optics and the laws of classical mechanics. He showed that the techniques, which he had developed for handling optical problems, are also very useful in mechanics. Today, these methods play a central role in analytical mechanics and quantum mechanics, while they are almost forgotten in light optics. *Newton* assumed that light consists of tiny particles, while *Huyghens* postulated

that light is a wave phenomenon. Moreover, Huyghens demonstrated, in 1690, that one could derive the concept of a *light ray* from the wave formalism without any contradictions. According to Huyghens' principle, each point of a wave surface at time t_0 acts as a source of an elementary wave. This wave is a spherical wave in the field-free region, as shown in Fig. 2.3. The summation of these waves performed at some later time $t = t_0 + \Delta t$ yields the new wave surface, which is the envelope of all elementary waves. The contributions of the backward propagating parts of the elementary waves cancel out by interference. The light rays are the orthogonal trajectories of the set of envelopes formed at times $t_\nu = t_0 + \nu \Delta t$, $\nu = 1, 2, \dots$. The wave description also accounts for diffraction effects, which one cannot explain in the frame of geometrical optics, which represents an approximation for the limit of very short wavelengths ($\lambda \rightarrow 0$).

According to the hypothesis of de Broglie [2], the electron has a particle and a wave property. We consider this duality in close relation to that of light by attributing a frequency $\nu = \omega/2\pi$ and a wavelength $\lambda = 2\pi/k$ to the electron. On account of this analogy, de Broglie postulated that the Einstein relation

$$E = h\nu = \hbar\omega \quad (2.30)$$

is also valid for the electron. Moreover, de Broglie assumed that the equivalent relation

$$\vec{p} = \hbar\vec{k} \quad (2.31)$$

exists between the canonical momentum $\vec{p}$ and the wave vector $\vec{k}$. By defining $k_4 = i\omega/c$ as the time-like component of a wave 4-vector $\vec{K} = (\vec{k}, k_4)$, we can write the two relations (2.30) and (2.31) as a single relativistic *covariant* equation

$$\vec{P} = \hbar\vec{K}. \quad (2.32)$$

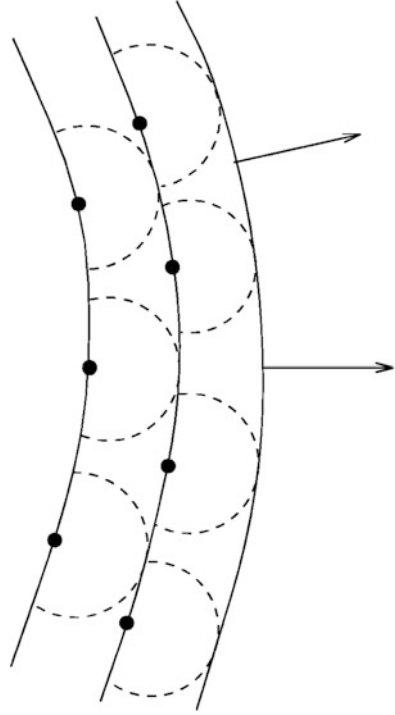
Hence, a matter wave with wave 4-vector $\vec{K}$ is attributed to an elementary particle with a canonical momentum 4-vector $\vec{P}$.

2.2.1 Eikonal and Fermat's Principle

The Hamilton–Jacobi equation is most appropriate for incorporating the wave nature of the electron. According to the rules of quantum mechanics, we must consider the components of the canonical momentum 4-vector as gradient operators

$$p_\mu = \frac{\hbar}{i} \frac{\partial}{\partial x_\mu}. \quad (2.33)$$

These operators act on the wave function $\psi_e = \psi_e(x_\mu) = \psi_e(\vec{r}, t)$. If we neglect the effect of the spin, the wave function is a single component complex function.

Fig. 2.3 Huyghens' principle

By substituting (2.33) for p_μ in the Hamilton–Jacobi equation (2.18), we readily obtain the *Klein–Gordon equation*

$$\sum_{\mu=1}^4 \left(\frac{\hbar}{i} \frac{\partial}{\partial x_\mu} + eA_\mu \right)^2 \psi_e + m_e^2 c^2 \psi_e = 0. \quad (2.34)$$

This four-dimensional wave equation describes correctly the relativistic motion of the electron if we ignore the negligibly small effect of the spin. In the absence of external fields ($A_\mu = 0$), the solutions are plane waves of the form

$$\psi_e = \psi_{e0} e^{iW/\hbar}. \quad (2.35)$$

The phase $W/\hbar = \Omega$ is the Lorentz-invariant scalar product formed by the four-dimensional position vector $\vec{R} = (\vec{r}, ict)$ and the wave 4-vector $\vec{K} = (\vec{k}, k_4 = i\omega/c)$, giving

$$W/\hbar = \sum_{\mu} x_{\mu} p_{k\mu} / \hbar = \sum_{\mu} k_{\mu} x_{\mu} = \vec{k} \vec{r} - \omega t. \quad (2.36)$$

By inserting the solution (2.35) into (2.34), we obtain the conservation of energy

$$\hbar^2(\vec{k}^2 - \omega^2/c^2) + m_e^2 c^2 = m^2 \vec{v}^2 - E^2/c^2 + E_0^2 = 0. \quad (2.37)$$

Here, we do not need to employ the gauge

$$\varphi = -icA_4 = 0, \quad \text{for } \vec{v} = 0. \quad (2.38)$$

In order to derive the eikonal equation (2.21), we assume stationary fields. Moreover, the form (2.35) of the field-free solution suggests the *WKB* ansatz

$$\psi_e = \psi_{e0} e^{i(S-Et)/\hbar}, \quad (2.39)$$

where $S = S(\vec{r})$ is a function of the position $\vec{r}$ of the electron. The WKB approximation of quantum mechanics is equivalent to the much older eikonal approximation of light optics. By inserting the ansatz (2.39) into the wave equation (2.34) and employing both the gauge (2.38) and the Lorentz gauge

$$\sum_{\mu=1}^4 \frac{\partial A_\mu}{\partial x_\mu} = \text{div } \vec{A} + \dot{\varphi}/c^2 = 0, \quad (2.40)$$

we eventually obtain

$$-i\hbar\Delta S + (\vec{\nabla}S + e\vec{A})^2 - 2m_e e\varphi^* = 0. \quad (2.41)$$

Here, $\Delta = \vec{\nabla}^2$ is the *Laplace operator*. In the classical limit $\hbar \rightarrow 0$, Eq. (2.41) reduces to the eikonal equation

$$(\vec{\nabla}S + e\vec{A})^2 = 2m_e e\varphi^*. \quad (2.21)$$

The solution of the eikonal equation (2.21) satisfies *Fermat's principle* (Fermat, 1679), which states that the optical path $L = S/q_0$ between the origin $\vec{r}_0$ and the point of observation $\vec{r}$ is an extremum:

$$L = S/q_0 = Ex \int_{\vec{r}_0}^{\vec{r}} n(\vec{r}) d\vec{s} = \frac{1}{k_0} \int_{\vec{r}_0}^{\vec{r}} \vec{k} d\vec{r}. \quad (2.42)$$

The use of variational principles dates back to the earliest Greek philosophers. They derived them on ground of their aesthetic and metaphysical ideal of simplicity for the laws of nature. Hero of Alexandria (125 B.C.) made the first rigorous use of an optical variational procedure when he proved that for a mirror, the angle of incidence equals the angle of reflection. He showed that in this case, the path taken by a ray from the object to the observer is the shortest of all possible paths. Fermat's principle is an extension of this principle for media with spatially varying index of refraction. We have chosen the normalization momentum

$$q_0 = \hbar k_0 = \sqrt{2em_e\Phi_0^*} \quad (2.43)$$

in such a way that the index of refraction for charged particles

$$n = n(\vec{r}) = k/k_0 = \sqrt{\frac{\varphi^*}{\Phi_0^*} - \frac{e}{q_0} \vec{A} \cdot \vec{e}_t} \quad (2.44)$$

is unity in the absence of an electromagnetic field ($\vec{A} = 0$, $\varphi^* = \varphi_0^* = \Phi_0^*$) in the space between the ray-defining points $\vec{r}$ and $\vec{r}_0$. Here, Φ denotes the electric potential on the optic axis, $\lambda_0 = 2\pi/k_0$ is the wavelength at the center of the starting plane $z = z_0$.

Our definition of the index of refraction corresponds to that of light optics because the optical path length (2.42) for charged particles degenerates to the geometrical distance $l = |\vec{r} - \vec{r}_0|$ between the ray-defining points in the absence of an electromagnetic field, as it is the case for light rays propagating in vacuum. From the point of view of wave optics, Fermat's principle is a direct consequence of the fact that the light rays are the orthogonal trajectories to the wave surfaces

$$k_0 l - \omega t = \text{const.} \quad (2.45)$$

To prove this behavior, we consider a set of wave surfaces $l_\nu = l_0 + \nu\lambda$, $\nu = 0, 1, \dots, n$, shown in Fig. 2.4. The separation between any two adjacent wave surfaces is chosen to be equal with the wavelength. We consider an arbitrary path connecting the origin P_0 with the endpoint P , as illustrated by the dashed curve. The solid curve represents the orthogonal trajectory. It readily follows from the figure that we can write the optical path length along the dashed curve as

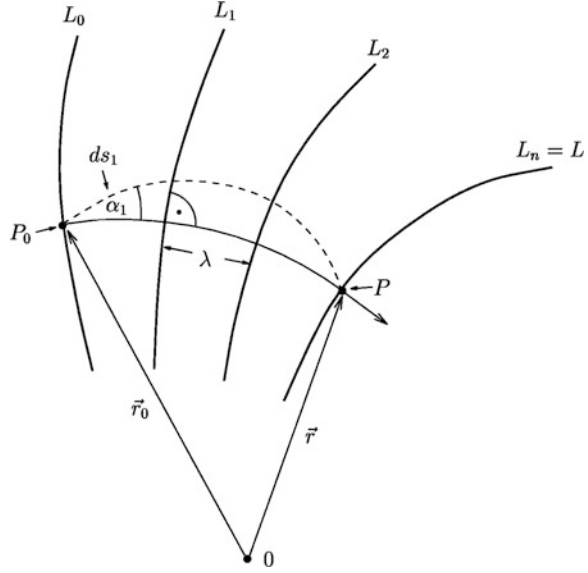
$$\int_{\vec{r}_0}^{\vec{r}} n \, ds = \frac{1}{k_0} \int_{\vec{r}_0}^{\vec{r}} k \, dz = \lambda_0 \int_{\vec{r}_0}^{\vec{r}} \frac{ds}{\lambda} \approx \lambda_0 \sum_{\nu=1}^n \frac{\Delta s_\nu}{\lambda_\nu} = \sum_{\nu=1}^n \frac{\lambda_0}{\cos \alpha_\nu}. \quad (2.46)$$

This length adopts a minimum if $\alpha_\nu = 0$. Hence, the true path is the trajectory, which is orthogonal to the wave surfaces. The second relation in Eq. (2.44) describes this behavior, as can be seen by taking the gradient.

2.2.2 Phase, Wavelength, Frequency, Phase and Group Velocity, and Index of Refraction

Already in 1828, Hamilton discussed the close “*formal*” relation between Fermat's principle of optics and Maupertuis's principle of mechanics. Owing to Hamilton's profound knowledge of optics and mechanics, it is very likely that he did not

Fig. 2.4 Fermat's principle
($L_n = l_n$)



consider this equivalence to be a meaningless coincidence. However, it took almost 200 years until de Broglie postulated that this equivalence is real for elementary particles reflecting the dualism between their wave and particle nature. Accordingly, we postulate as the *phase* of the matter wave

$$\Omega(\vec{r}, t) = \frac{W}{\hbar} = \frac{1}{\hbar} \sum_{\mu=1}^4 \int k_{\mu} dx_{\mu} = \int \vec{k} d\vec{r} - \omega t. \quad (2.47)$$

We know from electron microscopy that the phase of the scattered electron wave contains the information about the atomic structure of the object. Unfortunately, the geometrical path of the electron through the object is difficult to calculate, except for fast electrons passing through very thin objects (few atomic layers). In this special case, the electrons move approximately along straight lines through the object. These conditions are fulfilled in the electron microscope at medium voltages for thin amorphous objects, which behave like phase objects in light microscopy.

The energy E of a photon is related to its frequency ν by the Einstein relation $E = h\nu$. Both quantities are measurable. This is not the case for electrons and ions because we cannot unambiguously define their energy $E = E_0 + E_{\text{kin}} + E_{\text{pot}}$ since the electric potential φ and the related potential energy E_{pot} are not gauge invariant. Therefore, one can define the frequency of a charged-particle wave only up to an arbitrary constant. As a result, we can only measure differences of frequencies, as it is the case in any interference experiment.

The same behavior holds true for the wave vector $\vec{k} = (m\vec{v} - e\vec{A})/\hbar = \vec{\nabla}S/\hbar$. We cannot measure it because the magnetic vector potential is not gauge invariant.

Moreover, we confront the additional difficulty that the direction $\vec{e}_t = \vec{v}/v$ of the particle trajectory is not perpendicular to the wave surfaces in the presence of a magnetic field ($\vec{A} \neq 0$). Hence, the distance between any adjacent wave surfaces S_n and S_{n+1} measured along any trajectory does not represent the shortest distance $2\pi/k = \lambda$, as demonstrated in Fig. 2.5. The distance along the trajectory equals the wavelength λ only in the absence of a magnetic field.

In order to retain this convention, we define the wavelength of the electron wave in the same way as

$$\lambda = \frac{2\pi}{\vec{k}\vec{e}_t} = \frac{h}{mv - e\vec{A}\vec{e}_t} = \frac{2\pi}{k \cos \alpha}. \quad (2.48)$$

Here, α defines the angle between the direction of the actual path and the direction of the canonical momentum or wave vector, respectively. In the absence of a magnetic field ($\vec{A} = 0$), the wavelength

$$\lambda = \frac{h}{mv} = \frac{h}{m_e c} \frac{m_e c}{mv} = \lambda_C \sqrt{\frac{E_0}{2eU^*}} \approx \sqrt{\frac{1.5 \text{ V}}{U^*}} \text{ nm} \quad (2.49)$$

is a measurable quantity because the relativistic modified acceleration voltage $U^* = U(1 + eU/2E_0)$ is gauge invariant; $\lambda_C = 2\pi/k_C = 2.426 \text{ pm}$ denotes the *Compton wavelength*. For an accelerating voltage $U \approx U^* = 150 \text{ V}$ with respect to the cathode potential, the wavelength equals $1 \text{ \AA}$, which is roughly the diameter of a hydrogen atom. Therefore, the resolution limit

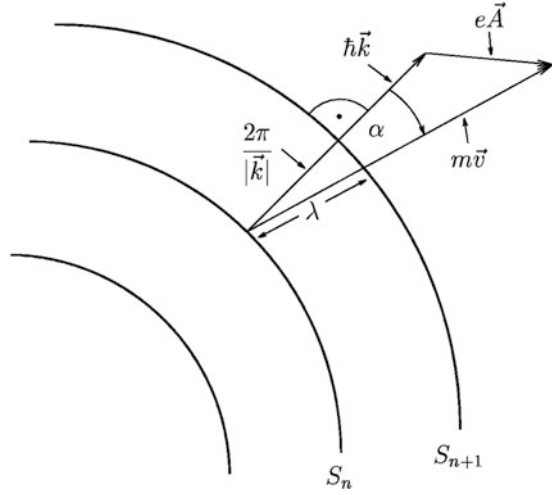
$$d \approx \lambda/\theta \quad (2.50)$$

of the electron microscope (EM) is very small. Unfortunately, the spherical aberration of the round lenses limits the maximum usable aperture angle $\theta \approx 0.01$ in conventional EMs. As a result, such EMs cannot achieve sub-angstrom resolution at voltages below about 1 MV. This behavior is the reason for the ongoing efforts to compensate for the unavoidable chromatic and spherical aberration of round lenses (*Scherzer theorem* [8]) by means of multipole or mirror correctors, respectively. We shall treat extensively the different correction methods in Chaps. 9 and 10.

One characterizes refracting media in light optics by their *index of refraction* $n = \lambda_v/\lambda$. We can use this definition also for the particle wave if we substitute λ_C for the vacuum wavelength λ_v of light. By employing the relation (2.48), we readily obtain the particle optics index of refraction as

$$n = \frac{\lambda_0}{\lambda} = \frac{m\vec{v} - e\vec{A}\vec{e}_t}{q_0} = \sqrt{\frac{\varphi^*}{\Phi_0^*}} - \frac{e}{q_0} \vec{A}\vec{e}_t. \quad (2.51)$$

Fig. 2.5 Definition of the wavelength of the electron wave in the presence of a magnetic vector potential



In analogy to light optics, the electromagnetic field represents an inhomogeneous *anisotropic* medium of refraction for the charged particles. The anisotropy stems from the directional dependence of n on the direction of flight of the particle in the presence of a magnetic field. Therefore, only electrostatic systems have an *isotropic* index of refraction. Using the terminology of light optics, all electron lenses represent *gradient-index* lenses because the electromagnetic potentials are continuous functions of the spatial coordinates, which cannot change abruptly at a given surface, as does the light-optical index of refraction at the surface of a lens.

The phase (2.47) of the electron wave cannot be measured because each component $k_\mu = (m\dot{x}_\mu - eA_\mu)/\hbar$ of the wave 4-vector depends on the 4-vector potential. Its component A_μ is only determined up to the derivative $\partial F/\partial x_\mu$ of an arbitrary scalar function $F = F(\vec{r}, x_4)$. By introducing this function, we change the gauge of the 4-vector potential ($\vec{A}, A_4 = i\varphi/c$) resulting in the phase

$$\Delta F = \sum_{\mu=1}^4 \int_{x_{\mu 0}}^{x_\mu} \frac{\partial F}{\partial x_\mu} dx_\mu = F - F_0. \quad (2.52)$$

Therefore, it is not possible to measure the absolute phase of the particle wave. This result is plausible because we must measure the phase by an interference experiment, which records phase differences or differences of wave vectors, respectively.

The frequency $\nu = \omega/2\pi$ of the electron wave relates to its energy E in the same way as in the case of light:

$$E = \hbar\omega = -icp_4 = -ic(m\dot{x}_4 - eA_4) = (mc^2 - e\varphi). \quad (2.53)$$

By employing the relation

$$\begin{aligned} \frac{1}{\hbar^2} \sum_{\mu=1}^4 m^2 \dot{x}_{\mu}^2 &= \frac{1}{\hbar^2} \sum_{\mu=1}^4 (p_{\mu} + eA_{\mu})^2 = \sum_{\mu=1}^4 (k_{\mu} + eA_{\mu}/\hbar)^2 \\ &= (\vec{k} + e\vec{A}/\hbar)^2 - (\omega + e\varphi/\hbar)^2/c^2 = -k_C^2, \end{aligned} \quad (2.54)$$

the frequency can be expressed in the form of a dispersion relation as

$$\omega = -e\varphi/\hbar + c\sqrt{(\vec{k} + e\vec{A}/\hbar)^2 + k_C^2}. \quad (2.55)$$

Since both the frequency and the wave vector depend on the gauge of the 4-vector potential, the phase velocity $v_p = \omega/k$ is not gauge invariant and, therefore, not measurable. Fortunately, this behavior is of no concern because it is not possible to transfer any information by means of a single monochromatic wave. We can transfer a signal only by means of a wave package formed by a superposition of waves with different wave vectors. This superposition produces a *beat*, which propagates with the measurable *group velocity*

$$\vec{v}_g = \vec{\nabla}_k \omega = c^2 \frac{\vec{k} + e\vec{A}/\hbar}{\omega + e\varphi/\hbar} = c^2 \frac{m\vec{v}}{mc^2} = \vec{v}. \quad (2.56)$$

The beat of the modulated particle wave propagates with the same measurable velocity as the corpuscular particle. In the presence of a magnetic field ($\vec{A} \neq 0$), the elementary Huyghens' waves are no longer spherical waves. They form elliptical waves in the case of constant vector potential. The corresponding wave surfaces are rotational ellipsoids where one of the two principal axes is located in the direction of the particle trajectory. One of the two focal points of the ellipsoid is located at the origin of the elementary wave. Using these elementary waves, Huyghens' construction of the wave surfaces is also applicable in the presence of an electromagnetic field. In this case, we must choose the distance between neighboring wave surfaces in such a way that the vector potential does not appreciably vary in the region between any two subsequent wave surfaces.

2.3 Aharanov–Bohm Effect

A very instructive example for the influence of the vector potential on the phase of the electron wave is the *Aharonov–Bohm effect* [41]. To demonstrate this effect, we consider the experimental arrangement of Moellenstedt and Dueker [42] shown in Fig. 2.6. It consists of a positively charged wire, forming an electron optical biprism, and a bifilar solenoid with an adjustable current placed below the wire. The current

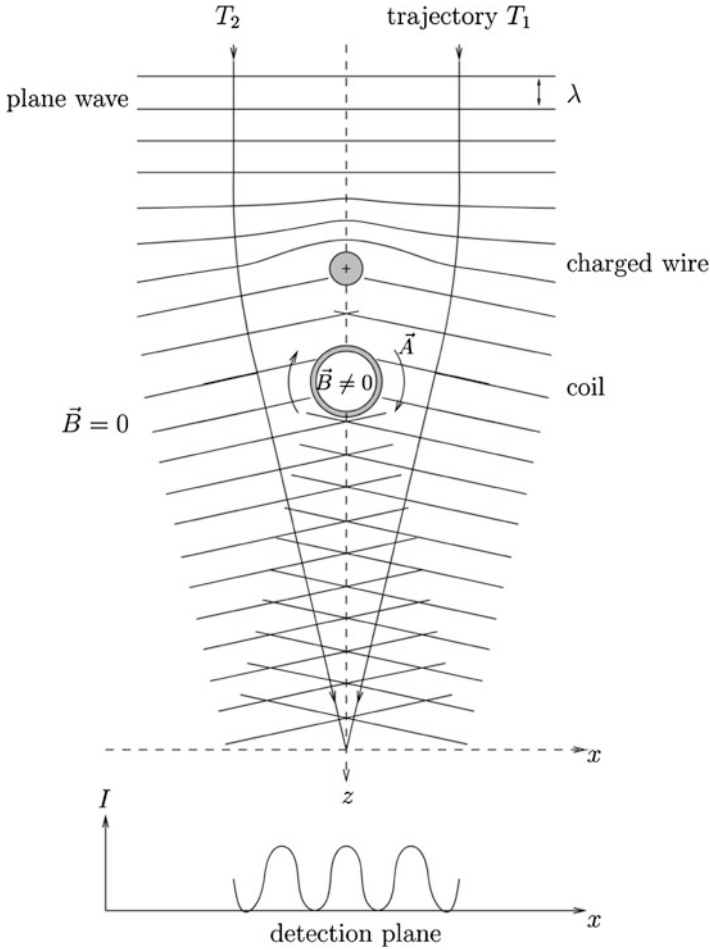


Fig. 2.6 Schematic arrangement of the Moellenstedt experiment in the case $\Theta = 0$ demonstrating the Aharonov–Bohm effect

produces a magnetic field only in the interior of the coil. Although the magnetic field vanishes in the region outside of the coil, the magnetic vector potential does not ($\vec{A} \neq 0$). We readily prove this condition by employing *Stoke's theorem*, giving

$$\oint \vec{A} d\vec{s} = \iint (\vec{\nabla} \times \vec{A}) d\vec{\sigma} = \iint \vec{B} d\vec{\sigma} = \Phi_m. \quad (2.57)$$

Here, $d\vec{\sigma}$ denotes the differential surface element of the area enclosed by the loop. If we put this contour in the field-free regime, we can move it without changing the value of the loop integral. If we vary the magnetic flux Φ_m by changing the current

in the coil, the vector potential must change in the entire outer space too. Hence, we cannot nullify the vector potential everywhere in this region by means of a gauge.

We further assume a plane wave for the incident electron, whose direction of propagation is parallel to the dashed line through the centers of the wire and the coil. The biprism splits the wave $\psi_e = \psi_{e1} + \psi_{e2}$ in two coherent partial waves

$$\psi_{e1} = \psi_{e0} e^{i\Omega_1(\vec{r},t)}, \quad \psi_{e2} = \psi_{e0} e^{i\Omega_2(\vec{r},t)} \quad (2.58)$$

which propagate in different directions and interfere in the region beneath the coil. The phases Ω_1 and Ω_2 are imaginary in the region where the intensity of the partial waves is negligibly small. In the detection plane, the overlapping parts of the waves form an interference pattern with intensity

$$I = \psi_e \bar{\psi} = 2 |\psi_{e0}|^2 [1 + \cos(\Omega_1 - \Omega_2)]. \quad (2.59)$$

The locations of the maxima and zeros of the intensity are given by the phase difference

$$\begin{aligned} \Delta\Omega &= \Omega_1 - \Omega_2 = \frac{1}{\hbar} \int_{T_1} (m\vec{v} - e\vec{A}) d\vec{s} - \frac{1}{\hbar} \int_{T_2} (m\vec{v} - e\vec{A}) d\vec{s} \\ &= \frac{mv}{\hbar} (l_1 - l_2) + \frac{e}{\hbar} \oint \vec{A} d\vec{s} = 2k_0 x \sin \vartheta \cos \Theta + \frac{e\Phi_m}{\hbar}. \end{aligned} \quad (2.60)$$

The biprism determines the deflection angle ϑ which is positive for the trajectory T_1 and negative for the trajectory T_2 . Electrons attributed to a single plane wave start from a common point source. The tilt angle Θ of the incident plane wave with respect to the optic z -axis defines the angular position of the point source.

In the case $\Theta = 0$, the path lengths l_1 and l_2 of the symmetric trajectories T_1 and T_2 coincide. Hence, the intensity at the center $x = 0$ of the detection plane depends only on the magnetic flux within the coil:

$$I = I_0 [1 + \cos(e\Phi_m/\hbar)]. \quad (2.61)$$

For an ideal point source, the fringe pattern extends over the entire detection plane. However, this ideal case is never realized in practice due to the finite extension of the effective source. Each point of this source radiates incoherently. For our purposes, it suffices to approximate the angular radiation characteristic of the source by the Gaussian distribution

$$R(\Theta) = \frac{1}{\Theta_0 \sqrt{\pi}} e^{-\Theta^2/\Theta_0^2}. \quad (2.62)$$

The integral over the normalized distribution (2.59) is unity. By employing this distribution, (2.59) and (2.60), we eventually obtain

$$\begin{aligned}
\langle I(x) \rangle / I_0 &= 1 + \int_{-\infty}^{\infty} R(\Theta) \cos \left(2k_0 x \sin \vartheta \cos \Theta + \frac{e\Phi_m}{\hbar} \right) d\Theta \\
&\approx 1 + \frac{\cos \left(2k_0 x \sin \vartheta \cos^2(\Theta_0/2) + \frac{e\Phi_m}{\hbar} \right)}{\sqrt[4]{1 + 4k_0^2 x^2 \sin^2(\Theta_0/2)}}.
\end{aligned} \tag{2.63}$$

Moellenstedt's experiment realizes this interference pattern. The fringes move when the current is changed, whereas the envelope stays fixed. The motion of the interference pattern results from the electric field $\vec{E} = -\dot{\vec{A}}$ induced by a temporal change $\dot{\vec{A}}$ of the vector potential. This relation is a direct consequence of Faraday's law

$$\dot{\Phi}_m = \frac{d}{dt} \iint \vec{B} \cdot d\vec{\sigma} = \frac{d}{dt} \oint \vec{A} d\vec{r} = - \oint \vec{E} d\vec{r}. \tag{2.64}$$

The induced electric field $\vec{E}$ in the outer region of the cylinder alters the classical path of the electron and the phase of its wave. The result of the experiment is of entirely quantum-mechanical nature because it originates from variations of the phases or wave surfaces, respectively. The stripes of the interference pattern move steadily if the current increases continuously because the magnetic flux has the form

$$\Phi_m(t) = \Phi_m(t_0) + (t - t_0)\dot{\Phi}_m = \Phi_m(t_0) - (t - t_0) \oint \vec{E} d\vec{r}. \tag{2.65}$$

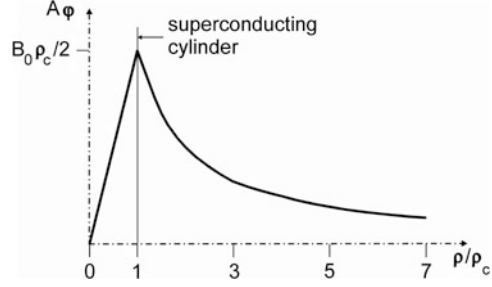
The result of the Moellenstedt experiment convincingly demonstrates the physical reality of the magnetic vector potential, contrary to the general belief that this quantity is a pure mathematical construct. The invention of the electron optical biprism and the development of highly coherent field emission electron guns gave birth to electron holography, which has become an important technique for determining electric and magnetic fields in solid objects on an atomic scale [43, 44].

In order to elucidate the Aharonov–Bohm effect in more detail, we assume an infinitely extended thin superconducting cylinder with radius ρ_c , as shown in Fig. 2.7. The axis of the cylinder defines the y -axis of the y, ρ, φ coordinate system. Moreover, we suppose that a current with constant current, I , per unit length of the cylinder is circulating in azimuthal φ direction. The corresponding current density is given by

$$\vec{j} = I\delta(\rho - \rho_c)\vec{e}_\varphi. \tag{2.66}$$

The magnetic flux density $\vec{B} = B_0\vec{e}_y$ is constant within the interior region of the cylinder and zero outside. Owing to the *Meissner effect*, the magnetic flux density is zero within the superconducting cylinder. We assume that the magnetic field lines are closed at infinity. The magnetic vector potential $\vec{A} = A_\varphi(\rho)\vec{e}_\varphi$ does not vanish in the outer region $\rho > \rho_c$ because $\vec{A}$ must be continuous at $\rho = \rho_c$. This behavior does not hold for its slope, which is discontinuous on the mantle of the cylinder, as

Fig. 2.7 Course of the magnetic vector potential $A_\varphi(\rho/\rho_c)$ produced by an azimuthal current within an infinitely thin superconducting cylinder with radius ρ_c



illustrated in Fig. 2.7. By considering this behavior, we write the magnetic vector potential as

$$\begin{aligned}\vec{A} &= \frac{1}{2} \vec{B} \times \vec{\rho} = A_\varphi \vec{e}_\varphi = \frac{\rho}{2} B_0 \vec{e}_\varphi, 0 \leq \rho < \rho_c, \\ \vec{A} &= A_\varphi \vec{e}_\varphi = \frac{B_0 \rho_c^2}{2\rho} \vec{e}_\varphi, \rho \geq \rho_c.\end{aligned}\quad (2.67)$$

To prove the validity of the relations (2.67), we calculate the magnetic flux density

$$\vec{B}(\rho) = \vec{\nabla} \times \vec{A} = \frac{\vec{e}_y}{\rho} \frac{\partial(\rho A_\varphi)}{\partial \rho} = \vec{e}_y B_0 \Theta(\rho/\rho_c). \quad (2.68)$$

The box function $\Theta(x)$ has the property

$$\Theta(x) = \begin{cases} 1 & \text{for } 0 \leq |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases}. \quad (2.69)$$

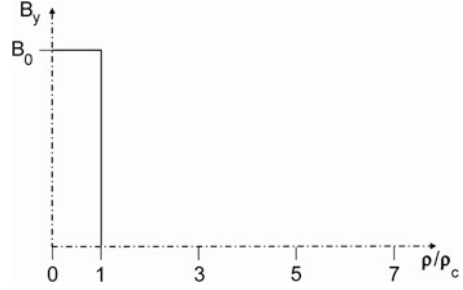
Accordingly, the magnetic field strength jumps at the boundary $\rho = \rho_c$ from its constant interior value B_0 to zero within the region $\rho \geq \rho_c$. The magnetic vector potential $A_\varphi(\rho)$ and the attributed magnetic flux density $B(\rho)$ are shown in Figs. 2.7 and 2.8, respectively, as functions of the normalized radial distance ρ/ρ_c .

We prove the validity of our representation of the magnetic vector potential by employing the relation $\vec{\nabla} \times \vec{B} = \mu_0 \vec{j}$. As a result, we obtain

$$\begin{aligned}\vec{\nabla} \times \vec{B} &= \vec{e}_\varphi \frac{\partial B_y(\rho)}{\partial \rho} \\ &= -\vec{e}_\varphi B_0 \delta(\rho - \rho_c) = \vec{e}_\varphi \mu_0 I \delta(\rho - \rho_c).\end{aligned}\quad (2.70)$$

The last relation readily gives $B_0 = -\mu_0 I$. To further prove the validity of the expressions (2.67) for the magnetic vector potential of our model, we calculate the circulation integral of the magnetic vector potential and the magnetic flux Φ_m contained within the circle with radius ρ , giving

Fig. 2.8 Magnetic flux density related to the magnetic vector potential depicted in Fig. 2.7



$$\oint \vec{A} \, d\vec{s} = \int_0^{2\pi} A_\phi \rho \, d\phi = \begin{cases} \pi \rho^2 B_0 & 0 \leq \rho \leq \rho_c \\ \pi \rho_c^2 B_0 & \rho \geq \rho_c \end{cases}, \quad (2.71)$$

$$\Phi_m = \int_0^{2\pi} \int_0^\rho B(\rho) \rho \, d\rho \, d\phi = \begin{cases} \pi \rho^2 B_0 & 0 \leq \rho \leq \rho_c \\ \pi \rho_c^2 B_0 & \rho \geq \rho_c \end{cases}.$$

The result demonstrates that the magnetic vector potential (2.67) satisfies Stoke's theorem

$$\oint \vec{A} \, d\vec{s} = \iint_\sigma (\vec{\nabla} \times \vec{A}) \cdot d\vec{\sigma} = \iint_\sigma \vec{B} \cdot d\vec{\sigma} = \Phi_m. \quad (2.72)$$

Moellenstedt's experiment proves convincingly that the fringes move when the current is changed. This change alters the phase of the electron wave but not the classical path of the electrons. Therefore, the result of the experiment is of entirely quantum-mechanical nature because it originates from variations of the phases or wave surfaces, respectively. The change of the phases results from the change of the vector potential, which depends on its boundary values determined by the current in the coil.

2.4 Ray Properties Associated with the Eikonal

Owing to the existence of the wave surfaces or action surfaces, respectively, the trajectories of particles originating from a common point can never mingle arbitrarily because the directions of their associated wave vectors remain always perpendicular to the wave surfaces. However, in the presence of an electromagnetic field, the initially *homocentric* bundle of rays will generally not be homocentric elsewhere such that the asymptotes intersect each other in a common point for all wave surfaces. This situation would only be the case for a rotationally symmetric ideal lens, which does not exist for charged particles. As a result, a spherical wave will not remain spherical if it propagates within an electromagnetic field. However, this behavior does not necessarily prevent an ideal imaging. We achieve such a point-to-point imaging because the ideal imaging system transfers an initially outgoing spherical wave located in the *object space* into a converging spherical wave in the *image space*. Then the optical path length $L = S/q_0$ between the object

point P_o and the image point P is the same for all rays connecting these *conjugate* points, as depicted in Fig. 2.9. This condition is less stringent since the bundle of rays needs not to be continuously homocentric in the region between the object and the image. We encounter approximately this situation in an aberration-corrected electron microscope.

The trajectories are perpendicular to the wave surfaces only in the absence of a magnetic field. In this case, the trajectories can never screw around each other. The magnetic field can produce such a twist only because in this case, the rays are not orthogonal to the wave surfaces. A measure for the “*screwing*” of the trajectories is the *circulation*

$$C = \oint m\vec{v} d\vec{s} = \oint \vec{\nabla} S d\vec{s} + e \oint \vec{A} d\vec{s} = e \iint \vec{B} d\vec{\sigma} = e \Phi_m. \quad (2.73)$$

The line integration has to be taken around a loop enclosing the boundary trajectories of a bundle of rays on a wave surface. We must perform the two-dimensional integration over the area enclosed by the loop. We obtain the last integral by applying *Stoke’s theorem* together with $\vec{\nabla} \times \vec{A} = \vec{B}$ and by considering that $\oint \vec{\nabla} S d\vec{s} = S - S = 0$. The result demonstrates that the screwing of the trajectories is proportional to the magnetic flux penetrating through the area of the wave surface formed by the loop, which encircles the bundle of rays. In hydrodynamics, the circulation defines the *curl strength* of a flow.

The curvature $\vec{\kappa}$ and torsion τ of the trajectory define the instantaneous change in the course of the particle at any given position. We obtain these quantities most conveniently by considering that the curl of the canonical momentum is zero:

$$\vec{\nabla} \times \vec{p} = \vec{\nabla} \times \vec{\nabla} S = \vec{\nabla} \times m\vec{v} - e\vec{\nabla} \times \vec{A} = \vec{\nabla} \times m\vec{v} - e\vec{B} = 0. \quad (2.74)$$

It should be noted that both $\vec{p} = \vec{p}(\vec{r}, \vec{r}_0)$ and $\vec{v} = \vec{v}(\vec{r}, \vec{r}_0)$ must be conceived as functions of the coordinates of the initial position $\vec{r}_0$ and the point of observation $\vec{r}$. This differs from the usual case where one fixes the trajectory by its position and slope at the origin.

The curvature $\vec{\kappa}$ and the torsion τ determine the rotation of the accompanying *Frenet–Serret trihedral* defined by the orthogonal unit vectors $\vec{e}_t, \vec{e}_n = \vec{\kappa}/\kappa$, and $\vec{e}_b = \vec{e}_t \times \vec{e}_n$ shown in Fig. 2.10. If we know the *tangential unit vector* $\vec{e}_t = \vec{v}/v$ and the electromagnetic field at a given position of the particle, both the curvature and the torsion of the trajectory can be readily determined from the relations

$$\begin{aligned} \vec{\kappa} &= \kappa \vec{e}_n = \frac{d\vec{e}_t}{ds} = (\vec{e}_t \vec{\nabla}) \vec{e}_t = -\vec{e}_t \times (\vec{\nabla} \times \vec{e}_t), \\ \tau &= \frac{d\vartheta_\tau}{ds} = -\vec{e}_n \frac{d\vec{e}_b}{ds} = \vec{e}_n \left(\frac{d\vec{e}_n}{ds} \times \vec{e}_t \right) = \frac{\vec{e}_t}{\kappa^2} \left(\vec{\kappa} \times \frac{d\vec{\kappa}}{ds} \right). \end{aligned} \quad (2.75)$$

The tangential vector and the *normal unit vector* $\vec{e}_n$ define the *tangential plane*, which embeds the differential path length. The reciprocal curvature $1/\kappa = \rho$

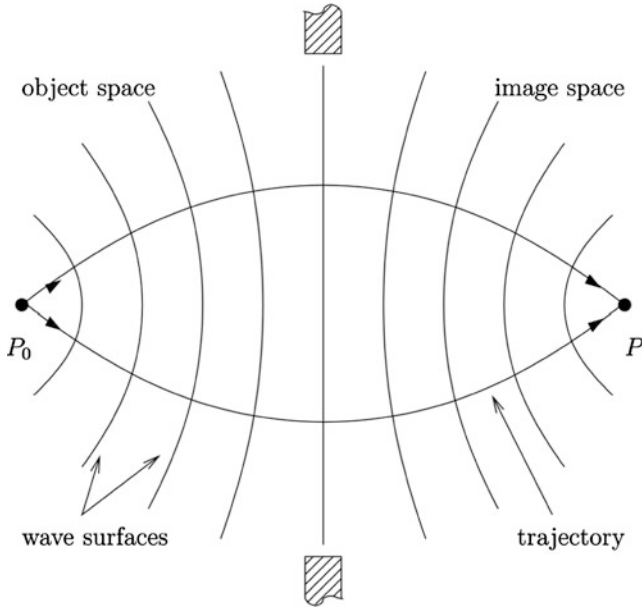


Fig. 2.9 Wave surfaces and particle trajectories in the case of ideal imaging

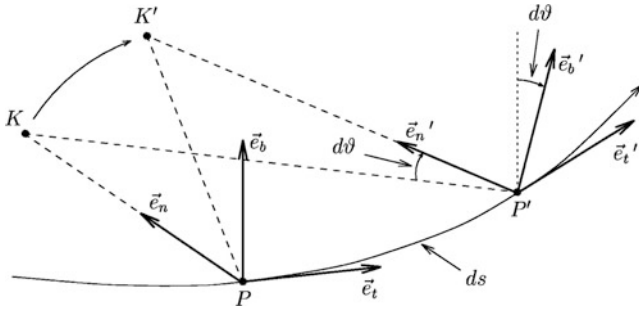


Fig. 2.10 Motion of the accompanying Frenet-Serret trihedral along a curved trajectory in the absence of a tangential component of the magnetic field ($B_t \equiv \vec{e}_t \vec{B} = 0$)

represents the *radius of curvature* whose origin K defines the *momentary center of curvature* of the trajectory at the point P , as illustrated in Fig. 2.10. The *normal* unit vector $\vec{e}_n$ points toward the center of curvature; the *binormal* unit vector $\vec{e}_b$ is perpendicular to the tangential plane. Both vectors rotate about the tangent by the differential angle $d\vartheta$ if the point P moves along the trajectory by the differential arc length ds .

We obtain the curl of the tangential unit vector from the last equation of (2.74) as

$$\vec{\nabla} \times (m\vec{v}\vec{e}_t) - e\vec{B} = m\nu(\vec{\nabla} \times \vec{e}_t) - \vec{e}_t \times \vec{\nabla} m\nu - e\vec{B} = 0. \quad (2.76)$$

Using this result, we derive from (2.75) the expression

$$\vec{\kappa} = \left[e v \vec{B} + \vec{v} \times \vec{\nabla} m v \right] \times \frac{\vec{v}}{m v^3} = \frac{e \vec{B} \times \vec{v}}{m v^2} - \frac{1 + e \varphi / E_0}{2} \frac{\vec{E}_\perp}{\varphi^*} \quad (2.77)$$

for the vector of curvature, where $\vec{E}_\perp = (\vec{e}_t \times \vec{E}) \times \vec{e}_t$ is the component of the electric field strength perpendicular to the direction of the particle velocity. Hence, if both $\vec{B}$ and $\vec{E}$ point in the direction of the velocity, the trajectory will not be curved. This behavior does not hold true for the torsion τ . Employing the relations (2.76) and (2.73), we eventually find from (2.74) after a lengthy calculation for the torsion the expression

$$\tau = \frac{e \vec{v} \vec{B}}{m v^2} - \frac{\vec{\kappa}}{\kappa^2} \left[(\vec{e}_t \vec{\nabla}) \frac{e \vec{B}}{m v} + \vec{e}_t \times (\vec{e}_t \vec{\nabla}) \frac{\vec{\nabla} m v}{m v} \right]. \quad (2.78)$$

The expression in the bracket vanishes for a constant electromagnetic field. In this case, the inverse torsion

$$\frac{1}{\tau} = \rho_L = \frac{m v}{e B_t} \quad (2.79)$$

coincides with the radius ρ_L of the *Larmor rotation*, where $B_t = \vec{B} \vec{e}_t = \vec{B} \vec{v} / v$ is the absolute value of the tangential component of the magnetic field in the direction of the velocity. The corresponding *angle of Larmor rotation* is given by

$$\vartheta_L = \int_{\vec{r}_0}^{\vec{r}} \frac{e \vec{B}}{m v} d\vec{s}. \quad (2.80)$$

It is important to note that the Larmor rotation does not affect the location of the center of curvature. Hence, in order to guarantee that the normal unit vector $\vec{e}_n$ of the accompanying triad always points in the direction of the center of curvature K , we must rotate the triad back by the angle ϑ_L . Although the torsion results primarily from the tangential component of the magnetic field, as expected from the relation (2.72) for the circulation, this component does not affect the curvature of the trajectory.

2.5 Propagation of the Electron Wave in Macroscopic Fields

According to the *Huyghens' principle* of light optics, we can consider each point of the wave front as the center of a secondary disturbance which gives rise to an elementary wave. The envelope of all elementary waves at any later instant $t - t_0$ forms the new wave front. This idea was put on a sound mathematical basis by Kirchhoff and Sommerfeld for treating diffraction problems in light optics [45, 46].

Diffraction effects arise in the neighborhood of shadows produced by apertures and in regions where a large number of rays intersect that start from a common point. Kichhoff and Sommerfeld applied their theory to the diffraction of scalar waves propagating in media with constant index of refraction. This approach is an approximation for the electromagnetic vector field of the light and applies well to the scalar electron waves. Unfortunately, the electromagnetic field of electron lenses represents an inhomogeneous and anisotropic medium for charged-particle waves. Therefore, we encounter the problem how to modify the light-optical approach in order to apply it to the propagation of waves in general media. This modification is possible with a sufficient degree of accuracy if the index of refraction does not change appreciably over a distance of several wavelengths. In this case, we can apply the so-called *WKB approximation* of quantum mechanics which is equivalent to the much older eikonal approximation of light optics. Owing to the short wavelength of fast charged particles, the assumptions of the WKB approach are well fulfilled for the macroscopic fields of electron optical elements.

2.5.1 Propagator

We consider an electron wave $\psi_e = \psi_e(\vec{r}, \vec{r}_s)$ which originates from a point source at position x_s, y_s, z_s . This elementary wave is defined in quantum mechanics as the *propagator* $\psi_e = P(\vec{r}, \vec{r}_s)$. In the absence of an electromagnetic field ($\vec{A} = 0, \phi = U$), the propagator is a spherical wave given by

$$\psi_e(\vec{r}, \vec{r}_s) = P = \frac{1}{4\pi |\vec{r} - \vec{r}_s|} e^{ik_0 |\vec{r} - \vec{r}_s|}, \quad k_0 = \sqrt{2em_e U^*}/\hbar. \quad (2.81)$$

In the presence of a field, the propagator is the solution of the inhomogeneous time-independent Klein–Gordon equation

$$\sum_{\mu=1}^3 \left(\frac{1}{i} \frac{\partial}{\partial x_\mu} + \frac{e}{\hbar} A_\mu \right)^2 P - \frac{2em_e \phi^*}{\hbar^2} P = -\delta^{(3)}(\vec{r} - \vec{r}_s). \quad (2.82)$$

The three-dimensional delta function $\delta^{(3)}(\vec{r} - \vec{r}_s)$ represents the radiation characteristic of the point source. The electromagnetic field steadily changes the form of the propagator if we move the point of observation away from the source. In the case of ideal imaging, the outgoing spherical wave is transformed into an incident spherical wave after propagating through the fields of the electron optical elements. In electron optical instruments, the wave traverses macroscopic electromagnetic fields between the source point $\vec{r}_s$ and the point of observation $\vec{r} = (x, y, z)$. In this case, the propagator has the form

$$P = \psi_e = a(\vec{r}, \vec{r}_s) e^{iS/\hbar}, \quad a = \bar{a}. \quad (2.83)$$

By substituting this ansatz for ψ in the time-independent Klein–Gordon equation (2.81), we obtain a complex equation for the real functions S and a . The separation of the complex equation in its real part and its imaginary part gives for $\vec{r} \neq \vec{r}_s$ the two real equations

$$\begin{aligned} (\vec{\nabla} S + e\vec{A})^2 - 2m_e e \varphi^* &= \hbar^2 \frac{\Delta a}{a}, \\ \operatorname{div} \left[a^2 (\vec{\nabla} S + e\vec{A}) \right] &= \operatorname{div} [a^2 m \vec{v}] = 0. \end{aligned} \quad (2.84)$$

We can neglect the term on the right-hand side of the first equation if

$$\left| \frac{\Delta a}{a} \right| \ll \frac{2m_e e \varphi^*}{\hbar^2}. \quad (2.85)$$

This condition is fulfilled in macroscopic electromagnetic fields except in the caustic regime where $\Delta a/a \rightarrow \infty$. If we exclude this region, we readily obtain the standard eikonal equation (2.21) with the solution

$$S = S(\vec{r}, \vec{r}_0) = \int_{\vec{r}_s}^{\vec{r}} (m\vec{v} - e\vec{A}) d\vec{r} = \int_{\vec{r}}^{\vec{r}_s} (m\vec{v} + e\vec{A}) d\vec{r}. \quad (2.86)$$

The integration must be performed along the true ray connecting the starting point $\vec{r}_s$ with the endpoint $\vec{r}$. In the absence of a magnetic field ($\vec{A} = 0$), the electron moving in opposite direction with velocity $-\vec{v}$ from the source point r_s to the point $\vec{r}$ takes the same path as that which travels from $\vec{r}$ to $\vec{r}_s$. In this case, the eikonal satisfies the relation

$$S(\vec{r}, \vec{r}_s) = S(\vec{r}_s, \vec{r}), \quad \vec{A} = 0. \quad (2.87)$$

In order that the electron travels the same path in both directions in the presence of a magnetic field ($\vec{A} \neq 0$), we must reverse the direction of the currents within the coils of the magnetic lenses ($\vec{A} \rightarrow -\vec{A}$). Only in this case is the Lorentz force the same for both directions.

The quantum-mechanical current density assigned to the elementary wave (2.83) is

$$\vec{j}_e = 2\hbar \operatorname{Im} \left(\bar{\psi}_e \vec{\nabla} \psi_e \right) + e \vec{A} \psi_e \bar{\psi}_e = a^2 (\vec{\nabla} S + e\vec{A}). \quad (2.88)$$

Accordingly, the second equation in (2.84) represents the law of conservation of current:

$$\operatorname{div} \vec{j}_e = 0, \quad \vec{j}_e = a^2 m \vec{v}. \quad (2.89)$$

We integrate this expression over the volume enclosed by the mantle formed by the outer rays of an infinitesimal pencil of rays originating from the source point $\vec{r}_s$ and by the cross sections $d\vec{\sigma}_s = -\vec{e}_r (\delta r_s)^2 d\Omega_s$ and $d\vec{\sigma} = \vec{e}_z dx dy$ of the pencil with the planes $z_s + \delta z_s$ and z , respectively, as illustrated in Fig. 2.11.

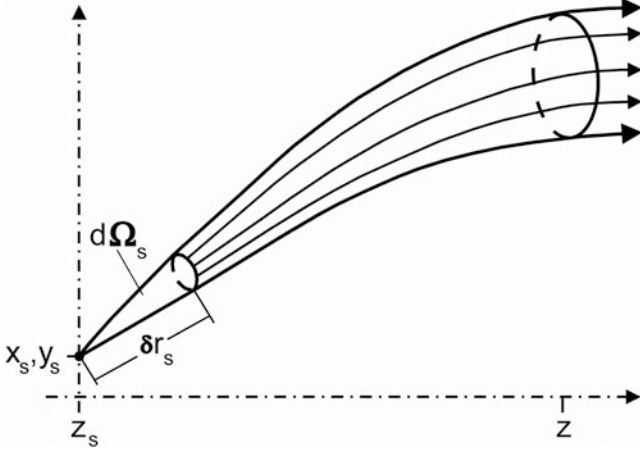


Fig. 2.11 Infinitesimal pencil of rays originating from the source point (x_s, y_s, z_s) within the differential solid angle $d\Omega_s$

Employing the Gauss theorem, we find

$$\iiint_{V_p} \text{div } \vec{j}_e dV = \oint_{\sigma_p} \vec{j}_e d\vec{\sigma} = -j_e(\delta r_s)^2 d\Omega_s + j_{ez}(\vec{r}) dx dy = 0. \quad (2.90)$$

The elementary wave (2.82) adopts the form of the spherical wave (2.81) with radius $\delta r_s = 1/4\pi a$ at the plane $z_s + \delta z_s$. At this plane, the differential solid angle $d\Omega_s$ is given by

$$d\Omega_s = \frac{dv_{xs} dv_{ys}}{v_s v_{zs}}. \quad (2.91)$$

Introducing this expression and $j_e(z_s + \delta z_s) = a^2 m_s v_s = m_s v_s / (4\pi \delta r_s)^2$ into (2.89), we obtain

$$j_{ez}(\vec{r}) = a^2 m v_z = \frac{m_s}{16\pi^2 v_{zs}} \frac{dv_{xs} dv_{ys}}{dx dy} = \frac{D_s}{16\pi^2 m_s v_{zs}}. \quad (2.92)$$

The determinant D_s is obtained as follows:

$$\begin{aligned} D_s &= \frac{dv_{xs} dv_{ys}}{dx dy} = \frac{\partial v_{xs}}{\partial x} \frac{\partial v_{ys}}{\partial y} - \frac{\partial v_{xs}}{\partial y} \frac{\partial v_{ys}}{\partial x} \\ &= \frac{1}{m_s^2} \left(\frac{\partial p_{xs}}{\partial x} \frac{\partial p_{ys}}{\partial y} - \frac{\partial p_{xs}}{\partial y} \frac{\partial p_{ys}}{\partial x} \right) = \frac{1}{m_s^2} \left(\frac{\partial^2 S}{\partial x \partial x_s} \frac{\partial^2 S}{\partial y \partial y_s} - \frac{\partial^2 S}{\partial x \partial y_s} \frac{\partial^2 S}{\partial x_s \partial y} \right). \end{aligned} \quad (2.93)$$

We have obtained the expressions of the second line by considering that (a) the vector potential in the expression $\vec{v}(\vec{r}_s, \vec{r}) = \vec{v}_s = [\vec{p}_s(\vec{r}_s, \vec{r}) + e\vec{A}(\vec{r}_s)]/m_s$ is solely

a function of the coordinates x_s, y_s, z_s of the point source and (b) that

$$p_{xs} = -\frac{\partial S(\vec{r}, \vec{r}_s)}{\partial x_s}, \quad p_{ys} = -\frac{\partial S(\vec{r}, \vec{r}_s)}{\partial y_s}. \quad (2.94)$$

The relation (2.92) yields for the amplitude of the elementary wave the expression

$$a = \frac{1}{4\pi} \sqrt{\frac{D_s}{mv_z m_s v_{zs}}}. \quad (2.95)$$

Introducing this result for a into (2.81), we obtain the semiclassical approximation of the propagator in the form

$$P(\vec{r}, \vec{r}_s) = \frac{1}{4\pi} \sqrt{\frac{D_s}{mv_z m_s v_{zs}}} e^{iS/\hbar}, \quad S = S(\vec{r}, \vec{r}_s). \quad (2.96)$$

In order that the propagator (2.96) satisfies the reciprocity relation

$$P(\vec{r}, \vec{r}_s, \vec{A}) = P(\vec{r}_s, \vec{r}, -\vec{A}). \quad (2.97)$$

Note that the expression (2.96) for the propagator (2.97) does not depend explicitly on the magnetic vector potential $\vec{A}$, yet implicitly because it is contained in the integrand of the eikonal (2.86).

The determinant (2.93) is a measure for the trajectory density. Hence, the curves with constant D_s represent curves of constant brightness. The trajectory density diverges ($D_s \rightarrow \infty$) if the point of observation is located on the caustic. In this case, all rays starting within a distinct differential solid angle from the common point source intersect each other in a conjugate point. The loci of all these points form the caustic. Strong interference effects show up in the vicinity of the caustic where an infinite number of adjacent trajectories of an initially homocentric bundle intersect each other.

2.5.2 Wave Propagation in the Infinite Half-Space

We aim to determine the propagation of a monochromatic wave within the half-space $z > z_0$ assuming that the wave is known at the arbitrary initial plane z_0 . This plane separates the entire space in a right half-space $z > z_0$ and a left half-space $z < z_0$. Because the direction of propagation points into the right half-space, we can assume any electromagnetic field for the left half-space. For our purpose, only the stationary field in the right half-space is of importance. In order to obtain the wave within this half-space, we have to solve the boundary-value problem for the

infinite half-space. We obtain the solution most conveniently by utilizing the *Green theorem*. The Green function is a special solution of the equation

$$\sum_{\mu=1}^3 \left(\frac{1}{i} \frac{\partial}{\partial x_{\mu}} - \frac{e}{\hbar} A_{\mu} \right)^2 G - \frac{2em_e \phi^*}{\hbar^2} G = -\delta^{(3)}(\vec{r} - \vec{r}_s), \quad (2.98)$$

which is the conjugate complex of Eq. (2.82). This procedure is equivalent to substituting $-\vec{A}$ for $\vec{A}$ in (2.82). The Green function $G = G(\vec{r}, \vec{r}_s)$ has a singularity at the point $\vec{r} = \vec{r}_s$ within the half-space $z \geq z_0$, and G vanishes at infinity and at the plane $z = z_0$:

$$G(\vec{r} = \vec{\infty}, \vec{r}_s) = G(x, y, z_0, \vec{r}_s) = 0. \quad (2.99)$$

The wave function $\psi = \psi(x, y, z)$ satisfies the equation

$$\sum_{\mu=1}^3 \left(\frac{1}{i} \frac{\partial}{\partial x_{\mu}} + \frac{e}{\hbar} A_{\mu} \right)^2 \psi - \frac{2em_e \phi^*}{\hbar^2} \psi = 0. \quad (2.100)$$

We multiply this equation by G and (2.98) by ψ . Subtracting the resulting expressions eventually gives

$$\text{div} \left(\psi \vec{\nabla} G - G \vec{\nabla} \psi + \frac{e}{\hbar} \vec{A} G \psi \right) = -\psi \delta^{(3)}(\vec{r} - \vec{r}_s). \quad (2.101)$$

We integrate this expression over the infinite volume of the right half-space. Applying the Gauss theorem, we obtain

$$\oint_{\sigma_{\infty}} \left(\psi \vec{\nabla} G - G \vec{\nabla} \psi + \frac{e}{\hbar} \vec{A} G \psi \right) d\vec{\sigma} = -\psi(\vec{r}_s). \quad (2.102)$$

The surface element $d\vec{\sigma}$ points in the direction of the outward normal. The surface σ_{∞} of integration consists of the plane $z = z_0$ and a hemisphere in that half-space, of infinitely large radius, centered at the origin ($x_0 = 0, y_0 = 0, z_0$). The Green function $G = G(\vec{r}, \vec{r}_s)$ is zero on this boundary. Moreover, we assume that $\psi = \psi(x, y, z)$ vanishes on the infinite hemisphere too. Because there are no contributions to the integral in (2.102) from the infinitely large hemisphere, we obtain the Sommerfeld diffraction formula

$$\psi(x_s, y_s, z_s) = \iint_{\infty} \psi(x_0, y_0, z_0) \frac{\partial G(\vec{r}_0, \vec{r}_s)}{\partial z_0} dx_0 dy_0. \quad (2.103)$$

For deriving this result, we have considered that the differentiation along the inward normal z_0 introduces a minus sign. By choosing the source point $\vec{r}_s$ of the Green function as the point of observation, $\vec{r}_s = \vec{r}$ we obtain the result

$$\psi(\vec{r}) = \iint_{\infty} \psi(x_0, y_0, z_0) \frac{\partial G(\vec{r}_0, \vec{r})}{\partial z_0} dx_0 dy_0, \quad z \geq z_0. \quad (2.104)$$

To evaluate this integral, we must first determine the Green function $G = G(\vec{r}_0, \vec{r}_s) = G(\vec{r}_0, \vec{r})$. In order to construct the Green function, we utilize the method of mirror charges employed in electrostatics for solving boundary problems. For our task, it suffices to introduce a mirror source located at position $x_s, y_s, 2z_0 - z_s$ in the left half-space, as shown in Fig. 2.12. In addition, we assume that the electromagnetic field in this regime is mirror symmetric with respect to the plane $z = z_0$ to that in the right half-space. This assumption is permissible without loss of generality because we do not consider the propagation of the electron wave in the region $z < z_0$.

We utilize the mirror approach and consider that we have derived (2.104) by introducing $-\vec{A}$ for $\vec{A}$ in (2.98). With these assumptions we obtain

$$G(\vec{r}, \vec{r}_s) = P(\vec{r}, \vec{r}_s, -\vec{A}) - P(\vec{r}, \vec{r}_{ms}, -\vec{A}). \quad (2.105)$$

By introducing the relation (2.97), we find

$$G(\vec{r}, \vec{r}_s) = P(\vec{r}_s, \vec{r}, \vec{A}) - P(\vec{r}_{ms}, \vec{r}, \vec{A}). \quad (2.106)$$

The expression on the right-hand side consists of elementary waves, one propagating from the point $(x, y, z > z_0)$ to the point (x_s, y_s, z_s) , the other from (x, y, z) to the mirror point $(x_{ms} = x_s, y_{ms} = y_s, z_{ms} = 2z_0 - z_s)$ within the actual electromagnetic field with potentials $\varphi = \varphi(x, y, z)$ and $\vec{A} = \vec{A}(x, y, z)$.

According to the reciprocity theorem, the reversal of the direction of flight together with the substitution ($\vec{A} \rightarrow -\vec{A}$) is equivalent with a permutation of the source point with the point of observation ($\vec{r} \leftrightarrow \vec{r}_s$). Therefore, we rewrite (2.106) in the form

$$G(\vec{r}_s, \vec{r}) = P(\vec{r}, \vec{r}_s, \vec{A}) - P(\vec{r}, \vec{r}_{ms}, \vec{A}). \quad (2.107)$$

It readily follows from (2.106) that the Green function vanishes at the symmetry plane $z = z_0$, whereas

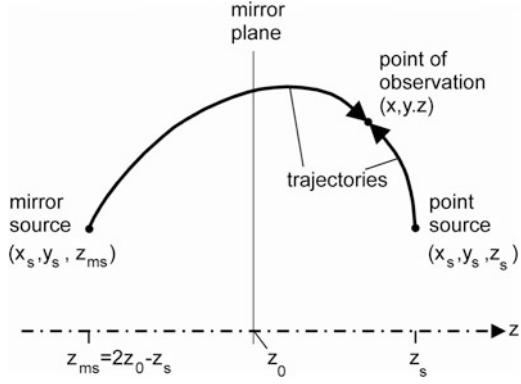
$$\left. \frac{\partial P(\vec{r}, \vec{r}_s)}{\partial z_s} \right|_{\vec{r}_s = \vec{r}_0} = - \left. \frac{\partial P(\vec{r}, \vec{r}_{ms})}{\partial z_s} \right|_{\vec{r}_{ms} = \vec{r}_0}. \quad (2.108)$$

If we make use of (2.107) and (2.108) in (2.109), we obtain the formula

$$\psi(\vec{r}) = 2 \iint_{\infty} \psi(x_0, y_0, z_0) \frac{\partial P(\vec{r}, \vec{r}_0)}{\partial z_0} dx_0 dy_0, \quad z > z_0. \quad (2.109)$$

This relation is valid for arbitrary macroscopic fields provided that $v_z > 0$. It can be considered as the generalization of the Rayleigh–Sommerfeld diffraction integral

Fig. 2.12 Scheme illustrating the construction of the Green function for the infinite half-space $z > z_0$ by means of the mirror method



which presupposes that the right half-space is field-free. We readily obtain the field-free propagator (2.81) by putting in (2.96)

$$\begin{aligned} S(\vec{r}, \vec{r}_s) &= mvR, \\ R &= \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}, \\ m_s &= m, v_{sz} = v_z = v(z - z_s)/R. \end{aligned} \quad (2.110)$$

In order to prove Huyghens' principle, we assume a field-free space and a point source with origin $x_s, y_s, z_s < z_0$. The point of observation is located at $x, y, z > z_0$. We aim to derive the wave function (2.81) at the point of observation by optical continuation going from the starting plane z_s to the plane z_0 and from there to the final plane z . By employing the formula (2.109) for the field-free space, we find the relation

$$\begin{aligned} \psi_e(\vec{r}) &= \frac{1}{8\pi^2} \iint_{\infty} \frac{e^{ik_0 R_{0s}}}{R_{0s}} \frac{\partial}{\partial z_0} \left(\frac{e^{ik_0 R_0}}{R_0} \right) dx_0 dy_0, \\ R_{0s} &= \sqrt{(z_0 - z_s)^2 + (\vec{\rho}_0 - \vec{\rho}_s)^2}, \\ R_0 &= \sqrt{(z - z_0)^2 + (\vec{\rho} - \vec{\rho}_0)^2}. \end{aligned} \quad (2.111)$$

We evaluate analytically the double integral by means of the Bessel integral representation of the spherical wave [47]

$$\frac{e^{ik_0 R_0}}{R_0} = \int_0^{\infty} J_0(|\vec{\rho} - \vec{\rho}_0|u) \frac{e^{-(z-z_0)\sqrt{u^2 - k_0^2}}}{\sqrt{u^2 - k_0^2}} u du, \quad (2.112)$$

where $J_0(|\vec{\rho} - \vec{\rho}_0|u)$ is the *Bessel function* of order zero. Differentiation of (2.101) with respect to z_0 gives

$$\frac{\partial}{\partial z_0} \left(\frac{e^{ik_0 R_0}}{R_0} \right) = \int_0^\infty J_0(|\vec{\rho} - \vec{\rho}_0|u) e^{-(z-z_0)\sqrt{u^2-k_0^2}} u \, du. \quad (2.113)$$

We introduce (2.112) for the first factor in the integrand of (2.111) and (2.113) for the second factor, giving

$$\begin{aligned} \psi_e(\vec{r}) = & \frac{1}{8\pi^2} \iint_{-\infty}^{\infty} \iint_{-\infty}^{\infty} J_0(|\vec{\rho} - \vec{\rho}_0|u) J_0(|\vec{\rho}_0 - \vec{\rho}_s|t) \frac{e^{(z_0-z)\sqrt{u^2-k_0^2}}}{\sqrt{t^2-k_0^2}} \\ & e^{-(z_0-z_s)\sqrt{u^2-k_0^2}} dx_0 dy_0 u \, dt \, dt. \end{aligned} \quad (2.114)$$

The integrations with respect to the variables t and u have to be taken from 0 to ∞ , whereas the integrations with respect to the variables x_0 and y_0 extend from $-\infty$ to ∞ . We evaluate the latter integrations by introducing polar coordinates ($x_0 - x_s = \tilde{\rho} \cos \phi$, $y_0 - y_s = \tilde{\rho} \sin \phi$), giving

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J_0(|\vec{\rho} - \vec{\rho}_0|u) J_0(|\vec{\rho}_0 - \vec{\rho}_s|t) dx_0 dy_0 &= \int_0^{2\pi} \int_0^\infty J_0(|\vec{\rho} - \vec{\rho}_s + \vec{\tilde{\rho}}|u) J_0(\tilde{\rho}t) \tilde{\rho} \, d\tilde{\rho} \, d\phi \\ &= 2\pi J_0(|\vec{\rho} - \vec{\rho}_s|u) \int_0^\infty J_0(\tilde{\rho}u) J_0(\tilde{\rho}t) \tilde{\rho} \, d\tilde{\rho} = 4\pi J_0(|\vec{\rho} - \vec{\rho}_s|u) \delta(u^2 - t^2). \end{aligned} \quad (2.115)$$

The last expression has been derived by utilizing the Bessel function representation of the delta function, which has the property

$$2 \int_0^\infty F(t^2) \delta(u^2 - t^2) t \, dt = F(u^2). \quad (2.116)$$

If we insert (2.115) into (2.114) and consider the property (2.116) of the delta function, we find that

$$\psi_e(\vec{r}) = \frac{1}{4\pi} \int_0^\infty J_0(|\vec{\rho} - \vec{\rho}_s|u) \frac{e^{-(z-z_s)\sqrt{u^2-k_0^2}}}{\sqrt{u^2-k_0^2}} u \, du = \frac{e^{ik_0 R}}{4\pi R}, \quad R = |\vec{r} - \vec{r}_s|. \quad (2.117)$$

This trivial result represents the field-free propagator (2.81), therewith demonstrating the validity of Huyghens' principle. To our knowledge, this proof has not yet been outlined in the literature. Owing to the spherical symmetry, the expression

(2.117) is valid for any location of the point of observation although the Sommerfeld formula (2.111) supposes that this point is located in the right half space.

2.5.3 Fundamentals of Diffraction and Image Formation

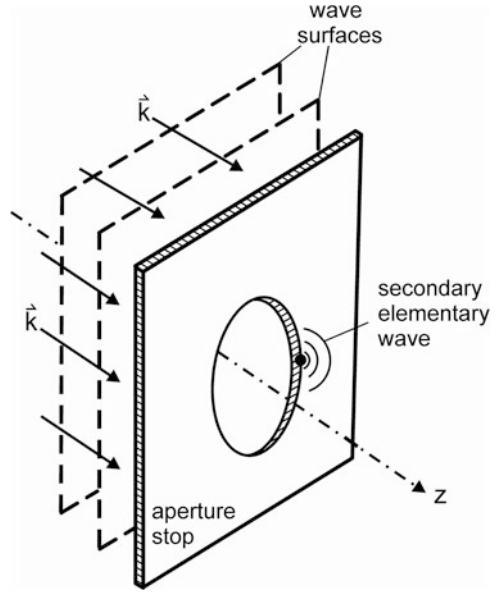
The semiclassical propagator (2.96) approximates the real propagator with a sufficient degree of accuracy as long as the source point $\vec{r}_s$ is connected with the point of observation $\vec{r}$ by a single trajectory. This is always the case for the field-free space. However, in the presence of an electromagnetic field, it may occur that two or more trajectories of an initially homocentric bundle of rays intersect each other at one or more points. If such a point is connected with the point source by n trajectories, the propagator consists of n elementary waves, each of which is defined by one of the classical paths connecting the point of observation with the point source:

$$P(\vec{r}, \vec{r}_s) = \frac{1}{4\pi \sqrt{m v_z m_s v_{zs}}} \sum_{v=1}^n \sqrt{D_{sv}} e^{i S_v(\vec{r}, \vec{r}_s)/\hbar}. \quad (2.118)$$

Since the connecting trajectories have different directions at their point of intersection, the associated waves propagate in different directions too. As a result, the superposition of these elementary waves produces pronounced interference phenomena near the point of intersection as observed in electron holography and phase-contrast electron microscopy. If the point of observation is located on the caustic, the determinant tends to infinity because an infinite number of trajectories, starting within a given differential pencil of rays, intersect each other in a common conjugate point. The loci of all these points form the caustic. We can avoid the divergence of the semiclassical approach in the caustic regime by subdividing the space between the caustic and the source into two half-spaces, one containing the source and the other the caustic. We choose the diffraction plane behind the objective lens as separating plane where a beam-limiting aperture is placed. Rays originating from the same object point do not intersect each other at this plane because it is located in front of the caustic. In the case of parallel illumination, the aperture plane z_a coincides with the back focal plane of the objective lens. This plane separates the source space from the space of observation which embeds the caustic. We describe the effect of the aperture on the wave by the aperture function $A(x_a, y_a)$. This function is unity in the region within the hole of the diaphragm and zero outside. Moreover, we suppose that the effect of the diaphragm on the transmitted wave is negligibly small at the aperture plane. However, the diffraction of the electron wave at the boundary of the aperture stop produces interference effects in the region behind the aperture. This behavior is demonstrated by employing Huyghens' principle, as shown schematically in Fig. 2.13.

In the following, we are concerned with the propagation of electron waves in macroscopic electromagnetic fields, which focus the electrons yet do not reflect any of them back to the initial plane. Because Huyghens' principle is also valid for the

Fig. 2.13 Diffraction of a plane incident wave at the annular boundary of an aperture stop causing interference effects in the regime behind the beam-limiting diaphragm



propagation of electron waves in macroscopic fields, we can subdivide the space between the source plane and the plane of observation in slices and continue from one slice to the other by reapplying the continuation formula (2.98). For describing the image formation in an electron microscope, it suffices to subdivide the space between the object plane z_o and the image plane z_i in two slices separated by the aperture plane z_a where the diffraction pattern of the object is formed. In the case of parallel illumination, the diffraction plane coincides with the back focal plane of the objective lens. One places generally the object plane at the back surfaces of the object which faces the objective lens. This plane defines the exit plane z_e of the incident electron wave having transmitted the object. In accordance with Abbe's theory of image formation in the light microscope, the aperture plane separates the image space $z > z_a$ from the object space $z < z_a$. We consider the propagation of the elementary wave, which originates from the point z_o, y_o in the object plane z_o , to the point x, y at the observation plane $z > z_a$, regardless if whether this point is located on the caustic or not. In order to avoid divergences, we propagate successively the elementary wave from the object plane z_o to the diffraction plane z_a and from there to the recording plane z . Thus, the propagator connecting the initial point $\vec{r}_o$ with the point of observation $\vec{r}$ in the image space is given by

$$P(\vec{r}, \vec{r}_o) = 2 \iint_{\infty} A(x_a, y_a) P(\vec{r}_a, \vec{r}_o) \frac{\partial P(\vec{r}, \vec{r}_a)}{\partial z_a} dx_a dy_a. \quad (2.119)$$

The square of the absolute value of the propagator $P(\vec{r}_i, \vec{r}_o)$ at the Gaussian image plane $z = z_i$ forms the so-called point spread function. The propagator (2.119) can

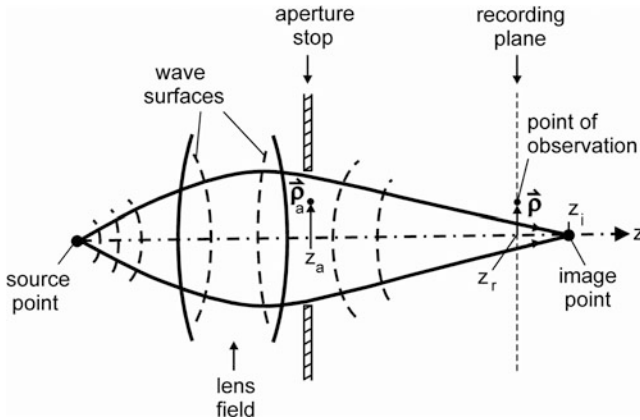


Fig. 2.14 Illustration of Abbe's theory of image formation in the microscope

be considered as the basis of Abbe's theory of image formation, as illustrated in Fig 2.14.

We obtain the wave function $\psi(\vec{r})$ in the image space ($z > z_a$) by introducing (2.119) for $P(\vec{r}, \vec{r}_0)$ into (2.109), giving

$$\psi(\vec{r}) = 4 \iint_{\infty} A(x_a, y_a) \frac{\partial P(\vec{r}, \vec{r}_a)}{\partial z_a} \left[\iint_{\infty} \psi(\vec{r}_o) \frac{\partial P(\vec{r}_a, \vec{r}_o)}{\partial z_o} dx_o dy_o \right] dx_a dy_a. \quad (2.120)$$

This formula is valid for arbitrary optical systems with straight axis except mirrors. The expression illustrates Abbe's reasoning. He assumes that the object is illuminated by a plane wave incident normally on the object. The wave is diffracted by the object and gives rise to a modified object wave $\psi(\vec{r}_o)$ at the exit plane $z_e = z_o$. Subsequently, the objective lens forms in the back focal plane a so-called Fraunhofer diffraction pattern of the object wave. Every point in this plane may be considered to be the center of a coherent secondary elementary wave, whose strength is proportional to the amplitude of the diffracted wave at this point. The coherent waves that proceed from the secondary sources will then interfere with each other forming the image of the object at the conjugate image plane. Owing to the imperfections of the lenses and the diffraction at the aperture, this image will not be sharp and affected by artifacts. The main task of electron optics is the reduction of these effects by correcting the aberrations of the electron lenses. In the ideal case, each spherical wave originating from any point at the object plane is transformed into an incident spherical wave whose origin is located in the image plane. The origin forms the image point which is conjugate to the corresponding object point. Real systems produce a spot instead of a point. The size of the radius of the spot determines the resolution.

2.5.4 Point Spread Function

The amplitude of the propagator (2.83) is only slightly affected by the aberrations of the objective lens. This does not hold true for the phase $S/\hbar$ because $\hbar$ is very small. By employing this standard approximation and a thin lens, the propagator (2.83) in front of the lens has the form

$$P(\vec{r}, \vec{r}_o) \approx \frac{e^{iS(\vec{r}, \vec{r}_o)/\hbar}}{4\pi |\vec{r} - \vec{r}_o|}. \quad (2.121)$$

The lens transfers this incident quasi-spherical wave into a converging distorted wave whose origin is located at the conjugate image point r_i . The object plane is located at the distance $g = z_i - z_o$ in front of the thin lens placed at the plane $z = z_l$, as depicted schematically in Fig. 2.15. The Gaussian image is formed at the plane z_i whose distance from the lens $b = z_i - z_l = Mg$ is generally very large compared with the object distance g ; M is the magnification of the image. We assume that the maximum aperture angle $\theta_0 \approx \rho_{l0}/g$ is limited by a circular aperture placed at the back focal plane of the lens.

Without loss of generality, we can replace in (2.119) the integration over the aperture plane by the integration over the midplane of the lens if we confine the total lens action to this plane. By introducing polar coordinates

$$x_l = g\theta \cos \phi, \quad y_l = g\theta \sin \phi, \quad (2.122)$$

we obtain

$$\begin{aligned} P(\vec{r}_i, \vec{r}_o) &= 2 \iint_{\infty} A(x_l, y_l) P(\vec{r}_l, \vec{r}_o) \frac{\partial P(\vec{r}_i, \vec{r}_l)}{\partial z_l} dx_l dy_l \\ &\approx \frac{1}{4\pi i \lambda} \frac{g}{b} \int_0^{\theta_0} \int_0^{2\pi} e^{i[S(\vec{r}_l, \vec{r}_o) + S(\vec{r}_i, \vec{r}_l)]/\hbar} \theta d\theta d\phi. \end{aligned} \quad (2.123)$$

In the absence of lenses located in the region between the aperture and the image plane, the total phase of the product of the two propagators has the form

$$(S(\vec{r}_l, \vec{r}_o) + S(\vec{r}_i, \vec{r}_l)) / \hbar = k(z_i - z_o) + k |\vec{\rho}_i / M - \vec{\rho}_o| \theta \cos(\phi - \phi_0) - \gamma_a(\vec{\theta}, \vec{\rho}_o). \quad (2.124)$$

The phase shift $\gamma_a(\vec{\theta}, \vec{\rho}_o)$ accounts for the aberrations of the objective lens. *Isoplanatic* imaging conditions prevail in the case $\gamma_a(\vec{\theta}, \vec{\rho}_o) = \gamma_a(\vec{\theta})$. The aberrations deform the spherical shape of ideal wave surfaces preventing stigmatic imaging, as illustrated in Fig. 2.16.

Fig. 2.15 Geometrical parameters defining the location of the object, image, and aperture plane with respect to the center z_l of the image-forming lens

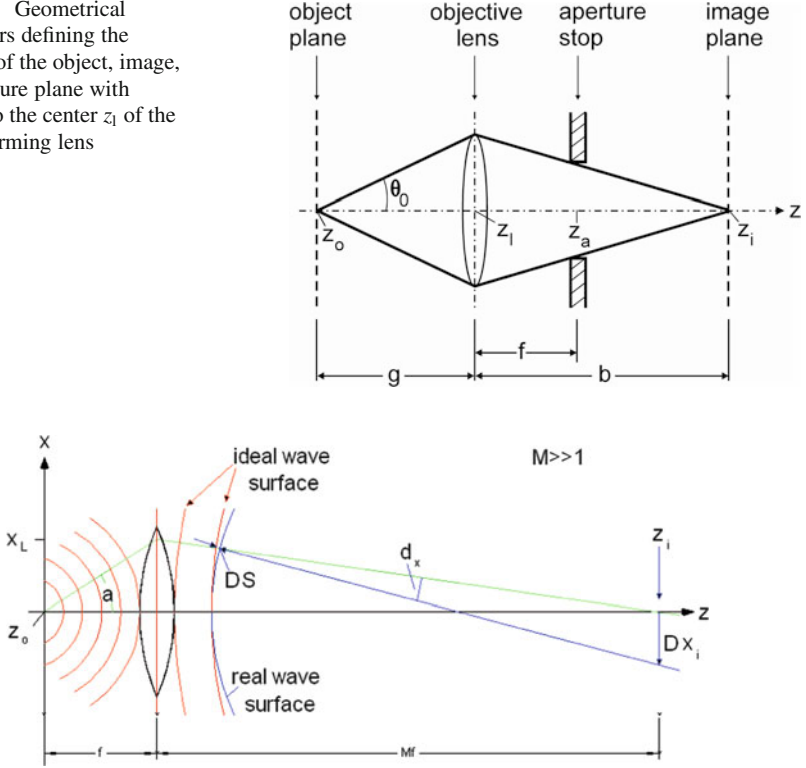


Fig. 2.16 Illustration of the ray path and the associated wave surfaces of electrons propagating through the field of an electron lens. The rays emanating from the center of the object plane z_o are orthogonal to the surfaces of constant phase of the associated outgoing spherical wave. The electron lens transforms this wave in a converging wave which is nonspherical in the presence of aberrations. The deviation DS of the wave surface from its ideal spherical shape induces the deviation Dx_i of the ray from its ideal position at the image plane z_i

In the case of an ideal lens ($\gamma_a = 0$), we can perform analytically the integrations over θ and ϕ in the expression (2.123), giving

$$\begin{aligned}
 P(\vec{r}_i, \vec{r}_o) &\approx \frac{e^{ik(z_i - z_o)}}{4\pi i \lambda M} \int_0^{\theta_0} \int_0^{2\pi} e^{ik|\vec{\rho}_i/M - \vec{\rho}_o| \theta \cos(\phi - \phi_0)} \theta d\theta d\phi \\
 &= \frac{e^{ik(z_i - z_o)}}{2i \lambda M} \int_0^{\theta_0} J_0(kR\theta) \theta d\theta = \frac{\theta_0}{4\pi i M} e^{ik(z_i - z_o)} \frac{J_1(kR\theta_0)}{R}, \\
 R &= |\vec{\rho}_i/M - \vec{\rho}_o|.
 \end{aligned} \tag{2.125}$$

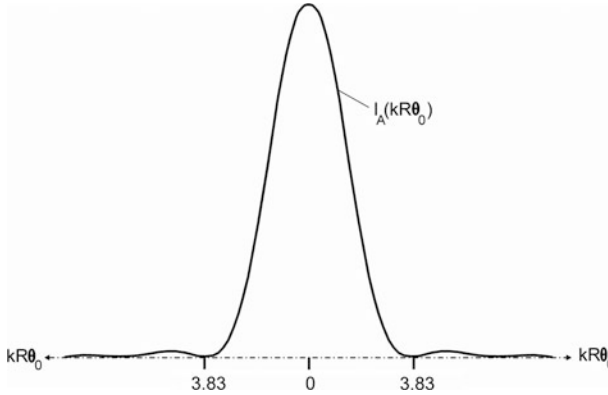


Fig. 2.17 Radial intensity distribution of the point spread function formed by an ideal lens (Airy pattern)

Here, $J_0(x)$ and $J_1(x)$ denote the Bessel functions of order zero and order one, respectively.

The squared absolute value of the point spread function (2.125) represents the Fraunhofer diffraction pattern of the circular aperture (Airy pattern) whose intensity distribution is given by

$$I_A(kR\theta_0) = |P(\vec{r}_i, \vec{r}_o)|^2 = \frac{\theta_0^4}{4\lambda^2 M^2} \left(\frac{J_1(kR\theta_0)}{kR\theta_0} \right)^2. \quad (2.126)$$

The intensity (2.126) of the *Airy pattern* consists of the central disc surrounded by concentric dark and bright rings with rapidly diminishing intensity, as illustrated in Fig. 2.17. The central intensity extends to the first zero at $kR\theta_0 = 3.83$ and comprises 83 % of the total intensity.

We define the resolution limit d of self-luminous object points by means of the *Rayleigh criterion* according to which two adjacent image points are resolved if the maximum of the Airy pattern of one image point is centered at the first zero of the Airy pattern of the other point, giving

$$d = \frac{3.83}{k\theta_0} = 0.61 \frac{\lambda}{\theta_0}. \quad (2.127)$$

This expression is the small-angle approximation of the Abbe formula for the diffraction-limited resolution of the light microscope.

Chapter 3

Multipole Expansion of the Stationary Electromagnetic Field

Abstract Multipole expansions of the electrostatic potential and the stationary magnetic vector potential about the arbitrarily curved optic axis are derived in complex notation. Utilizing analytical solutions of the Laplace equations, simple electron optical elements with straight optic axis are discussed. The charge simulation method is introduced for precisely calculating the electromagnetic potentials of systems with given geometry of the electrodes and pole pieces, respectively.

Constant currents form *stationary* magnetic fields. In the *static* limit, the field-producing charges are at rest, and the currents are zero. In this case, the magnetic field vanishes. Therefore, within the frame of our definition, static fields are purely electrostatic. We rarely encounter time-dependent fields in charged-particle optics because in most cases, the reciprocal transition time of the particle through the system is significantly smaller than the maximum frequencies of the fields. Hence, we can consider these fields as stationary with a sufficient degree of accuracy. In most cases, charged-particle optics is concerned with the propagation of a confined ensemble of particles through a system. Examples are the electron microscope, accelerators, spectrometers, and beam-guiding systems. For these systems, it is advantageous to choose the central trajectory as the z -axis of an orthogonal coordinate system, as schematically illustrated in Fig. 3.1. In order that we can develop the curved sections into a plane, the torsion of the curved axis must be zero. In this case, all sections, which contain the centers of curvature of the optic axis, are *plane sections*.

Charged particles must propagate in vacuo. The beam-guiding electromagnetic fields are formed by the voltages applied to the electrodes and the currents within the coils of the magnets. The spatial distribution of the electric and magnetic potentials is determined by their boundary values on the surfaces of the electrodes and pole pieces, respectively. The task of electron optics is an inverse problem because we must determine the geometry of the electrodes and pole pieces, which will provide the required imaging or propagation. Unfortunately, we cannot directly solve this delicate problem.

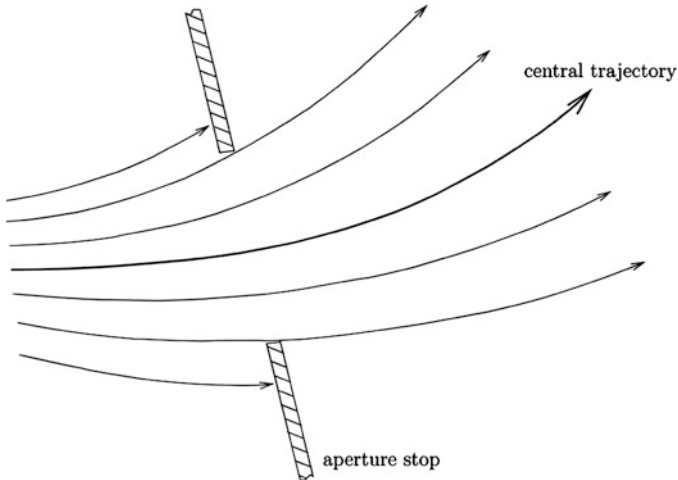


Fig. 3.1 Realization of a confined bundle of trajectories by an aperture stop

In systems with a straight axis of symmetry, this axis coincides with the central trajectory and forms the *optic axis* of the system. Examples are systems with rotational symmetry or with at least two sections of symmetry about a common axis. For systems with a curved axis, such as spectrometers, beam separators, and storage rings, it is advantageous to choose the trajectory formed by the central particle with nominal energy as the proper optic axis, which usually represents the z -axis of the curved coordinate system. In order to attribute the optimum optic axis to a system without a well-defined axis of symmetry, we assume at the outset that we can define the z -axis arbitrarily.

The off-axis position of the particle is defined by its coordinates x and y . We consider these coordinates as *dependent* variables of the z -coordinate, which we choose as the *independent* variable. However, this choice is only appropriate as long as the ray gradients are sufficiently small. This condition is fulfilled as long as the kinetic energy of the particles is large compared to the energy spread of the beam. Hence, for mirrors and electron sources, we must retain the time as the independent variable in order to avoid divergences. Choosing the z -axis as independent variable has the advantage of defining a trajectory by the coordinates of its intersections through special planes.

3.1 Scalar Potentials

In order that the charged particles of the beam do not interact with other particles, they must propagate in high vacuum. Therefore, it is necessary to place the coils of the magnets and the electrodes outside of this region. Since we consider only beams

with low current densities, we can neglect the effect space charge resulting from the particles of the beam. Hence, we assume that only the *external* currents and charges produce the electromagnetic field within which the charged particles propagate. In the region of the beam, we have

$$\vec{j} = 0, \quad \rho_e = 0, \quad \mu = \mu_0, \quad \varepsilon = \varepsilon_0, \quad (3.1)$$

where $\vec{j}$ and ρ_e are the internal current density and the internal charge density, respectively. Considering further the stationary condition $\partial/\partial t = 0$, the *Maxwell equation* adopts the simple form

$$\vec{\nabla} \times \vec{E} = 0, \quad \vec{\nabla} \times \vec{B} = 0, \quad \vec{\nabla} \cdot \vec{E} = \varepsilon_0 \vec{\nabla} \cdot \vec{E} = 0, \quad \vec{\nabla} \cdot \vec{B} = 0. \quad (3.2)$$

Considering the relation $\vec{\nabla} \times \vec{\nabla} = 0$, we satisfy the first two equations by expressing the electric field strength and the magnetic field strength each as the gradient of a scalar potential:

$$\vec{E} = -\vec{\nabla}\varphi, \quad \vec{B} = -\vec{\nabla}\psi. \quad (3.3)$$

Both the electric potential φ and the *scalar magnetic potential* ψ satisfy the *Laplace equation*

$$\vec{\nabla}^2 \varphi = \Delta \varphi = 0, \quad \vec{\nabla}^2 \psi = \Delta \psi = 0. \quad (3.4)$$

We readily verify the validity of these equations by substituting the expressions (3.3) for $\vec{E}$ and $\vec{B}$ in the third and fourth equation of (3.2). The values of the potentials on the boundary surfaces define the solutions of these equations. In the case of high electric conductivity of the electrodes and high permeability ($\mu \rightarrow \infty$) of the pole pieces of the magnets, the surfaces of these elements are also surfaces of constant electric and/or scalar magnetic potential. As a result, the spatial distributions of the electric and the magnetic fields are the same if the geometry of the electrodes and the magnets coincide. This behavior facilitates considerably the determination of the magnetic field because we can reduce the problem to an equivalent electrostatic boundary-value problem. In order to avoid saturation effects, the curvature of the pole pieces facing the beam must be sufficiently small.

3.1.1 Complex Variables

For mathematical simplicity and for obtaining a good physical comprehension of the formulas describing the properties of electron optical systems, it is advantageous to combine the x - and the y -components of any vector to a single complex quantity. The standard notation for specific two-dimensional vectors is

$$w = x + iy = \rho e^{i\theta}, \quad p = p_x + ip_y, \quad A = A_x + iA_y. \quad (3.5)$$

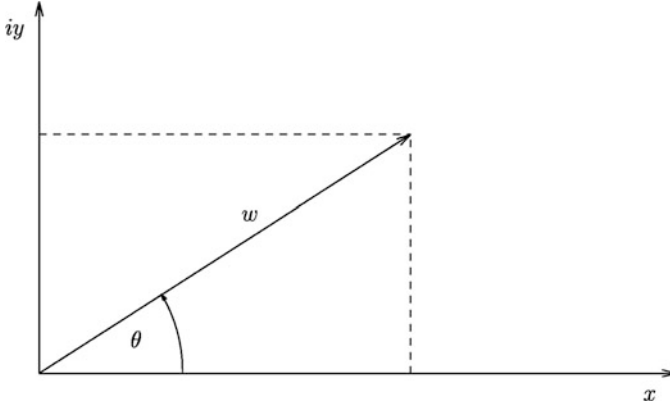


Fig. 3.2 Representation of the complex position vector $w = x + iy = \rho e^{i\theta}$, which defines the lateral distance of a particle from the optic axis

We indicate the corresponding conjugate complex quantities by a bar, for example, $\bar{w} = x - iy = \rho e^{-i\theta}$. The polar coordinates $\rho = \rho(z) = |w|$, $\theta = \theta(z) = \arctan(y/x)$ define the distance and the azimuthal position of the particle at a given plane, as depicted in Fig. 3.2. We further introduce the *complex curvature* of the optic axis

$$\Gamma = \Gamma(z) = (\kappa_x + i\kappa_y) = |\kappa| e^{i\vartheta}. \quad (3.6)$$

The twist angle ϑ relates with the torsion τ via

$$\vartheta = \vartheta(z) = \int_{-\infty}^z \tau_a(z') dz' = \int_{-\infty}^z \tau(z') dz' - \vartheta_L. \quad (3.7)$$

We define the torsion τ and the angle ϑ_L by requiring that the optic axis is representing a particle trajectory. In this case, τ represents the torsion of the accompanying triad and ϑ_L the *angle of Larmor rotation*. This rotation results from the longitudinal component $B_z = \vec{B} \vec{e}_t = \vec{B} \vec{e}_z$ of the magnetic field along the optic axis ($x = 0$, $y = 0$) and does not depend on the curvature of the trajectory. The *twist angle* ϑ is referred to a curved orthogonal coordinate system whose z -axis coincides with the space curve and whose lateral coordinates x and y are rotated back with respect to this angle. Accordingly, the y -axis remains fixed in space along the entire curve. Hence, the y - z plane is evolvable. We note that the torsion τ_a of the optic axis must not necessarily coincide with that of the accompanying triad of a particle trajectory.

Within the frame of complex notations, scalar and vector products of any two-dimensional vectors $\vec{a} \rightarrow a = a_x + ia_y$ and $\vec{b} \rightarrow b = b_x + ib_y$ are expressed as

$$\vec{a} \vec{b} = \text{Re}(a\bar{b}) = \text{Re}(\bar{a}b), \quad (3.8)$$

$$\vec{a} \times \vec{b} = \vec{e}_z(a_x \bar{b}_y - a_y \bar{b}_x) = \vec{e}_z \text{Im}(\bar{a}b). \quad (3.9)$$

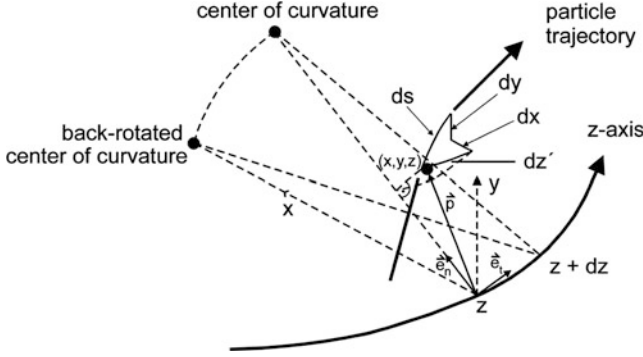


Fig. 3.3 Illustration of the differential path length element ds and its components dx , dy , and dz referred to the curved x , y , z -coordinate system. The center of curvature denotes the momentary center of curvature, which is rotated back by the twist angle in order that the y -direction remains fixed

Here, Re and Im denote the *real part* and the *imaginary part*, respectively; $\vec{e}_z$ is the unit vector in the direction of the optic axis.

By employing the complex notation, the expression (2.64) for the torsion of a trajectory adopts the simple for

$$\tau = \frac{\partial \vartheta_L}{\partial z} + \frac{\partial \vartheta}{\partial z} = \frac{1}{\rho_L} + \text{Im} \frac{d}{dz} \ln \Gamma. \quad (3.10)$$

The torsion is zero if the sum of the Larmor rotation and the imaginary part of the complex curvature vanishes. Only in this special case the moving trihedral forms an orthogonal coordinate system.

3.1.2 Laplace Equation

In the following, we consider an orthogonal curvilinear x - y - z coordinate system and choose the arc length of the reference curve as the z -coordinate. The y -axis points in the direction of the back-rotated binormal of the accompanying trihedral, as shown in Fig. 3.3. The metric coefficients g_1, g_2, g_3 are obtained most conveniently by expressing the differential arc length $d\vec{s}$ of the particle trajectory by means of its components dx, dy , and dz , which are referred to the curvilinear coordinate system. The connection of these quantities is illustrated in Fig. 3.3 and given by

$$ds^2 = g_1^2 dx^2 + g_2^2 dy^2 + g_3^2 dz^2 = dx^2 + dy^2 + dz^2. \quad (3.11)$$

The component dz' of the infinitesimal curve element $d\vec{s}$ (3.11) differs from dz due to the curvature of the z -axis:

$$g_3 dz = dz' = dz(1 - \vec{\kappa} \vec{\rho}). \quad (3.12)$$

We readily obtain the metric coefficients from the relations (3.11) and (3.12) as

$$g_1 = g_2 = 1, \quad g_3 = 1 - \bar{\kappa}\vec{\rho} = 1 - \text{Re}\{\Gamma\bar{w}\}. \quad (3.13)$$

The representation of Laplace equation in arbitrary orthogonal curvilinear coordinates is listed in textbooks on vector analysis [48]. For the metric coefficients (3.13), the equation for the electric potential φ adopts the form

$$\Delta\varphi = \frac{1}{g_3} \left[\frac{\partial}{\partial x} \left(g_3 \frac{\partial\varphi}{\partial x} \right) + \frac{\partial}{\partial y} \left(g_3 \frac{\partial\varphi}{\partial y} \right) + \frac{\partial}{\partial z} \left(\frac{1}{g_3} \frac{\partial\varphi}{\partial z} \right) \right] = 0. \quad (3.14)$$

We obtain the corresponding equation for the scalar magnetic potential by substituting ψ for φ . In order to rewrite this equation in terms of the complex off-axial coordinates w and $\bar{w}$, we express the x - and the y -component of the gradient by means of the complex coordinates w and $\bar{w}$:

$$\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} = \frac{\partial}{\partial w} \frac{\partial w}{\partial x} + \frac{\partial}{\partial \bar{w}} \frac{\partial \bar{w}}{\partial x} + i \left(\frac{\partial}{\partial w} \frac{\partial w}{\partial y} + \frac{\partial}{\partial \bar{w}} \frac{\partial \bar{w}}{\partial y} \right) = \frac{\partial}{\partial w} + \frac{\partial}{\partial \bar{w}} - \frac{\partial}{\partial w} + \frac{\partial}{\partial \bar{w}} = 2 \frac{\partial}{\partial \bar{w}}. \quad (3.15)$$

Using this relation together with its conjugate complex, we eventually derive the complex representation of Eq. (3.15) as

$$g_3 \Delta\varphi = 4\text{Re} \left[\frac{\partial}{\partial w} \left(g_3 \frac{\partial\varphi}{\partial \bar{w}} \right) \right] + \frac{\partial}{\partial z} \left(\frac{1}{g_3} \frac{\partial\varphi}{\partial z} \right) = 0. \quad (3.16)$$

In the case of a straight optic axis, the metric coefficient is $g_3 = 1$. This differential equation has been treated extensively in electrical engineering. Many analytical solutions are listed in Ollendorff [49]. Numerical methods are discussed extensively by Hawkes and Kasper [15] and by Munro [50]. Therefore, we restrict our considerations to series expansions, specific analytical solutions which are of importance for charged-particle optics, and to the semi-analytical charge-simulation method.

3.1.3 Planar Fields

Planar fields are two-dimensional fields such that the potential is independent of one of the three spatial coordinates. These fields represent special cases of three-dimensional fields obtained by neglecting the fringing fields in one direction. We can realize fields of this type approximately in the case of slit lenses and extended multipoles such that their extension along the axis of symmetry is large compared with the distance of their pole faces from this axis. In this case, we can approximate the field within the multipoles with a sufficient degree of accuracy by that of a *plane* (two-dimensional) multipole.

Without loss of generality, we can choose any of the three coordinates as the coordinate, which does not affect the potential. However, in order to stay within the convention of charged-particle optics, the z -axis always represents the optic axis, which coincides with the axis of symmetry in the case of multipoles. To describe the potential of plane multipoles, we put $g_3 = 1$, $\partial/\partial z = 0$, while for slit lenses, we have $\partial/\partial x = 0$ since in this case it is common practice to place the x -axis along the direction of the infinitely extended slits or wires. In the former case, the Laplace equation adopts the two-dimensional form

$$\Delta\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} = 4 \frac{\partial^2\varphi}{\partial w \partial \bar{w}} = 0. \quad (3.17)$$

It readily follows from the representation in complex coordinates that the general solution has the form

$$\varphi = \text{Re}F(w). \quad (3.18)$$

Here, $F(w)$ is an arbitrary analytical function of the complex variable w . For example, the potential of a plane multipole with *multiplicity* m is given by the *harmonic polynomial*

$$\varphi = \varphi_m = \text{Re}(\Phi_m \bar{w}^m) = \rho^m [\Phi_{mc} \cos m\theta + \Phi_{ms} \sin m\theta] = |\Phi_m| \rho^m \cos m(\theta - \alpha_m). \quad (3.19)$$

Each of the multipole strengths

$$\Phi_m = \Phi_{mc} + i \Phi_{ms} = |\Phi_m| e^{im\alpha_m} \quad (3.20)$$

is generally complex. The phase α_m defines the orientation of the axes of symmetry of the multipole m with respect to the x - and y -coordinates. In the terminology of accelerator physics, multipoles with azimuthal orientation $\alpha_m = 0$ ($\Phi_{ms} = 0$) are called *regular multipoles*, while those with orientation $\alpha_m = \pi/2m$ ($\Phi_{mc} = 0$) are called *skew multipoles*.

As an example, we consider the potential of a plane multipole with multiplicity $m = 2$. This multipole represents a quadrupole illustrated in Fig. 3.4. The equipotentials

$$\varphi_2 = \text{Re}\{\Phi_2 \bar{w}^2\} = |\Phi_2| \rho^2 \cos 2(\theta - \alpha_2) = \Phi_{2c}(x^2 - y^2) + 2\Phi_{2s}xy = \text{const} \quad (3.21)$$

form hyperbolas. The complex electric field strength

$$E_x + iE_y = -2 \frac{\partial \varphi_2}{\partial \bar{w}} = -2\Phi_2 \bar{w} \quad (3.22)$$

is proportional to the distance from the axis $w = 0$. In the special case $\alpha_2 = \Phi_{2s} = 0$, the electrodes are centered along the coordinate axes, and the components of the electric field strength are found from (3.22) as

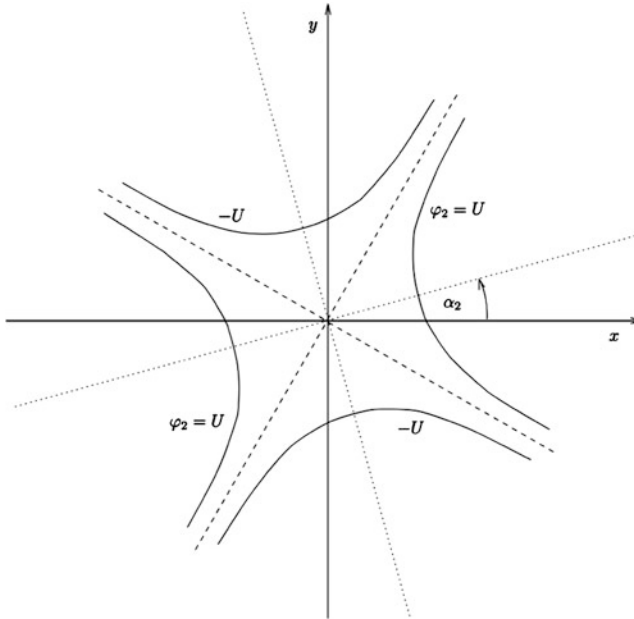


Fig. 3.4 Equipotentials of a plane quadrupole whose principal sections (*dotted lines*) are rotated by the angle α_2 with respect to the x - and y -axis, respectively

$$E_x = -2\Phi_{2c}x, \quad E_y = 2\Phi_{2c}y. \quad (3.23)$$

These relations reveal that a charged particle which propagates in one of the two symmetry sections x, z ($y = 0$) or y, z ($x = 0$) does not experience a force perpendicular to these sections. Hence, a particle, which initially propagates in the z -direction on one of these so-called *principal sections*, will remain in this section along its entire path. Since the components (3.23) of the electric field strength have opposite signs, it *focuses* the charged particles in one principal section and *defocuses* them in the other.

3.2 Systems with Straight Axis

Systems with a straight axis are formed by special arrangements of the electrodes and/or magnets such that the system possesses a symmetry axis. We choose this axis as the *optic axis* of the system because it forms a special trajectory along which the lateral forces vanish. Hence, the external fields do not deflect particles, which initially travel along this straight axis. Examples of such systems are the arrangements of round lenses in the electron microscope, of quadrupoles in linear accelerators, and of multipoles in aberration correctors. Since the curvature of the straight axis vanishes ($\Gamma = 0, g_3 = 0$), the proper coordinate system is Cartesian.

3.2.1 Multipole Expansion of the Scalar Potential

We have shown that the harmonic polynomials are special solutions of the two-dimensional Laplace equation. In this case, the multipole strengths Φ_v are constant along the optic axis. If the extension of the multipoles along this axis is finite, the multipole strengths become functions of the z -coordinate due to the inevitable fringing fields. Since the corresponding potential must satisfy Laplace's equation, higher-order terms in the off-axial square distance $\rho^2 = w\bar{w}$ also arise. It is noteworthy that these terms do not affect the multiplicity of the multipole field exhibiting a well-defined symmetry about the optic axis. Hence, it is possible to realize a “pure” multipole field of finite extension contrary to statements found in the literature.

The decomposition of the electric potential φ in a sum of multipole terms φ_v ,

$$\varphi = \sum_{v=1}^{\infty} \varphi_v, \quad (3.24)$$

corresponds to a Fourier series expansion with respect to the azimuthal angle θ about the optic axis. Owing to these considerations, the power series expansion of the component φ_v of the electric potential must have the form

$$\varphi_v = \operatorname{Re} \sum_{\lambda=0}^{\infty} a_{v\lambda}(z)(w\bar{w})^{\lambda} \bar{w}^v. \quad (3.25)$$

The coefficients $a_{v\lambda}(z)$ are generally complex, as in the planar case. The first coefficient

$$a_{v0} = a_{v0}(z) = \Phi_v(z) \quad (3.26)$$

is arbitrary and defines the *complex multipole strength*, which determines the spatial distribution of the potential near the optic axis. The z -dependence of this coefficient solely depends on the geometry of the multipole electrodes.

The other coefficients $a_{v\lambda}$ with $\lambda > 0$ are proportional to derivatives of $\Phi_v(z)$. They are obtained by inserting the series representation (3.25) for φ_v in the Laplace equation

$$4 \frac{\partial^2 \varphi_v}{\partial w \partial \bar{w}} + \frac{\partial^2 \varphi_v}{\partial z^2} = 0. \quad (3.27)$$

As a result, we find

$$\operatorname{Re} \left[4 \sum_{\lambda=1}^{\infty} a_{v\lambda}(\lambda + v) \lambda \bar{w}^v (w\bar{w})^{\lambda-1} + \sum_{\lambda=0}^{\infty} a''_{v\lambda} \bar{w}^v (w\bar{w})^{\lambda} \right] = 0, \quad (3.28)$$

where the dashes denote differentiations with respect to the z -coordinate. If we replace in the first sum the summation index λ by $\lambda + 1$ and consider the factor of each monomial must vanish owing to the linear independence of different powers

of $w\bar{w}$, we readily derive the recurrence formulas

$$4a_{v,\lambda+1}(\lambda + v + 1)(\lambda + 1) = -a''_{v\lambda}, \quad \lambda = 0, 1, 2, \dots \quad (3.29)$$

Starting with the given complex coefficient $a_{v0} = \Phi_v = \Phi_v(z)$, we obtain by means of successive insertion

$$\begin{aligned} a_{v1} &= -\frac{1}{4} \frac{1}{v+1} a''_{v0} = -\frac{1}{4} \frac{1}{v+1} \Phi''_v, \\ a_{v2} &= -\frac{1}{4} \frac{1}{2} \frac{1}{v+2} a''_{v1} = \frac{1}{4^2} \frac{1}{2!} \frac{v!}{(v+2)!} \Phi_v^{[4]}, \\ &\vdots \\ a_{v\lambda} &= (-)^{\lambda} \frac{1}{4^{\lambda} \lambda!} \frac{v!}{(\lambda + v)!} \Phi_v^{[2\lambda]}. \end{aligned} \quad (3.30)$$

Hence, each coefficient $a_{v\lambda}$, $\lambda \neq 0$ is given by the 2λ th differentiation of the complex multipole strength $\Phi_v(z) = \Phi_{vc}(z) + i\Phi_{vs}(z)$ multiplied by a specific constant factor. By substituting the last expression for $a_{v\lambda}$ in the formula (3.25), we obtain for the multipole component $\varphi_v = \varphi_v(w, \bar{w}; z)$ of the electric potential (3.24) the power series expansion

$$\varphi_v = \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{v!}{\lambda!(\lambda + v)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} \operatorname{Re} \{ \Phi_v^{[2\lambda]}(z) \bar{w}^v \}. \quad (3.31)$$

In order that the z -axis coincides with a particle trajectory, the lateral force $F_{\perp} = F_x + iF_y = -eE_{\perp}$ must vanish along this axis. Using the expansion (3.24) for the electric potential together with (3.31), we obtain the condition

$$E_{\perp}(x = 0, y = 0; z) = -2 \frac{\partial \varphi}{\partial \bar{w}} \Big|_{w=0} = -\Phi_1(z) = 0. \quad (3.32)$$

Hence, the z -axis forms only a straight optic axis if the dipole component φ_1 of the electric potential φ vanishes along this axis.

The rotationally symmetric scalar potential φ_0 is the most important special case since it describes electrostatic round lenses. In order to determine the corresponding electrostatic potential $\varphi_0 = \varphi_0(w, \bar{w}; z) = \varphi_0(\rho, z)$ in the entire space, it suffices to know its axial potential

$$\varphi(0, 0; z) = \Phi_0(z) = \Phi_{0c}(z) =: \Phi(z). \quad (3.33)$$

Hence, the Laplace equation restricts the shape of the equipotentials and, as a consequence, the spatial distribution of the electron optical index of refraction (2.51).

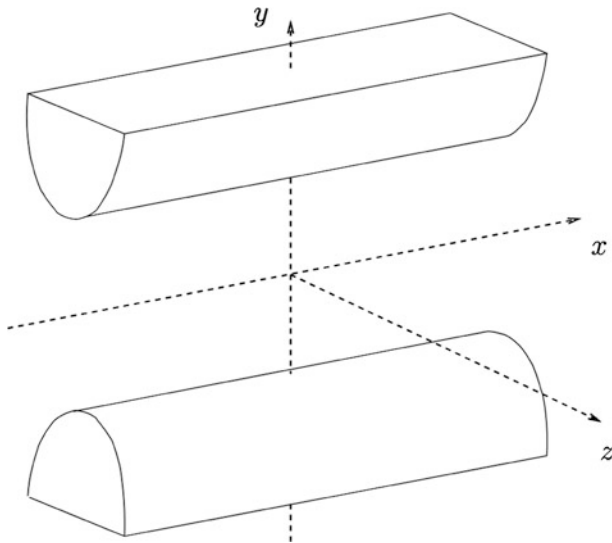


Fig. 3.5 Schematic arrangement of the electrodes of an electrostatic cylinder lens

The azimuthal orientation of any multipole field with multiplicity $\nu = m$ with respect to the direction of the x -axis is given by the angle

$$\alpha_m = \alpha_m(z) = \frac{1}{m} \arctan (\Phi_{ms}(z)/\Phi_{mc}(z)). \quad (3.34)$$

This angle defines the location of one of the m principal sections of the multipole field. These sections are plane sections only if the ratio Φ_{ms}/Φ_{mc} is constant. If this condition is not fulfilled, the principal sections are “screwed.”

3.2.2 Electrostatic Cylinder Lenses

Electrodes, which extend infinitely in a direction perpendicular to the straight optic axis, form electrostatic “cylinder” lenses. In light optics, such lenses are glass cylinders whose index of refraction does not depend on the direction of the cylinder axis. We choose this axis as the x -axis of the rectilinear x, y, z coordinate system. The potential and the arrangement of the infinitely extended electrodes of electrostatic cylinder lenses must be symmetric with respect to the x - z plane, as it is the case for slit lenses and the electrodes shown in Fig. 3.5. In practice, it suffices if the length of the electrodes is large compared to the distance in the y -direction between any two electrodes placed symmetrically about the plane section $y = 0$. Since the x -axis points in the direction of the electrodes, the potential of electrostatic cylinder lenses does not depend on this coordinate. Hence, the potential of electrostatic cylinder

lenses $\varphi = \varphi_c = \varphi_c(y, z)$ satisfies the two-dimensional Laplace equation

$$\frac{\partial^2 \varphi_c}{\partial y^2} + \frac{\partial^2 \varphi_c}{\partial z^2} = 0. \quad (3.35)$$

In Sect. 3.1.3, we have introduced *planar* solutions of the Laplace equation, which are analytical functions of the complex variable $w = x + iy$. In this case, the infinitely extended electrodes lie in line with the x -axis. Hence, the solutions of Eq. (3.35) are analytical functions of the complex variable $z + iy$.

The solutions for the potential of electrostatic cylinder lenses must have *even* mirror symmetry with respect to the section $y = 0$. Therefore, the potential φ_c must depend on the square of the variable y , as can be seen from the Taylor series expansion

$$\varphi_c = \text{Re} \Phi_c(z + iy) = \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{y^{2\lambda}}{(2\lambda)!} \Phi_c^{[2\lambda]}(z). \quad (3.36)$$

Here, $\Phi_c(z) = \bar{\Phi}_c(z) = \varphi_c(y=0, z) = \Phi(z)$ is the potential along the optic axis. Since we can expand each potential distribution in a series of multipole potentials, the representation (3.36) must also be obtainable from the multipole expansion

$$\varphi = \sum_{v=0}^{\infty} \varphi_v = \sum_{v=0}^{\infty} \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{v!}{\lambda!(v+\lambda)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} \text{Re} \{ \Phi_v^{[2\lambda]}(z) \bar{w}^v \}. \quad (3.37)$$

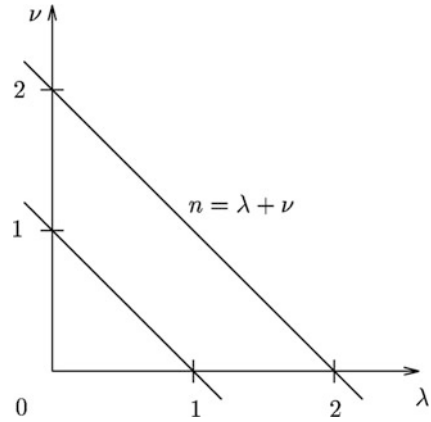
The multipole strength $\Phi_v(z)$ near the optic axis defines the multipole potential $\varphi_v = \varphi_v(x, y, z)$ in the entire space. It follows from the condition

$$\varphi = \varphi(w, \bar{w}, z) = \varphi_c(y, z) = \varphi_c(-y, z) \quad (3.38)$$

that the multipole strengths with *odd* index $v = 2n + 1$ must vanish and that the multipole strengths $\Phi_{2n} = \bar{\Phi}_{2n} = \Phi_{2n,c}(z)$ with *even* index must be real. For determining these functions, we rewrite the expansion (3.36) in terms of the complex coordinate $w = x + iy$:

$$\begin{aligned} \varphi_c &= \sum_{n=0}^{\infty} \frac{\Phi_c^{[2n]}}{(2n)!} \left(\frac{w - \bar{w}}{2} \right)^{2n} = \sum_{n=0}^{\infty} \sum_{\mu=0}^{2n} (-)^{\mu} \frac{\Phi_c^{[2n]}}{4^n (2n)!} \binom{2n}{\mu} w^{2n-\mu} \bar{w}^{\mu} \\ &= \sum_{n=0}^{\infty} \frac{\Phi_c^{[2n]}}{4^n (2n)!} \left\{ \sum_{\mu=0}^n (-)^{\mu} \binom{2n}{\mu} w^{2n-\mu} \bar{w}^{\mu} + \sum_{\mu=n}^{2n} (-)^{\mu} \binom{2n}{\mu} w^{2n-\mu} \bar{w}^{\mu} \right. \\ &\quad \left. - (-)^n \binom{2n}{n} (w\bar{w})^n \right\}. \end{aligned} \quad (3.39)$$

Fig. 3.6 Change of the summation sequence by substituting $n = \nu + \lambda$ for either ν or λ and vice versa



We reorder the summations over μ in the last expression by replacing this index in the first sum by $n - \nu$ and in the second sum by $n + \nu$. The *binomial coefficients* satisfy the relation

$$\binom{2n}{n - \nu} = \binom{2n}{n + \nu} = \frac{(2n)!}{(n - \nu)!(n + \nu)!}. \quad (3.40)$$

Using this relation, we readily derive the representation

$$\begin{aligned} \varphi_c &= \sum_{n=0}^{\infty} \frac{\Phi_c^{[2n]}}{4^n (2n)!} \left\{ \sum_{\nu=0}^n \binom{2n}{n - \nu} (-)^{n-\nu} w^{n+\nu} \bar{w}^{n-\nu} \right. \\ &\quad \left. + \sum_{\nu=0}^n \binom{2n}{n + \nu} (-)^{n+\nu} w^{n-\nu} \bar{w}^{n+\nu} - (-)^n \binom{2n}{n} (w\bar{w})^n \right\} \\ &= \sum_{n=0}^{\infty} \sum_{\nu=0}^n \frac{2(-)^{n-\nu}}{1 + \delta_{0\nu}} \frac{\Phi_c^{[2n]}}{(n - \nu)!(n + \nu)!} \left(\frac{w\bar{w}}{4} \right)^n \operatorname{Re} \left(\frac{\bar{w}}{w} \right)^\nu. \end{aligned} \quad (3.41)$$

Here, $\delta_{0\nu}$ denotes the *Kronecker symbol* which is 1 for $\nu = 0$ and zero else.

In the last step, we reorder the summations by changing the summation over n by means of a summation over the index $\lambda = n - \nu$, as illustrated in Fig. 3.6. The resulting summations over the indices ν and λ must be taken from 0 to ∞ . By means of this substitution, we finally obtain for φ_c the representation

$$\varphi_c = \sum_{\nu=0}^{\infty} \sum_{\lambda=0}^{\infty} \frac{2(-)^\lambda}{1 + \delta_{0\nu}} \frac{1}{\lambda!(\lambda + 2\nu)!} \left(\frac{w\bar{w}}{4} \right)^\lambda \frac{1}{4^\nu} \operatorname{Re} \{ \Phi_c^{[2\nu+2\lambda]} \bar{w}^{2\nu} \}. \quad (3.42)$$

The comparison of this representation with the multipole expansion (3.37) reveals that (3.42) indeed represents a multipole expansion consisting of multipoles with even multiplicity $m = 2\nu$ whose strengths are given by

$$\Phi_{2\nu}(z) = \bar{\Phi}_{2\nu}(z) = \frac{2}{4^\nu(1 + \delta_{0\nu})} \frac{\Phi_c^{[2\nu]}(z)}{(2\nu)!}. \quad (3.43)$$

This result demonstrates that we can conceive the cylinder lens as a special superposition of a round lens with a quadrupole, octopole, etc., in such a way that the sum of their refraction powers cancel out in the x -direction.

3.3 Systems with Curved Axis

The central trajectory of any ensemble of rays is curved if the electromagnetic field has a nonvanishing dipole field in the domain of the particles. Examples are deflection elements in spectrometers and monochromators and the dipole magnets in an accelerator or storage ring generating a closed quasi-circular central trajectory. In this case, it is advantageous to match the z -coordinate of the coordinate system to this trajectory, which forms the optic axis. Nevertheless, it is not necessary that the optic axis is representing a trajectory. In order to study the most general case, we assume an arbitrary skew axis with a given complex curvature (3.6).

3.3.1 Recurrence Formula for the Coefficients of the Power Series Expansion

It is widely believed that systems with arbitrary skew axis are rather unmanageable and that the mathematical treatment does not yield detailed information on the imaging properties of these systems. However, this conjecture does not hold true if we expand the potential in a series of multipoles about the skew axis [51, 52]. By employing complex notation, we represent the potential by the power series expansion of the form

$$\varphi = \sum_{\lambda=0}^{\infty} \sum_{\mu=0}^{\infty} b_{\lambda\mu}(z) w^\lambda \bar{w}^\mu = \sum_{\lambda=0}^{\infty} \sum_{\mu=\lambda}^{\infty} (2 - \delta_{\lambda\mu}) \operatorname{Re} (b_{\lambda\mu} w^\lambda \bar{w}^\mu). \quad (3.44)$$

We derive the second series by considering that the potential is a real function. Therefore, the complex expansion coefficients must satisfy the relation

$$b_{\lambda\mu} = \bar{b}_{\mu\lambda}. \quad (3.45)$$

For determining these coefficients, we assume in accordance with the special case of a straight optic axis that we know the complex strength $\Phi_\mu = \Phi_\mu(z)$ of each multipole component:

$$b_{0\mu} = \bar{b}_{\mu 0} = \frac{1 + \delta_{\mu 0}}{2} \Phi_\mu(z). \quad (3.46)$$

The coefficient $b_{00} = \bar{b}_{00} = \Phi_0(z) = \Phi(z) = \varphi(x=0, y=0, z)$ defines the axial potential along the curved optic axis. For determining the recurrence relation for the coefficients $b_{\lambda\mu}$, we first expand the inverse metric coefficient in the Laplace equation (3.16) in a power series giving

$$\frac{1}{g_3} = \left(1 - \frac{\Gamma}{2}\bar{w} - \frac{\bar{\Gamma}}{2}w\right)^{-1} = \sum_{m,n=0}^{\infty} \binom{m+n}{n} \frac{\bar{\Gamma}^m}{2^m} \frac{\Gamma^n}{2^n} w^m \bar{w}^n. \quad (3.47)$$

Subsequently, we insert this expansion and the series (3.44) for the potential φ into the Laplace equation (3.16) giving

$$\begin{aligned} g_3 \Delta \varphi &= 4 \operatorname{Re} \left\{ \frac{\partial}{\partial w} \left(g_3 \frac{\partial \varphi}{\partial \bar{w}} \right) \right\} + \frac{\partial}{\partial z} \left(\frac{1}{g_3} \frac{\partial \varphi}{\partial z} \right) \\ &= 4 \sum_{\lambda, \mu=0}^{\infty} \operatorname{Re} \frac{\partial}{\partial w} \left[\left(1 - \frac{\bar{\Gamma}}{2}w - \frac{\Gamma}{2}\bar{w} \right) (\mu+1) b_{\lambda, \mu+1} w^\lambda \bar{w}^\mu \right] \\ &\quad + \sum_{\lambda, \mu=0}^{\infty} \frac{\partial}{\partial z} \sum_{m,n=0}^{\infty} \binom{m+n}{n} \frac{\bar{\Gamma}^m}{2^{m+n}} \frac{\Gamma^n}{2^n} \frac{db_{\lambda\mu}}{dz} w^{m+\lambda} \bar{w}^{n+\mu}. \end{aligned} \quad (3.48)$$

In order to derive the recurrence formula, we must rearrange the four summations of the term. We achieve this by changing the summations with respect to the indices λ and m and those with respect to μ and n . As illustrated in Fig. 3.6, we substitute λ' for $\lambda+m$ and μ' for $\mu+n$, respectively. Due to these rearrangements, the upper limit of the sum over m changes from infinity to $m = \lambda'$ and that of the sum over n from infinity to $n = \mu'$. After performing this rearrangement, we can drop the dash at the indices because the substitutions $\lambda' \rightarrow \lambda$ and $\mu' \rightarrow \mu$ are merely a change of notation. By performing the differentiation with respect to w in the first term, we obtain

$$\begin{aligned} &\sum_{\lambda, \mu=0}^{\infty} 4(1+\mu) \operatorname{Re} \left[b_{\lambda, \mu+1} \frac{\partial}{\partial w} \left(1 - \frac{\bar{\Gamma}}{2}w - \frac{\Gamma}{2}\bar{w} \right) w^\lambda \bar{w}^\mu \right] \\ &= \sum_{\lambda, \mu=0}^{\infty} (1+\mu) b_{\lambda, \mu+1} \bar{w}^\mu (2\lambda w^{\lambda-1} - \bar{\Gamma}[\lambda+1]w^\lambda - \Gamma\lambda w^{\lambda-1}) \\ &\quad + \sum_{\lambda, \mu=0}^{\infty} (1+\mu) \bar{b}_{\lambda, \mu+1} w^\mu (2\lambda \bar{w}^{\lambda-1} - \Gamma[\lambda+1]\bar{w}^\lambda - \bar{\Gamma}\lambda \bar{w}^{\lambda-1}) \\ &= \sum_{\lambda, \mu=0}^{\infty} [4(\lambda+1)(\mu+1)b_{\lambda+1, \mu+1} - (\lambda+1)(2\mu+1)\Gamma b_{\lambda+1, \mu} \\ &\quad - (\mu+1)(2\lambda+1)\bar{\Gamma} b_{\lambda, \mu+1}] w^\lambda \bar{w}^\mu. \end{aligned} \quad (3.49)$$

The last expression is derived by considering (a) the relation (3.45), (b) by proper changes ($\lambda - 1 \rightarrow \lambda, \mu - 1 \rightarrow \mu$) of the indices for specific terms in order to obtain corresponding monomials, and (c) by exchanging the indices λ and μ in the last term of the second double sum. By employing these manipulations, the relation (3.48) adopts the form

$$\begin{aligned}
 g_3 \Delta \varphi = & \sum_{\lambda, \mu=0}^{\infty} \{4(\lambda+1)(\mu+1)b_{\lambda+1, \mu+1} - (\lambda+1)(2\mu+1)\Gamma b_{\lambda+1, \mu}\} w^\lambda \bar{w}^\mu \\
 & + \sum_{\lambda, \mu=0}^{\infty} \left\{ \sum_{m=0}^{\lambda} \sum_{n=0}^{\mu} \frac{(m+n)!}{2^{m+n} m! n!} \frac{d}{dz} \left(\bar{\Gamma}^m \Gamma^n \frac{db_{\lambda-m, \mu-n}}{dz} \right) \right. \\
 & \left. - (\mu+1)(2\lambda+1)\bar{\Gamma} b_{\lambda, \mu+1} \right\} w^\lambda \bar{w}^\mu = 0. \tag{3.50}
 \end{aligned}$$

Since this relation must be satisfied for all values of w and $\bar{w}$, the factor in front of each monomial $w^\lambda \bar{w}^\mu$ must be zero. This requirement yields the recurrence formula

$$\begin{aligned}
 4(\lambda+1)(\mu+1)b_{\lambda+1, \mu+1} = & (\lambda+1)(2\mu+1)\Gamma b_{\lambda+1, \mu} + (\mu+1)(2\lambda+1)\bar{\Gamma} b_{\lambda, \mu+1} \\
 & - \sum_{m=0}^{\lambda} \sum_{n=0}^{\mu} \frac{(m+n)!}{2^{m+n} m! n!} \frac{d}{dz} \left(\bar{\Gamma}^m \Gamma^n \frac{db_{\lambda-m, \mu-n}}{dz} \right). \tag{3.51}
 \end{aligned}$$

In the special case of a straight optic axis ($\Gamma = \bar{\Gamma} = 0$), all terms vanish on the right-hand side except the first term ($m=0, n=0$) $b''_{\lambda\mu}$ of the double sum. The resulting coefficients $b_{\lambda\mu}$ relate to the equivalent coefficients $a_{v\lambda}$ of the multipole expansion (3.25) via

$$a_{v\lambda} = (2 - \delta_{0v})b_{\lambda, v+\lambda}. \tag{3.52}$$

We verify this relation by replacing the summation over the index μ by the summation with respect to the index $v = \mu - \lambda$ in the second expression of the power series expansion (3.44). The index v defines the *multiplicity* of each multipole component.

3.3.2 Power Series Expansion of the Electric Potential

The recurrence formula (3.51) is especially suited for a computer-assisted algebraic determination of the higher-order coefficients. We derive the coefficients for the lower-order terms relatively easily direct from (3.51) because only few multipole

strengths $\Phi_\mu = 2b_{0\mu}/(1+\delta_{0\mu})$, $\mu = 0, 1, 2, \dots$, contribute. Since the coefficients are obtained successively with increasing order $n = \lambda + \mu$ from the recurrence formula (3.51), we must start with $\lambda = 0$ and $\mu = 0$ yielding

$$4b_{11} = \Gamma b_{10} + \bar{\Gamma} b_{01} - b''_{00} = \text{Re}(\bar{\Gamma} \Phi_1) - \Phi''_{00}. \quad (3.53)$$

In the next step, we put $\mu = 1$ which gives

$$\begin{aligned} 8b_{12} = 8\bar{b}_{21} &= 3\Gamma b_{11} + 2\bar{\Gamma} b_{02} - \frac{d}{dz} (b'_{01} + \Gamma b'_{00}/2) \\ &= -\frac{1}{4} [2\Phi''_1 - 4\bar{\Gamma} \Phi_2 - 3\text{Re}(\bar{\Gamma} \Phi_1) + 5\Gamma \Phi'' + 2\Gamma' \Phi']. \end{aligned} \quad (3.54)$$

Putting subsequently $\mu = 2$ and $\lambda = 1$, respectively, we obtain the next-order coefficients $b_{13} = \bar{b}_{31}$ and $b_{22} = \bar{b}_{22}$. The relations (3.53) and (3.54) suffice to determine the power series expansion of the electric potential up to the fourth-order terms inclusively. Without recapitulating the rather lengthy derivation of these coefficients, we give the result of the expansion:

$$\begin{aligned} \varphi = \text{Re} \left\{ \Phi + \Phi_1 \bar{w} + \Phi_2 \bar{w}^2 - \frac{1}{4} (\Phi'' - \bar{\Gamma} \Phi_1) w \bar{w} + \Phi_3 \bar{w}^3 \right. \\ + \frac{1}{16} (4\Phi_2 \bar{\Gamma} - 2\Phi''_1 + 3\Gamma \text{Re}[\bar{\Gamma} \Phi_1] - 5\Phi'' \Gamma - 2\Phi' \Gamma') w \bar{w}^2 + \Phi_4 \bar{w}^4 \\ + \frac{1}{4} \left(\Phi_3 \bar{\Gamma} - \frac{1}{3} \Phi''_2 + \frac{5}{12} \Phi_2 \Gamma \bar{\Gamma} - \frac{3}{8} \Phi'_1 \Gamma - \frac{1}{6} \Phi'_1 \Gamma' + \frac{5}{16} \Gamma^2 \text{Re}[\Phi_1 \bar{\Gamma}] \right. \\ \left. - \frac{11}{16} \Phi'' \Gamma^2 - \frac{13}{24} \Phi' \Gamma' \Gamma \right) w \bar{w}^3 \\ \left. + \frac{1}{64} \left(\Phi^{[4]} + 6\Phi_2 \Gamma^2 - 6\Phi'_1 \bar{\Gamma} - 4\Phi'_1 \bar{\Gamma}' - \Phi_1 \bar{\Gamma}'' + \frac{9}{2} \Phi_1 \Gamma \bar{\Gamma}^2 - \frac{19}{2} \Phi'' \Gamma \bar{\Gamma} \right. \right. \\ \left. \left. - 7\Phi' \bar{\Gamma}' \Gamma \right) w^2 \bar{w}^2 + \dots \right\}. \end{aligned} \quad (3.55)$$

We readily obtain from (3.55) the equivalent expansion for the scalar magnetic potential by means of the substitutions $\varphi \rightarrow \psi$, $\Phi \rightarrow \Psi$, and $\Phi_\mu \rightarrow \Psi_\mu$. In the case of a straight axis ($\Gamma = 0$), the series reduces considerably. The result coincides with that obtained from the multipole expansion (3.37) for systems with straight axis:

$$\begin{aligned} \varphi = \text{Re} \left\{ \Phi - \frac{1}{4} \Phi'' w \bar{w} + \frac{1}{64} \Phi^{[4]} w^2 \bar{w}^2 + \Phi_1 \bar{w} \right. \\ \left. - \frac{1}{8} \Phi'_1 w \bar{w}^2 + \Phi_2 \bar{w}^2 - \frac{1}{12} \Phi'_2 w \bar{w}^3 + \Phi_3 \bar{w}^3 + \Phi_4 \bar{w}^4 + \dots \right\}. \end{aligned} \quad (3.56)$$

Systems with *plane midsection symmetry* are an important class of curvilinear systems. In these systems, the optic axis forms a plane curve ($\Gamma = \bar{\Gamma}$) which lies in the symmetry section $y = 0$. The potential is either symmetric or antisymmetric with respect to this plane. Since the curvature Γ of the optic axis is real, all complex multipole strengths $\Phi_\mu = \Phi_{\mu c} + i\Phi_{\mu s}$ must be purely imaginary ($\Phi_\mu = -\bar{\Phi}_\mu = i\Phi_{\mu s}$) in the antisymmetric case

$$\varphi(\bar{w}, w; z) = -\varphi(w, \bar{w}; z) \rightarrow \varphi(x, -y; z) = -\varphi(x, y; z). \quad (3.57)$$

The strengths $\Phi_\mu = \bar{\Phi}_\mu = \Phi_{\mu c}$ are real in the symmetric case

$$\varphi(\bar{w}, w; z) = \varphi(w, \bar{w}; z) \rightarrow \varphi(x, -y; z) = \varphi(x, y; z). \quad (3.58)$$

One realizes the symmetric potential in electrostatic deflection systems and monochromators, while the antisymmetric potential is of importance for magnetic systems such as deflecting prisms, accelerators, and imaging energy filters.

3.3.3 Index of Refraction

In the case of stationary electromagnetic fields, the optic z -axis is always chosen as the independent variable regardless if this axis represents a particle trajectory or not. With this choice, we obtain most conveniently the path equations by means of Fermat's principle $\delta l = 0$, where we write the optical path length (2.42) in the form

$$L = S/q_0 = \int_{z_0}^z n \frac{ds}{dz} dz = \int_{z_0}^z \mu dz. \quad (3.59)$$

Considering the relation $d\vec{s} = \vec{e}_z g_3 dz + \vec{e}_x dx + \vec{e}_y dy$ and employing the complex quantities (3.5), we obtain

$$\mu = \mu(w, \bar{w}, w', \bar{w}'; z) = n \frac{ds}{dz} = \frac{\vec{k}}{k_0} \frac{d\vec{s}}{dz} = \frac{1}{q_0} \left[mv \frac{ds}{dz} - e\vec{A} \frac{d\vec{s}}{dz} \right] = \mu_e + \mu_m. \quad (3.60)$$

The *variational* function [53, 54] $\mu = \mu_e + \mu_m$ consists of an electrostatic part μ_e and a magnetic part μ_m .

Employing the relation (2.44), we find the *electric* term as

$$\mu_e = \sqrt{\frac{\varphi^*}{\Phi_0^*} \frac{ds}{dz}} = \sqrt{\frac{\varphi^*}{\Phi_0^*} (g_3^2 + w' \bar{w}')}. \quad (3.61)$$

The *magnetic* term relates to the components of the magnetic vector potential via

$$\mu_m = -\frac{e}{q_0} [g_3 A_z + \text{Re}(\bar{A}w')] . \quad (3.62)$$

Primes denote differentiations with respect to the z -coordinate. The complex lateral distances $w = w(z)$ and $\bar{w} = \bar{w}(z)$ must be conceived as *independent* from each other for the variation of the optical path length (3.59).

3.4 Magnetic Vector Potential

In order to obtain a power series expansion for the variational function (3.60), we need to know the power series expansion of the electrostatic potential and of the magnetic vector potential. Since we may assume the permeability of unsaturated iron pole pieces as infinite, their faces represent surfaces of constant scalar magnetic potential. Therefore, it is advantageous to express the magnetic vector potential in terms of the multipole components of the scalar magnetic potential. The series expansion of this potential is readily obtained from that (3.55) of the electric potential, as outlined in Sect. 3.2.2. The magnetic vector potential $\vec{A}$ is connected with the scalar magnetic potential ψ via $\vec{B} = -\text{grad } \psi = \vec{\nabla} \times \vec{A}$. For the curved coordinate system, the second relation has the detailed form

$$\begin{aligned} \vec{e}_x \frac{\partial \psi}{\partial x} + \vec{e}_y \frac{\partial \psi}{\partial y} + \vec{e}_z \frac{\partial \psi}{\partial z} \\ = \frac{\vec{e}_x}{g_3} \left[\frac{\partial A_y}{\partial z} - \frac{\partial(g_3 A_z)}{\partial y} \right] - \frac{\vec{e}_y}{g_3} \left[\frac{\partial A_x}{\partial z} - \frac{\partial(g_3 A_z)}{\partial x} \right] + \vec{e}_z \left[\frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x} \right]. \end{aligned} \quad (3.63)$$

By employing the complex quantities (3.5), we can write this three-component vector equation as a set of two equations, one real and the other complex:

$$\frac{1}{g_3} \frac{\partial \psi}{\partial z} = -2\text{Im} \frac{\partial A}{\partial w} = -2\text{Re} \left(i \frac{\partial \bar{A}}{\partial \bar{w}} \right), \quad (3.64)$$

$$2ig_3 \frac{\partial \psi}{\partial w} = -\frac{\partial \bar{A}}{\partial z} + 2 \frac{\partial(g_3 A_z)}{\partial w}. \quad (3.65)$$

The solution of the Laplace equation for the scalar magnetic potential has the form

$$\psi = \text{Re} \Pi = \text{Re} \sum_{\lambda=0}^{\infty} \sum_{\nu=0}^{\infty} (2 - \delta_{0\nu}) (w\bar{w})^\lambda b_{\lambda, \nu + \lambda} \bar{w}^\nu. \quad (3.66)$$

Since the Laplace equation (3.14) is a linear differential equation for ψ , the imaginary part of the complex potential Π is also a solution of this equation. We obtain

the power series expansion for the real part $\psi = \text{Re}\Pi$ of the complex potential $\Pi = \psi + i\Omega$ from the expansion (3.55) for the electric potential by substituting the magnetic multipole strengths for the corresponding electric strengths, as indicated in the text beneath the formula. The imaginary part Ω must also satisfy the Laplace equation. Therefore, we can assume without loss of generality that the multipole expansion of Ω has the same structure as that of ψ or φ , respectively. Accordingly, we can choose the multipole coefficients $\Omega_\nu = \Omega_\nu(z)$ of the expansion for Ω arbitrarily. The best choice is $\Omega_\nu = -i\Psi_\nu$, $\Omega_0 = 0$. Using this relation, we derive

$$\begin{aligned} \Pi = \psi + i\Omega = \Psi + \Psi_1\bar{w} + \Psi_2\bar{w}^2 + \frac{1}{4}(\Psi_1\bar{\Gamma} - \Psi''_1)\bar{w}w + \Psi_3\bar{w}^3 + \frac{1}{8}(2\Psi_2\bar{\Gamma} - \Psi''_2)\bar{w}^2w \\ + \frac{3}{16}\Psi_1\bar{\Gamma}\text{Re}[\Gamma\bar{w}^2w] - \frac{1}{16}\text{Re}\left[(5\Psi''_1\Gamma + 2\Psi'_1\Gamma')\bar{w}^2w\right] + \dots \end{aligned} \quad (3.67)$$

Since both the real part ψ and the imaginary part Ω of (3.67) satisfy the Laplace equation, Π is also a solution of this equation. We can obtain this multipole expansion much easier from the expansion (3.55) of the electric potential by substituting Ψ_ν for Φ_ν and by dropping the “Re” sign in front of all terms containing the multipole strength Ψ_ν with $\nu \neq 0$. For terms, which are products of two real parts, we must retain the Re sign unchanged for the factor containing the complex curvature Γ . Considering further that $\Psi_0 = \Psi$ is real, we readily find the expression (3.67) for the complex magnetic potential [55].

The external currents and magnets determine the magnetic vector potential only up to the gradient of an arbitrary scalar function. Accordingly, we can impose on $\vec{A}$ a constraint, which defines the gauge. The most common choices are the *Coulomb* gauge $\text{div } \vec{A} = 0$ and $A_\mu = 0$, where the index μ refers to one of the three indices 1, 2, and 3. These gauges are most favorable for systems with a straight optic axis. However, they are not the optimum choices for systems with a curved axis. In this case, we aim for another gauge, which largely simplifies the power series expansion of the magnetic term (3.62) of the variational function (3.60). This is the case if the relations between the components $A = A_x + iA_y$, A_z of the magnetic vector potential and the complex scalar potential Π are most simple [54]. The corresponding gauge is

$$\frac{1}{g_3} \frac{\partial(\text{Im}\Pi)}{\partial z} = -2\text{Re}\left(\frac{\partial A}{\partial w}\right) = -2\text{Im}\left(i \frac{\partial \bar{A}}{\partial \bar{w}}\right). \quad (3.68)$$

To demonstrate the advantage of this choice, we multiply (3.68) by i and add Eq. (3.63) to give

$$\frac{1}{g_3} \frac{\partial(\text{Re}\Pi)}{\partial z} + \frac{i}{g_3} \frac{\partial(\text{Im}\Pi)}{\partial z} = -2\text{Re}\left(i \frac{\partial \bar{A}}{\partial \bar{w}}\right) - 2i\text{Im}\left(i \frac{\partial \bar{A}}{\partial \bar{w}}\right). \quad (3.69)$$

By combining the real part and imaginary part on both sides, we obtain

$$\frac{\partial \bar{A}}{\partial \bar{w}} = \frac{i}{2g_3} \frac{\partial \Pi}{\partial z}. \quad (3.70)$$

This rather simple relation connects the conjugate complex lateral component $\bar{A}$ of the magnetic vector potential with the *complex* magnetic potential $\Pi = \Pi(w, \bar{w}, z)$. Integration with respect to the coordinate $\bar{w}$ readily yields

$$\bar{A} = \frac{i}{2} \int_0^{\bar{w}} \frac{1}{g_3} \frac{\partial \Pi}{\partial z} d\bar{w}. \quad (3.71)$$

The integration can be performed analytically by expanding both $1/g_3 = [1 - \text{Re}\{\Gamma \bar{w}\}]^{-1}$ and Π in a power series with respect to w and $\bar{w}$. As a result, we obtain an expansion of $\bar{A}$ in terms of the complex multipole strengths of the scalar magnetic potential and the complex curvature Γ . In order to obtain the relation between these quantities and the longitudinal component A_z of the magnetic vector potential, we differentiate (3.64) with respect to the variable $\bar{w}$. Subsequently, we subtract the resulting expression

$$2 \frac{\partial}{\partial \bar{w}} \left(g_3 \frac{\partial \psi}{\partial w} \right) = \frac{\partial}{\partial \bar{w}} \left(g_3 \frac{\partial \Pi}{\partial w} + g_3 \frac{\partial \bar{\Pi}}{\partial w} \right) = i \frac{\partial^2 \bar{A}}{\partial \bar{w} \partial z} - 2i \frac{\partial^2 (g_3 A_z)}{\partial \bar{w} \partial w} \quad (3.72)$$

from the Laplace equation of the complex magnetic potential, giving

$$\frac{\partial}{\partial w} \left(g_3 \frac{\partial \Pi}{\partial \bar{w}} \right) + \frac{\partial}{\partial \bar{w}} \left(g_3 \frac{\partial \Pi}{\partial w} \right) = - \frac{\partial}{\partial z} \left(\frac{1}{2g_3} \frac{\partial \Pi}{\partial z} \right) = i \frac{\partial^2 \bar{A}}{\partial z \partial \bar{w}}. \quad (3.73)$$

We obtain the second relation on the right-hand side by making use of Eq. (3.70). Integration of the resulting differential equation for $g_3 A_z$,

$$\frac{\partial}{\partial w} \left(g_3 \frac{\partial \Pi}{\partial \bar{w}} \right) - \frac{\partial}{\partial \bar{w}} \left(g_3 \frac{\partial \bar{\Pi}}{\partial w} \right) = 2i \frac{\partial^2 (g_3 A_z)}{\partial w \partial \bar{w}}, \quad (3.74)$$

with respect to w and $\bar{w}$ yields

$$g_3 A_z = \frac{1}{2i} \left\{ \int_0^{\bar{w}} g_3 \frac{\partial \Pi}{\partial \bar{w}} d\bar{w} - \int_0^w g_3 \frac{\partial \bar{\Pi}}{\partial w} dw \right\} = \text{Im} \int_0^{\bar{w}} g_3 \frac{\partial \Pi}{\partial \bar{w}} d\bar{w} = \text{Im} \int_0^{\bar{w}} \left([1 - \text{Re}\{\Gamma \bar{w}\}] \frac{\partial \Pi}{\partial \bar{w}} \right) d\bar{w}. \quad (3.75)$$

Partial integration of this integral gives

$$\int_0^{\bar{w}} g_3 \frac{\partial \Pi}{\partial \bar{w}} d\bar{w} = g_3 \Pi|_{\bar{w}=0} + \frac{\Gamma}{2} \int_0^{\bar{w}} \Pi d\bar{w} = g_3 \Pi - [1 - \bar{\Gamma}w/2] \Pi(w, \bar{w} = 0, z) + \frac{\Gamma}{2} \int_0^{\bar{w}} \Pi d\bar{w}. \quad (3.76)$$

Here, $\Pi(w, \bar{w} = 0, z) = \Pi(w = 0, \bar{w} = 0, z) = \Pi_0 = \Psi(z) = \bar{\Psi}(z)$ is the *real* scalar magnetic potential along the optic axis. Considering that $\text{Im}\Psi = 0$, we derive the connection between the longitudinal component A_z of the magnetic vector potential $\vec{A}(w, \bar{w}, z)$ and the complex scalar magnetic potential $\Pi = \Pi(w, \bar{w}, z)$ as

$$g_3 A_z = g_3 \text{Im}\Pi + \frac{1}{2} \text{Im}\Gamma \int_0^{\bar{w}} (\Pi - \Pi_0) d\bar{w}. \quad (3.77)$$

In the special case of a straight optic axis ($\Gamma = 0$), we obtain the simple result

$$A_z = \text{Im}\Pi. \quad (3.78)$$

The chosen gauge makes the magnetic vector potential zero along the optic axis ($\vec{A}(w = 0, \bar{w} = 0, z) = 0$). This behavior follows readily from the representations (3.71) and (3.77) for the components $\vec{A} = A_x - iA_y$ and A_z , respectively.

3.4.1 Rectilinear Systems

Many electron optical systems are composed of elements with a symmetry axis. If we center these axes along a straight axis, the arrangement forms a rectilinear system. For these systems, the gauge (3.68) represents the Coulomb gauge

$$\text{div } \vec{A} = \frac{\partial A_z}{\partial z} + 2\text{Re} \left(\frac{\partial A}{\partial w} \right) = \frac{\partial \text{Im}\Pi}{\partial z} - \frac{\partial \text{Im}\Pi}{\partial z} = 0. \quad (3.79)$$

The last expression has been obtained by substituting (3.78) for A_z and (3.68) for the last term of the second expression in (3.79). In the case of a curved axis, the gauge (3.68) differs from the Coulomb gauge.

In the current-free domain of the electron beam, we can represent the magnetic field by the negative gradient of the real part of the complex magnetic potential

$$\Pi = \sum_{\nu=0}^{\infty} \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{\nu!}{\lambda!(\lambda + \nu)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} \Psi_{\nu}^{[2\lambda]}(z) \bar{w}^{\nu}. \quad (3.80)$$

We derive this multipole representation by substituting Ψ_{ν} for Φ_{ν} in the expression (3.31) for the electric multipole components. By taking the imaginary part of (3.80) and by considering the relation (3.78), we obtain

$$A_z = \text{Im}\Pi = \sum_{\nu=0}^{\infty} \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{\nu!}{\lambda!(\lambda + \nu)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} \text{Im} \{ \Psi_{\nu}^{[2\lambda]}(z) \bar{w}^{\nu} \}. \quad (3.81)$$

In order to derive the equivalent series representation for the conjugate complex lateral component $\bar{A}$, we employ the relation (3.71). By putting $g_3 = 1$ and inserting the series representation (3.80) for Π , we can perform the integrations with respect to $\bar{w}$ analytically, giving

$$\bar{A} = \frac{i}{2} \bar{w} \sum_{\nu=0}^{\infty} \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{\nu!}{\lambda!(\lambda + \nu + 1)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} \Psi_{\nu}^{[2\lambda+1]}(z) \bar{w}^{\nu}. \quad (3.82)$$

The representations (3.81) and (3.82) yield a decomposition of the magnetic vector potential in terms of the multipole components of the scalar magnetic potential. The representation is especially useful for systems composed of magnetic multipole elements.

3.4.2 Magnetic Fields with Special Symmetry

The focusing properties of magnetic systems strongly depend on the symmetry of their fields. Elements with *rotational symmetry* form an important class of lenses because they affect the charged particles in the same way as the round glass lenses affect the light rays. Moreover, the number of aberrations is a minimum for rotationally symmetric fields because only the terms with $\nu = 0$ contribute to the lateral component (3.82) of the magnetic vector potential. The *longitudinal* component (3.81) vanishes since the axial component of the magnetic flux density

$$B(z) = B_z(x = 0, y = 0, z) = -\Psi'(z) = -\bar{\Psi}'(z) \quad (3.83)$$

along the optic axis is real resulting in $\text{Im}\Pi = A_z = 0$. The lateral component

$$A = i \frac{w}{2} \sum_{\lambda=0}^{\infty} \frac{(-)^{\lambda}}{\lambda!(\lambda + 1)!} \left(\frac{w\bar{w}}{4} \right)^{\lambda} B^{[2\lambda]}(z) \quad (3.84)$$

points in the azimuthal direction

$$\vec{e}_{\theta} = -\vec{e}_x \sin \theta + \vec{e}_y \cos \theta \rightarrow -\sin \theta + i \cos \theta = iw/|w|. \quad (3.85)$$

We realize the *antisymmetric* potential approximately in the fringing field domains of sector magnets with plane entrance and exit faces. The potential of magnetic cylinder lenses represents the *symmetric* type. In the simplest case, such a lens is formed by two parallel infinite wires located at the same distance a above and

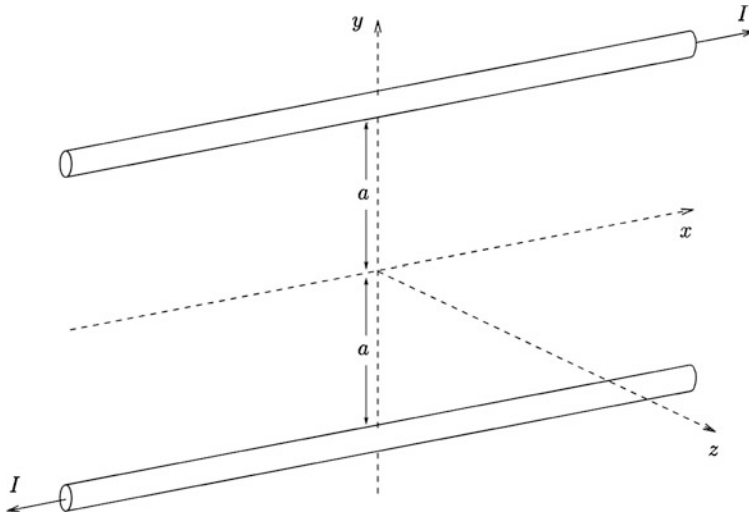


Fig. 3.7 Schematic arrangement of the conductors and the direction of the currents in a magnetic cylinder lens

beneath the symmetry plane, respectively, as illustrated in Fig. 3.7. The wires lie in the x -direction. The direction of the current I in the upper wire is opposite with respect to its direction in the lower wire. This behavior reflects the fact that circular currents about the optic axis produce the rotationally symmetric field. It also demonstrates that the chosen gauge for the vector potential is reasonable because $\vec{A}$ has the same direction as the field-producing currents.

We encounter elements with *planar* magnetic fields with a sufficient degree of accuracy in deflection magnets with plane-parallel surfaces and in magnetic cylinder lenses. For these elements, the scalar magnetic potential is independent of one of the three Cartesian coordinates x , y , z . Without loss of generality, we choose x as this coordinate. Moreover, we differentiate between planar magnetic potentials with *odd* mirror symmetry and those with *even* mirror symmetry with respect to the plane midsection $y = 0$ of the elements. Considering that $\partial/\partial x = 0$ for planar fields, any analytical function

$$\Psi(z + iy) = \psi_s(z, y) + i\psi_a(z, y) = \text{Re}\psi(z + iy) + i\text{Im}\Psi(z, iy) \quad (3.86)$$

is a solution of the two-dimensional Laplace equation. If we require that $\Psi(z)$ is a real function, the real part of $\Psi(z + iy)$ represents the symmetric solution ψ_s with respect to the midplane $y = 0$, while the imaginary part represents the antisymmetric solution ψ_a . This behavior can be readily verified by expanding $\Psi(z + iy)$ in a Taylor series with respect to iy .

In the case of the Coulomb gauge, the direction of the magnetic vector potential coincides with the direction of the external currents. Hence, for magnetic cylinder

lenses, we have

$$A_y = 0, \quad A_z = 0. \quad (3.87)$$

We obtain the remaining component A_x most conveniently from the relations

$$B_z = -\frac{\partial A_x}{\partial y} = -\frac{\partial \psi}{\partial z}. \quad (3.88)$$

Considering the even symmetry of the scalar magnetic potential with respect to y , the second equation gives

$$A_x = \int \frac{\partial \psi}{\partial z} dy = \operatorname{Re} \int \frac{\partial \Psi(z + iy)}{\partial z} dy = \operatorname{Re} \int \frac{\partial \Psi}{\partial iy} dy = -\operatorname{Re} [i \Psi(z + iy)] = \operatorname{Im} \Psi(z + iy). \quad (3.89)$$

Hence, the magnetic vector potential of a cylinder lens $A_x(z, -y) = -A_x(z, y)$ has odd symmetry with respect to y , which accounts for the opposite direction of the current above and beneath the symmetry plane $y = 0$. By expanding A_x in a Taylor series, we find that the vector potential of a magnetic cylinder lens

$$\vec{A} = \vec{e}_x A_x = -\vec{e}_x \sum_{\lambda=0}^{\infty} (-)^{\lambda} \frac{y^{2\lambda+1}}{(2\lambda+1)!} B^{[2\lambda]}(z) \quad (3.90)$$

is entirely determined by the magnetic flux density

$$B(z) = B_z(y = 0, z) = -\left. \frac{\partial \psi}{\partial z} \right|_{y=0} = -\Psi'(z) \quad (3.91)$$

along the optic axis, as in the case of round lenses.

3.4.3 Systems with Curved Axis

Systems with curved axis must contain dipole components with respect to the optic axis in order that this axis is a curved axis. In the most general case, these systems are composed of multipole terms with odd and even multiplicity. For determining the primary geometrical aberrations of these systems, only terms up to the third order in the expansion of the variational function need taken into account. Hence, it suffices to outline the expansion of the magnetic potentials up to the third order inclusively.

By inserting the expansion (3.67) of the complex magnetic potential Π in the expressions (3.71) and (3.77) for the components of the magnetic vector potential, we obtain

$$\begin{aligned}
-2i\bar{A} = & \Psi'\bar{w} + \frac{1}{2}\Psi'\bar{\Gamma}w\bar{w} + \frac{1}{4}(2\Psi'_1 + \Psi'\Gamma)\bar{w}^2 + \frac{1}{4}\Psi'\bar{\Gamma}^2w^2\bar{w} \\
& + \frac{1}{8}\left(\Psi_1\bar{\Gamma}' + 3\Psi'_1\bar{\Gamma} + 2\Psi'\Gamma\bar{\Gamma} - \Psi'''\right)w\bar{w}^2 + \frac{1}{12}(4\Psi'_2 + 2\Psi'_1\Gamma + \Psi'\Gamma^2)\bar{w}^3 + \dots,
\end{aligned} \tag{3.92}$$

$$\begin{aligned}
g_3A_z = & \text{Im}\left\{\Psi_1\bar{w} + \left(\Psi_2 - \frac{1}{4}\Psi_1\Gamma\right)\bar{w}^2 - \frac{1}{4}\Psi_1\bar{\Gamma}w\bar{w} + \left(\Psi_3 - \frac{1}{3}\Psi_2\Gamma\right)\bar{w}^3\right\} \\
& - \frac{1}{32}\text{Im}\left(8\Psi_2\bar{\Gamma} + 4\Psi''_1 + 2\Psi''\Gamma - \Psi_1\Gamma\bar{\Gamma} - \bar{\Psi}_1\Gamma^2\right)w\bar{w}^2 + \dots.
\end{aligned} \tag{3.93}$$

To check the validity of our general approach, we assume a constant electric potential and require that the chosen optic axis is representing a trajectory. By employing the expansion (3.67) and considering the relation $\vec{B} = -\text{grad}(\text{Re}\Pi)$, we obtain the magnetic flux density along the optic axis as

$$\vec{B}_0 = \vec{B}(x=0, y=0, z) = B\vec{e}_z - \Psi_{1c}\vec{e}_x - \Psi_{1s}\vec{e}_y. \tag{3.94}$$

If we insert this trivial result into Eqs. (2.66) and (2.67) for the curvature and the torsion, respectively, and consider that the velocity of the axial particle is $\vec{v} = v\vec{e}_z$, we find

$$\vec{\kappa} = \frac{e}{mv}(\Psi_{1c}\vec{e}_y - \Psi_{1s}\vec{e}_x). \tag{3.95}$$

Hence, the dipole component of the magnetic field determines entirely the vector of curvature $\vec{\kappa}$. Employing complex notation, (3.95) takes the simple form

$$\kappa = i\frac{e}{mv}\Psi_1. \tag{3.96}$$

We derive the torsion of the axis as follows:

$$\begin{aligned}
\tau = & \frac{eB}{mv} + \frac{e}{mv}\frac{\vec{\kappa}}{\kappa^2}(\Psi'_{1c}\vec{e}_x + \Psi'_{1s}\vec{e}_y) = \frac{eB}{mv} + \frac{\Psi_{1c}\Psi'_{1s} - \Psi_{1s}\Psi'_{1c}}{\Psi_{1c}^2 + \Psi_{1s}^2} \\
= & \frac{eB}{mv} + \frac{d}{dz}\arctan\left(\frac{\Psi_{1s}}{\Psi_{1c}}\right) = \frac{d\vartheta_L}{dz} + \frac{d\alpha_1}{dz}.
\end{aligned} \tag{3.97}$$

The comparison of this result with the relation (3.7) reveals that the twist angle $\vartheta = \vartheta(z)$ coincides with the angle $\alpha_1(z)$ enclosed by the x -axis and the direction of the dipole field at the plane z , whereas the *torsion angle* ϑ_τ of the accompanying Frenet–Serret trihedral is $\vartheta_\tau = \vartheta_L + \alpha_1$. So far, the influence of the Larmor rotation on the torsion angle has not been considered in the literature.

3.5 Integral Representation of the Multipole Components of the Potential

The electric and the scalar magnetic potentials within the domain of the particle beam are entirely determined by their values at the boundary surfaces. However, in many cases, we confront the inverse problem to find the geometry of the electrodes or the iron pole pieces for a given multipole strength $\Phi_\nu = \Phi_\nu(z)$ along the optic axis. We must find these strengths from the conditions imposed on the path of rays for achieving the required imaging properties. To determine the geometry of the boundary surfaces in the case of a straight axis, it is often advantageous to express the multipole potential (3.31) in the integral form [54]

$$\phi_\nu = \frac{1}{2\pi} \frac{4^\nu (\nu!)^2}{(2\nu)!} \operatorname{Re} \left\{ \bar{w}^\nu \oint \Phi_\nu(z + i\rho \sin \alpha) (\cos \alpha)^{2\nu} d\alpha \right\}. \quad (3.98)$$

This representation seems to be of little use because the complex integral can be evaluated analytically only for a few rather simple functions. However, if $\Phi_\nu(z)$ can be approximated with a sufficient degree of accuracy by a sum of these functions, the integration yields an analytical representation of the potential $\phi_\nu(x, y, z)$ in the entire space. The potential (3.98) describes a *pure* multipole field produced by a distinct electrode arrangement with 2ν -fold symmetry about the optic axis. We must choose the surfaces of the electrodes in such a way that they form surfaces of constant potential. The French mathematician Laplace has first derived the formula for the rotationally symmetric component ($\nu = 0$). One can readily show the equivalence of the representations (3.31) and (3.98) by expanding $\Phi_\nu(z + i\rho \sin \alpha)$ in a Taylor series with respect to $i\rho \sin \alpha$. In the resulting series

$$\phi_\nu = \frac{1}{2\pi} \frac{4^\nu (\nu!)^2}{(2\nu)!} \operatorname{Re} \left\{ \bar{w}^\nu \sum_{n=0}^{\infty} \frac{1}{n!} \Phi_\nu^{[n]}(z) \int_0^{2\pi} (i\rho \sin \alpha)^n (\cos \alpha)^{2\nu} d\alpha \right\}, \quad (3.99)$$

the integrals with odd powers of n vanish owing to the antisymmetry of the corresponding integrands. The remaining integrals for $n = 2\lambda$ can be evaluated analytically, giving

$$\int_0^{2\pi} \sin^{2\lambda} \alpha \cos^{2\nu} \alpha d\alpha = 2\pi \frac{(2\lambda)!(2\nu)!}{4^{\lambda+\nu} \lambda! \nu! (\lambda + \nu)!}. \quad (3.100)$$

By inserting this result into (3.99), we obtain directly the power series expansion (3.31) for ϕ_ν .

We can evaluate analytically the integral (3.98) for multipole strengths with exponential distribution

$$\Phi_\nu(z) = \Phi_{\nu 0} e^{-z/a}, \quad \Phi_{\nu 0} = |\Phi_{\nu 0}| e^{i\nu\alpha_\nu}. \quad (3.101)$$

Inserting this expression into the integrand of (3.98) and considering the integral representation

$$J_\nu(x) = \frac{(2x)^\nu \nu!}{2\pi(2\nu)!} \int_0^{2\pi} e^{-ix\sin\alpha} \cos^{2\nu}\alpha \, d\alpha, \quad x = \rho/a, \quad (3.102)$$

for the Bessel function of integer order ν , we obtain for the multipole potential the expression

$$\begin{aligned} \phi_\nu &= \phi_\nu(\rho, \theta) = 2^\nu \nu! e^{-z/a} J_\nu(\rho/a) \operatorname{Re} \left(\Phi_{\nu 0} \frac{\bar{w}^\nu}{(\rho/a)^\nu} \right) \\ &= (2a)^\nu \nu! |\Phi_{\nu 0}| e^{-z/a} J_\nu(\rho/a) \cos \nu(\theta - \alpha_\nu). \end{aligned} \quad (3.103)$$

The result shows that the multipole component ϕ_ν in terms of cylindrical coordinates (ρ, θ) represents the harmonic component of multiplicity ν with respect to the azimuth θ of the multipole series expansion. The angle α_ν defines the azimuthal orientation of the symmetry axes of the multipole with respect to the fixed Cartesian coordinate system.

3.6 Potentials of Simple Systems

The external charges and currents produce electromagnetic fields in the domain of the particle beam. The charges are located on the surfaces of the electrodes, whereas the coils for the currents are placed within the casing of the iron pole pieces. The surfaces of the electrodes and those of unsaturated pole pieces form surfaces of constant electric and magnetic scalar potential, respectively. However, the potentials applied to the boundary surfaces do not define unambiguously the spatial distribution of the external charges and currents. The reason for this ambiguity is due to the fact that we can realize a given distribution of the electrostatic potential in a defined domain by different external charge distributions, as utilized by the method of *mirror charges*. For example, we can describe the potential produced by a point charge located in front of an infinite conducting plate as a superposition of the potential of the charge either with the potential of the induced charges on the plate or with that of the mirror charge placed at the conjugate position within the other half-space.

Owing to this behavior and the linearity of the potential equation, it is possible, at least in principle, to superpose potentials of any charges in such a way that the resulting potential coincides with the given potential on the boundaries [56]. This

procedure is especially useful if the constituent potentials are given in analytical form. The analytical representation enables one to calculate precisely the paths of the particles in regions where the particle velocity is very small, as it is the case for a mirror. Small inaccuracies of the field in the region of the turning points may result in large directional changes of the trajectories. The range of application of this “*charge-simulation procedure*” will be the larger, the more appropriate analytical solutions of the Laplace equation are available [57]. One can then quickly calculate the linear combination, which yields the optimum geometry, position, and voltage of the electrodes for given imaging properties of the optical system. Moreover, the analytical model fields yield a good survey of the focusing properties of the corresponding elements and their dependence on the adjustable parameters. Another numerical approach utilizes finite element methods for determining the fields and the optical properties of electron lenses [58]. Kasper established a numerical procedure for calculating static multipole fields [59]. Here, we aim to find simple analytical solutions of the Laplace equation, which satisfy specific boundary conditions.

3.6.1 Laplace Equation for Oblate Spheroidal Coordinates

The potentials of relatively simple electrode arrangements are derived most conveniently by employing curved orthogonal coordinates [48]. For this purpose, it is advantageous to use *oblate spheroidal coordinates* (u, v, θ) , where u and v are defined by the transformations

$$z = uv, \quad \rho^2 = (1 + u^2)(1 - v^2), \quad -\infty \leq u \leq \infty, \quad 0 \leq v \leq 1. \quad (3.104)$$

The surfaces $u = \text{const}$ are confocal oblate spheroids, which degenerate to the unit disc for $u = 0$, as depicted in Fig. 3.8. The surfaces $v = \text{const}$ form confocal hyperboloids which are orthogonal to the surfaces $u = \text{const}$. The optic axis ($v = 1$) and the plane $z = 0$, $\rho \geq 1$ ($v = 0$) are degenerate special case of the hyperboloids. In these new coordinates, the Laplace equation takes the form

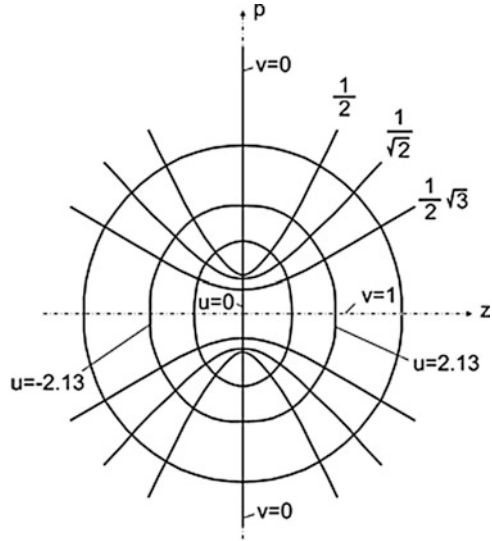
$$\Delta\varphi = \frac{1}{u^2 + v^2} \left\{ \frac{\partial}{\partial u} \left((1 + u^2) \frac{\partial\varphi}{\partial u} \right) + \frac{\partial}{\partial v} \left((1 - v^2) \frac{\partial\varphi}{\partial v} \right) \right\} + \frac{1}{(1 + u^2)(1 - v^2)} \frac{\partial^2\varphi}{\partial\theta^2}. \quad (3.105)$$

Particular solutions of this equation are the harmonics

$$\varphi_\nu = F_\nu(u, v) \cos \nu(\theta - \alpha_\nu), \quad \nu = 0, 1, \dots, \quad (3.106)$$

Each of these solutions describes a pure multipole field with multiplicity ν about the straight optic axis. The angle α_ν defines the azimuthal orientation of the multipole.

Fig. 3.8 Representation of the spheroidal coordinates $u = \text{const}$ and $v = \text{const}$ in the ρ, z -coordinate system, the degenerate hyperboloid $v = 0$ forms an aperture with an opening radius $\rho = 1$ in the plane $z = 0$



The general solution of (3.105) can be decomposed in a sum of the harmonics (3.106) representing the multipole expansion (3.24). The equation for the function $F_v = F_v(u, v)$ is derived by inserting (3.106) into the Laplace equation (3.105), giving

$$\frac{1}{u^2 + v^2} \left\{ \frac{\partial}{\partial u} (1 + u^2) \frac{\partial F_v}{\partial u} + \frac{\partial}{\partial v} (1 - v^2) \frac{\partial F_v}{\partial v} \right\} - \frac{v^2}{(1 + u^2)(1 - v^2)} F_v = 0. \quad (3.107)$$

The most important special cases of this equation are $v = 0$ and $v = 2$. They relate to systems with rotational symmetry and twofold symmetry, respectively.

3.6.2 Solutions with Rotational Symmetry

Rotationally symmetric elements are applied predominately in electron microscopes, electron ray tubes, and in electron lithography instruments. By considering $v = 0$ and $\varphi = \varphi_0 = F_0(u, v)$, the potential equation (3.107) reduces to the equation

$$\frac{\partial}{\partial u} \left((1 + u^2) \frac{\partial \varphi}{\partial u} \right) + \frac{\partial}{\partial v} \left((1 - v^2) \frac{\partial \varphi}{\partial v} \right) = 0. \quad (3.108)$$

A simple particular solution of this equation is

$$\varphi = \Phi_0 + (\Phi_L - \Phi_0) \frac{v}{u^2 + v^2}, \quad (3.109)$$

where Φ_0 can be conceived as the acceleration voltage and Φ_L as the potential at the center ($u = 0, v = 1$) of the lens.

We formally realize this round lens by a conducting aperture put at potential Φ_0 and an infinitely thin uniformly charged ring placed within the hole of the aperture close to the edge. The ring charge induces an opposite charge at the inner edge of the aperture forming a circular dipole. The resulting potential distribution resembles that of an electrostatic “*einzel lens*” or *unipotential lens* in the vicinity of the optic axis, as convincingly demonstrated by the equipotentials depicted in Fig. 3.9. In practice, three spatially separated electrodes form such an einzel lens where the two outer electrodes are at potential Φ_0 . The model lens acts as a mirror for $\Phi_L < 0$. We find the solution (3.109) by assuming that the axial potential of the lens has the bell-shaped distribution

$$\varphi(z, \rho = 0) = \Phi(z) = \Phi_0 + \frac{\Phi_L - \Phi_0}{1 + z^2}. \quad (3.110)$$

By inserting this axial potential in the integral formula (3.98) for $v = 0$, and considering the relations

$$\frac{1}{1 + z^2} = \operatorname{Re} \left(\frac{1}{1 + iz} \right), \quad v + iu = \sqrt{(1 + iz)^2 - \rho^2}, \quad (3.111)$$

we obtain

$$\begin{aligned} \varphi(u, v) &= \Phi_0 + \frac{\Phi_L - \Phi_0}{2\pi} \operatorname{Re} \int_0^{2\pi} \frac{d\alpha}{1 + iz - \rho \sin \alpha} \\ &= \Phi_0 + \operatorname{Re} \left\{ \frac{\Phi_L - \Phi_0}{\sqrt{(1 + iz)^2 - \rho^2}} \right\} = \Phi_0 + (\Phi_L - \Phi_0) \frac{v}{u^2 + v^2}. \end{aligned} \quad (3.112)$$

This result coincides exactly with the particular solution (3.109) of the Laplace equation (3.108). The electron optical properties of this model field have been explored by Glaser and Schiske [60, 61].

We derive other solutions of Eq. (3.108) by employing Bernoulli’s separation of variables

$$\varphi(u, v) = U(u)V(v). \quad (3.113)$$

Writing the separation constant as $l(l+1)$, we find that U and V satisfy the *Legendre* differential equations

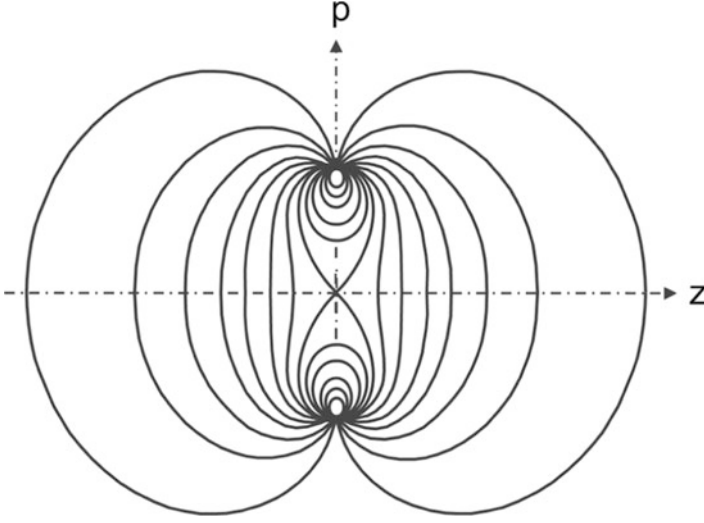


Fig. 3.9 Equipotentials $\varphi = \Phi_0 + (\Phi_L - \Phi_0)v/(u^2 + v^2)$ of the model einzel lens

$$(1 + u^2) \frac{d^2 U}{du^2} + 2u \frac{dU}{du} - l(l + 1)U = 0, \quad (3.114)$$

$$(1 - v^2) \frac{d^2 V}{dv^2} - 2v \frac{dV}{dv} + l(l + 1)V = 0. \quad (3.115)$$

We consider only solutions for integer l . Equation (3.114) reduces to the standard Legendre differential equation (3.115) if we substitute iu for u . Hence, we can represent the solution $U(u)$ by the equivalent solution $V(iu)$ with imaginary argument. Two linearly independent solutions of (3.115) are the *Legendre functions* $P_l(v)$ and $Q_l(v)$. These functions are represented by elementary functions for integer *degree* l . In this case, $P_l(v)$ is given by the *Legendre polynomial* of degree l . The functions $Q_l(v)$ have a logarithmic singularity on the optic axis $v = 1$, while the corresponding functions $Q_l(iu)$ with imaginary argument are regular [62]. Therefore, we take into account only these physically reasonable solutions. Then the particular solutions of the rotationally symmetric potential equation (3.108) have the form

$$\varphi = \varphi^{(l)} = a_l + (b_l P_l(iu) + c_l Q_l(iu)) P_l(v), \quad (3.116)$$

where a_l , b_l and c_l are arbitrary constants which must be chosen in such a way that the solution (3.116) satisfies the boundary conditions. Explicit expressions of the Legendre functions for $l = 0, 1, 2$ are

$$P_0(v) = P_0(iu) = 1, \quad Q_0(iu) = \frac{1}{2} \ln \frac{1+iu}{1-iu} = i \arctan u, \quad (3.117)$$

$$P_1(v) = v, \quad P_1(iu) = iu, \quad Q_1(iu) = -u \arctan u - 1, \quad (3.118)$$

$$P_2(v) = (3v^2 - 1)/2, \quad P_2(iu) = -(3u^2 + 1)/2, \quad Q_2(iu) = iP_2(iu) \arctan u - i3u/2. \quad (3.119)$$

Using these expressions, one can show that each of the three particular solutions represents the potential of a special electron optical element. The most simple solution ($l = 0$) is realized in the magnetic case by a flat coil with current density

$$j(\rho) = \begin{cases} \frac{2I}{\pi\rho\sqrt{\rho^2-1}}, & \text{for } \rho > 1, \\ 0, & \text{for } \rho \leq 1, \end{cases} \quad (3.120)$$

where

$$I = \int_1^\infty j(\rho) d\rho \quad (3.121)$$

is the total current about the optic axis. We derive the corresponding scalar magnetic potential from the expression (3.116) by substituting ψ for φ by and by choosing the constants as $a_0 = b_0 = 0$, $c_0 = i\mu_0 I/\pi$, giving

$$\psi = -\frac{\mu_0 I}{\pi} \arctan u. \quad (3.122)$$

The potential has a discontinuity in the plane $z=0$ where the coil is placed. By traversing the coil at lateral distance $\rho > 1$ from $z = -\varepsilon$ to $z = \varepsilon$ ($\varepsilon \ll 1$), the scalar magnetic potential jumps from

$$\psi_0 = \psi(\rho, z = -\varepsilon) = \frac{\mu_0 I}{\pi} \arctan \sqrt{\rho^2 - 1} \quad (3.123)$$

to $-\psi_0$ at the other side of the discontinuity surface. This so-called *magnetic sheet* represents a dipole layer with variable dipole strength. These dipoles are equivalents of the current density, which produces the discontinuity surface. Hence, the potential of the sheet is evidently double-valued.

The coordinates z and ρ are normalized with respect to the inner radius ρ_0 of the coil. By considering this normalization, we find the axial distribution of the magnetic flux density as

$$B = B_z(\rho = 0, z) = -\frac{1}{\rho_0} \left. \frac{\partial \psi}{\partial z} \right|_{\rho=0} = -\frac{1}{\rho_0} \left. \frac{\partial \psi}{\partial u} \right|_{u=z} = \frac{B_0}{1+z^2}, \quad B_0 = \frac{\mu_0 I}{\pi \rho_0}. \quad (3.124)$$

W. Glaser introduced this field distribution in 1940 as a model for a magnetic electron lens. It is known as *Glaser's bell-shaped model* and yields analytical solutions for the paraxial path equation in terms of circular functions [9]. We should notice that our solution for the current density is not unique if we restrict the domain to $\rho < 1$. Glaser has determined the winding distribution $n_w(z, \rho = 1)$ in axial direction for a thin solenoid, which also produces the bell-shaped axial field (3.124). However, his solution for n_w is a very complicated Fourier integral, which cannot be evaluated analytically.

The solution for $l = 1$ is

$$\varphi = \varphi^{(1)} = b_1 uv + c_1 v(u \arctan u + 1). \quad (3.125)$$

This solution describes the potential within two domains with different asymptotic electric field strengths $E_{-\infty}$ and E_{∞} . A conducting circular aperture with normalized radius $\rho = 1$ separates the domains from each other.

The conducting aperture $v = 0$ is a surface of constant potential $a_1 = \Phi_a$. By imposing the constraints

$$-\left. \frac{\partial \varphi}{\partial z} \right|_{z=-\infty} = E_{-\infty}, \quad -\left. \frac{\partial \varphi}{\partial z} \right|_{z=\infty} = E_{\infty}, \quad (3.126)$$

and considering the asymptotic form

$$\lim_{u \rightarrow \pm \infty} \varphi = \Phi_a + b_1 z + c_1 |z| \frac{\pi}{2} \quad (3.127)$$

of (3.125), we find

$$\varphi = \Phi_a - \frac{z}{2} (E_{-\infty} + E_{\infty}) + \frac{z}{\pi} (E_{-\infty} - E_{\infty}) \left(\arctan u + \frac{1}{u} \right). \quad (3.128)$$

We have depicted the equipotentials in Fig. 3.10 for (a) $E_{-\infty} = 0$ and (b) $E_{\infty} = -E_{-\infty}$. In the latter case, the potential exhibits a saddle point at the center $z = 0$, $\rho = 0$. The electric field strength $E_z(0, 0) = -\varphi'(0, 0) = -\Phi'(0)$ vanishes at this point, which forms the apex of a double cone tangential to the equipotentials intersecting the center. The cone angle 2γ defined in Fig. 3.10b differs from 90° .

The angle γ does not depend on the configuration of the rotationally symmetric system and is obtained by expanding the potential in a power series about the saddle point, giving

$$\varphi(\rho, z) \approx \Phi(z) - \frac{1}{4} \Phi''(z) \rho^2 + \dots \approx \Phi(0) + \frac{1}{2} \Phi''(0) z^2 - \frac{1}{4} \Phi''(0) \rho^2 + \dots. \quad (3.129)$$

By imposing the condition that (3.129) represents the equipotential $\varphi = \Phi(0)$, we readily obtain

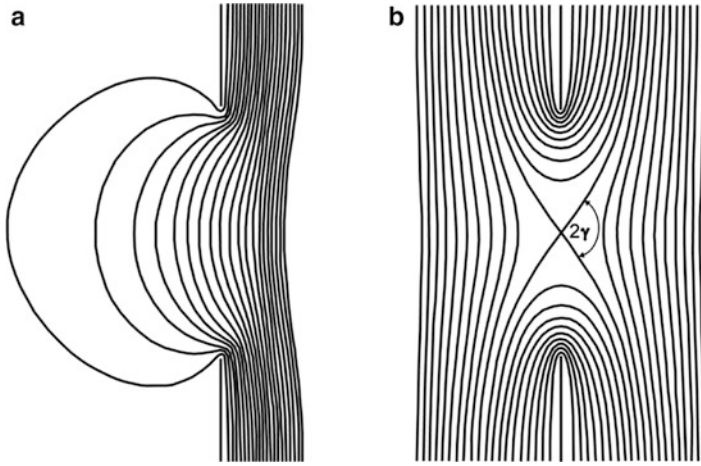


Fig. 3.10 Equipotentials of the potential distribution (3.138) for (a) $E_{-\infty} = 0$ and (b) $E_{\infty} = -E_{-\infty}$

$$\tan \gamma = \rho/z = \sqrt{2} \rightarrow \gamma = 54.83^\circ. \quad (3.130)$$

In the case $E_{-\infty} = 0$, the potential (3.129) describes the penetration of the electric field into the left half-space through the hole of the aperture.

Equation (3.128) shows that the potential forms a cylindrical quadrupole field in the region of the apex. One has realized such a field in ion traps formed by a hyperbolic toroid and two hyperboloids. The ions are confined in the interior region if a proper voltage $U_0 - U \cos \omega t$ is applied between the toroid and the two-sheet hyperboloid. This voltage consists of a static part U_0 and a high-frequency (time-dependent) part with amplitude U and frequency ω . If we neglect radiation effects, we obtain for the potential the solution $\varphi = \Phi_0 + (U_0 - U \cos \omega t)(\rho^2/2 - z^2)/\rho_0^2$. This potential yields *Mathieu's* differential equation for the motion of the particles; ρ_0 is the shortest distance of the electrodes from the axis. Stable solutions exist for distinct values of U_0/ω^2 and U/ω^2 . We shall treat extensively the properties of this ion trap in Chap. 13.

We encounter an equivalent behavior in the case of strong focusing and for the propagation of electrons within the periodic atomic potential of crystalline objects. Such stability problems always arise if the periodicity or eigenfrequency of a system interacts with the periodicity of an external force. (To make the apples fall down from a young tree, one must joggle it with its eigenfrequency.)

The solution (3.116) with $l = 2$ can be used for describing the potential of a convex diode mirror. We require that the region in front of the mirror is field-free apart from the fringing field penetrating through the hole of its entrance electrode. Accordingly, we impose the constraints $\varphi(z = -\infty) = \Phi_0$ and $\varphi'(z = -\infty) = 0$ on the particular solution

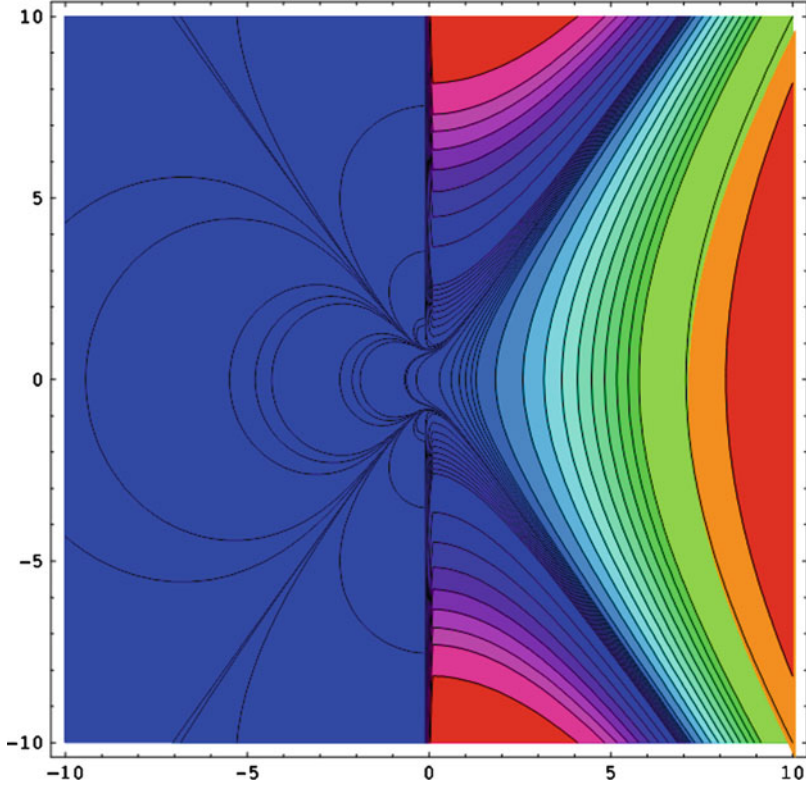


Fig. 3.11 Equipotentials of the model potential (3.131) representing a convex diode mirror, which we can realize by means of a toroidal aperture and a hyperboloid

$$\varphi = \varphi^{(2)} = a_2 + (3v^2 - 1)(3u^2 + 1) \left\{ b_2 + c_2 \left(\arctan u + \frac{3u}{3u^2 + 1} \right) \right\}. \quad (3.131)$$

Considering $(3v^2 - 1)(3u^2 + 1) = 6z^2 - 3\rho^2 + 2$ and introducing the dimensionless parameter $\sigma = -c_2/\Phi_0$ defined by the location and the potential of the second hyperbolic mirror electrode, we eventually obtain for the mirror model the potential distribution shown in Fig. 3.11

$$\varphi/\Phi_0 = 1 - \sigma(6z^2 - 3\rho^2 + 2) \left[\frac{1}{2} + \frac{1}{\pi} \left(\arctan u + \frac{3u}{3u^2 + 1} \right) \right]. \quad (3.132)$$

The normalized potential is 1 at $z = -\infty$ and on the hyperbolic surface $\rho = \sqrt{2z^2 + 2}/3$. We may choose this equipotential as the surface of the entrance electrode, while any equipotential $\varphi < -\Phi_0$ represents a proper reflecting electrode for the actual mirror.

Any linear combination of N model potentials also represents a solution of the Laplace equation. We can construct such multielectrode configurations in many different ways because we can place the constituent elements at arbitrary positions $z_i, i = 1, 2, \dots, N$ along the optic axis, as outlined and illustrated in Sect. 3.7. Since this method provides the potential distribution in analytical form, it enables a fast and very precise numerical evaluation of the path equations. Moreover, the model potentials avoid sharp edges of the electrodes, which may result in unduly large electric field strengths.

3.6.3 Multipoles

We write the solutions of Eq. (3.106) in the most general case as a sum of harmonics representing multipoles of the form (3.107). By multiplying Eq. (3.107) with the factor $u^2 + v^2 = u^2 + 1 + v^2 - 1$, we derive the form

$$\frac{\partial}{\partial u} \left((1 + u^2) \frac{\partial F_v}{\partial u} \right) + \frac{\partial}{\partial v} \left((1 - v^2) \frac{\partial F_v}{\partial v} \right) + \frac{v^2}{1 + u^2} F_v - \frac{v^2}{1 - v^2} F_v = 0. \quad (3.133)$$

We obtain particular solutions of this equation by employing Bernoulli's ansatz $F_v = U_v(u) V_v(v)$ for the separation of variables, giving the differential equations of the *associated Legendre functions* with real and imaginary argument:

$$(1 - v^2) \frac{d^2 V_v}{dv^2} - 2v \frac{dV_v}{dv} + \left[l(l + 1) - \frac{v^2}{1 - v^2} \right] V_v = 0, \quad (3.134)$$

$$(1 + u^2) \frac{d^2 U_v}{du^2} + 2u \frac{dU_v}{du} - \left[l(l + 1) - \frac{v^2}{1 + u^2} \right] U_v = 0. \quad (3.135)$$

Linearly independent solutions of these equations are the so-called *spherical harmonics* of *first* and *second kind* $V_v = P_l^v(v)$, $Q_l^v(v)$ and $U_v = P_l^v(iu)$, $Q_l^v(iu)$, respectively. If the multiplicity v of the multipole is an integer, we can derive the associated Legendre functions from the Legendre functions $P_l(v)$ and $Q_l(v)$ by means of the relations

$$P_l^v(v) = (-1)^v (1 - v^2)^{v/2} \frac{d^v P_l(v)}{dv^v}, \quad (3.136)$$

$$Q_l^v(v) = (-1)^v (1 - v^2)^{v/2} \frac{d^v Q_l(v)}{dv^v}. \quad (3.137)$$

For integer l , these formulas give nontrivial solutions $P_l^v(v) U_v(u)$ only if $l \geq v$.

As an example, we investigate the potential of a quadrupole ($v = 2$) for the separation coefficients $l = 1$ and $l = 2$. In the latter case, we can employ the relations

(3.136) and (3.137), while this is not possible for $l = 1$ since it gives for $P_1^2(v)$ the trivial result zero.

The solution $Q_2^2(v)$ cannot be taken into account because it diverges on the optic axis $v = 1$. Therefore, the realistic solution for $l = 2$ has the form

$$\begin{aligned} F_2 &= (1 - v^2) \left\{ C_1(1 + u^2) + C_2 \left[3(1 + u^2) \arctan u + 3u + \frac{2u}{1 + u^2} \right] \right\} \\ &= C_1 \rho^2 + C_2 \rho^2 \left\{ 3 \arctan u + \frac{u(5 + 3u^2)}{(1 + u^2)^2} \right\}. \end{aligned} \quad (3.138)$$

The first term describes an infinitely extended plane quadrupole, which corresponds to the homogeneous part of the electric field (3.128) of the rotationally symmetric aperture field. In order to obtain a model field for a quadrupole with a finite extension along the optic axis, we choose $C_1 = 0$ and superpose two potentials of the remaining type, where we have centered the first quadrupole at the position $z_1 = 0$ and the second at position $z_2 > 0$. The corresponding coordinates $u_i, i = 1, 2$ are given by

$$u_i = \frac{\text{sgn}(z - z_i)}{\sqrt{2}} \sqrt{(z - z_i)^2 + \rho^2 - 1 + \sqrt{[(z - z_i)^2 + \rho^2 - 1]^2 + 4(z - z_i)^2}}, \quad (3.139)$$

where $\text{sgn}(x)$ defines the sign of x . By employing the expression (3.106), we obtain the model potential

$$\varphi_2 = U \left\{ \arctan u_1 - \arctan u_2 + \frac{u_1(u_1^2 + 5/3)}{(1 + u_1^2)^2} - \frac{u_2(u_2^2 + 5/3)}{(1 + u_2^2)^2} \right\} \rho^2 \cos 2(\theta - \alpha_2). \quad (3.140)$$

This solution describes the potential of a quadrupole with *effective* length $l_{\text{eff}} = z_2 - z_1$. Since we have used normalized dimensionless coordinates, the coefficient U has the dimension of a voltage. The potential vanishes in the limit $z \rightarrow \pm\infty$. The potential differs appreciably from zero only in the region $z_1 \leq z \leq z_2$.

As an example for $l < v$, we consider the special case $l = 1, v = 2$ which gives a model for a short quadrupole. For these coefficients, the differential equation (3.134) has the simple particular solution

$$V_2 = Q_1^2(v) = \frac{1}{2}(1 - v^2) \frac{d^2}{dv^2} \left(v \ln \frac{1 + v}{1 - v} \right) = \frac{2}{1 - v^2}. \quad (3.141)$$

To obtain the other linearly independent solution of the corresponding differential equation, we employ the “*product ansatz*”

$$V_2 = g(v) \frac{1}{1 - v^2}, \quad (3.142)$$

giving the differential equation

$$g'' + \frac{2v}{1-v^2} g' = 0. \quad (3.143)$$

Twofold integration of this equation readily yields

$$g = C_1 \left(v - \frac{v^3}{3} \right) + C_2. \quad (3.144)$$

The coefficients C_1 and C_2 are the constants of the two integrations. By substituting (3.144) for g in the relation (3.142) and considering that $U_2(u) = V_2(iu)$ apart from the value of the coefficients, we eventually derive the general solution

$$F_2 = U_2(u)V_2(v) = \frac{C_1(v - v^3/3) + C_2}{1-v^2} \frac{C_3(u + u^3/3) + C_4}{1+u^2}. \quad (3.145)$$

Two of the four coefficients C_i , $i = 1, 2, 3, 4$, are uniquely specified by the conditions $F_2(v, u = \pm \infty) = 0$ and $F_2(v = 1, u) \neq \infty$. The first condition gives $C_3 = 0$ and the second $C_2 = -2C_1/3$. The second relation guarantees that the potential stays finite on the axis $v = 1$. By introducing the new coefficient $\Phi_{20} = -C_1 C_4/4$ and considering $2 - 3v + v^3 = (1 - v)^2(2 + v)$, we find

$$F_2 = -\frac{C_1 C_4}{3} \frac{2 - 3v + v^3}{(1 - v^2)(1 + u^2)} = \frac{4}{3} \Phi_{20} \frac{2 + v}{(1 + v)^2(1 + u^2)^2} \rho^2. \quad (3.146)$$

The corresponding potential $\varphi_2 = F_2 \cos 2(\theta - \alpha_2)$ has a bell-shaped axial distribution in the paraxial domain, as can be seen by putting $v = 1$ and $u = z$ in the expression (3.146) for $F_2(u, v)$. The result

$$\varphi_2 \approx \Phi_2(z) \rho^2 \cos 2(\theta - \alpha_2) = \frac{\Phi_{20}}{(1 + z^2)^2} \rho^2 \cos 2(\theta - \alpha_2) \quad (3.147)$$

demonstrates convincingly that (3.146) represents the two-dimensional distribution of the potential of a short quadrupole within any plane section $\theta = \text{const}$. The result (3.146) can also be verified by means of the integral representation (3.98) of the multipole potential by putting $v = 2$ and $\Phi_2(z) = \Phi_{20}(1 + z^2)^{-2}$, although the analytical integration is rather lengthy. Summarizing, we can state that it is possible to produce *pure* multipole fields with realistic geometry of the electrodes or magnets.

3.7 Charge-Simulation Method

The charge-simulation method approximates the electrostatic potential φ by a proper superposition of analytical solutions of the Laplace equation (3.4). In order to keep the number of elements as small as possible and to avoid inaccuracies at the boundaries, it is advantageous to place the charges inside the electrodes [56, 63]. We consider an arrangement of electrodes, which we put at different potentials. The disappearance of the electric field in the metallic interior of the electrodes results in a charge distribution on their surfaces. These surface charges produce a static electric field in the region outside of the electrodes, as it is the case for electrostatic electron lenses. The charge-simulation method utilizes the fact that the electrostatic potential within a finite domain can be produced by different charge distributions outside this regime. Since the electric field within the electrodes is of no interest, it is possible to produce the exterior field by interior charges whose distribution is considerably simpler than that of the surface charges. If the geometry of the electrodes exhibits a given symmetry, the charge distribution must satisfy this symmetry too.

The analytic potential fields are superposed in such a way that the potential at the boundaries is approximated with the required degree of accuracy. The number of necessary elements strongly depends on the geometry of the electrodes and on the proper choice and location of the individual charge elements. Therefore, the efficient use of the charge-simulation method requires some skill in optimally locating and selecting appropriate charge distribution elements. The accuracy of the approximated potential improves with increasing distance of the point of observation from the boundaries. Since the potential is given in analytical form, we also obtain analytically the components of the electric field strength, thus avoiding inaccuracies resulting from numerical differentiations. Accordingly, we can readily calculate the electron trajectories with high accuracy and with the least amount of computing time. High accuracy of the electric field is necessary for very small electron velocities, as it happens for mirrors or in the region near the tip of a field emission gun. In the following, we assume a system consisting of N_e electrodes, each of which is at constant potential V_k .

We choose the charge elements in such a way that they generate distinct equipotentials, which have approximately the same shape and the same potential as the surfaces of the electrodes. The number and types of charge elements, which must be taken into account for a given degree of accuracy, depend on (a) the symmetry properties of the boundaries and (b) on the position of the charge elements in the interior of the electrodes. For fixed number of charge elements, the approximated potential has the form

$$\varphi(x, y, z) = \sum_{i=1}^N \sum_{j=1}^{n_i} \varphi_{ij}(x, y, z). \quad (3.148)$$

The charge distribution of type i centered at position $\vec{r}_{ij}$ produces the potential φ_{ij} . The upper summation limits n_i and N indicate the number of elements of each type and the number of different types, respectively. Hence, the total number of charge elements is

$$N_t = \sum_{i=1}^N n_i. \quad (3.149)$$

For mathematical convenience, we write the potential φ_{ij} of each charge element as the product of the dimensionless spatial distribution function $G_{ij}(x, y, z)$ and the potential strength

$$\varphi_{0ij} = \frac{q_{ij}}{4\pi\epsilon_0 a_{ij}}. \quad (3.150)$$

Here, q_{ij} denotes the total charge of the charge element (i, j) and a_{ij} its characteristic length, for example, the radius of a charged ring. With this procedure the potential (3.147) adopts the form

$$\varphi(x, y, z) = \sum_{i=1}^N \sum_{j=1}^{n_i} \varphi_{0ij} G_{ij}(x, y, z). \quad (3.151)$$

Because the functions $G_{ij}(x, y, z)$ are given in analytical form, we can readily differentiate analytically the expression (3.150), giving the electric field

$$\vec{E} = -\text{grad}\varphi(x, y, z) = -\sum_{i=1}^N \sum_{j=1}^{n_i} \varphi_{0ij} \text{grad } G_{ij}(x, y, z). \quad (3.152)$$

The approximated potential must coincide with the exact potential at distinct control points (boundary points) $\vec{r}_{c\mu\nu}$ located on the surface of the electrode μ with constant potential V_μ . We use this condition to determine the unknown quantities φ_{0ij} and $\vec{r}_{ij}$. Unfortunately, no analytical method exists for determining the necessary number and types of charge elements and their strength φ_{0ij} and location $\vec{r}_{ij}$. Even if the number and types of charge elements are given, it is impossible to determine reliably their location because the dependence of each function $G_{ij}(x, y, z)$ on the position $\vec{r}_{ij}$ of the corresponding charge element is nonlinear. To avoid this difficulty, we fix the positions $\vec{r}_{ij}$ and restrict our calculations to the determination of the strengths φ_{0ij} . Because the electrodes are conductors, the total charge of all electrodes of the electron optical element must be zero:

$$\sum_{i=1}^N \sum_{j=1}^{n_i} q_{ij} = 0. \quad (3.153)$$

Due to this constraint, the number of control points is $N_t - 1$. We account for the additional condition (3.152) by using $n_k - 1$ control points for the element k instead

of n_k . In this case, we obtain the set of linear equations

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \varphi_{0ij} G_{ij}(\vec{r}_{c\mu\nu}, \vec{r}_{ij}) = V_\mu, \quad \mu = 1 \dots N, \quad \nu = 1 \dots n_\mu - \delta_{k\mu}. \quad (3.154)$$

Here, $\delta_{k\mu}$ is the Kronecker symbol. We can readily solve the system of linear equations (3.153) by means of the Gauss algorithm. By employing a trial-and-error iteration procedure for the positions $\vec{r}_{ij}$, we find an arrangement of the charge elements which fits the applied potential on the electrodes with the required accuracy. Once the solution of (3.153) is known, the potential of the electrostatic device is given by

$$\varphi(x, y, z) = \sum_{i=1}^N \sum_{j=1}^{n_i} \varphi_{0ij} G_{ij}(x, y, z, \vec{r}_{ij}). \quad (3.155)$$

Because the position of the charge elements is fixed, a change of the electrode potentials V_k only alters the strengths φ_{0ij} of the charges. The charge-simulation method is very versatile and has been applied successfully to electrostatic round lenses, deflection units, and two-dimensional arrays of conductors. Difficulties will only arise in the vicinity of sharp edges. However, sharp edges are avoided in realistic systems in order to prevent electrical breakdowns. An exemption is the cathode tip of field emission guns.

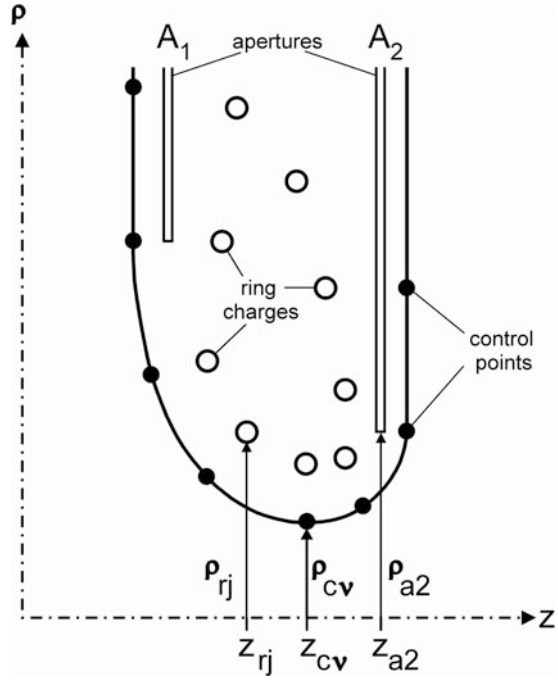
3.7.1 Systems of Rotational Symmetry

Rotationally symmetric elements such as electron lenses and cathodes are important elements of numerous instruments, electron microscopes in particular. To simulate the electrostatic potential of their constituent elements, we utilize the rotationally symmetric elements given in Sect. 3.6.2 together with charged rings and line charges. Because the rings and the apertures are centered along the optic axis, the coordinates of their central position vectors $\vec{r}_{rj}$ and $\vec{r}_{aj}$ are

$$z_{rj}, x_{rj} = 0, \quad y_{rj} = 0, \quad z_{aj}, x_{aj} = 0, \quad y_{aj} = 0. \quad (3.156)$$

Figure 3.12 shows schematically the appropriate arrangement of these elements for the simulation of the potential of a single rotationally symmetric electrode.

Fig. 3.12 Locations z_{rj} and z_{av} of the centers of the charged rings and plane apertures, respectively, introduced for simulating the equipotential surface of a rotationally symmetric electrode. The control points (ρ_{cv}, z_{cv}) are located on the surface of the electrode formed by rotating the solid curve about the optic axis



The potential of a charged ring with radius ρ_j is given by the expression

$$\varphi_{rj} = \frac{q_{rj}}{4\pi\epsilon_0\rho_j} G_{rj}(z, \rho), \quad G_{rj}(z, \rho) = \frac{\rho_j}{2\pi} \oint \frac{d\phi}{\sqrt{(z - z_{rj})^2 + \rho^2 + \rho_j^2 - 2\rho\rho_j \cos \phi}}, \quad (3.157)$$

For evaluating the integral, we use the relations

$$\begin{aligned} \rho^2 + \rho_j^2 - 2\rho\rho_j \cos \phi &= (\rho + \rho_j)^2 - 4\rho\rho_j \cos^2(\phi/2), \\ k^2 &= \frac{4\rho\rho_j}{(z - z_{rj})^2 + (\rho + \rho_j)^2}. \end{aligned} \quad (3.158)$$

In addition, we introduce the angle $\psi = (\phi + \pi)/2$, giving

$$\begin{aligned} G_{rj}(z, \rho) &= \frac{2\rho_j}{\pi \sqrt{(z - z_{rj})^2 + (\rho + \rho_j)^2}} K(k_j) \\ K(k_j) &= \int_0^{\pi/2} \frac{d\psi}{\sqrt{1 - k_j^2 \sin^2 \psi}}. \end{aligned} \quad (3.159)$$

The function $K(k_j)$ is the complete elliptic integral of the first kind [62]. We obtain the components of the electric field strength (3.152) by differentiating the expression (3.157) with respect to the coordinates z and ρ , respectively. Using the same evaluation procedure as for the potential, we eventually obtain

$$\begin{aligned}
 E_{jz} &= \frac{q_{rj}}{2\pi^2\epsilon_0} \frac{z\sqrt{(z-z_{rj})^2 + (\rho + \rho_{rj})^2}}{\left((z-z_{rj})^2 + \rho^2 + \rho_{rj}^2\right)^2 - 4\rho^2\rho_{rj}^2} E(k_j), \\
 E_{\rho j} &= \frac{q_{rj}}{4\pi^2\epsilon_0\rho\sqrt{(z-z_{rj})^2 + (\rho + \rho_{rj})^2}} \left[K(k_j) + \frac{\rho^2 - \rho_j^2 - z^2}{(z-z_{rj})^2 + (\rho - \rho_{rj})^2} E(k_j) \right], \\
 E(k_j) &= \int_0^{\pi/2} \sqrt{1 - k_j^2 \sin^2 \psi} d\psi.
 \end{aligned} \tag{3.160}$$

The function $E(k_j)$ is the complete elliptic integral of the second kind.

Line charges are useful for simulating the potential of multipoles and of the cathode of field emission electron guns. In the later case, the line charge is placed along the optic axis. The potential φ_{lj} of the line charge j with length $2l_j$ placed parallel to the optic axis at the lateral distance $\vec{\rho}_j$ is given by the Coulomb integral expression

$$\begin{aligned}
 \varphi_{lj} &= \frac{q_j}{8\pi\epsilon_0 l} \int_{-l}^l \frac{dz'}{\sqrt{(z' - z + z_j)^2 + (\vec{\rho} - \vec{\rho}_j)^2}} \\
 &= \frac{q_j}{8\pi\epsilon_0 l} \ln \frac{\sqrt{(z-l-z_j)^2 + (\vec{\rho} - \vec{\rho}_j)^2} + l - z + z_j}{\sqrt{(z+l-z_j)^2 + (\vec{\rho} - \vec{\rho}_j)^2} - l - z + z_j}.
 \end{aligned} \tag{3.161}$$

We have placed the center of the line charge in the plane $z = z_j$. Kasper and coworkers have calculated the potential of numerous rotationally symmetric systems by superposing the potentials of apertures and charged rings. Several results are listed in [15]. The method has also been applied successfully by Preikszas and Rose for optimizing the electric and magnetic objective lens of a low-voltage electron microscope [64].

Chapter 4

Gaussian Optics

Abstract The Gaussian method for describing the imaging of optical lenses in paraxial approximation is applied to charged-particle optics by linearizing the equations of motion. The consequences resulting from the Lagrange–Helmholtz relations are discussed, and Newton’s equation of paraxial image formation is derived for the thin-lens approximation and for the Glaser model field. The imaging properties of cylinder lenses and quadrupole systems are discussed in detail with the impetus on strong focusing, stigmatic imaging, and their usefulness as aberration correctors.

The electromagnetic field forms an *inhomogeneous* and *anisotropic* medium of refraction for charged particles in the most general case. Hence, in the terminology of light optics, electron lenses are *gradient-index* lenses. The anisotropy results from the magnetic force, which depends on the direction of the particle velocity. Therefore, only electrostatic systems have an isotropic index of refraction. Since all realistic electromagnetic fields are inhomogeneous, the equations for the x - and y -coordinates of a trajectory form a system of two coupled nonlinear differential equations. The solutions of such systems are very involved and exhibit chaotic behavior in many cases. The deleterious effect of the nonlinear terms of the forces remains sufficiently small if the particle beam is confined to the vicinity of the axis, which may be straight or curved. One achieves this in practice by means of apertures, which remove particles with large ray gradients from the beam.

Paraxial conditions prevail approximately if the diameter of the beam within the region of the external fields stays smaller than about one fifth of the diameters of the inner faces of the electrodes and/or magnetic pole pieces. In this case, we can describe the propagation of the particles with a sufficient degree of accuracy by neglecting the nonlinear terms in the trajectory equations. The famous mathematician *C.F. Gauss* has first introduced this *paraxial* or *Gaussian* approximation in light optics. His approximation only considers terms up to the second order in the expansion of the variational function (3.60) with respect to the complex coordinates w , $\bar{w}$ and their derivatives. The resulting path equations are

then two complex second-order linear differential equation whose general solutions are linear combinations of four arbitrary linear independent particular solutions of the system. This behavior enables one to describe the optical properties of various elements in a simple way by characteristic quantities such as focal length and principal planes. At the beginning of electron optics, one has explored primarily the paraxial properties and the aberrations of axially symmetric systems [13, 65, 66]. Later, Melkich [67] investigated the paraxial properties of quadrupole lenses, which became important elements of accelerators due to their strong focusing properties [68]. In electron optics, quadrupoles have first been introduced as *stigmators* [69] and later as substitutes for axially symmetric lenses. Kawakatsu et al. developed a quadrupole quadruplet, which they substituted for the projector system of an electron microscope [70]. Bauer designed and tested experimentally a quadrupole objective compound lens consisting of a symmetric quadrupole triplet and an antisymmetric doublet [71]. The experiments proved the feasibility of the lens for paraxial imaging. However, the resolution was rather poor due to the extremely large spherical aberration of the compound lens. In 1947, Scherzer demonstrated that one can eliminate the unavoidable aberrations of round lenses by introducing quadrupoles and octopoles into the system [16]. This finding initiated extensive studies on the properties of quadrupole–octopole systems [72–75]. Hawkes has summarized the results of these investigations in his book on quadrupole optics [76].

4.1 Paraxial Path Equation

Fermat's principle is the most convenient procedure for deriving the equations of the Gaussian or paraxial rays and of the deviations from these *ideal* trajectories. Paraxial conditions provide ideal imaging because, within the frame of validity of this approximation, any rotationally symmetric electromagnetic field yields a perfect image of the object plane $z = z_o$ at a distinct plane $z = z_i$. These two planes are termed *conjugate planes*. Unfortunately, the ideal course of the paraxial rays is limited to a narrow axial domain, which does not suffice in many cases to achieve the required imaging properties. For example, the unavoidable nonlinear terms in the path equations prevent atomic resolution in a conventional 100-kV electron microscope. Although it is not possible to eliminate the deviation of the true path from its paraxial approximation everywhere, it is possible to nullify the resulting aberrations up to a given order at a distinct plane by means of so-called *stigmators* and *correctors*. For reasons of generality, we first derive the general paraxial ray equations for arbitrary systems with a curved optic axis. This axis does not need to be a possible ray. Subsequently, we reduce these ray equations by considering systems with special symmetry.

It is most appropriate to expand the variational function (3.60) in a series of homogeneous polynomials:

$$\mu = \mu(w, \bar{w}, w', \bar{w}'; \kappa; z) = \sum_{n=0}^{\infty} \mu^{(n)}(w, \bar{w}, w', \bar{w}'; \kappa; z). \quad (4.1)$$

Each polynomial $\mu^{(n)}$ comprises all terms of degree n of the power series expansion of μ with respect to $w, \bar{w}, w', \bar{w}'$, and the relative energy deviation of the particle:

$$\kappa = \frac{\Delta E}{E_0} = \frac{\Delta \Phi}{\Phi_0}. \quad (4.2)$$

To obtain a dimensionless *chromaticity* parameter, we normalize the deviation $\Delta E = e\Delta\Phi$ of the particle energy from the mean energy by the nominal energy $E_0 = e\Phi_0$ at a distinct plane $z = z_0$. This plane is the object plane in the case of an electron microscope.

For obtaining the Gaussian approximation, we only need to consider terms $\mu^{(n)} = \mu_e^{(n)} + \mu_m^{(n)}$ up to the second degree inclusively. We readily derive the magnetic parts of the polynomials by inserting the expressions (3.92) and (3.93) for the component $\bar{A}$ and the term $g_3 A_z$, respectively, into the relation (3.62) for μ_m . As a result, we find

$$\begin{aligned} \mu_m^{(0)} &= 0, \quad \mu_m^{(1)} = -\frac{e}{q_0} \text{Im}(\Psi_1 \bar{w}), \\ \mu_m^{(2)} &= \frac{e}{q_0} \text{Im} \left\{ \frac{1}{2} \Psi' w' \bar{w} + \frac{1}{4} \Psi_1 \bar{w} (\Gamma \bar{w} + \bar{\Gamma} w) - \Psi_2 \bar{w}^2 \right\}. \end{aligned} \quad (4.3)$$

The corresponding polynomials of the electric part (3.61) of the variational function are more involved due to the square root. In order to avoid lengthy expressions, we introduce the standard abbreviations

$$\varepsilon = \frac{e}{2E_0} = \frac{e}{2m_e c^2} \approx 1/MV, \quad \gamma_0 = \gamma_0(z) = 1 + 2\varepsilon\Phi = \frac{1}{\sqrt{1 - v_0^2/c^2}}, \quad (4.4)$$

where $v_0 = v(x = 0, y = 0, z)$ is the axial velocity:

$$v_0 = \frac{1}{\gamma_0} \sqrt{\frac{2e}{m_e} \Phi^*}. \quad (4.5)$$

Here $\Phi^* = \varphi^*(x = 0, y = 0, z)$ is the relativistic modified axial potential defined by the expression (2.22), and γ_0 represents the relativistic factor for a particle moving along the optic axis ($w = 0$). Introducing (4.4) in the expression (3.60) for the electric part of the variational function, it takes the form

$$\mu_e = 2\varepsilon^{1/2} \sqrt{\varphi + \varepsilon\varphi^2 + (1 + 2\varepsilon\varphi)\Delta\Phi + \varepsilon(\Delta\Phi)^2} \sqrt{g_3^2 + w'\bar{w}'}. \quad (4.6)$$

We expand the two square roots up to second-order terms inclusively by using the power series expansion (3.54) for φ and the expression (3.13) for the metric coefficient g_3 , giving

$$(g_3^2 + w' \bar{w}')^{1/2} = \left[1 - 2\text{Re}(\Gamma \bar{w}) + \{\text{Re}(\Gamma \bar{w})\}^2 + w' \bar{w}' \right]^{1/2} \approx 1 - \text{Re}(\Gamma \bar{w}) + w' \bar{w}'/2 + \dots, \quad (4.7)$$

$$\begin{aligned} (\varphi + \varepsilon \varphi^2)^{1/2} \approx & \sqrt{\Phi^*} \left\{ 1 + \frac{\gamma_0}{2\Phi^*} \text{Re} \left[\Phi_1 \bar{w} + \Phi_2 \bar{w}^2 - \frac{1}{4}(\Phi'' - \Phi_1 \bar{\Gamma}) w \bar{w} \right]^2 \right\} \\ & - \frac{1}{8\Phi^{*3/2}} [\text{Re}(\Phi_1 \bar{w})] + \frac{\gamma_0 \Phi_o}{2\sqrt{\Phi^*}} \kappa \\ & - \frac{\Phi_o}{4\Phi^{*3/2}} \kappa \text{Re}(\Phi_1 \bar{w}) - \frac{\Phi_o^2}{8\Phi^{*3/2}} \kappa^2 + \dots. \end{aligned} \quad (4.8)$$

Multiplying the two expansions and disregarding higher-order terms, we obtain for the polynomials $\mu_e^{(n)}$, $n = 0, 1, 2$, the expressions:

$$\mu_e^{(0)} = \sqrt{\frac{\Phi^*}{\Phi_0^*}}, \quad \mu_e^{(1)} = \sqrt{\frac{\Phi^*}{\Phi_0^*}} \text{Re} \left\{ \frac{\gamma_0}{2\Phi^*} \Phi_1 \bar{w} - \Gamma \bar{w} \right\} + \sqrt{\frac{\Phi^*}{\Phi_0^*}} \frac{\gamma_0 \Phi_o}{2\Phi^*} \kappa, \quad (4.9)$$

$$\begin{aligned} \mu_e^{(2)} = & \frac{1}{2} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \text{Re} \left\{ w' \bar{w}' - \frac{w \bar{w}}{2} \left[\frac{\Phi'' + \Phi_1 \bar{\Gamma}}{\Phi^*} + \frac{1}{8} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} \right] + \left[\gamma_0 \frac{2\Phi_2 - \Phi_1 \Gamma}{\Phi^*} - \frac{1}{8} \frac{\Phi_1^2}{\Phi^{*2}} \right] \bar{w}^2 \right\} \\ & - \sqrt{\frac{\Phi^*}{\Phi_0^*}} \kappa \text{Re} \left\{ \frac{\Phi_o}{8\Phi^*} \left(\frac{\Phi_1}{\Phi^*} + 2\gamma_0 \Gamma \right) \bar{w} \right\} - \frac{1}{2} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \frac{\Phi_o^2}{\Phi^{*2}} \kappa^2. \end{aligned} \quad (4.10)$$

We derive the paraxial path equation by substituting the Gaussian approximation

$$\mu_G = \sum_{n=0}^2 \mu^{(n)}, \quad \mu^{(n)} = \mu_e^{(n)} + \mu_m^{(n)} \quad (4.11)$$

for the variational function (3.60) in the Euler–Lagrange equation:

$$\frac{d}{dz} \frac{\partial \mu}{\partial \bar{w}'} - \frac{\partial \mu}{\partial \bar{w}} = 0. \quad (4.12)$$

The term $\mu^{(0)}$ does not contribute to the ray equation because it does not depend on the off-axis coordinates. The linear polynomial $\mu^{(1)}$ accounts for the inhomogeneous part and the bilinear term $\mu^{(2)}$ for the homogeneous part and the chromatic inhomogeneous part of the resulting complex paraxial ray equation:

$$\begin{aligned}
w'' + \frac{\gamma_0}{2\Phi^*}(\Phi' - i\nu_0 B)w' + \frac{\gamma_0}{4\Phi^*} \left[\Phi'' - i\nu_0 B' + \text{Re}(\bar{\Gamma}(\Phi_1 + i\nu_0 \Psi_1)) + \frac{\Phi_1 \bar{\Phi}_1}{2\gamma_0 \Phi^*} \right] w \\
- \frac{\gamma_0}{\Phi^*} \left[\Phi_2 + i\nu_0 \Psi_2 - \frac{\Gamma}{4}(2\Phi_1 + i\nu_0 \Psi_1) - \frac{1}{8\gamma_0} \frac{\Phi_1^2}{\Phi^*} \right] \bar{w} \\
= -\frac{\Phi_0}{4\Phi^*} \left[\frac{\Phi_1}{\Phi^*} + 2\gamma_0 \Gamma \right] \kappa + \frac{\gamma_0}{2\Phi^*} (\Phi_1 + i\nu_0 \Psi_1) - \Gamma.
\end{aligned} \tag{4.13}$$

This expression represents the most general relativistic correct Gaussian path equation for the motion of charged particles in arbitrary electromagnetic fields referred to a curved coordinate system with optional complex curvature. Within the nomenclature of electron optics, $B_z(x = 0, y = 0, z) = -\Psi'(z) = B(z)$ denotes the axial component of the magnetic flux density along the optic axis. In order that this axis represents a trajectory, the inhomogeneous term on the right-hand side of (4.13) must vanish:

$$\Gamma - \frac{\gamma_0}{2\Phi^*}(\Phi_1 + i\nu_0 \Psi_1) = 0. \tag{4.14}$$

Only in this case is the optic axis ($w = 0$) a solution of the Gaussian path equation. In the case of a straight optic axis ($\Gamma = 0$), we have

$$\Phi_1 + i\nu_0 \Psi_1 = 0. \tag{4.15}$$

This relation represents in complex notation the well-known *Wien condition* $\vec{E} + \vec{v} \times \vec{B} = 0$ for an axial electron ($w = 0$). The lateral electric force and the lateral magnetic force compensate each other in this case for electrons with nominal velocity $\vec{v} = \vec{v}_0 = v_0 \vec{e}_z$. The Wien filter utilizes this behavior. In order that the straight optic axis forms a trajectory for any velocity, the dipole components Φ_1 and Ψ_1 of the electromagnetic field must both vanish. We assume in the following that the optic axis represents a trajectory. Substituting Γ from (4.14) into (4.13), we find the homogeneous differential equation

$$\begin{aligned}
w'' + \frac{\gamma_0}{2\Phi^*}(\Phi' - i\nu_0 B)w' + \frac{\gamma_0}{4\Phi^*} \left[\Phi'' - i\nu_0 B' + \frac{\Phi_1 \bar{\Phi}_1}{2\gamma_0 \Phi^*} + \frac{\gamma_0}{2\Phi^*} |\Phi_1 + i\nu_0 \Psi_1|^2 \right] w \\
+ \left[\frac{\gamma_0^2}{32} \frac{(3\Phi_1 + 2i\nu_0 \Psi_1)^2}{\Phi^{*2}} + \frac{4 - \gamma_0^2}{32} \frac{\Phi_1^2}{\Phi^{*2}} - \gamma_0 \frac{\Phi_2 + i\nu_0 \Psi_2}{\Phi^*} \right] \bar{w} \\
= -\frac{\Phi_0}{4\Phi^*} \left[(1 + \gamma_0^2) \frac{\Phi_1}{\Phi^*} + i \frac{\gamma_0^2 v_0}{\Phi^*} \Psi_1 \right] \kappa.
\end{aligned} \tag{4.16}$$

This complex Gaussian path equation represents a set of two real inhomogeneous linear second-order differential equations, which are generally coupled with respect

to the off-axis coordinates x and y . The inhomogeneous term on the right-hand side accounts for the *dispersion*. The dispersion vanishes in the case of a straight axis ($\Phi_1 = 0, \Psi_1 = 0$). The structure of (4.16) convincingly demonstrates that the paraxial properties of any electron optical system depend only on the axial, dipole, and quadrupole components of the electromagnetic field with respect to the optic axis. To survey the effect of these components, it is advantageous to investigate the imaging properties of this equation for distinct symmetries of the electromagnetic field. The paraxial equations for w and $\bar{w}$ decouple if the last term on the left-hand side of (4.16) vanishes. This is the case if the dipole and the quadrupole strengths satisfy the relation

$$\Phi_2 + i\nu_0\Psi_2 = \frac{\gamma_0}{32\Phi^*} [(3\Phi_1 + 2i\nu_0\Psi_1)^2 + (4/\gamma_0^2 - 1)\Phi_1^2]. \quad (4.17)$$

Hence, it is possible to achieve in paraxial approximation rotationally symmetric focusing properties by superposing the quadrupole and dipole components appropriately. By considering (4.13) in the special case of a straight optic axis ($\Gamma = 0$), we readily derive the condition for the *stigmatic* Wien filter [77]:

$$\Phi_2 + i\nu_0\Psi_2 = \frac{\Phi_1^2}{8\gamma_0\Phi^*}. \quad (4.18)$$

Another important special case of (4.17) is realized in circular accelerators for a field index $n_f = 1/2$ of the wedge-shaped bending magnets. Their dipole and quadrupole field components satisfy the relation

$$\Psi_2 = i \frac{\gamma_0\nu_0}{8\Phi^*} \Psi_1^2 = \frac{i}{8} \sqrt{\frac{2e}{m_e\Phi^*}} \Psi_1^2. \quad (4.19)$$

This complex relation connects both the strength and the azimuthal orientation α_1 of the dipole component of the magnetic field with those of the quadrupole component. Its azimuthal angle is $\alpha_2 = \alpha_1 + \pi/4$. Within the frame of complex notation, the magnetic field index has the form

$$n_m = -\frac{2}{\Gamma} \frac{\Psi_2}{\Psi_1} = -\frac{1}{\Gamma} \frac{\partial^2 \psi}{\partial \bar{w}^2} \bigg/ \frac{\partial \psi}{\partial \bar{w}} \bigg|_{\bar{w}=0}. \quad (4.20)$$

In the special case of a plane circular axis, this definition coincides with that used in the context of accelerators. Within the frame of our definition, the field index may become complex. Although the path equation is decoupled with respect to w and $\bar{w}$ if the field components satisfy (4.17), this does *not* imply that the real part and the imaginary part of the equation are decoupled with respect to the variables x and y . This behavior results from the complex coefficients of the differential equation (4.13) in the case $B \neq 0$. However, we can obtain a differential equation

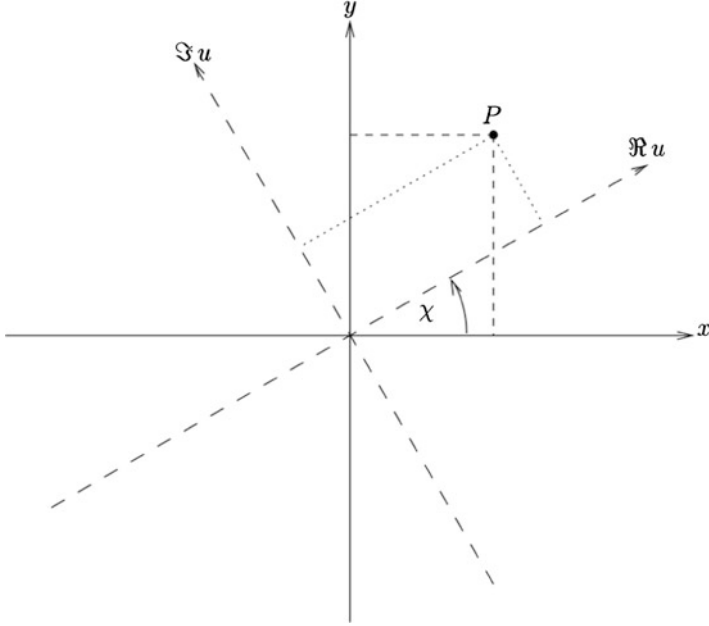


Fig. 4.1 Interrelationship between the fixed x, y -coordinate system and the rotating u_r, u_i -coordinate system ($u_r = \text{Re} u, u_i = \text{Im} u$)

with real coefficients by introducing a rotating coordinate system, which rotates with the angle $\chi = \chi(z)$ about the optic axis.

As illustrated in Fig. 4.1, we obtain the components of the new complex u -coordinate by the transformation

$$u = \text{Re} u + i \text{Im} u = w e^{-i\chi} = x \cos \chi + y \sin \chi - i[x \sin \chi - y \cos \chi]. \quad (4.21)$$

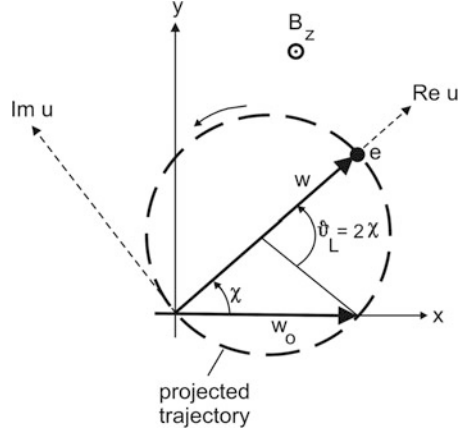
The rotation angle is determined by the condition that the coefficients of the paraxial ray equation in the u, z -coordinate system are real. This is possible despite the fact that two coefficients of the initial equation are complex. In order to prove this behavior, we differentiate $w = u \exp(i\chi)$ with respect to z , giving

$$\begin{aligned} w' &= (u + i\chi' u) e^{i\chi}, \\ w'' &= [u'' + 2i\chi' u' + (i\chi'' - \chi'^2)u] e^{i\chi}. \end{aligned} \quad (4.22)$$

Substituting w' and w'' from (4.22) into (4.13), we find that the imaginary parts of the resulting coefficients for u' and u vanish, if we select

$$\chi' = \frac{\hbar_0 v_0}{4\Phi^*} B = \sqrt{\frac{e}{8m_e \Phi^*}} B = \frac{1}{2} \frac{eB}{m v_0} = \frac{1}{2} \tau_L(z, w = 0) = \frac{1}{2} \vartheta_L'(z). \quad (4.23)$$

Fig. 4.2 Connection between the rotation angle χ of the rotating u -coordinate system and the angle ϑ_L of the Larmor rotation



Therefore, we get the surprising result that the twist angle

$$\chi(z) = \sqrt{\frac{e}{8m_e}} \int_{z_0}^z \frac{B}{\sqrt{\Phi^*}} dz = \frac{1}{2} \vartheta_L(z) \quad (4.24)$$

of the rotating coordinate system is *half* the angle of axial Larmor rotation. To elucidate this apparent discrepancy, we consider the motion of a particle in a homogeneous axial magnetic field. Then the trajectory of the particle is a helix. The projection of the helix onto the $x - y$ plane forms a circle, as shown in Fig. 4.2 for a particle intersecting the optic axis.

A particle, which intersects the optic axis, propagates in the u, z -coordinate system in a *plane* section, while this section is *twisted* in the fixed x, y, z -coordinate system. The particle is initially at the position $w_0 = w(z_0)$. Since the Larmor rotation of the particle refers to the center of the circle, the associated angle of Larmor rotation ϑ_L is *twice* as large as the twist angle χ of the rotating coordinate system. Surprisingly, this peculiar behavior is unnoticed in the literature.

By employing the relation (4.23) together with the transformations (4.21) and (4.22), the paraxial path equation (4.16) takes the form

$$u'' + \frac{\gamma_0 \Phi'}{2\Phi^*} u' + \left[\frac{\gamma_0 \Phi''}{4\Phi^*} + \frac{e}{8m_e} \frac{B^2}{\Phi^*} + \frac{1}{8} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} + \frac{1}{8} \left| \frac{\gamma_0 \Phi_1}{\Phi^*} + i \sqrt{\frac{2e}{m_e \Phi^*}} \Psi_1 \right|^2 \right] u - G \bar{u} = D\kappa, \quad (4.25)$$

$$G = \left[\frac{\gamma_0 \Phi_2}{\Phi^*} + i \sqrt{\frac{2e}{m_e \Phi^*}} \Psi_2 - \frac{1}{32} \left(\frac{3\gamma_0 \Phi_1}{\Phi^*} + 2i \sqrt{\frac{2e}{m_e \Phi^*}} \Psi_1 \right)^2 - \frac{4 - \gamma_0^2}{32} \frac{\Phi_1^2}{\Phi^{*2}} \right] e^{-2i\chi}, \quad (4.26)$$

$$D = -\frac{\Phi_0}{4\Phi^*} \left[(1 + \gamma_0^2) \frac{\Phi_1}{\Phi^*} + i \sqrt{\frac{2e}{m_e \Phi^*}} \Psi_1 \right] e^{-i\chi}. \quad (4.27)$$

The transformation to the rotating coordinate system is extremely helpful in the case $G = 0$ because the two remaining coefficients of the homogeneous part of the differential equation are then real functions of z . Moreover, the transformation has removed the magnetic induction B from the coefficient of u' . One knows from the theory of linear differential equations that it is possible to remove the term with the second highest derivative by means of a proper transformation of the variable. For the paraxial path equation (4.25), we employ the simple transformation

$$u = \sqrt[4]{\frac{\Phi_o^*}{\Phi^*}} U = \left(\frac{\Phi_o^*}{\Phi^*} \right)^{1/4} e^{-i\chi} (X + iY). \quad (4.28)$$

The resulting transformed path equation is

$$U'' + TU - G\bar{U} = \sqrt[4]{\frac{\Phi^*}{\Phi_o^*}} D\kappa. \quad (4.29)$$

In order that U has the same dimension as u , we have normalized in (4.28) the relativistic modified axial potential Φ^* by its value Φ_o^* at the object plane $z = z_o$. We can separate the complex differential equation in two decoupled real second-order differential equations only if $(G/\bar{G})' = 0$. The rotationally symmetric *focusing power*

$$T(z) = \bar{T}(z) = \frac{2 + \gamma_o^2}{16} \frac{\Phi^2}{\Phi^{*2}} + \frac{eB^2}{8m_e\Phi^*} + \frac{1}{8} \left| \frac{\Phi_1}{\Phi^*} \right|^2 + \frac{1}{8} \left| \frac{\gamma_o\Phi_1}{\Phi^*} + i \sqrt{\frac{2e}{m_e\Phi^*}} \Psi_1 \right|^2 \quad (4.30)$$

is real and positive definite regardless if the axis is curved or straight. Therefore, the curvature U'' and the lateral distance U of the transformed trajectory have always opposite sign in the case of paraxial rotational symmetry ($G = 0$). Hence, a short rotationally symmetric field acts always as a *convex* lens on the paraxial rays, as depicted in Fig. 4.3a.

In order to obtain a divergent lens, the trajectory must intersect the optic axis at least once within the extended field of the “*thick*” lens (Fig. 4.3b). Contrary to light optics, *short concave* electron einzel lenses do not exist. This behavior is a direct consequence of the constraint imposed on the index of refraction by the Laplace equation for the electromagnetic potentials. As a result, the spatial distribution of the electron optical refraction index is not adjustable arbitrarily as in light optics. A round electron lens is “*short*” if the extension l of its field satisfies the condition

$$T_{\max} l^2 < 1. \quad (4.31)$$

It is always possible to subdivide a thick lens in a sequence of short lenses.

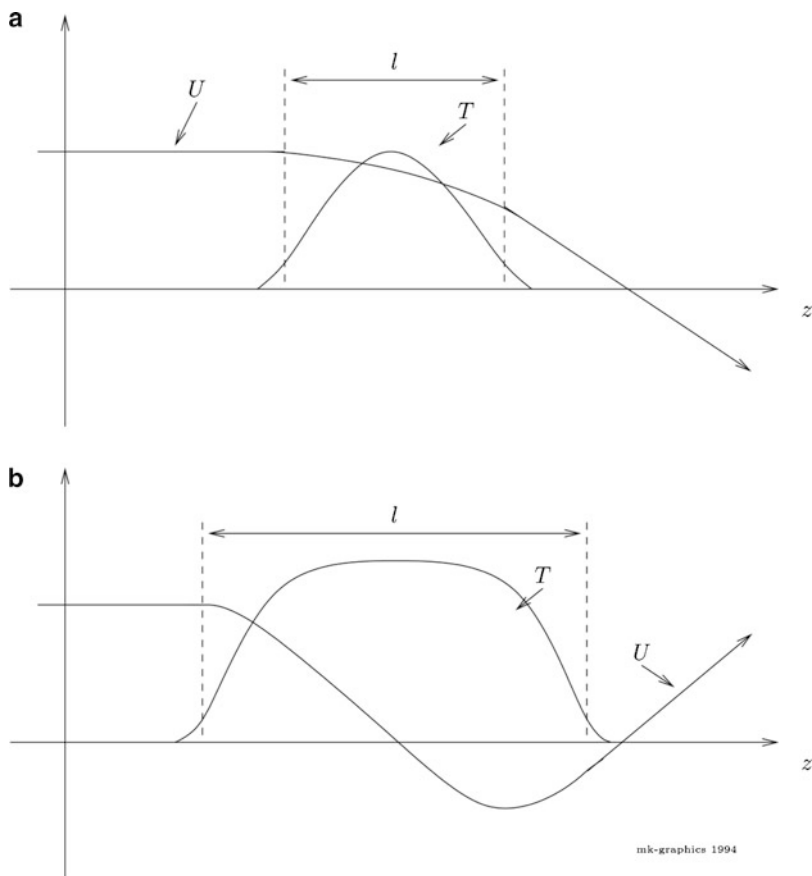


Fig. 4.3 Illustration of (a) the converging focusing effect of a short charged-particle lens ($T_{\max} l^2 \ll 1$) and (b) the diverging effect of an extended (thick) lens ($T_{\max} l^2 \geq 1$)

A “*thin*” lens is a lens with a very short axial extension such that the trajectory changes its direction but not appreciably its off-axis distance within the domain of the lens fields.

4.2 Orthogonal Systems with Midsection Symmetry

Midsection symmetry is widely employed in systems with a curved axis. Typical examples are accelerators, storage rings, spectrometers used in particle physics, and the beam separators and imaging energy filters incorporated in electron

microscopes. The complex path equation (4.16) separates in two decoupled real equations for systems with *midsection* symmetry

$$\varphi(x, -y, z) = \varphi(x, y, z), \quad \psi(x, -y, z) = -\psi(x, y, z). \quad (4.32)$$

These conditions are satisfied for

$$\Phi_v = \overline{\Phi}_v = \Phi_{vc}, \quad \Psi_v = -\overline{\Psi}_v = i\Psi_{vs}, \quad B = 0. \quad (4.33)$$

In this case, the optic axis is located on the plane midsection $y = 0$. The electric potential φ is symmetric with respect to this section, while the scalar magnetic potential is antisymmetric. Inserting (4.33) into (4.16), we find

$$\begin{aligned} x'' + \frac{\gamma_0 \Phi'}{2\Phi^*} x' + \left[\frac{\gamma_0 \Phi''}{4\Phi^*} + \frac{3\gamma_0^2 + 2}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} + \eta^2 \frac{\Psi_{1s}^2}{\Phi^*} - \frac{5\gamma_0 \Phi_{1c}}{4\Phi^*} \frac{\eta \Psi_{1s}}{\sqrt{\Phi^*}} - \gamma_0 \frac{\Phi_{2c}}{\Phi^*} + 2\eta \frac{\Psi_{2s}}{\sqrt{\Phi^*}} \right] x \\ = -\frac{\Phi_0}{\Phi^*} \left[\frac{1 + \gamma_0^2}{4} \frac{\Phi_{1c}}{\Phi^*} - \frac{\gamma_0}{2} \eta \frac{\Psi_{1s}}{\sqrt{\Phi^*}} \right] \kappa, \end{aligned} \quad (4.34)$$

$$y'' + \frac{\gamma_0 \Phi'}{2\Phi^*} y' + \left[\frac{\gamma_0 \Phi''}{4\Phi^*} - \frac{\gamma_0^2}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} + \frac{\gamma_0 \Phi_{1c}}{4\Phi^*} \frac{\eta \Psi_{1s}}{\sqrt{\Phi^*}} + \gamma_0 \frac{\Phi_{2c}}{\Phi^*} - 2\eta \frac{\Psi_{2s}}{\sqrt{\Phi^*}} \right] y = 0, \quad \eta = \sqrt{\frac{e}{2m_e}}. \quad (4.35)$$

The dispersion term on the right-hand side of (4.34) accounts for the *chromaticity* in first approximation. The dispersion term is of first degree in the chromatic parameter and of zero order with respect to the geometrical position coordinates x , y and their derivatives. The dispersion is not equivalent to the primary chromatic aberration of an electron lens. This aberration is of second *rank*.

In dispersive systems, the coupling between geometric and chromatic effects is very strong. Therefore, it is important to adopt an unambiguous terminology that prevents confusion. We define the *order* n of an aberration as the sum of the exponents of the geometric ray parameters (typically object coordinates and slope components). The exponent of the chromatic parameter is called the *degree* of the aberration. The sum of *order* + *degree* defines the *rank*. This terminology is widely used in electron optics in the context of imaging energy filters and monochromators. For example, the dispersion is a first-rank aberration, whereas the primary chromatic aberration of a lens is a second-rank aberration because it is of first order and of first degree. The rank is a measure for the magnitude of an aberration. The larger the rank, the smaller is the influence of the corresponding aberration.

Purely magnetic systems with midsection symmetry are realized predominantly in practice. In this case, we have $\Phi' = \Phi'' = 0$, $\Phi_v = 0$. Then the paraxial path equations (3.34) and (3.45) reduce considerably, to give

$$\begin{aligned}
x'' + \left(\eta^2 \frac{\Psi_{1s}^2}{\Phi^*} + 2\eta \frac{\Psi_{2s}}{\sqrt{\Phi^*}} \right) x &= \kappa^* \frac{\eta \Psi_{1s}}{2\sqrt{\Phi^*}}, \\
y'' - 2\eta \frac{\Psi_{2s}}{\sqrt{\Phi^*}} y &= 0, \\
\kappa^* &= \frac{2\gamma_0}{1 + \gamma_0} \kappa = 2 \frac{\Delta p_{\text{kin}}}{p_{\text{kin}}}.
\end{aligned} \tag{4.36}$$

Here κ^* is the relativistic modified chromatic parameter. The reduced equations (4.36) represent the basic equations for the paraxial trajectories in circular accelerators. Replacing $\eta \Psi_{1s}/\Phi^{*1/2} = -\Gamma = -1/R$ by the radius of curvature $R = R(z)$ and substituting $k_2 = k_2(z)$ for $-2\eta \Psi_{2s}/\Phi^{*1/2}$, the equations adopt the familiar form used in accelerator physics:

$$\begin{aligned}
x'' + \left[\frac{1}{R^2} - k_2 \right] x &= -\frac{1}{R} \frac{\Delta p_{\text{kin}}}{p_{\text{kin}}}, \\
y'' + k_2 y &= 0.
\end{aligned} \tag{4.37}$$

The same set of equations holds for the Gaussian optics of energy filters, magnetic monochromators and beam separators [78–81]. The minus sign on the right-hand side of the first equation accounts for the fact that the direction of the x -axis points *toward* the center of curvature. Accordingly, a particle with a somewhat larger kinetic momentum ($\Delta p_{\text{kin}} > 0$) will be deflected away from the optic axis in the negative x -direction.

Conical sector magnets with tapered pole faces form superposed dipole and quadrupole fields. The inner faces of these magnets are equipotentials of the form

$$\psi = \Psi_{1s}y + 2\Psi_{2s}xy + \Psi_{3s}y(3x^2 - y^2) + (2\Gamma\Psi_{2s} - \Psi_{1s}'')(x^2 + y^2)y/8 + \dots = \text{const}. \tag{4.38}$$

The multipole strengths are constant within the inner region of the multipoles. In order that these strengths relate to each other in a given way, it is necessary to shape the pole faces appropriately. To produce a magnetic field with a given field index (4.20) in the paraxial domain $\rho \ll R$, the cross sections of the pole faces must be hyperbolas of the form

$$\frac{y}{R} \left(1 - n_m \frac{x}{R} \right) = \pm \frac{\psi}{\Psi_{1s}R} = \pm \text{const}. \tag{4.39}$$

The *apex* of these “skew” hyperbolas is located on the x -axis at the position $x_a = n_m R$.

Other important examples of systems with midsection symmetry are magnets with plane-parallel inner surfaces. These magnets exert a quadrupole action in the region of their fringing fields if the optic axis is inclined with respect to the normal of the *isoinduction* lines $B_y(x = 0, y = 0, z) = -\Psi_{1s}(z)$ by the angle $\phi = \phi(z) \neq 0$.

The strength of the fringe-field quadrupole

$$\Psi_{2s} = -\frac{1}{2}\Psi'_{1s}(z) \tan \phi(z) \quad (4.40)$$

is proportional to the derivative of the dipole strength Ψ_{1s} and the tangent of the local angle $\phi(z)$ enclosed by the direction of the optic axis and the normal to the magnetic induction lines. Accordingly, the quadrupole component of the fringing field with respect to the optic axis vanishes only if this axis is perpendicular to the isoinduction lines at any point within the domain of the fringing fields. Imaging energy filters and beam separators are important examples of systems composed of homogeneous magnets.

Monochromators and electric energy analyzers are examples of purely electric systems ($B = 0, \Psi_v = 0$) with midsection symmetry ($\Phi_v = \bar{\Phi}_v = \Phi_{vc}$). In these systems, the electrodes are centered about the plane symmetry section $y = 0$, which embeds the curved optic axis. We readily derive the paraxial path equations for these systems from (4.34) and (4.35) as

$$\begin{aligned} x'' + \frac{\gamma_0}{2} \frac{\Phi'}{\Phi^*} x' + \left[\frac{\gamma_0}{4} \frac{\Phi''}{\Phi^*} + \frac{3\gamma_0^2 + 2}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} - \gamma_0 \frac{\Phi_{2c}}{\Phi^*} \right] x &= -\frac{1 + \gamma_0^2}{4} \frac{\Phi_o}{\Phi^*} \frac{\Phi_{1c}}{\Phi^*} \kappa, \\ y'' + \frac{\gamma_0}{2} \frac{\Phi'}{\Phi^*} y' + \left[\frac{\gamma_0}{4} \frac{\Phi''}{\Phi^*} - \frac{\gamma_0^2}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} + \gamma_0 \frac{\Phi_{2c}}{\Phi^*} \right] y &= 0. \end{aligned} \quad (4.41)$$

In the special case of a spherical analyzer or monochromator [82], the axial potential is constant ($\Phi' = \Phi'' = 0, \Phi = \Phi_o$) and the electric quadrupole and dipole strengths are interrelated:

$$\Phi_{2c} = \frac{3}{4} \Gamma_0 \Phi_{1c}. \quad (4.42)$$

Here Γ_0 is the curvature of the chosen circular axis. If we require that this axis represents the path of a particle, its curvature

$$\Gamma = \frac{\gamma_0}{2} \frac{\Phi_{1c}}{\Phi^*} \quad (4.43)$$

must coincide with Γ_0 , giving

$$\Phi_{2c} = \frac{3\gamma_0}{8} \frac{\Phi_{1c}}{\Phi^*}. \quad (4.44)$$

Substituting (4.44) for Φ_{2c} in (4.41), these equations take the form

$$\begin{aligned} x'' + \frac{1}{4} \frac{\Phi_{1c}^2}{\Phi^{*2}} x &= -\frac{1}{2} \frac{1 + \gamma_0^2}{1 + \gamma_0} \frac{\Phi_{1c}}{\Phi^*} \kappa, \\ y'' + \frac{\gamma_0^2}{4} \frac{\Phi_{1c}^2}{\Phi^{*2}} y &= 0. \end{aligned} \quad (4.45)$$

The comparison of the two equations reveals that precise rotationally symmetric focusing prevails only in the nonrelativistic limit $\gamma_0 \rightarrow 1$. Fortunately, the electric field index

$$n_e = \frac{1}{\Gamma} \frac{\Phi_{2c}}{\Phi_{1c}} \quad (4.46)$$

can be adjusted by employing toroidal electrodes whose center of vertical curvature may differ from that of the horizontal curvature. By employing toroidal sector electrodes, it is possible to adjust the vertical and the horizontal focusing properties in a given way. One has utilized this possibility in the design of optimum electrostatic monochromators for quasi-monochromatic electron guns [27]. We shall further examine the paraxial properties of systems with curved axis after we have investigated the solutions of the Gaussian path equation for systems with a straight optic axis.

4.3 Systems with a Straight Optic Axis

Systems with straight optic axis are composed of elements, which do not introduce dipole fields. Each of these elements has a straight axis of symmetry, which represents a possible trajectory. By aligning the axes of these elements along a common straight line, we obtain a system with a straight optic axis. Within the frame of Gaussian optics, such systems are composed of round lenses and quadrupoles. The electron microscope is the most important example of a system with a straight optic axis. Since conventional electron microscopes are solely composed of round lenses, apart from the stigmator compensating for misalignment errors, we first investigate the electron optical properties of these lenses.

4.3.1 Systems with an Axis of Rotational Symmetry

Systems with rotational symmetry are composed of electrostatic and magnetic round lenses. The most common electrostatic lens is the *einzel* lens or *unipotential* lens, which consists of three circular electrodes centered on a common axis. The two outer electrodes are at earth potential, whereas the central electrode is either at a higher or at a lower potential. In the former case, the lens forms an *accelerating* einzel lens and in the latter case, a *retarding* einzel lens. If the constant potential in front of the lens differs from that behind it, the lens is termed *immersion lens* in analogy to light microscopy where the object one immerses in a liquid with a high index of refraction. One uses retarding immersion lenses predominantly as objective lenses in low-voltage scanning electron microscopes. In order to reduce beam broadening resulting from stochastic Coulomb interactions

between the probe-forming electrons, one uses voltages of about 10 kV within the column. The electrons are decelerated within the last lens close to the object. The objective lens of a photoemission electron microscope is a typical example of an accelerating immersion lens. The properties of axially symmetric electrostatic lenses are discussed extensively in books by Heddle [83], Harting and Read [84], and in the review article by Baranova and Yavor [85].

A round magnetic lens consists of a solenoid, usually enclosed in a rotational symmetric iron casing with a narrow gap in order to confine the axial extension of the magnetic field. Strong electromagnetic fields with short axial extension act on charged particles like glass lenses on light rays. This is the reason for terming devices, which produce such fields, as charged-particle lenses.

In the case of rotational symmetry, all multipole components of the electric and the magnetic potentials vanish except the terms with multiplicity $\nu = 0$:

$$\Phi_\nu = 0, \quad \Psi_\nu = 0, \quad \nu = 1, 2, \dots \quad (4.47)$$

With these constraints, the general path equation (4.25) collapses to the rather simple equation:

$$u'' + \frac{\gamma_0}{2} \frac{\Phi'}{\Phi^*} u' + \left(\frac{\gamma_0}{4} \frac{\Phi''}{\Phi^*} + \frac{e}{8m_e} \frac{B^2}{\Phi^*} \right) u = 0. \quad (4.48)$$

Since the coefficients of this equation are real, particular solutions are the same for both the real part and the imaginary part of the complex off-axis coordinate u . Therefore, it suffices to find two linearly independent real solutions $u_1(z) = \bar{u}_1(z)$ and $u_2(z) = \bar{u}_2(z)$ of the second-order linear differential equation (4.48) for forming the most general solution:

$$u(z) = C_1 u_1(z) + C_2 u_2(z). \quad (4.49)$$

The *complex* constants C_1 and C_2 define the trajectory. Hence, *four* real parameters are necessary and sufficient for describing a distinct trajectory. These parameters are usually the coordinates x_0 , y_0 and the slope components x'_0 , y'_0 of the ray at the object plane z_0 . For convenience, we assume that the angle χ of the rotating coordinate system (4.21) is zero at the object plane ($\chi(z = z_0) = 0$), giving

$$\begin{aligned} u_0 &= u(z_0) = w_0 = x_0 + iy_0, \\ u'_0 &= u'(z_0) = w'_0 - i\chi'_0 w_0 = x'_0 + iy'_0 - i\chi'_0 w_0. \end{aligned} \quad (4.50)$$

The derivative χ'_0 of the twist angle (4.23) at the object plane is nonzero in the presence of a magnetic field. Although analytical solutions of the paraxial ray equation (4.48) are available only in a few simple cases, it is possible to find important general paraxial properties from the solution (4.49) and the interrelation,

which exists between any two linearly independent solutions $u_1(z)$, $u_2(z)$ and their derivatives. The most important results concern *Busch's theorem* and the “*theorem of alternating images*”. The theorem of Busch states as follows: “*Each rotationally symmetric electromagnetic field acts in paraxial approximation as a converging lens forming a distortion-free stigmatic image.*” The theorem of alternating images accounts for the fact that an image of the source is located in the region between any two subsequent images of the object plane and vice versa.

In order to easily demonstrate stigmatic image formation, we choose the *axial ray* $u_\varepsilon = u_\alpha(z)$ and the *principal ray* $u_\pi = u_\pi(z)$ as the pair of linearly independent solutions u_1 and u_2 . These *fundamental* trajectories satisfy at the object plane the initial conditions

$$\begin{aligned} u_\alpha(z_0) = u_{\alpha 0} = 0, \quad u'_\alpha(z_0) = u'_{\alpha 0} = 1, \\ u_\pi(z_0) = u_{\pi 0} = 1, \quad u'_\pi(z_0) = u'_{\pi 0} = 0. \end{aligned} \quad (4.51)$$

By choosing these rays as the pair of linearly independent particular solutions u_1 and u_2 , we obtain for the trajectory of a particle, which intersects the object plane at the point w_0 with slope w'_0 , the solution

$$u = u'_0 u_\alpha(z) + u_0 u_\pi(z). \quad (4.52)$$

The course of this trajectory and that of fundamental rays u_α and u_π is schematically shown in Fig. 4.4 for different slopes $u'_0 = u'_{0i}$, $i = 1, 2, 3, 4$. If the electromagnetic field is sufficiently strong, the slope of the axial ray reverses its sign. As a result, this ray intersects the optic axis again at some plane $z = z_i$. Moreover, the various members of the pencil of rays emanating from any point u_0 at the object plane z_0 intersect each other again in the point

$$u_i = u(z_i) = u_\pi(z_i)u_0 = Mu_0. \quad (4.53)$$

Since this relation holds true for all points of the object plane, a stigmatic image of this plane will be formed at the plane $z = z_i$. The magnification $M = u_{\pi i}$ is referred to the u, z -coordinate system. In this coordinate system, M is negative if the axial ray intersects the optic axis $2N$ -times between z_0 and the final image plane z_i . Intermediate images are then located at the planes of intersection. The magnification is *positive* if an odd number of intermediate images are located between z_0 and z_i . If we return to the fixed w, z -Cartesian system, the image is rotated with respect to the object by the twist angle χ_i in the presence of a magnetic field. Negative magnification implies that the image of the object is “upside down” within the frame of the rotating coordinate system.

If the lens is too weak to bend the axial ray back toward the axis, a real image does not exist. However, we can then define a *virtual image* in analogy to light optics. The intersection of the asymptote of the axial ray with the optic axis defines the location

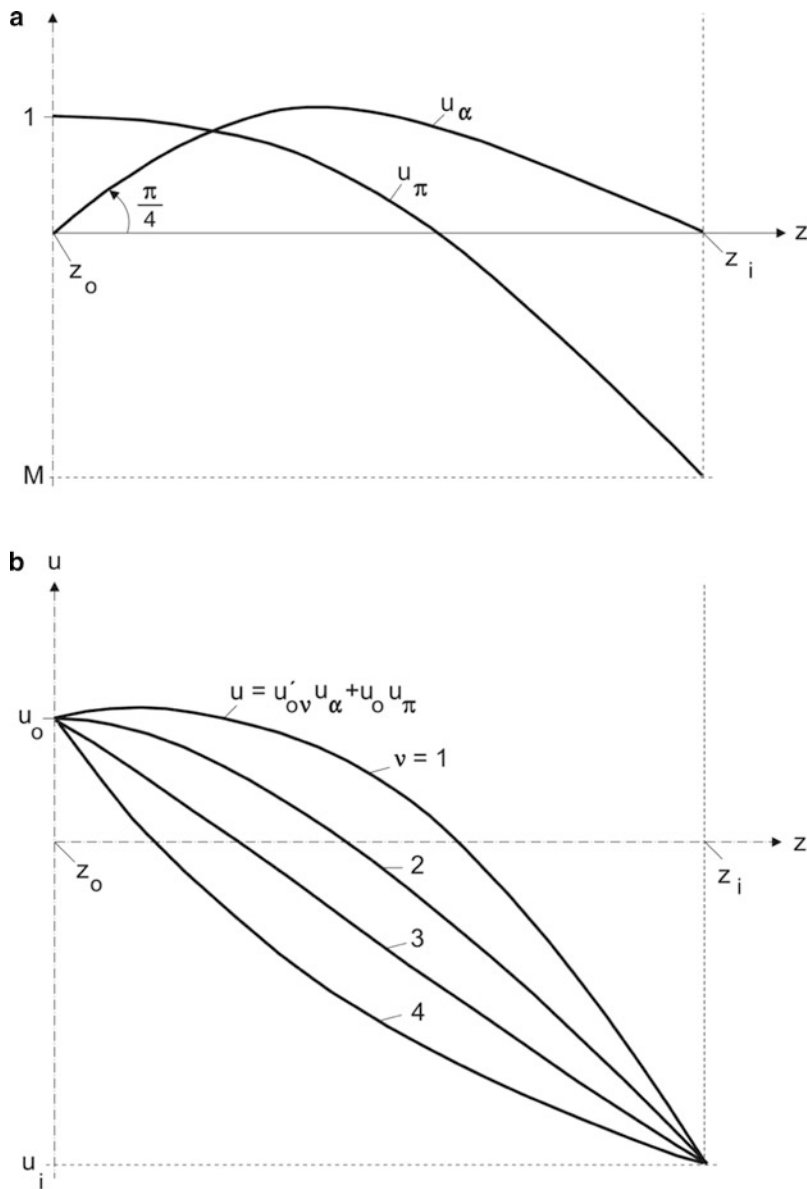


Fig. 4.4 (a) Course of the axial ray u_α and the principal ray u_π between the object plane z_o and the image plane z_i ($B \rightarrow i$) and (b) course of the trajectories of a pencil of rays in the case of stigmatic imaging

z_i of the virtual image plane, as illustrated in Fig. 4.5. The point of intersection of the asymptote of u_π with this plane gives the magnification of the virtual image. Accordingly, we can say that the object plane z_o will be virtually imaged in the plane $z = z_i$. Hence, a round electron lens forms a stigmatic image for any location

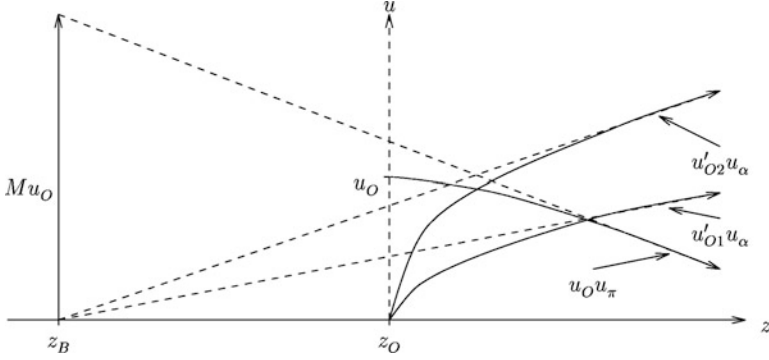


Fig. 4.5 Virtual stigmatic imaging in the case of a weak round lens

of the object plane in front of the lens, regardless of its strength. This finding is a direct consequence of Busch's theorem.

4.3.2 Wronski Determinant

The eikonal connects the position and slopes of any two trajectories with each other. The *Lagrange invariants* describe this connection. They collapse to a single invariant in the case of rotational symmetry. In Gaussian approximation, this invariant is the *Wronski determinant* or *Wronskian* of the paraxial path equation (4.47). In order to derive this determinant, we consider two linearly independent solutions u_1 and u_2 of this equation, so that

$$\begin{aligned}\sqrt{\Phi^*} \frac{d}{dz} (\sqrt{\Phi^*} u'_1) + \left(\frac{\gamma_0}{4} \Phi'' + \frac{e}{8m_e} B^2 \right) u_1 &= 0, \\ \sqrt{\Phi^*} \frac{d}{dz} (\sqrt{\Phi^*} u'_2) + \left(\frac{\gamma_0}{4} \Phi'' + \frac{e}{8m_e} B^2 \right) u_2 &= 0.\end{aligned}\quad (4.54)$$

Multiplying the first equation by u_2 and the second by u_1 and subtracting yields

$$\sqrt{\Phi^*} \left\{ u_1 \frac{d}{dz} (\sqrt{\Phi^*} u'_2) - u_2 \frac{d}{dz} (\sqrt{\Phi^*} u'_1) \right\} = \sqrt{\Phi^*} \frac{d}{dz} [\sqrt{\Phi^*} (u_1 u'_2 - u_2 u'_1)] = 0. \quad (4.55)$$

Integration of the second differential readily gives the Wronskian:

$$\sqrt{\Phi^*} (u_1 u'_2 - u_2 u'_1) = \text{const.} \quad (4.56)$$

Several important optical laws follow from this invariant.

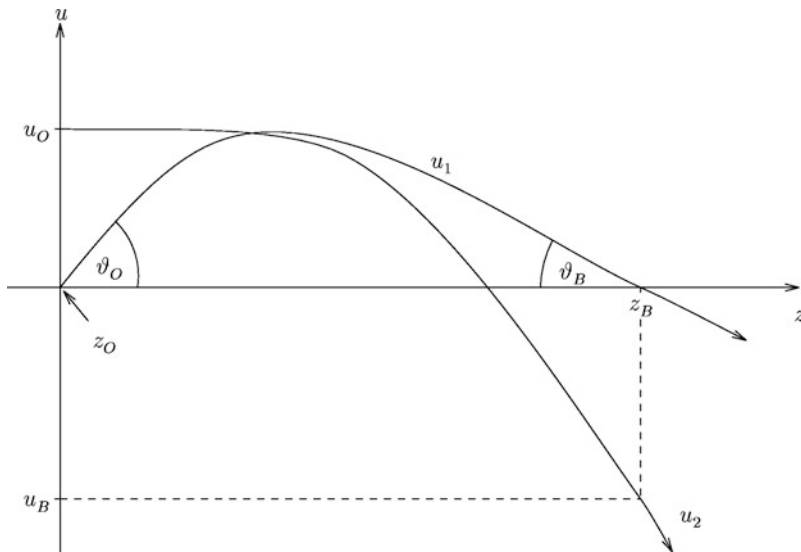


Fig. 4.6 Paraxial rays u_1 and u_2 used for deriving the Lagrange–Helmholtz relation of light optics

4.3.3 Lagrange–Helmholtz Relation

The Lagrange–Helmholtz relation of light optics connects the magnification of a lens with the slope components of the ray u_1 , which intersects the optic axis at the object and image planes. Within the frame of validity of Gaussian approximation, we fix the slopes of this ray at the planes z_o and z_i , respectively, by the angles $u'_1(z_o) = u'_{1o} \approx \vartheta_o$ and $u'_{1i} \approx \vartheta_i$, as shown in Fig. 4.6.

The magnification $M = u_i/u_o = u_{2i}/u_{2o}$ is determined by the ray u_2 , which interests the object and image planes at the points u_o and u_i , respectively. By evaluating the Wronskian of the two trajectories at these planes, we find

$$\sqrt{\Phi_o^*} u_{2o} u'_{1o} = \sqrt{\Phi_i^*} u_{2i} u'_{1i}. \quad (4.57)$$

Considering that $\Phi^{*1/2}$ is proportional to the electron optical index of refraction n_0 on the optic axis (2.50), we may rewrite (4.57) in the form of light optics as

$$n_{0i} M \vartheta_i = n_{0o} \vartheta_o. \quad (4.58)$$

This relation represents the Lagrange–Helmholtz formula. This formula proves that the magnification in paraxial approximation is entirely defined by the slope of the axial ray and the index of refraction at the object and image plane.

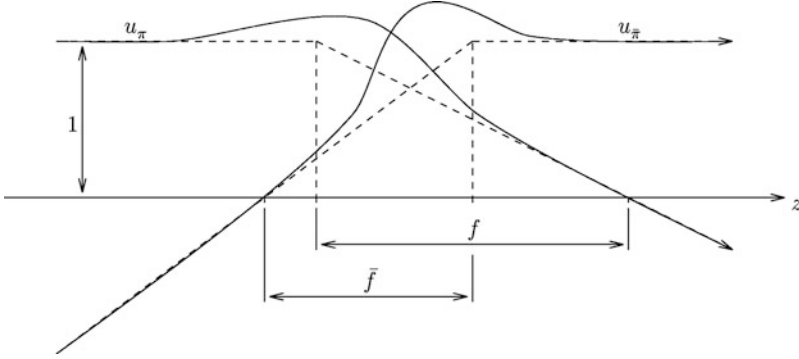


Fig. 4.7 Principal rays u_π and $u_{\bar{\pi}}$ defining the focal lengths f and $\bar{f}$ of a confined rotationally symmetric electromagnetic field

The relation (4.58) is only valid within the frame of Gaussian approximation, which implies small slopes $\tan \vartheta \approx \vartheta$. Using the tangent, which is the case in many textbooks of light optics, violates the assumptions of Gaussian approximation. We derive another important relation from (4.56) by choosing the *principal* rays u_π and $u_{\bar{\pi}}$ as the two linearly independent solutions. These rays satisfy the relations

$$\begin{aligned} u_\pi(z = -\infty) &= 1, & u'_\pi(-\infty) &= 0, & u'_\pi(\infty) &= -1/f, \\ u_{\bar{\pi}}(z = \infty) &= 1, & u'_{\bar{\pi}}(\infty) &= 0, & u'_{\bar{\pi}}(-\infty) &= 1/\bar{f}. \end{aligned} \quad (4.59)$$

Here we have adopted the conventional notation of light optics, where the bar refers to the cardinal elements of the lens located on the object side. The *object principal ray* $u_{\bar{\pi}}$ defines the *object* focal length $\bar{f}$, while the *image principal ray* u_π defines the *image* focal length f , as illustrated in Fig. 4.7. Considering (4.59), the Wronski determinant for these two rays taken at $z = -\infty$ and $z = \infty$ give

$$-\frac{u'_\pi(\infty)}{u'_\pi(-\infty)} \frac{u_{\bar{\pi}}(\infty)}{u_{\bar{\pi}}(-\infty)} = \frac{\bar{f}}{f} = \sqrt{\frac{\Phi_{-\infty}^*}{\Phi_\infty^*}} = \frac{n_{0,-\infty}}{n_{0,\infty}}. \quad (4.60)$$

Accordingly, the two conjugate focal lengths of an einzel lens coincide, while they differ for an immersion lens, as it is the case for the light-optical equivalents.

4.3.4 Theorem of Alternating Images

Within a multistage system, such as the electron microscope, each plane will be imaged repeatedly. As an example, we consider the formation of the images of two apertures A and C located at planes $z = z_\alpha$ and $z = z_\gamma$, as depicted in Fig. 4.8. Typical locations are the object plane and the back focal plane of a lens. This plane

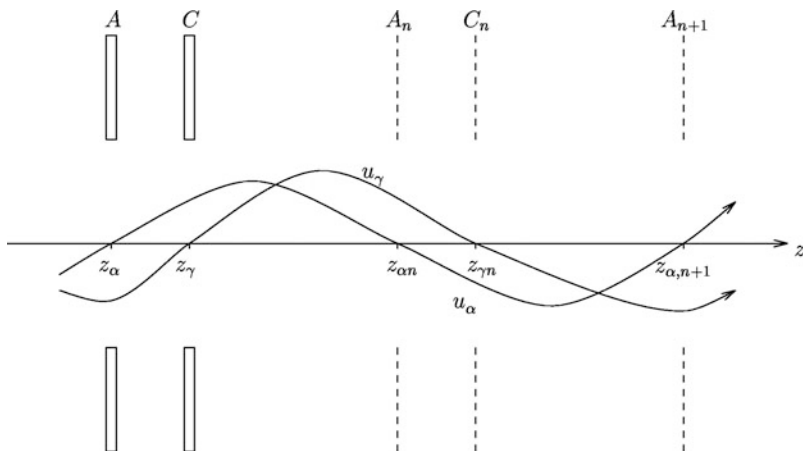


Fig. 4.8 Theorem of alternating images

is an image plane of the source for parallel illumination. Without loss of generality, we center one aperture at the object plane $z_\alpha = z_o$ and the other at an image $z_\gamma = z_s$ of the *effective* source. As the pair of linearly independent trajectories, we select the fundamental rays

$$u_1 = u_\alpha, \quad u_\alpha(z_\alpha) = 0, \quad u_2 = u_\gamma, \quad u_\gamma(z_\gamma) = 0, \quad (4.61)$$

where u_α and u_γ intersect the optic axis at the center of the aperture A and C, respectively. The aperture A is imaged in the planes $z_{\alpha n}$ and the aperture C in the planes $z_{\gamma n}$, $n = 1, 2, \dots$. We take the Wronskian of the two rays (4.61) at any two subsequent images A_n and A_{n+1} of the aperture A, giving

$$\left(\sqrt{\Phi^*} u_\gamma u'_\alpha \right)_{z_{\alpha n}} = \left(\sqrt{\Phi^*} u_\gamma u'_\alpha \right)_{z_{\alpha, n+1}}. \quad (4.62)$$

Since $u_\alpha(z_{\alpha n}) = u_\alpha(z_{\alpha, n+1}) = 0$, the slopes of u_α at the planes $z_{\alpha n}$ and $z_{\alpha, n+1}$ must have *opposite* sign, as demonstrated in Fig. 4.8. Considering this behavior, it readily follows from (4.62) that u_γ must change its sign in the region between two subsequent images of the aperture A. This is only possible if an image ($u_\gamma(z_{\gamma n}) = 0$) of the aperture C is located this domain. Accordingly, we can state, “An optical system always forms an image of the source in the domain between any two subsequent images of the object plane.”

Figure 4.9 illustrates the consequences of this theorem for the image formation in an ideal electron microscope. The crossover of the cathode defines the effective source, which is located at some distance from the surface of the emitter. For a field emission gun, the crossover is generally virtual and located inside the tip of the emitter. The condenser system adjusts the illumination of the object. In order to achieve an ideal illumination system, the condenser should consist of

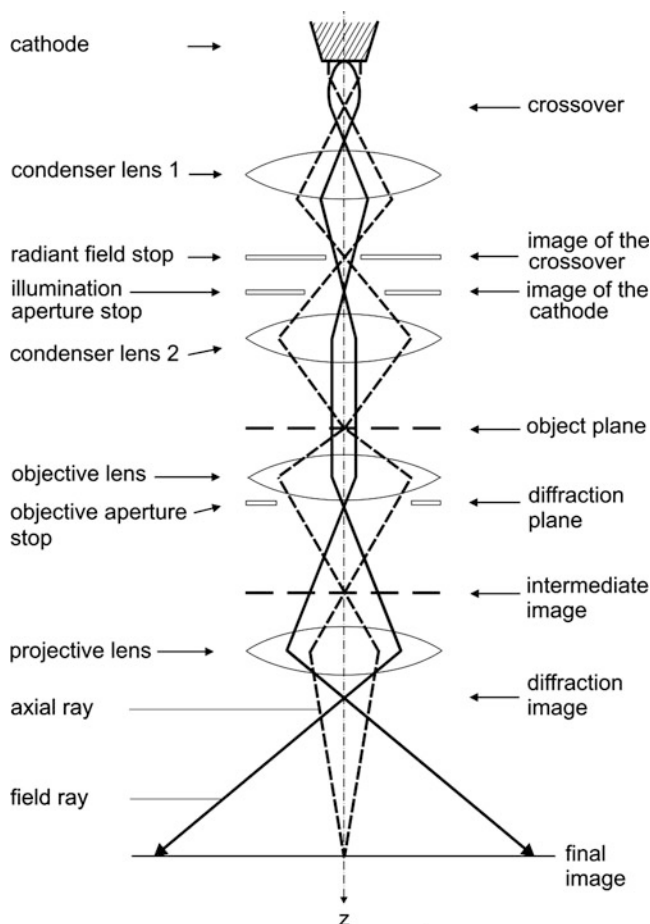


Fig. 4.9 Path of the fundamental paraxial trajectories and location of the images and beam-limiting apertures in a transmission electron microscope illustrating the theorem of alternating images

two lenses and two apertures, one placed at the image of the crossover, the other (illumination aperture) at an image of the cathode surface. The condenser lenses image the crossover aperture onto the object plane. This aperture limits the field of illumination, whereas the illumination aperture determines the maximum angle of illumination.

A very suitable kind of illumination is the “*Koehler illumination*,” which is largely utilized in light microscopy. We have chosen this illumination in Fig. 4.9. In this case, an image of the surface of the cathode is located in the back focal plane of the objective lens. The Koehler illumination has the advantage that local variations of the electron emission on the cathode surface do not show up as artifacts

in the image of the object. We can vary the location of the crossover image by changing the illumination mode. For Koehler illumination, the back focal plane of the objective lens is also the diffraction plane of the object. In accordance with the famous optician E. Abbe, one defines the diffraction pattern at this plane as the “*primary image*.”

Owing to the spherical aberration of the objective lens, the large-angle scattered electrons miss the Gaussian image point and blur the image. In order to remove these electrons from the beam and for obtaining the so-called scattering absorption contrast, one places an objective aperture at the back focal plane of this lens.

Each intermediate image of the object is also an image of the illumination-field aperture, and each image of the illumination-angle aperture coincides with that of the objective aperture. The special locations of the two illumination apertures allow one to vary the illuminated area in the object plane without affecting the angular illumination and vice versa. The characteristic planes in an electron microscope are, therefore, real and virtual images of the object and the crossover plane or images of the two characteristic apertures, respectively. No trick enables one to form two subsequent images of one of these two planes without having an image of the other plane located between them.

4.3.5 Longitudinal Magnification

Apart from the lateral magnification, we define the longitudinal magnification M_l . This magnification is a measure for the shift d_{zi} of the image if we move the object by a small distance d_{zo} . To obtain the relation between these distances, we assume that the aperture C shown in Fig. 4.8 represents the new position of the aperture A after being shifted in negative z -direction by the distance

$$d_{zo} = z_\alpha - z_\gamma = z_o - z_\gamma = \frac{u_{\gamma o}}{u'_{\gamma o}}. \quad (4.63)$$

As illustrated in Fig. 4.10, the image of A is then shifted by the distance

$$d_{zi} = z_{\alpha i} - z_{\gamma i} = \frac{u_{\gamma i}}{u'_{\gamma i}}. \quad (4.64)$$

By employing these expressions and the relation (4.62) at the planes $z_{\alpha n} = z_{\alpha 0} = z_o$, $z_{\alpha, n+1} = z_{\alpha 1} = z_i$, the longitudinal magnification takes the form

$$M_l = \frac{d_{zi}}{d_{zo}} = \frac{u_{\gamma i} u'_{\gamma o}}{u_{\gamma o} u'_{\gamma i}} = \sqrt{\frac{\Phi_1^*}{\Phi_o^*}} \frac{u_{\gamma i}^2}{u_{\gamma o}^2} = M^2 \frac{n_{0i}}{n_{0o}}. \quad (4.65)$$

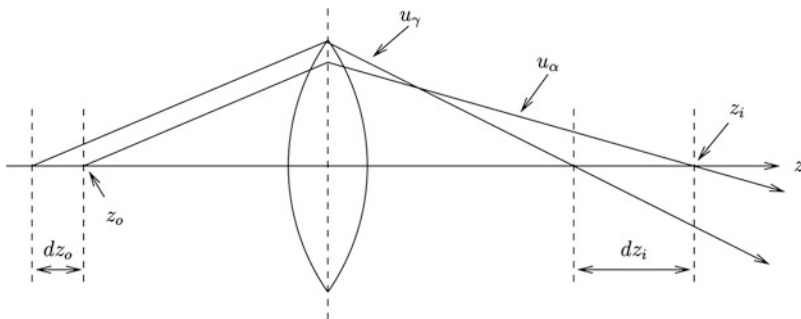


Fig. 4.10 Longitudinal magnification

This result demonstrates that the longitudinal magnification and the lateral magnification M largely differ from each other in the case $M \gg 1$. They coincide only in the case $M = n_{0o}/n_{0i}$.

4.3.6 Characteristic Paraxial Rays

In an electron microscope, the direction of flight of the electrons changes rather abruptly within the object due to scattering with the constituent atoms. The imaging system behind the object and the illumination system in front of the object determine the course of the scattered electrons, which form an image of the object in paraxial approximation. The aberrations in the image plane also depend on the mode of illumination and the location of beam-limiting apertures. In order to survey this influence, it is advantageous to select the *fundamental rays* u_α and u_γ as the pair of linearly independent solutions of the paraxial ray equation. The *axial ray* u_α starts from the center of the object plane with unit slope (4.51), whereas the *field ray* u_γ intersects the center of the diffraction plane z_d , which is also an image of the effective source. This ray satisfies the boundary conditions

$$u_\gamma(z_d) = 0, \quad (4.66)$$

$$u_\gamma(z_o) = u_{\gamma o} = 1. \quad (4.67)$$

Considering the boundary conditions (4.51), (4.66), and (4.67) for the fundamental rays at the object plane, we find that for these rays the constant of the Lagrange–Helmholtz relation (4.56) adopts the value $\text{const} = \Phi_o^{*1/2}$.

The principal rays u_π and $u_{\bar{\pi}}$ are best suited for calculating the *cardinal elements* of a lens, which are the focal lengths and the locations of the focal planes and principal planes, respectively. These rays satisfy the boundary conditions (4.59). Usually, one defines the trajectory by its slope and position components (4.50) at the object plane. In this case, we find the coefficients of the paraxial trajectory

$$u = C_\alpha u_\alpha + C_\gamma u_\gamma, \quad (4.68)$$

$$C_\alpha = \omega = u'_o - u'_{\gamma o} u_o, \quad C_\gamma = u_o. \quad (4.69)$$

The position u_d of the trajectory (4.68) at the diffraction plane is $u_d = C_\alpha u_{\alpha d} = \omega u_{\alpha d}$. Hence, in paraxial approximation, we have

$$u = \omega u_\alpha + u_o u_\gamma, \quad \omega = u_d / u_{\alpha d} = u'_o - u'_{\gamma o} u_o. \quad (4.70)$$

If we fix the ray by its positions in the object and diffraction planes, the second relation of (4.70) defines the complex slope ω of the trajectory at the object plane in paraxial approximation. We shall use this relation for investigating the dependence of the primary aberrations in an electron microscope on the illumination and the objective aperture.

4.3.7 Thin-Lens Approximation

We define lenses whose axial extension is smaller than their focal length either as *thin* lenses or as *short* lenses. A *thin lens* alters primarily the direction of the particle, whereas the lateral distance of an axis-parallel incident particle remains almost unchanged within the lens field. If the distance changes appreciably, the lens is termed a *short lens*. For a *thick lens*, the focal length is smaller than the extension of the lens, so that the trajectory intersects the optic axis within the domain of the field, as depicted in Fig. 4.3.

We can improve considerably the accuracy of the thin-lens approximation by considering that the modified principal ray $U_\pi = \Phi^{*-1/4} e^{-i\chi} w_\pi$ in the rotating coordinate system satisfies the assumption of being constant within the lens much better than the principal ray w_π . The focusing strength in the reduced complex path equation (4.29) has the form

$$T = \frac{2 + \gamma_0^2}{16} \frac{\Phi'^2}{\Phi^{*2}} + \frac{e}{8m_e} \frac{B^2}{\Phi^*}. \quad (4.71)$$

The expression is quadratic in the axial electric and magnetic field strengths $-\Phi'$ and B , respectively. This quadratic dependency results from the fact that the axial fields exert a deflecting force on an electron only if its velocity has a lateral component. Since this component is very small in the case of high energies, we cannot use round lenses for electron energies higher than about 1 MeV. In this case, one employs quadrupoles because their deflecting force acts directly on the electron, regardless of its direction of flight.

The electric part of the focusing strength (4.71) does not depend on the mass of the particle, whereas the magnetic part is inversely proportional to the mass. This

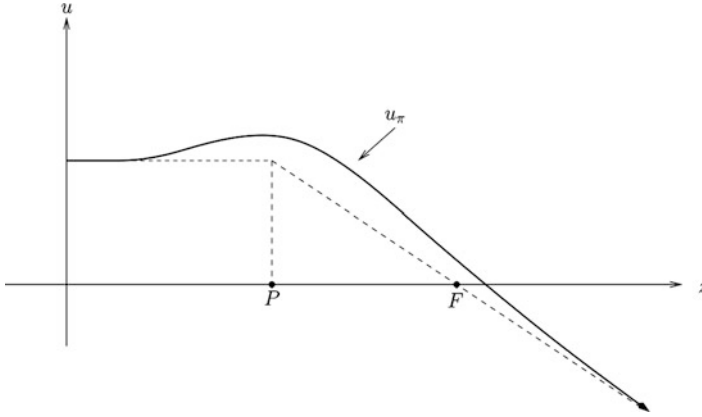


Fig. 4.11 Definition of the image focal point F and the image principal point P by means of the asymptotes of the image principal ray u_π

is the reason why one employs electrostatic round lenses for focusing ions with medium energies (≤ 100 keV).

The asymptotes of the principal rays u_π and $u_{\bar{\pi}}$ determine the location of the four *cardinal planes* z_F , z_P , $z_{\bar{F}}$ and $z_{\bar{P}}$ of any round lens. The intersection of the two asymptotes of the image principal ray u_π defines the *image principal plane* z_P , as depicted in Fig. 4.11. The emergent asymptote intersects the optic axis at the center F of the *image focal plane* z_F . We denote this point as the *image focus* and the center P of the image principal plane z_P as the *image principal point*. A decelerating electric einzel lens first deflects the incident ray u_π away from the optic axis. However, the reduced rays are always refracted toward this axis. The asymptotes of the object cardinal ray $u_{\bar{\pi}}$ define the locations $z_{\bar{F}}$ and $z_{\bar{P}}$ of the conjugate *object cardinal points* $\bar{F}$ and $\bar{P}$. If we know the location of the cardinal planes of a lens, we can readily obtain the image point of any given object point by means of the Gauss construction shown in Fig. 4.12. These planes also define the cardinal elements of the lens. These are the image and object focal lengths

$$f = z_F - z_P, \quad \bar{f} = z_{\bar{P}} - z_{\bar{F}} \quad (4.72)$$

and the separation of the principal planes

$$\Delta = z_P - z_{\bar{P}}. \quad (4.73)$$

The two principal planes shown in Fig. 4.12 are crossed in the sense that the image principal plane z_P is located in front of the object principal plane $z_{\bar{P}}$. This inversion of the principal planes holds for all short electron lenses. Once we know the foci and the focal lengths, we find the relation between object plane z_o and image plane z_i from the Newton lens equation

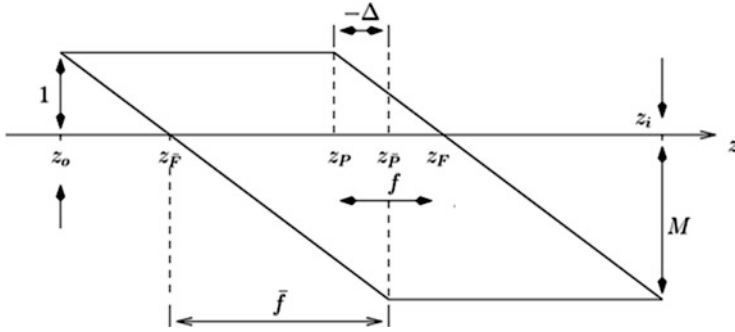


Fig. 4.12 Gaussian construction of the image point by means of the cardinal points of a lens

$$(z_{\bar{F}} - z_o)(z_i - z_F) = \bar{f}f. \quad (4.74)$$

We can readily verify this formula by means of the image construction shown in Fig. 4.12.

Apart from the principal rays, the nodal ray

$$u_v = f u_\pi - \bar{f} u_{\bar{\pi}} \quad (4.75)$$

is another cardinal ray, which is utilized in the *telescopic* case where the principal rays run both parallel to the optic axis in the field-free object and image spaces. These rays are then linearly dependent and can no longer describe an arbitrary ray. If the principal rays are symmetric in the telescopic limit, the nodal ray serves as the other linearly independent cardinal ray. Only in this case the nodal ray stays finite in the telescopic limit $f \rightarrow \infty$, $\bar{f} \rightarrow \infty$. Its incident and emergent asymptotes are

$$\begin{aligned} u_{v\bar{as}} &= f - (z - z_{\bar{F}}) = -(z - z_{\bar{N}}), & z_{\bar{N}} &= z_{\bar{F}} + f = z_{\bar{P}} + f - \bar{f}, \\ u_{vas} &= -(z - z_F) - \bar{f} = -(z - z_N), & z_N &= z_F - \bar{f} = z_P + f - \bar{f}. \end{aligned} \quad (4.76)$$

The *nodal planes* $z_{\bar{N}}$ and z_N coincide with the principal planes $z_{\bar{P}}$ and z_P for unipotential lenses ($\Phi_{-\infty}^* = \Phi_{\infty}^*$, $f = \bar{f}$). The two asymptotes (4.76) of the nodal ray are parallel to each other having a unit negative slope. The incident asymptote $u_{v\bar{as}}$ of the nodal ray (4.75) intersects the optic axis at the object nodal point $\bar{N}$; the emergent asymptote u_{vas} intersects this axis at the image nodal point N .

The asymptotes of the nodal ray are antisymmetric with respect to the central plane midway between the nodal planes z_N and $z_{\bar{N}}$. These planes move to infinity in the telescopic limit if the degenerated principal rays are also antisymmetric. This happens for an odd number of foci within the telescopic system. In this case, we must use the *symmetric ray*

$$u_\sigma = f u_\pi + \bar{f} u_{\bar{\pi}} \quad (4.77)$$

as the other linearly independent ray. Its incident asymptote $u_{\sigma\bar{a}s}$ and its emergent asymptote $u_{\sigma as}$ are given by

$$\begin{aligned} u_{\sigma\bar{a}s} &= z - z_{\bar{U}}, & z_{\bar{U}} &= f - z_F, \\ u_{\sigma as} &= -(z - z_U), & z_U &= \bar{f} + z_F. \end{aligned} \quad (4.78)$$

The two conjugate “unit” planes z_U and $z_{\bar{U}}$ define the locations of the object plane and the image plane, respectively, for negative unit magnification $M = -1$.

For determining the cardinal elements of a thin lens, we transform the reduced paraxial differential equation (4.29) with $G = 0$ in an integral equation by integrating the former equation twice. This transformation has the advantage that we can incorporate the boundary conditions. These are

$$u_{\pi}(z = -\infty) = u_{\pi, -\infty} = 1, \quad u'_{\pi, -\infty} = 0 \quad (4.79)$$

for the image principal ray u_{π} and

$$U_{\pi}(-\infty) = U_{\pi, -\infty} = \Phi_{-\infty}^{*1/4}, \quad U'_{\pi, -\infty} = \Phi_{-\infty}^{*1/4} \left[u'_{\pi} + \frac{\gamma_0}{4} \frac{\Phi'}{\Phi^*} u_{\pi} \right]_{z=-\infty} = 0 \quad (4.80)$$

for the modified image principal ray U_{π} , where we have assumed that the electric field of the lens vanishes at infinity ($\Phi'(-\infty) = 0$). Considering these boundaries and integrating the double integral by parts yields

$$U_{\pi} = \Phi_{-\infty}^{*1/4} - \int_{-\infty}^z \int_{-\infty}^{\zeta'} T(\zeta) U_{\pi}(\zeta) d\zeta d\zeta' = \Phi_{-\infty}^{*1/4} - z \int_{-\infty}^z T U_{\pi} d\zeta + \int_{-\infty}^z \zeta T U_{\pi} d\zeta. \quad (4.81)$$

This inhomogeneous integral equation for U_{π} is valid for an arbitrary static electromagnetic round lens since we have not made any approximations. We can solve this equation by the method of successive iteration. In order to achieve convergence the focusing strength T must decrease for $|z| \rightarrow \infty$ faster than z^{-3} .

The emergent asymptote

$$u_{\pi, as} = \Phi_{\infty}^{*-1/4} \left\{ \Phi_{-\infty}^{*1/4} - z \int_{-\infty}^{\infty} T U_{\pi} dz + \int_{-\infty}^{\infty} z T U_{\pi} dz \right\} \quad (4.82)$$

and the incident asymptote $u_{\pi, as} = u_{\pi}(-\infty) = 1$ of the principal ray u_{π} define the locations of the image cardinal planes z_P and z_F . We obtain the asymptote (4.82) in first approximation by substituting the zero-order approximation $U_{\pi}^{(0)} = \Phi_{-\infty}^{*1/4}$ for U_{π} in the integrands. The emergent asymptote intersects the optic axis at the image focal plane. We derive its location from the condition $u_{\pi, as}(z_F) = 0$ in first-order approximation as

$$z_F = \frac{1 + \int_{-\infty}^{\infty} z T dz}{\int_{-\infty}^{\infty} T dz}. \quad (4.83)$$

The incident and emergent asymptotes of the image principal ray intersect at the image principal plane. Hence, its location is defined and is given by $u_{\pi,as}(z_P) = 1$ yielding in first-order approximation

$$z_P = \frac{1 - (\Phi_{\infty}^*/\Phi_{-\infty}^*)^{\frac{1}{4}} + \int_{-\infty}^{\infty} z T dz}{\int_{-\infty}^{\infty} T dz}. \quad (4.84)$$

By substituting (4.83) and (4.84) for z_F and z_P in (4.72), we obtain for the image focal length the expression

$$f = z_F - z_P = \left(\frac{\Phi_{\infty}^*}{\Phi_{-\infty}^*} \right)^{\frac{1}{4}} \frac{1}{\int_{-\infty}^{\infty} T dz}. \quad (4.85)$$

Since the focusing strength T is positive definite (4.71), we have $f > 0$. Hence, all thin round electron lenses are convergent. We obtain the corresponding object quantities relatively easily from the formulas of the image cardinal elements, be considering that the object principal ray $u_{\bar{\pi}}$ is parallel to the optic axis on the field-free image-space side of the lens. Accordingly, we can derive the corresponding cardinal elements from (4.83)–(4.85) by replacing $z \rightarrow -z$ and $\infty \rightarrow -\infty$. The results are

$$z_{\bar{F}} = -\frac{1 - \int_{-\infty}^{\infty} z T dz}{\int_{-\infty}^{\infty} T dz}, \quad z_{\bar{P}} = -\frac{1 - (\Phi_{\infty}^*/\Phi_{-\infty}^*)^{\frac{1}{4}} - \int_{-\infty}^{\infty} z T dz}{\int_{-\infty}^{\infty} T dz}, \quad (4.86)$$

$$\bar{f} = z_{\bar{P}} - z_{\bar{F}} = \left(\frac{\Phi_{-\infty}^*}{\Phi_{\infty}^*} \right)^{\frac{1}{4}} \frac{1}{\int_{-\infty}^{\infty} T dz} = \sqrt{\frac{\Phi_{-\infty}^*}{\Phi_{\infty}^*}} f. \quad (4.87)$$

The second relation of (4.87) coincides with the exact relation (4.60), which connects the object and image focal lengths. This result is surprising in so far that we have derived (4.84) by means of the first-order approximation of the asymptotes of the principal rays. The quantity

$$\langle z \rangle = z_C = \frac{\int_{-\infty}^{\infty} z T dz}{\int_{-\infty}^{\infty} T dz} \quad (4.88)$$

defines the location of the “center of gravity” of the lens. For convenience, we can set the origin of the coordinate system at this center yielding $\langle z \rangle = 0$. To first-order approximation, the principal planes coincide in this plane for a thin einzel lens ($\Phi_{-\infty}^* = \Phi_{\infty}^*$). For an immersion lens ($\Phi_{-\infty}^* \neq \Phi_{\infty}^*$), the principal planes may be located at considerable distances away from the center z_C at positions

$$\begin{aligned}
z_P &= z_C + \left[(\Phi_{-\infty}^* / \Phi_{\infty}^*)^{\frac{1}{4}} - 1 \right] f, \\
z_{\bar{P}} &= z_C - \left[(\Phi_{\infty}^* / \Phi_{-\infty}^*)^{\frac{1}{4}} - 1 \right] \bar{f}.
\end{aligned} \tag{4.89}$$

These planes are separated by the distance

$$\begin{aligned}
\Delta &= z_P - z_{\bar{P}} = \frac{2 - (\Phi_{\infty}^* / \Phi_{-\infty}^*)^{\frac{1}{4}} - (\Phi_{-\infty}^* / \Phi_{\infty}^*)^{\frac{1}{4}}}{\int_{-\infty}^{\infty} T \, dz} \\
&= - \frac{\left[(\Phi_{\infty}^* / \Phi_{-\infty}^*)^{\frac{1}{8}} - (\Phi_{-\infty}^* / \Phi_{\infty}^*)^{\frac{1}{8}} \right]^2}{\int_{-\infty}^{\infty} T \, dz} < 0.
\end{aligned} \tag{4.90}$$

Since this distance is negative, we have an inversion of the principal planes of immersion lenses. In order to find if this holds also true for einzel lenses, we must employ the second-order approximation for the emergent asymptote of the principal ray. We obtain this approximation in the second step of the iteration by substituting the first-order approximation of the modified principal ray

$$U_{\pi}^{(1)} = \Phi_{-\infty}^{*1/4} \left\{ 1 - z \int_{-\infty}^z T(\zeta) d\zeta + \int_{-\infty}^z \zeta T(\zeta) d\zeta \right\} \tag{4.91}$$

for U_{π} in the formula (4.82) for the emergent asymptote of the image principle ray. We do not need to calculate the denominator of the expression (4.90) to second order because the denominator stays finite in the limit $\Phi_{-\infty}^* = \Phi_{\infty}^*$. After a rather lengthy yet straightforward calculation, we find that the separation distance

$$\begin{aligned}
\Delta &= \frac{[\int_{-\infty}^{\infty} zT \, dz]^2 - \int_{-\infty}^{\infty} z^2 T \, dz \cdot \int_{-\infty}^{\infty} T \, dz}{\int_{-\infty}^{\infty} T \, dz} \\
&= (\langle z \rangle^2 - \langle z^2 \rangle) \int_{-\infty}^{\infty} T \, dz = - \frac{\langle (z - \langle z \rangle)^2 \rangle}{f} < 0
\end{aligned} \tag{4.92}$$

is negative for thin einzel lenses too. Accordingly, we can state that the inversion of the principal planes holds true for all short electron lenses.

The focal length is the dominant cardinal element of a lens. We obtain the focal lengths of short electric lenses and short magnetic lenses from (4.87) and (4.71) by setting $B = 0$ in the former case and $\Phi' = 0$, $\Phi_{-\infty}^* = \Phi_{\infty}^* = \Phi_0$ in the latter case, giving

$$\frac{1}{f_e} = \frac{1}{\bar{f}_e} \sqrt{\frac{\Phi_{-\infty}^*}{\Phi_{\infty}^*}} = \left(\frac{\Phi_{-\infty}^*}{\Phi_{\infty}^*} \right)^{\frac{1}{4}} \int_{-\infty}^{\infty} \frac{2 + (1 + 2\varepsilon\Phi)^2}{16(1 + \varepsilon\Phi)^2} \frac{\Phi'^2}{\Phi^2} dz \tag{4.93}$$

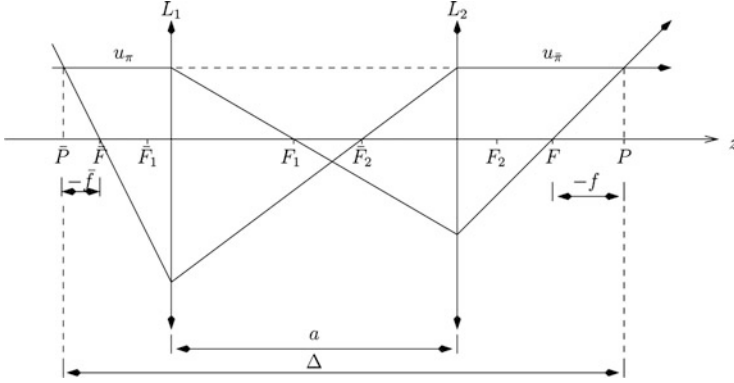


Fig. 4.13 Construction of the cardinal points of an immersion compound lens by means of the cardinal planes of the constituent thin lenses, which are separated by the distance a

for the focal lengths of electric lenses and

$$\frac{1}{f_m} = \frac{1}{\bar{f}_m} = \frac{e}{8m_e\Phi_0^*} \int_{-\infty}^{\infty} B^2 dz \quad (4.94)$$

for magnetic lenses. Nowadays, one considers (4.94) for the focal length of magnetic lenses, first derived by H. Busch (1926), as the “*foundation stone*” of geometrical electron optics.

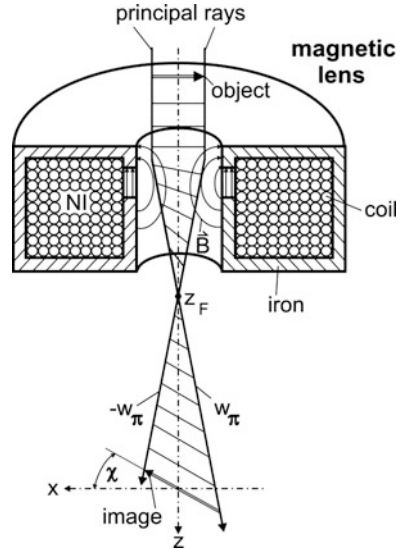
High-resolution electron microscopes use strong magnetic objective lenses such that the intersections z_π and $z_{\bar{\pi}}$ of the principal rays with the optic axis are immersed in the magnetic field. In the case of high magnification, one places the object at the plane $z = z_o = z_{\bar{\pi}}$, which represents the *actual* front focal plane of the objective lens. The corresponding *objective focal length* is

$$f_o = \frac{1}{u'_{\bar{\pi}}(z_{\bar{\pi}})} = -\frac{1}{u'_\pi(z_\pi)}. \quad (4.95)$$

One always uses this definition for the focal length of objective lenses of electron microscopes.

We can conceive thick lenses as *compound lenses* composed of thin lenses. The focal length of a compound lens may be positive or negative depending on the number of intersections of the principal rays with the optic axis. For such lenses, the separation Δ of the principal planes may become positive. We can construct a divergent thick lens with positive Δ already by a combination of two thin lenses separated by a proper distance a , as demonstrated in Fig. 4.13. We denote the focal points of the two lenses as $F_1, \bar{F}_1$ and $F_2, \bar{F}_2$, respectively, while $F, \bar{F}, P, \bar{P}$ and $f, \bar{f}$ characterize the cardinal elements of the resulting immersion compound lens.

Fig. 4.14 Focusing and Larmor rotation χ by the axially symmetric field $\vec{B} = \vec{B}(\rho, z)$ of an iron-shielded magnetic electron lens; NI defines the ampere-turns of the coil



4.4 Magnetic Model Lens

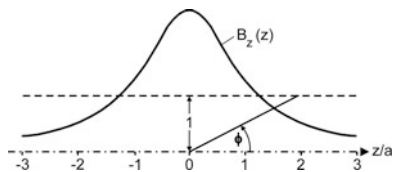
Axially symmetric magnetic electron lenses are produced by iron-shielded coils. The pole pieces confine the axial extension of the magnetic field to the region of the circular slit in the iron shield, as shown in Fig. 4.14. Because these lenses are the basic elements of electron microscopes and probe-forming systems, one needs to know their imaging properties regardless if the lens is short or thick. Busch's formula (4.94) for the focal length of a magnetic lens applies to the paraxial focusing in very short fields only. This assumption is valid if the focal length is large in comparison with the axial extension of the magnetic field. However, in particular the objective lenses of electron microscopes are thick magnetic lenses whose fields do not provide simple analytical formulas for both the paraxial rays and the aberration coefficients, with the exception of the Glaser bell-shaped field

$$B_z(z) = B(z) = \frac{B_0}{1 + z^2/a^2}. \quad (4.96)$$

Here B_0 denotes the maximum magnetic flux density and $2a$ the full width at half maximum [86]. The model field (4.96) approximates sufficiently well the field of a symmetric magnetic lens, although it decreases more slowly than the field of iron-shielded magnetic lenses. Glaser's treatment of the bell-shaped field has provided a qualitative understanding of the properties of magnetic lenses, especially at a time when computers were not available.

The Glaser field has the advantage that it enables one to express its paraxial properties and the coefficients of the primary aberrations in analytical form as

Fig. 4.15 Geometrical relation between the angle ϕ and the z -coordinate and course of the Glaser field (4.96)



functions of the object plane z_0 and the dimensionless lens strength

$$k^2 = \frac{eB_0^2 a^2}{8m_e \Phi_0^*} = \frac{e\mu_0^2 N^2 I^2}{8\pi^2 m_e \Phi_0^*}. \quad (4.97)$$

Here μ_0 and NI denote the permeability of the vacuum and the number of ampere-turns of the lens coil, respectively.

The specific feature of the bell-shaped field allows us to investigate comprehensively the general properties of a magnetic lens without the need of large numerical calculations for many different lens parameters. Substituting (4.96) and $\Phi' = \Phi'' = 0$ in the paraxial path equation (4.48), we obtain

$$u'' + \frac{k^2 a^2}{(a^2 + z^2)^2} u = 0. \quad (4.98)$$

This equation can be transformed in a differential equation with constant coefficients by means of the substitutions

$$z = a \cot \phi, \quad u = \frac{\tilde{u}(\phi)}{\sin \phi}, \quad \sigma = \sqrt{1 + k^2}. \quad (4.99)$$

The resulting equation has the simple form

$$\frac{d^2 \tilde{u}}{d\phi^2} + \sigma^2 \tilde{u} = 0. \quad (4.100)$$

Any solution of this equation is a linear combination of the two basic solutions

$$\tilde{u}_1(\phi) = \sin \sigma \phi, \quad \tilde{u}_2(\phi) = \cos \sigma \phi. \quad (4.101)$$

The relation between the auxiliary angle ϕ and the z -coordinate is shown in Fig. 4.15 together with the bell-shaped field (4.96).

The image principal rays $u_\pi(\phi)$ and the object principal ray $u_{\bar{\pi}}(\phi)$ satisfy the boundary conditions $u_\pi(\pi) = 1$, $u_{\bar{\pi}}(0) = 1$ and have the form

$$u_\pi(\phi) = \frac{\sin \sigma(\pi - \phi)}{\sigma \sin \phi}, \quad u_{\bar{\pi}}(\phi) = \frac{\sin \sigma \phi}{\sigma \sin \phi}. \quad (4.102)$$

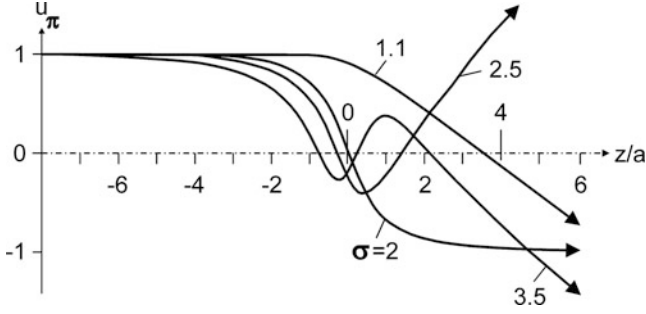


Fig. 4.16 Course of the image principle ray $u_{\pi}(z)$ within the region of the bell-shaped field (4.96) for increasing values of the lens parameter σ

The course of the image principal ray $u_{\pi}(z)$ is shown in Fig. 4.16 for different values of the parameter σ . The zeros $\phi_{\pi n} = \phi_{Fn} = \pi(1 - n/\sigma)$, $\sigma \geq n = 1, \dots$, of the ray $u_{\pi}(z)$ define the positions of the real image focal planes, whereas the zeros $\phi_{\bar{\pi}n} = \phi_{\bar{F}n} = n\pi/\sigma$ of the object principal ray $u_{\bar{\pi}}(\phi)$ represent the real object focal planes. The focal planes are located at positions

$$z_{Fn} = -z_{\bar{F}n} = -a \cot(n\pi/\sigma), \quad \sigma \geq n = 1, \dots \quad (4.103)$$

The object and image focal planes are arranged symmetrically with respect to the center plane $z = 0$ of the bell-shaped field. The number of focal planes is given for each kind by the integer value of the parameter σ . There will be only one image focal plane and one object focal plane for $0 < k^2 < 3$ or $1 < \sigma < 2$, respectively. This is the case for most lenses used in practice.

Apart from the principal rays, we also need the axial ray $u_{\alpha}(\phi)$ and the field ray $u_{\gamma}(\phi)$ for describing the paths of the electrons. We assume that the field ray intersects the center of the aperture plane $z_a = a \cot \phi_a$. The axial ray $u_{\alpha}(\phi)$ starts from the center of the object plane $z_o = a \cot \phi_o$ with slope $u'_{\alpha}(z_o) = -(1/a)\dot{u}_{\alpha}(\phi_o) \sin^2 \phi_o = 1$, where the dot denotes differentiation with respect to the angle ϕ . The field ray intersects the object plane at $u_{\gamma}(\phi_o) = 1$. By applying these boundary conditions, we find

$$\begin{aligned} u_{\alpha}(\phi) &= \frac{a \sin \sigma(\phi_o - \phi)}{\sigma \sin \phi_o \sin \phi}, \\ u_{\gamma}(\phi) &= \frac{\sin \phi_o}{\sin \sigma(\phi_o - \phi_a)} \frac{\sin \sigma(\phi - \phi_a)}{\sin \phi}. \end{aligned} \quad (4.104)$$

The zeros of the axial ray define the locations of the image planes $z_{in} = a \cot \phi_{in}$ which are conjugate to the object plane. These planes are located at angular positions

$$\phi_{in} = \phi_o - n \frac{\pi}{\sigma} \geq 0. \quad (4.105)$$

Most lenses used in practice form only a single image ($n = 1$); multiple images will occur for strong lenses ($\sigma > 2$). We obtain the real focal length from (4.104) by substituting $\phi = 0$ and $\phi_o = \phi_{\bar{F}} = n\pi/\sigma$ in the expression for the axial ray $u_\alpha(\phi)$, giving

$$f_n = \bar{f}_n = -\frac{a \cos n\pi}{\sin \frac{n\pi}{\sigma}} = (-1)^{n+1} \frac{a}{\sin \frac{n\pi}{\sigma}}. \quad (4.106)$$

The expression (4.105) furnishes the angular connection between the object plane and the conjugate image plane. Since $z = a \cot \phi$, we can rewrite (4.105) such that we obtain a relation between the position z_o of the object plane and the location of the conjugate image planes z_{in} . By taking the cotangent of the relation (4.105), we obtain

$$\begin{aligned} z_o &= a \cot \phi_o = a \cot(\phi_{in} + n\pi/\sigma) \\ &= a \frac{\cot \phi_i \cot(n\pi/\sigma) - 1}{\cot \phi_i + \cot(n\pi/\sigma)} = \frac{z_{in} z_{Fn} + a^2}{z_{Fn} - z_{in}}. \end{aligned} \quad (4.107)$$

Writing $a^2 = a^2 / \sin^2(n\pi/\sigma) - a^2 \cot^2(n\pi/\sigma) = f_n \bar{f}_n + z_{Fn} z_{\bar{F}n}$, we readily derive Newton's equation of geometrical optics:

$$(z_{in} - z_{Fn})(z_o - z_{\bar{F}n}) = -f_n \bar{f}_n. \quad (4.108)$$

The Glaser model field has the peculiar property that its cardinal elements do not depend on the location of the object plane within the field. Fields with this property are named “Newtonian” fields. The magnification M_n is determined by the value of the field ray at the image plane. We readily obtain from (4.104) and (4.105) the expression

$$M_n = u_\gamma(\phi_{in}) = \frac{\sin \phi_o}{\sin \phi_{in}} \frac{\sin \sigma(\phi_{in} - \phi_a)}{\sin \sigma(\phi_o - \phi_a)} = \frac{(-1)^n \sin \phi_o}{\sin(\phi_o - n\pi/\sigma)}. \quad (4.109)$$

The angle of the Larmor rotation (4.24) between the object and the image is given by

$$\chi(z_i) = -k \int_{\phi_o}^{\phi_{in}} d\phi = \frac{nk\pi}{\sigma}. \quad (4.110)$$

Because the sign of the parameter k is given by the sign of B_0 , the image rotation can be compensated for by changing appropriately the sign of the currents in the lenses of an electron microscope. Since the image position z_i , the focal length f and the magnification M are functions of k^2 and z_o ; any two of the five quantities z_o, z_i, f, k^2, M can in principle be used as the two independent parameters defining the imaging conditions.

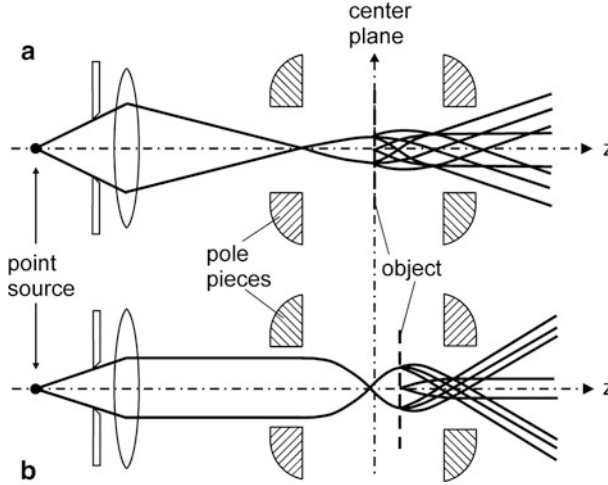


Fig. 4.17 Trajectories of the incident electrons and of the scattered electrons within a thick objective lens operating (a) in the symmetric condenser-objective mode and (b) in the telescopic second-zone mode

4.4.1 Condenser-Objective Lens

The Glaser lens with excitation $k^2 = 3$ represents a so-called single-field condenser-objective lens [87]. This lens is telescopic as a whole, whereas each half of the lens acts as a lens with shortest focal length $f = a$. We can consider the thick condenser-objective lens as a symmetric compound lens composed of a pre-lens in the regime $z < z_0$ and of a post-lens for the region $z \geq z_0$. In an electron microscope, the pre-field of the compound lens acts as the condenser lens and the post-field as the objective lens. In the case of very high magnification ($M \rightarrow \infty$), the specimen is positioned at the symmetry plane $z = 0$ so that the pre-lens and the post-lens are mirror symmetric, as illustrated in Fig. 4.17a. Then the back focal plane z_{F1} of the condenser lens (pre-lens) and the front focal plane z_{F2} of the objective lens (post-lens) coincide with the center plane $z_s = z_{F1} = z_{F2} = 0$ of the system. The condenser lens focuses an incident parallel beam at the specimen. In order to obtain parallel illumination, we must image the effective source into the front focal plane of the pre-lens located at position $z = -a$, ($\phi = 3\pi/4$). This plane represents the so-called entrance pupil. Owing to the symmetry of the system, the back focal plane of the post-lens is at position $z = a$, ($\phi = \pi/4$) where the objective aperture is placed. The object plane $z_0 = z_{F2}$ is shifted beyond the center plane $z = 0$ in the so-called second-zone mode ($3 < k^2 \leq 8, n = 2$). In the special case $k^2 = 21/4$, $\sigma = 5/2$ and high magnification ($M \rightarrow \infty$), the object is positioned at the plane $z_0 = z_{F2} = 0.325a$. For this location, the condenser pre-lens operates in the telescopic mode, as depicted in Fig. 4.17b. The focal length of the post-lens is $f_2 = 1.05a$, and the image focal point (4.83) is positioned at $z_{F2} = 1.38a$.

4.5 Quadrupoles

At high energies above a few MeV, the focusing strength of rotationally symmetric lenses is too weak to focus the beam within a sufficiently short distance from the lens. As a result, the diameter of the beam may become unduly large. Apart from round lenses, quadrupoles affect the paraxial path of rays. Since their electromagnetic field is transversal inside the quadrupole, its refraction power is proportional to the field strength. This is the reason why these “strong focusing” elements are exclusively employed in high-energy accelerators and storage rings. However, the interest in quadrupoles started much earlier primarily in the context to compensate for the aberrations of the round lenses in the electron microscope. Owing to the ability of quadrupoles to produce a line focus, they are components of correctors, which eliminate the unavoidable spherical aberration of round lenses by means of octopoles.

Unfortunately, quadrupoles have the property to refract charged particles toward the optic axis in one of its two orthogonal *principal sections* and deflect them away from the axis in the other. In order to achieve an overall focusing effect, at least two quadrupoles with opposite polarity of their corresponding electrodes or magnetic poles are necessary. Hence, the total focusing power of the system is only the difference between the strong focusing and defocusing of the two elements.

An electron moving on a *principal section* does not experience any force perpendicular to this section. Therefore, the electron will stay on this section and eventually intersect the optic axis along its course. The electrodes and pole pieces of conventional quadrupole elements are centered symmetrically about two orthogonal plane sections such that the geometry of the arrangement does not change with respect to an azimuthal rotation of 90° . Since the polarity of adjacent electrodes or magnetic poles is opposite, the potential changes its sign for this rotation. In order that the principal sections of the electric quadrupoles coincide with those of the magnetic quadrupoles, one must center the electrodes about sections rotated by 45° with respect to those of the magnetic pole pieces, as shown in Fig. 4.18. In this arrangement, the quadrupoles are “regular” in the sense that their principal sections coincide with the $x - z$ and the $y - z$ sections of the Cartesian coordinate system.

In the ideal case of “pure” quadrupoles, the transversal cross sections of the electrodes and magnetic poles form hyperbolas. In practice, the form of the electrodes and pole faces differs from this ideal shape. The deviation produces additional higher multipole components of the electric and/or magnetic potential. We assume that the deviations satisfy the symmetry conditions so that the complex multipole strengths fulfill the relation

$$\Phi_\nu e^{i\nu\frac{\pi}{2}} = -\Phi_\nu, \quad \Psi_\nu e^{i\nu\frac{\pi}{2}} = -\Psi_\nu. \quad (4.111)$$

The resulting requirement $e^{i\nu\frac{\pi}{2}} = -1$ shows that the deviations induce higher-order multipole components of the potentials with multiplicity

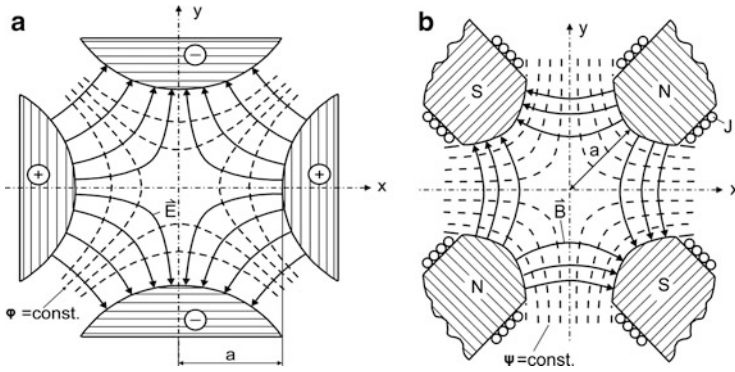


Fig. 4.18 Cross section of (a) an electric quadrupole and (b) a magnetic quadrupole whose principal sections coincide with the $x - z$ section and $y - z$ section, respectively

$$\nu = 4m + 2, \quad m = 1, 2, \dots \quad (4.112)$$

Hence, the lowest higher-order multipole component of a quadrupole consisting of four identical electrodes or pole pieces, respectively, is a dodecapole component ($\nu = 6$).

Quadrupoles are lenses with two-section symmetry and field-free axis. For systems composed of these elements, we have in the frame of Gaussian optics

$$\Gamma = 0, \quad \Phi' = 0, \quad B = 0, \quad \Phi_1 = 0, \quad \Psi_1 = 0, \quad \Phi = \Phi_0. \quad (4.113)$$

Considering these relations, the paraxial path equation (4.16) reduces to the simple form

$$w'' - G\bar{w} = 0, \quad (4.114)$$

$$G = G(z) = \gamma_0 \frac{\Phi_{2c}}{\Phi_0^*} - \sqrt{\frac{2e}{m_e \Phi_0^*}} \Psi_{2s} + i \left\{ \gamma_0 \frac{\Phi_{2s}}{\Phi_0^*} + \sqrt{\frac{2e}{m_e \Phi_0^*}} \Psi_{2c} \right\}. \quad (4.115)$$

Although the structure of the complex differential equation (4.114) looks rather simple, its solution is rather complicated because the two real equations for the x - and y -components of the Gaussian trajectory are coupled if the quadrupole strength G is complex. This strength is a real function of z for “regularly” oriented quadrupoles such that their *skew* components vanish:

$$\Phi_{2s} = 0, \quad \Psi_{2c} = 0. \quad (4.116)$$

Then the principal sections of the electric and magnetic quadrupoles coincide, and Eq. (4.114) yields the two uncoupled real equations:

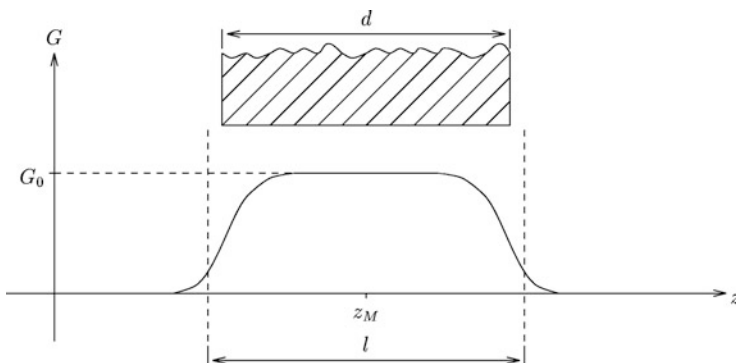


Fig. 4.19 Course of the strength $G(z)$ of a freestanding quadrupole

$$\begin{aligned} x'' - Gx &= 0, \\ y'' + Gy &= 0, \end{aligned} \quad (4.117)$$

$$G = \bar{G} = \gamma_0 \frac{\Phi_{2c}}{\Phi_0^*} - \sqrt{\frac{2e}{m_e \Phi_0^*}} \Psi_{2s}. \quad (4.118)$$

The two equations (4.117) show that the quadrupole field focuses the ray in one of the two principal sections and defocuses it in the other depending on the sign of the quadrupole strength G .

4.5.1 Imaging Properties of a Single Quadrupole

The quadrupole strength (or field function) $G = G(z)$ of a single freestanding quadrupole is symmetric with respect to the midplane z_M of the quadrupole, as depicted in Fig. 4.19. The function of thick quadrupole lenses has an almost rectangular shape apart from the fringe fields at the entrance and exit regions.

The axial width d of a thick quadrupole is large compared with its “radius” a . Since the paraxial focusing strength G does not depend on the derivatives of Φ_2 and Ψ_2 , we can approximate the field function $G(z)$ of a thick quadrupole with a sufficient degree of accuracy by the box-shaped function

$$G(z) = \begin{cases} 0 & \text{for } |z - z_M| > l/2 \\ G_0 & \text{for } |z - z_M| \leq l/2 \end{cases}. \quad (4.119)$$

Here G_0 denotes the constant interior quadrupole strength. The *effective length* of the quadrupole field

$$l = \frac{1}{G_0} \int_{-\infty}^{\infty} G(z) dz \quad (4.120)$$

is somewhat larger than the thickness d of the electrodes or pole pieces, respectively. Experiments have shown that this length is connected with d and a by the “rule of thumb”

$$l \approx d + 0.15a. \quad (4.121)$$

In order to obtain analytical solutions of the path equations (4.117), we employ the rectangular field model (4.119) and introduce the dimensionless quadrupole strength

$$k^2 = G_0 l^2. \quad (4.122)$$

Moreover, we place the origin of the coordinate system at the center of the quadrupole ($z_M = 0$). In this case, the path equations (4.117) have simple analytical solutions. In particular, we derive for the principal rays within the field of the quadrupole the expressions

$$\begin{aligned} y_\pi &= \cos k(z/l + 1/2), & x_\pi &= \cosh k(z/l + 1/2), \\ y_{\bar{\pi}} &= \cos k(z/l - 1/2), & x_{\bar{\pi}} &= \sinh k(z/l - 1/2). \end{aligned} \quad (4.123)$$

In the field-free region $|z - z_M| = |z| < l/2$, the rays form straight lines, which coincide with the asymptotes. They define the cardinal elements, as illustrated in Fig. 4.20. Without repeating the straightforward calculations, we merely state the result for the focal lengths, the locations of the focal planes, and those of the principal planes:

$$z_{Fx} = -z_{\bar{F}x} = l \left[\frac{1}{2} - \frac{\coth k}{k} \right], \quad z_{Fy} = -z_{\bar{F}y} = l \left[\frac{1}{2} + \frac{\cot k}{k} \right], \quad (4.124)$$

$$z_{Px} = -z_{\bar{P}x} = l \left[\frac{1}{2} - \frac{\tanh k/2}{k} \right], \quad z_{Py} = -z_{\bar{P}y} = l \left[\frac{1}{2} - \frac{\tan k/2}{k} \right], \quad (4.125)$$

$$f_x = \bar{f}_x = -\frac{l}{k \sinh k}, \quad f_y = \bar{f}_y = \frac{l}{k \sin k}. \quad (4.126)$$

For the design of quadrupole systems, it is useful to employ the short-lens approximation in the first step. This approximation gives a rough survey of the imaging properties of a given system and enables one to find arrangements, which meet the best the requirements imposed on the system. The formulas (4.126) demonstrate that a weak or short quadrupole focuses the rays in one principal section and defocuses them in the other principal section. Both sections are perpendicular to each other. For the divergent $x - z$ section, the locations of the focal planes are always reversed in the sense that the image focal plane lies in the object space and the object focal plane in the image space, as illustrated for a short quadrupole ($k \ll 1$) in

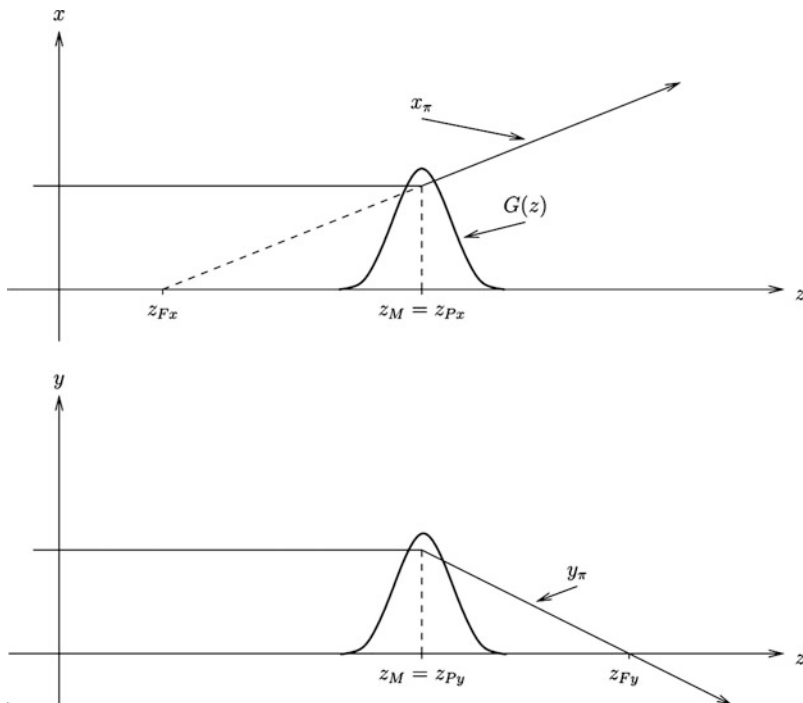


Fig. 4.20 Determination of the image cardinal points of a short quadrupole by means of the cardinal rays x_π and y_π

Fig. 4.20. Accordingly, the corresponding focal length f_x is always negative. We readily obtain the focal lengths of a weak short quadrupole with a non-rectangular field distribution by integrating the equations (4.117), yielding the approximation

$$\frac{1}{f_x} = -\frac{1}{\bar{f}_x} = -\frac{1}{f_y} = \frac{1}{f_y} = \int_{-\infty}^{\infty} G(z) dz. \quad (4.127)$$

The lens action in the $y - z$ section is convergent for weak lenses, while it may become divergent for thick quadrupoles ($k > \pi$). The relations (4.125) demonstrate that the focal lengths for the divergent and the convergent sections can never precisely coincide.

Unlike round lenses, two sets of cardinal elements define the imaging properties of quadrupoles. One set characterizes the divergent $x - z$ section, the other the convergent $y - z$ section. Hence, we need eight cardinal elements to characterize unambiguously the imaging properties of quadrupoles. The separation of the principal planes $\Delta_x = z_{Px} - z_{\bar{P}x}$ for the divergent section is always positive, while that for the convergent section is negative for not too thick quadrupoles. This opposite behavior poses a major obstacle for constructing quadrupole systems with round-lens imaging properties.

4.5.2 Quadrupole Multiplets

A single short quadrupole forms a real line focus and a virtual line focus located on opposite sides of the midplane z_M . A strong quadrupole may form two virtual astigmatic images, which can never coincide to form a virtual stigmatic image. In order to obtain stigmatic imaging between a distinct pair of planes, at least two spatially separated quadrupoles with opposite polarity are necessary. The system forms an antisymmetric quadrupole doublet if the two elements have the same geometry and are exited antisymmetrically. In order to survey the properties of a quadrupole doublet, we consider the formula for the focal length f of a system consisting of two thin lenses separated by a distance D :

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{D}{f_1 f_2}. \quad (4.128)$$

For quadrupoles, we must employ this formula separately for the $x - z$ section and the $y - z$ section. For the antisymmetric doublet shown in Fig. 4.21, the focal lengths of the constituent quadruples are

$$f_{1x} = -f_{2x} = f_0, \quad f_{1y} = -f_{2y} = -f_0. \quad (4.129)$$

By inserting the focal length for the $x - z$ section and that for the $y - z$ section separately into (4.128), we find that for the focal lengths of the antisymmetric quadrupole doublet coincide:

$$f_x = \bar{f}_x = f = \frac{f_0^2}{D}, \quad f_y = \bar{f}_y = f = \frac{f_0^2}{D}. \quad (4.130)$$

At first glance, this result seems to show that the system behaves like a round lens. Unfortunately, this conjecture does not hold true because the principal planes for the $x - z$ section and the $y - z$ section are located at different positions, as demonstrated convincingly for the special case $D = f_0 = f$ in Fig. 4.21. The image principal ray x_π intersects then the center of the second quadrupole and the object principal ray y_π the center of the first quadrupole. Although the sequence and the separation of the cardinal points coincide for the two principal sections, the cardinal points of the $y - z$ section are displaced by a distance $2f$ with respect to the corresponding points in the $x - z$ section. Therefore, the antisymmetric quadrupole doublet does not have the imaging properties of a round lens.

In order to obtain a quadrupole system equivalent to a round lens, each pair of corresponding cardinal points of the two sections must be located in a common cardinal plane. Since we have four cardinal points for each section, at least four quadrupoles are necessary to match their locations. We can readily achieve this, for example, by placing an identical second doublet at a distance $2f$ behind the doublet shown in Fig. 4.21. The resulting antisymmetric quadrupole quadruplet acts like a

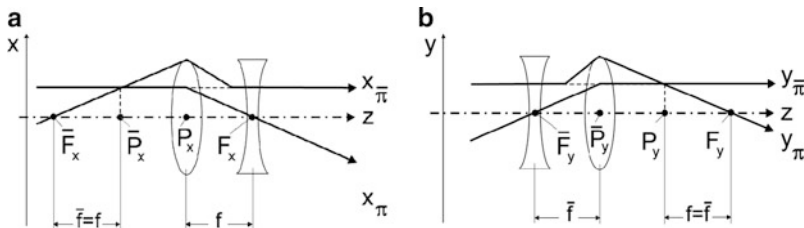


Fig. 4.21 Determination of the cardinal points of an antisymmetric quadrupole doublet by the cardinal rays $x_\pi, x_{\bar{\pi}}$ for (a) the $x-z$ section and (b) by the cardinal rays $y_\pi, y_{\bar{\pi}}$ for the $y-z$ section, respectively

thick telescopic round lens, although it forms two orthogonal astigmatic line images inside the system. Owing to the imposed symmetry, the second doublet compensates for the astigmatism introduced by the first quadrupole doublet. Such a quadrupole quadruplet enables the correction of the unavoidable spherical aberration of round lenses by means of three octopoles, two of which are centered at the line foci.

A quadrupole system yields in general astigmatic imaging such that the system images each point of the object in two orthogonal image lines, each of which is located at one of the two *astigmatic* image planes. If we move the object, the two images may move toward each other or away from each other depending on the direction of motion of the object plane. Hence, the two images will coincide for a specific location of the object plane forming a stigmatic image.

The antisymmetric quadrupole doublet only enables stigmatic imaging for a single object plane. The resulting image exhibits a strong first-order *elliptical* distortion because the magnifications M_x and M_y are different for the two principal sections, as illustrated in Fig. 4.22 by means of two axial rays. In this case, the system images a circle in an ellipse. To obtain a distortion-free image, one needs at least three quadrupoles. If the ratio $M_x/M_y = \vartheta_{iy}/\vartheta_{ix}$ differs from one, the imaging is *anamorphic*. We call a system to be *pseudo-stigmatic* if it produces a point-to-point image only for distinct isolated points in the object space. In order to obtain a *stigmatic* quadrupole system with variable focal length acting like a round lens, one needs at least five quadrupoles because one adjustable parameter is required for varying the focal length and the other four to match the four cardinal points of the $x-z$ section with the corresponding points of the $y-z$ section. Bauer's quadrupole objective lens satisfies this condition [71]. The general belief and assertions in the literature that four quadrupoles are sufficient for this purpose are erroneous because we cannot vary the specific focal length of such a system without introducing an elliptical distortion. Dymnikov and Yavor have extensively studied the quadrupole quadruplet as a substitute for an axially symmetric lens [88]. Their antisymmetric system is known as the *Russian quadruplet*.

We define quadrupole systems with round-lens properties as quadrupole *anastigmats* (non-astigmatic). Calculations have shown that one obtains a highly versatile quadrupole anastigmat free of chromatic aberration of magnification and distortion by appropriately combining two quadrupole triplets [89]. This system represents an

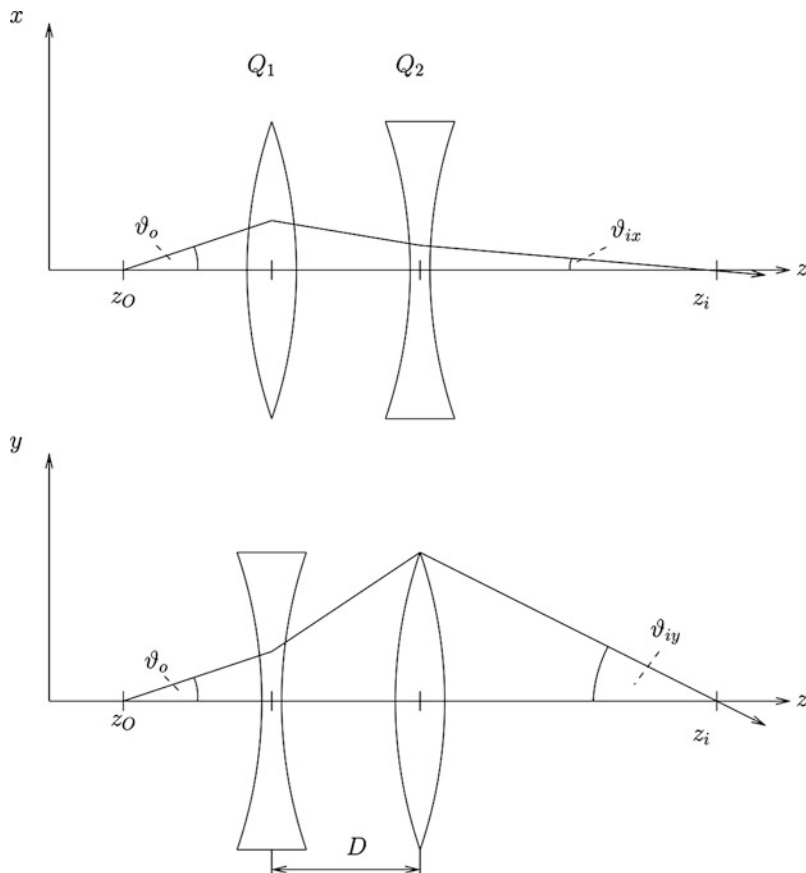


Fig. 4.22 Formation of first-order distortion of the stigmatic image formed by a quadrupole doublet

excellent substitute for the projector lens system of an electron microscope because it images the intermediate image of either the object plane or the diffraction plane with variable magnification in a fixed detection plane. The range of magnification depends on the separation distance between the two triplets. The system can also provide an astigmatic image. One applies this mode for the imaging of the stigmatic energy-loss spectrum in order to avoid unduly large intensities at the CCD camera.

4.5.3 Strong Focusing

The maximum achievable electric field strength $E_{\max}$ and/or the maximum magnetic flux density $B_{\max}$ limit the focusing power of charged-particle lenses. Since their focal length increases with increasing accelerating voltage for fixed field strengths,

it is important for the focusing of relativistic particles to know which elements yield strongest focusing for given maximum field strengths. In order to answer this question, we consider periodic arrangements of electric and magnetic round lenses and quadrupoles. We first investigate a system of electrostatic round lenses formed by a sequence of equally spaced thick aperture electrodes, which are at alternating potential $\varphi_a = \Phi_0 \pm \Phi_a$. The distance between the midplanes of any two adjacent apertures is a , so that a system consisting of $n + 1$ apertures has the total length $l = na$. Neglecting the somewhat different fringing fields at the entrance and exit apertures, we may assume that the axial potential within the system has the form

$$\varphi(\rho = 0, z) = \Phi(z) = \Phi_0 + U \cos \pi \frac{z}{a}. \quad (4.131)$$

By employing the integral formula (3.3 for the case of rotational symmetry ($\nu = 0$), we find

$$\varphi(\rho, z) = \frac{1}{2\pi} \int_0^{2\pi} \Phi(z + i\rho \sin \alpha) d\alpha = \Phi_0 + U I_0(\pi\rho/a) \cos \pi \frac{z}{a}, \quad (4.132)$$

where

$$I_0(\rho/a) = \frac{1}{2\pi} \int_0^{2\pi} e^{-\pi \frac{\rho}{a} \sin \alpha} d\alpha \quad (4.133)$$

is the modified Bessel function of zero order. This function has the asymptotic form

$$I_0(x) \approx \begin{cases} 1 + \frac{x^2}{4} & \text{for } x \ll 1, \\ \frac{e^x}{\sqrt{2\pi x}} & \text{for } x \gg 1. \end{cases} \quad (4.134)$$

We suppose that the bore radius of the circular aperture holes r_a is appreciably smaller than a , and we center the apertures at the planes $z = ka$, $k = 0, 1, \dots, n$. Setting $z = 0$ and $\rho = r_a$, we obtain

$$\varphi(r_a, 0) = \Phi_0 + \Phi_a = \Phi_0 + U I_0\left(\pi \frac{r_a}{a}\right) \approx \Phi_0 + U. \quad (4.135)$$

Hence, in the case $r_a \leq a/2\pi$, the axial potential at the center of the aperture is roughly equal to the potential at the aperture electrode ($U \approx \Phi_a$). We further assume that the aperture electrodes adopt the shape of plane plates with thickness $a/2$ at radial distances $\rho \geq 0.4a$. Then the largest electric field strength is $E_a = 4\Phi_a/a$ between any two adjacent electrodes. In the relativistic case, the acceleration voltage is large compared with the voltage applied to the electrodes. Therefore, we can employ (4.108) for the focal length of a weak lens and substitute Φ_0 for Φ in the denominator, giving

$$\begin{aligned}
\frac{1}{f_{\text{re}}} &\approx \frac{2 + (1 + 2\varepsilon\Phi_0)^2}{16\Phi_0^2(1 + \varepsilon\Phi_0)^2} \int_0^l \Phi'^2 dz = \frac{\Phi_a^2}{\Phi_0^2} \frac{2 + (1 + 2\varepsilon\Phi_0)^2}{16(1 + \varepsilon\Phi_0)^2} \frac{\pi^2}{a^2} n \int_0^a \sin^2(\pi z/a) dz \\
&= \frac{\pi^2}{128} \frac{l E_a^2}{\Phi_0^2} \frac{2 + (1 + 2\varepsilon\Phi_0)^2}{4(1 + \varepsilon\Phi_0)^2}.
\end{aligned} \tag{4.136}$$

Now, we consider a sequence of n electrostatic quadrupoles. The polarity of the electrodes of subsequent quadrupoles alternates along the optic axis. The thickness of each element and the spacing between two neighboring elements are both $a/2$. Their bore radius is r_q , and the gap width between two electrodes of a quadrupole is $g = r_q\pi/4$. Applying a voltage $\pm U$ to these electrodes, we obtain with a sufficient degree of accuracy the quadrupole strength

$$\Phi_2 \approx \frac{U}{r_q^2} = \frac{\pi U}{4gr_q} = \frac{\pi E_q}{8r_q}. \tag{4.137}$$

Here $E_q = 2U/g$ is the electric field strength between two neighboring electrodes having voltages U and $-U$, respectively. To obtain a convergent element, we combine two elements to a quadrupole doublet whose focal length (4.127) is the same for the $x - z$ section and the $y - z$ section:

$$\frac{1}{f_D} = \frac{a}{f_x^2} = \frac{a}{f_y^2}, \quad \frac{1}{f_y} = -\frac{1}{f_x} \approx \frac{1 + 2\varepsilon\Phi_0}{(1 + \varepsilon\Phi_0)\Phi_0} \int_0^{a/2} \Phi_2 dz \approx \frac{\pi}{16} \frac{a E_q}{r_q \Phi_0} \frac{1 + 2\varepsilon\Phi_0}{1 + \varepsilon\Phi_0}. \tag{4.138}$$

By employing these results, the total focal length of the $n/2$ doublets is given by

$$\frac{1}{f_{\text{qe}}} \approx \frac{n}{2f_D} \approx \frac{\pi^2}{128} \frac{la^2}{r_q^2} \frac{E_q^2}{\Phi_0^2} \frac{(1 + 2\varepsilon\Phi_0)^2}{4(1 + \varepsilon\Phi_0)^2}. \tag{4.139}$$

If we assume maximum tolerable electric field strength $E_a = E_q = E_{\text{max}}$ for both the rotationally symmetric system and the quadrupole system, we find for the ratio of their total focal lengths (4.136) and (4.139) the expression

$$\frac{f_{\text{re}}}{f_{\text{qe}}} \approx \frac{a^2}{r_q^2} \frac{(1 + 2\varepsilon\Phi_0)^2}{2 + (1 + 2\varepsilon\Phi_0)^2}. \tag{4.140}$$

This ratio does not depend on the total length of the system. The axial extension a of a thick quadrupole is large compared with its bore radius r_q . Therefore, we can make the ratio (4.140) much larger than 1. In this case, the quadrupole system has a significantly larger refraction power than the equivalent round-lens system. The formulas (4.136) and (4.139) are valid for weak focusing $f_{\text{re}} > l$, $f_{\text{qe}} > l$, whereas the ratio (4.140) does not have this restriction because it does not depend on the

total length l of the system. This ratio only requires that the constituent elements are short lenses, which is always the case for relativistic particles ($\varepsilon\Phi_0 \gg 1$). If we perform the same considerations for the magnetic case, we find that the resulting ratio

$$\frac{f_{\text{rm}}}{f_{\text{qm}}} \approx \frac{a^2}{r_q^2} \quad (4.141)$$

does not depend on the acceleration voltage. This ratio coincides with that for the electrostatic case in the relativistic limit. By employing the formulas (4.127) and (4.130) together with (4.118), we eventually find the ratio of the focal length of a short antisymmetric electric quadrupole doublet and that of the equivalent magnetic doublet as

$$\frac{f_{\text{De}}}{f_{\text{Dm}}} \approx \frac{1 + \varepsilon\Phi_0}{(1 + 2\varepsilon\Phi_0)^2} \frac{2e\Phi_0 B_q^2}{m_e E_q^2} = v_0^2 \frac{B_q^2}{E_q^2} = \beta_0^2 \frac{c^2 B_q^2}{E_q^2}. \quad (4.142)$$

To survey this ratio, we assume maximum achievable field strengths:

$$\begin{aligned} B_q &= B_{\text{max}} \approx 2 \text{ Tesla} = 2 \cdot 10^{-4} \frac{\text{Vs}}{\text{cm}^2}, \\ E_q &= E_{\text{max}} \approx 10 \frac{\text{kV}}{\text{mm}} = 10^5 \frac{\text{V}}{\text{cm}}. \end{aligned} \quad (4.143)$$

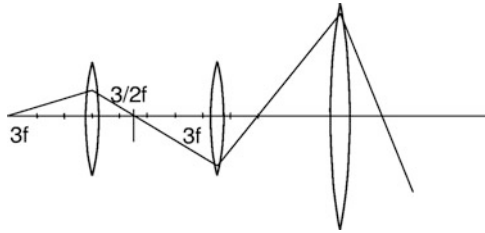
Using these values together with the value $c \approx 3 \cdot 10^{10} \text{ cm/s}$ for the velocity of light, we find

$$\frac{f_{\text{De}}(E_{\text{max}})}{f_{\text{Dm}}(B_{\text{max}})} \approx 3,600\beta_0^2. \quad (4.144)$$

This result demonstrates that alternating magnetic quadrupoles focus relativistic particles most efficiently. Therefore, one characterizes this focusing of relativistic particles as “strong focusing.” For nonrelativistic particles such as heavy ions at voltages smaller than several 100 kV, electric quadrupoles focus more strongly than magnetic quadrupoles. The same holds true for round lenses. This is the reason why one employs primarily magnetic round lenses for focusing electrons in electron microscopes, whereas one employs electrostatic lenses for focusing ions in ion microprobes operating at voltages below 50 kV.

However, strong focusing does not necessarily imply that the trajectories remain confined in the paraxial domain. Although the particles oscillate about the optic axis, the amplitude of their oscillations may build up. This phenomenon is quite general and occurs if two oscillations entangle with each other. In the case of beam-guiding systems, the oscillations of the particles about the optic axis may be entangled with the periodicity of the lens sequence. One utilizes the buildup of the amplitude in the electron microscope to realize large magnifications for a given length of the microscope column. Figure 4.23 shows an example of this buildup for a sequence of thin round lenses separated by a distance $4.5f$, where f is the focal length of each

Fig. 4.23 Buildup of the amplitude of the divergent trajectory in an unstable system



lens. The chosen trajectory starts at a distance $3f$ in front of the first lens. In the electron microscope, this trajectory represents a field ray $u = u_0 u_\gamma$, which intersects the optic axis at the diffraction plane. The amplitude of the corresponding axial rays decreases along the optic axis because this ray intersects the axis at the distance $3f/2$ in front of the first lens. This behavior is a consequence of the Helmholtz–Lagrange relation for the two linearly independent fundamental rays.

We must avoid increasing amplitudes of the oscillations in accelerators and storage rings to prevent that particles hit the boundaries and are lost. In order to achieve stability, the arrangement and the excitation of the lenses must satisfy specific conditions. Then paraxial trajectories of particles with nominal energy remain confined within the paraxial domain along their entire course. Nevertheless, even in this case, particles may escape from the stable paraxial domain due to buildup effects induced by nonlinear forces and chromaticity.

To investigate the stability requirements, we start with a system of identical round lenses, each two separated by the same distance L representing the periodicity length of the system. We suppose that the complex trajectory u has the initial values $u_0 = u(z_0)$ and $u'_0 = u'(z_0)$ at an arbitrary initial plane z_0 , which we choose as the object plane. By employing the form (4.52) for the Gaussian trajectory, the position and slope of the trajectory at the plane $z_1 = z_0 + L$ are then

$$\begin{aligned} u(z_1) &= u_1 = u_0 u_{\pi 1} + u'_0 u_{\alpha 1}, \\ u'(z_1) &= u'_1 = u_0 u'_{\pi 1} + u'_0 u'_{\alpha 1}. \end{aligned} \quad (4.145)$$

We rewrite these equations in matrix form, giving

$$\begin{pmatrix} u_1 \\ u'_1 \end{pmatrix} = M_r \begin{pmatrix} u_0 \\ u'_0 \end{pmatrix}, \quad M_r = M_r(z_1, z_0) = \begin{pmatrix} u_{\pi 1} & u_{\alpha 1} \\ u'_{\pi 1} & u'_{\alpha 1} \end{pmatrix}. \quad (4.146)$$

The elements of the round-lens transfer matrix M_r are the values of the position and slope of the axial ray u_α and the image principal ray u_π (4.51) taken at the plane z_1 . Hence, the transfer matrix at a distance of N period lengths is

$$M_r(NL + z_0) = (M_r(z_1, z_0))^N. \quad (4.147)$$

In order to achieve stability, the elements of this matrix must not exceed a given limit. To find a criterion for this condition, we consider the eigenvalue equation of the matrix M_r :

$$\lambda \begin{pmatrix} u_0 \\ u'_0 \end{pmatrix} = M_r \begin{pmatrix} u_0 \\ u'_0 \end{pmatrix}. \quad (4.148)$$

This set of two equations has nontrivial solutions for u_0 and u'_0 only if $\det(M_r - I\lambda) = 0$, yielding

$$\lambda^2 - \lambda(u_{\pi 1} + u'_{\alpha 1}) + u_{\pi 1}u'_{\alpha 1} - u_{\alpha 1}u'_{\pi 1} = 0. \quad (4.149)$$

Considering the Helmholtz–Lagrange relation $u_{\pi}u'_{\alpha} - u_{\alpha}u'_{\pi} = 1$ in the case of constant electric potential ($\Phi_0 = \Phi_1$), the two solutions of (4.149) are

$$\lambda_{1,2} = \frac{u_{\pi 1} + u'_{\alpha 1}}{2} \pm \sqrt{\left(\frac{u_{\pi 1} + u'_{\alpha 1}}{2}\right)^2 - 1}. \quad (4.150)$$

The solutions λ_1 and λ_2 are real for $|u_{\pi 1} + u'_{\alpha 1}| \geq 2$; they are complex or imaginary else.

The system shown in Fig. 4.23 has diverging properties. To verify this behavior, we determine the eigenvalues of the corresponding transfer matrix for a single period L . We readily obtain the elements of this matrix by assuming thin lenses and by considering the starting values (4.51) of the rays u_{α} and u_{π} at the plane $z_0 = z_o$. Employing the Gaussian construction for the trajectories, we find $u_{\alpha 1} = 0$, $u_{\pi 1} = -1/2$, $u'_{\alpha 1} = -2$, $u'_{\pi 1} = -1/f$, giving eigenvalues

$$\lambda_1 = \lambda_{\lambda} = -1/2, \quad \lambda_2 = \lambda_{\pi} = -2. \quad (4.151)$$

Since these values are real and differ from one, the system does not have stable solutions. The amplitude of the axial ray u_{α} decreases by a factor $1/2$ for each period L , whereas that of the principal ray u_{π} increases by a factor of 2. These trajectories have this property in all magnifying electron microscopes.

Without loss of generality we may define

$$\cos \mu = (u_{\pi 1} + u'_{\alpha 1})/2. \quad (4.152)$$

The parameter μ is either real, or complex or imaginary depending on the value of $(u_{\pi 1} + u'_{\alpha 1})/2$. By substituting $\cos \mu$ for this quantity in (4.150), its solutions adopt the simple form

$$\lambda_{1,2} = \cos \mu \pm i \sin \mu = e^{\pm i\mu}. \quad (4.153)$$

Hence, if μ is real, the motion will be stable. The eigensolutions u_{e1} and u_{e2} are generally complex functions, which satisfy the periodicity relation

$$u_{e1,2}(z + NL) = e^{\pm iN\mu} u_{1,2}(z). \quad (4.154)$$

Owing to the linearity of the Gaussian path equation, each of the two linearly independent eigensolutions can be written as a linear combination of the two fundamental rays u_α and u_π .

In order that the trajectories adopt their initial values after passing N cells, the parameter μ must satisfy the condition

$$\mu = 2\pi \frac{n}{N}, \quad n = 1, 2, \dots \quad (4.155)$$

The motion of the particles becomes instable for $\mu = \pm\pi$, which corresponds to $|u_{\pi 1} + u'_{\alpha 1}| = 2$. In this case, an infinitely small disturbance may cause a broadening of the beam.

In the case of a periodic arrangement of quadrupoles, we must treat the stability considerations for the horizontal $x - z$ section and the vertical $y - z$ section separately. We suppose that the principal sections of the quadrupoles coincide. Then we can adjust the coordinate system in such a way that the quadrupole strength $G(z) = \bar{G}(z)$ is real so that the complex path equation decouples in the two real equations (4.117) for the x - and the y -coordinate, respectively. The general solutions of these equations

$$x(z) = x_0 x_\pi(z) + \alpha x_\alpha(z), \quad y(z) = y_0 y_\pi(z) + \beta y_\beta(z) \quad (4.156)$$

are linear combinations of one of the two axial rays $x_\alpha = x_\alpha(z)$, $y_\beta = y_\beta(z)$ and one of the two image principal rays $x_\pi = x_\pi(z)$, $y_\pi = y_\pi(z)$. The initial conditions for these rays at the plane $z_0 = z_0$ are the same as those (4.51) for rotational symmetry. In this case, the two trajectories of each pair degenerate:

$$x_\alpha = y_\beta = u_\alpha, \quad x_\pi = y_\pi = u_\pi. \quad (4.157)$$

The focusing at a given plane in a quadrupole system is convergent in one of the two principal sections and divergent in the other. Therefore, the paths of rays differ for these sections, as illustrated in Fig. 4.24. In order to achieve stability for the quadrupole system, the transfer matrices for the $x - z$ section and the $y - z$ section

$$M_x(z_1, z_2) = \begin{pmatrix} x_{\pi 1} & x_{\alpha 1} \\ x'_{\pi 1} & x'_{\alpha 1} \end{pmatrix}, \quad M_y(z_1, z_0) = \begin{pmatrix} y_{\pi 1} & y_{\beta 1} \\ y'_{\pi 1} & y'_{\beta 1} \end{pmatrix} \quad (4.158)$$

must both have imaginary eigenvalues

$$\begin{aligned} \lambda_{x1,2} &= e^{\pm i\mu_x}, & \cos \mu_x &= (x_{\pi 1} + x'_{\alpha 1})/2, \\ \lambda_{y1,2} &= e^{\pm i\mu_y}, & \cos \mu_y &= (y_{\pi 1} + y'_{\beta 1})/2. \end{aligned} \quad (4.159)$$

Fig. 4.24 Course of the paraxial trajectory of an electron moving (a) in the $x - z$ section and (b) in the $y - z$ section of a system composed of alternating quadrupoles with strength $\pm G_0$

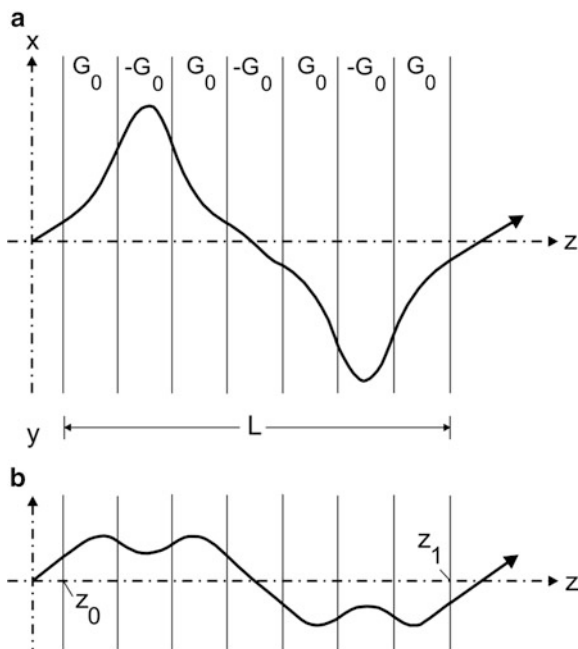
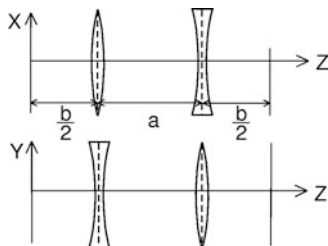


Fig. 4.25 FODO cell consisting of two thin quadrupoles with opposite polarity; the total length of the cell is $L = a + b$



Hence, we achieve paraxial stability in all sections if

$$|x_{\pi 1} + x'_{\alpha 1}| < 2, \quad |y_{\pi 1} + y'_{\beta 1}| < 2. \quad (4.160)$$

To satisfy these conditions, we must choose the free parameters of a cell appropriately. The simplest double focusing unit cell of a quadrupole system is the so-called FODO (focusing-drift-defocusing-drift) cell [90], which consists of two spatially separated quadrupoles with opposite polarity, as depicted schematically in Fig. 4.25. We assume thin quadrupoles with focal lengths $-f_{x1} = -f_{y1} = f_1$ and $-f_{x2} = f_{y2} = f_2$, respectively. The period length of the cell $L = a + b$ consists of the distance a between the two quadrupoles of the cell and the drift space b between the quadrupoles of any two adjacent cells.

We obtain the components of the two transfer matrices in thin-lens approximation most conveniently by multiplication of the matrices of the constituent elements of the cell or by successive application of the Gauss construction. The transfer matrices

of a thin quadrupole with focal lengths $\pm f$ are

$$M_{qc} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}, \quad M_{qd} = \begin{pmatrix} 1 & 0 \\ 1/f & 1 \end{pmatrix}, \quad (4.161)$$

where the subscripts “c” and “d” indicate the *convergent* section and the *divergent* section of the quadrupole, respectively. The transfer matrix for a distance l in field-free space has the form

$$M_{0l} = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \quad (4.162)$$

for both principal sections. After a lengthy yet straightforward calculations we eventually derive

$$\begin{aligned} x_{\pi 1} &= 1 - \frac{b+2a}{2f_1} + \frac{b}{2f_2} - \frac{ab}{2f_1 f_2}, & x'_{\alpha 1} &= 1 + \frac{b+2a}{2f_2} - \frac{b}{2f_1} - \frac{ab}{2f_1 f_2}, \\ y_{\pi 1} &= 1 + \frac{b+2a}{2f_1} - \frac{b}{2f_2} - \frac{ab}{2f_1 f_2}, & y'_{\beta 1} &= 1 - \frac{b+2a}{2f_2} + \frac{b}{2f_1} - \frac{ab}{2f_1 f_2}. \end{aligned} \quad (4.163)$$

We introduce the new parameters

$$v_1 = \frac{a+b}{2f_1} = \frac{L}{2f_1}, \quad v_2 = \frac{L}{2f_2}, \quad \varepsilon = \frac{2ab}{(a+b)^2} \quad (4.164)$$

and add the two expressions of each line, giving

$$\begin{aligned} x_{\pi 1} + x'_{\alpha 1} &= 2[1 + v_2 - v_1 - \varepsilon v_1 v_2], \\ y_{\pi 1} + y'_{\beta 1} &= 2[1 + v_1 - v_2 - \varepsilon v_1 v_2]. \end{aligned} \quad (4.165)$$

The geometry parameter ε can only take values in the range $0 \leq \varepsilon \leq 1/2$. The permissible values of v_1 and v_2 for stable motion lie inside the domain bounded by the four curves $v_2 = v_2^{(n)}(v_1)$, $n = 1, 2, 3, 4$, each of which satisfies one of the conditions

$$x_{\pi 1} + x'_{\alpha 1} = \pm 2, \quad y_{\pi 1} + y'_{\beta 1} = \pm 2. \quad (4.166)$$

These constraints together with the relations (4.165) yield the four branches of the boundary curve, which encloses the stability domain in the v_1, v_2 plane:

$$v_2^{(1)} = \frac{v_1}{1 - \varepsilon v_1}, \quad v_2^{(2)} = \frac{v_1 - 2}{1 - \varepsilon v_1}, \quad v_2^{(3)} = \frac{v_1}{1 + \varepsilon v_1}, \quad v_2^{(4)} = \frac{v_1 + 2}{1 + \varepsilon v_1}. \quad (4.167)$$

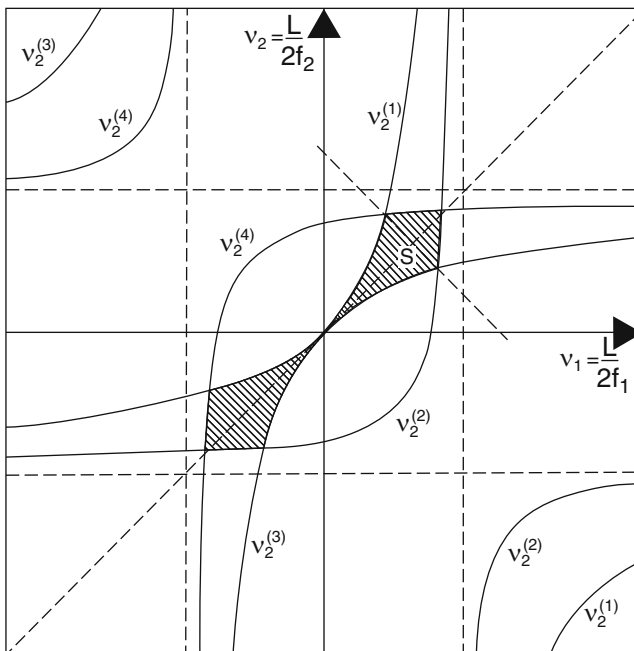


Fig. 4.26 Regions of stability in terms of the parameters v_1 and v_2 of the quadrupole doublet shown in Fig. 4.25 for a geometry parameter $\varepsilon = 0.4$

The stability domain consists of two sheets centered about the diagonal, as illustrated in Fig. 4.26. Each domain has the shape of a “necktie.” The intersection points $v_{1t} = v_{2t} = \pm\sqrt{2/\varepsilon}$ of the two curves $v_2^{(2)}$ and $v_2^{(4)}$ form the tips of the two stability diagrams. The point S of maximum stability has the largest distances from the boundaries of the stability domain. This point lies on the diagonal $v_1 = v_2 = v$ and has coordinates

$$v_{1s} = v_{2s} = v_s = \pm \frac{1}{2} \sqrt{1 + \frac{4}{\varepsilon}}. \quad (4.168)$$

The phase parameters μ_x and μ_y of the eigenvalues (4.159) coincide for points located on the diagonal:

$$\mu_x = \mu_y = \arccos(1 - \varepsilon v^2). \quad (4.169)$$

If we operate the system at the most stable point S with coordinates (4.1568), the range

$$(\Delta v_1)^2 + (\Delta v_2)^2 \leq 1/4 \quad (4.170)$$

of the permissible deviations $\Delta v_1 = v_1 - v_c$ and $\Delta v_2 = v_2 - v_c$ is a maximum.

Usually the point

$$\nu_1 = \nu_2 = \nu_s = \sqrt{1/\varepsilon} \quad (4.171)$$

is chosen as the stability point. Since for this point the phase parameters are $\mu_x = \mu_y = \mu = \pi/2$, we obtain periodicity of the trajectories (4.155) in four cells. As an example, we consider the quadrupole doublet shown in Fig. 4.21 for $b = 2a = 2f$, giving $L = a + b = 3a$, $\varepsilon = 4/9$, and $f_1 = f_2 = f_s = \nu_s L/2 = L/3 = a$. This result reveals that the telescopic quadrupole quadruplet discussed in Sect. 4.6.2 is composed of two of these cells.

Strongest focusing occurs if (4.168) adopts a minimum. This is the case for $\varepsilon = 1/2$ resulting in $a = b = L/2$ and $f_1 = f_2 = f_{\text{opt}} = \pm L/3$. For these values, each of the two boundary curves $\nu_2^{(2)}$ and $\nu_2^{(4)}$ degenerates in two orthogonal straight lines:

$$\nu_2^{(1)} = \frac{2\nu_1}{2 - \nu_1}, \quad \nu_2^{(2)} = -2, \quad \nu_1^{(2)} = 2, \quad \nu_2^{(3)} = \frac{2\nu_1}{2 + \nu_1}, \quad \nu_2^{(4)} = 2, \quad \nu_1^{(4)} = -2. \quad (4.172)$$

Each pair of orthogonal straight lines forms the tip of one of the two neckties.

The use of the more realistic rectangular model for extended quadrupoles does not change the general findings of the thin-lens approximation because the paraxial focusing strength (4.145) does not depend on derivatives of the quadrupole strengths. Therefore, this approximation is well suited to survey rapidly the focusing and stability properties of cells consisting of any number of quadrupoles and dipoles and to find optimum beam-guiding systems. We need the realistic field distribution only for precisely designing the final system.

4.6 Electrostatic Cylinder Lenses

Electrostatic cylinder lenses are arrangements of electrodes perpendicular to the optic axis that form planar electrostatic fields, as outlined in Sect. 3.1.3. The electrodes of cylinder lenses extend infinitely parallel to the x -axis. Figure 4.27 shows the cross section of a typical einzel cylinder lens. Within the frame of Gaussian optics, we can conceive an electric cylinder lens as a superposition of a round lens with a quadrupole.

According to the condition (3.43), the quadrupole strength $\Phi_2 = \Phi_2(z) = \Phi_{2c}(z)$ and the axial potential $\Phi = \Phi(z)$ are related which each other by

$$\Phi_2 = \frac{1}{4}\Phi''. \quad (4.173)$$

By employing this condition together with $\Gamma = 0$, $\Phi_1 = 0$, $\Psi_1 = 0$, $B = 0$, and $\Psi_2 = 0$, the complex path equation (4.13) reduces to

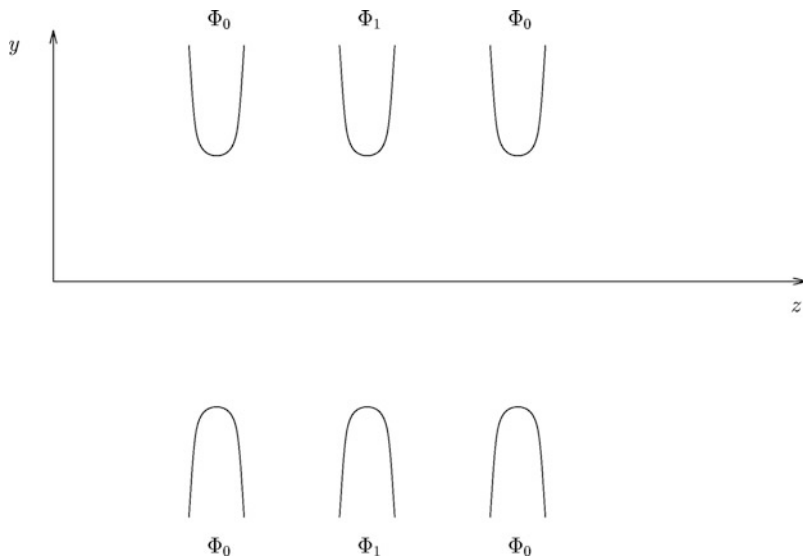


Fig. 4.27 Vertical $y - z$ cross section of the electrodes of an electrostatic cylinder einzel lens, $\Phi_1 = \Phi_m$ is the potential of the middle electrode and Φ_0 the potential of the column

$$w'' + \frac{\gamma_0 \Phi'}{2\Phi^*} w' + \frac{\gamma_0 \Phi''}{4\Phi^*} (w - \bar{w}) = 0. \quad (4.174)$$

The real part and the imaginary part of this equation form the decoupled set of differential equations

$$\begin{aligned} \sqrt{\Phi^*} \frac{d}{dz} \left(\sqrt{\Phi^*} x' \right) &= 0, \\ y'' + \frac{\gamma_0 \Phi'}{2\Phi^*} y' + \frac{\gamma_0 \Phi''}{2\Phi^*} y &= 0 \end{aligned} \quad (4.175)$$

for the x -component and the y -component of the paraxial ray $w = x(z) + iy(z)$. We can directly integrate the equation for the x -component, yielding

$$x' \sqrt{\Phi^*} = x'_0 \sqrt{\Phi_0^*} = \text{const.} \quad (4.176)$$

Hence, the slope of the x -component of the trajectory never changes sign. Accordingly, an electrostatic cylinder lens does not focus in the $x - z$ section. Therefore, such a lens cannot form a stigmatic image.

4.6.1 Modified Paraxial Equation

To survey easily the imaging properties in the $y - z$ section, it is useful to employ modified coordinates Y and Z such that the paraxial equation for the y -component adopts a form equivalent to that for round lenses. However, it is not possible to achieve this for the electrostatic cylinder lens with the transformation of only the lateral coordinate. Since we aim to eliminate the term with y' and that containing the second derivative of Φ , we need two free parameters or two transformations, respectively. In accordance with the transformation in the case of round lenses, we propose the transformations

$$y = \Phi^{*n} Y, \quad dz = \Phi^{*m} dZ. \quad (4.177)$$

The exponents n and m are free parameters. Using the abbreviation

$$\dot{Y} = \frac{dY}{dZ} \quad (4.178)$$

and the relations (4.177), we obtain

$$\begin{aligned} y' &= n\Phi^{*n-1}\gamma_0\Phi'Y + \Phi^{*n-m}\dot{Y}, \\ y'' &= \Phi^{*n-2m}\ddot{Y} + (2n-m)\gamma_0\Phi^{*n-m-1}\Phi'\dot{Y} + n\gamma_0\Phi^{*n-1}\Phi''Y \\ &\quad + n[2\varepsilon\Phi^* + (n-1)\gamma_0^2]\Phi^{*n-2}\Phi'^2Y. \end{aligned} \quad (4.179)$$

By setting $n = m = -1/2$, the two equations (4.175) adopt the reduced form

$$\Phi^*\ddot{Y} + QY = 0, \quad Q = \frac{1 + \gamma_0^2}{4} \frac{\Phi'^2}{\Phi^{*2}}, \quad (4.180)$$

$$\dot{x} = \dot{x}_0. \quad (4.181)$$

The cylinder lens is always convergent in the focusing $y - z$ section because the focusing strength Q is positive definite. Although the y -component can be bend away from the optic axis within a limited region of the lens field, this is not possible for the modified Y -component. Nevertheless, the total focusing is always convergent in the $y - z$ section. The modified path equation for the Y -component has the same form as the equation for the modified coordinate U of a round lens. We readily demonstrate this behavior by substituting T for Q/Φ^* in Eq. (4.180). However, the courses of the corresponding paraxial trajectories will be different in the $x - y - z$ coordinate system due to the different transformations of the paraxial path equation for the round lens and the cylinder lens.

The x -component of the paraxial ray of a cylinder lens forms the straight line

$$x = x_0 + \dot{x}_0(Z - Z_0) \quad (4.182)$$

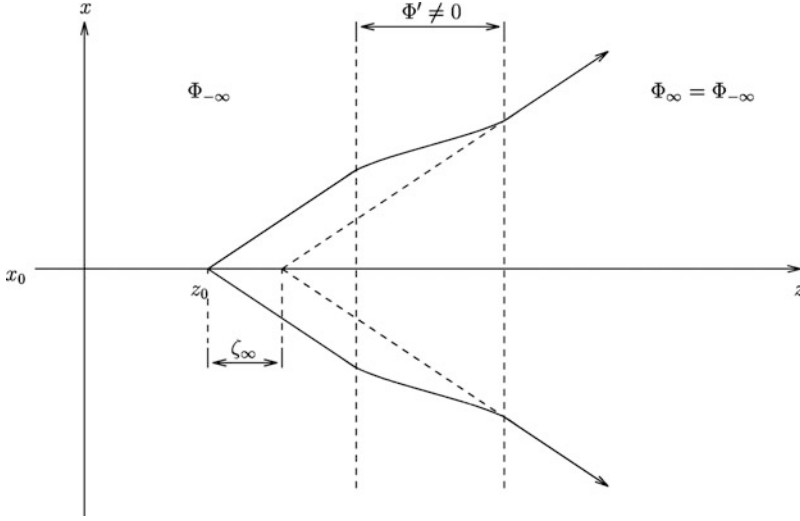


Fig. 4.28 Trajectory displacement in the horizontal $x-z$ section of an acceleration cylinder einzel lens ($\Phi_m > \Phi_0 = \Phi_{-\infty} = \Phi_{\infty}$)

in the $x-Z$ section, whereas its course is wavy in the $x-z$ section of the laboratory system, as illustrated in Fig. 4.28. We obtain the course of the x -component of the paraxial trajectory in this coordinate system by considering the second relation in (4.162) as

$$x = x_0 + x'_0 \int_{z_0}^z \sqrt{\frac{\Phi_0^*}{\Phi^*}} dz = x_0 + x'_0 \sqrt{\frac{\Phi_0^*}{\Phi^*}} (z - z_0 + \zeta), \quad (4.183)$$

where

$$\zeta = z_0 \left(\sqrt{\frac{\Phi_0^*}{\Phi^*}} - 1 \right) + \frac{\sqrt{\Phi_0^*}}{2} \int_{z_0}^z \frac{z \gamma_0 \Phi'}{\Phi^{*3/2}} dz \quad (4.184)$$

represents the axial offset of the ray tangent from its initial intersection point z_0 with the optic axis. In the case of an einzel lens ($\Phi_{\infty} = \Phi_{-\infty} = \Phi_0$), the cylinder lens acts in the horizontal $x-z$ section in the same way as a plane-parallel glass plate on light rays. Figure 4.28 shows schematically this situation.

The electrostatic cylinder einzel lens acts like a prism for the rays propagating in the horizontal $x-z$ section. We derive the total parallel displacement at the far side of the einzel lens from (4.184) as

$$\zeta_{\infty} = \frac{\sqrt{\Phi_{\infty}^*}}{2} \int_{-\infty}^{\infty} \frac{z \gamma_0 \Phi'}{\Phi^{*3/2}} dz. \quad (4.185)$$

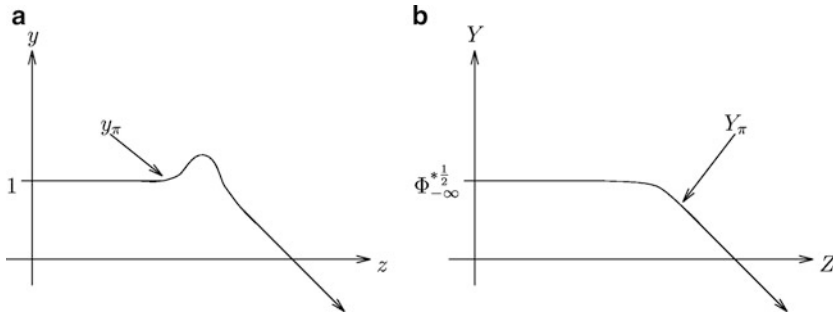


Fig. 4.29 Course (a) of the vertical image principal ray y_π and (b) its transformation Y_π in the $Y - Z$ section of the transformed coordinate system

Since Φ' is antisymmetric with respect to the midplane of a symmetric einzel lens, the shift of the horizontal ray components is unavoidable for such lenses.

4.6.2 Short Cylinder Lenses

To derive appropriate approximations for the cardinal elements of short cylinder lenses, we transform the paraxial path equation (4.180) in an integral equation, as in the case of round lenses. The resulting equation for the $Y - Z$ section has the form

$$Y = Y_0 + \dot{Y}_0(Z - Z_0) - \int_{Z_0}^Z \left\{ \int_{Z_0}^Z \frac{Q}{\Phi^*} Y d\tilde{Z} \right\} dZ. \quad (4.186)$$

This integral equation defines a trajectory with initial values $Y(Z_0) = Y_0$ and $\dot{Y}(Z_0) = \dot{Y}_0$ at the starting plane Z_0 .

Employing the transformed coordinates Y and Z is especially useful for obtaining in the first step of the iteration already a rather good approximation for the cardinal elements because for these coordinates, Q is quadratic in the electric field strength. Moreover, since (4.180) does not contain a term with $\dot{Y}$, the lateral distance of the transformed principal ray Y_π does not change appreciably within the field region of a short cylinder lens. This is not the case for the actual principal ray y_π , as demonstrated convincingly in Fig. 4.29.

For determining the cardinal elements of a short cylinder lens, we need to know the asymptotes of the principal rays.

The initial values of the transformed image principal ray $Y_\pi = \Phi^{*1/2} y_\pi$ at the starting plane $z_0 = -\infty$ are

$$Y_\pi(z_0 = -\infty) = Y_{\pi, -\infty} = \Phi_{-\infty}^{*1/2}, \quad \dot{Y}_{\pi, -\infty} = y'_{\pi, -\infty} + \frac{\gamma_{-\infty} \Phi'_{-\infty}}{2\Phi_{-\infty}^*} = 0. \quad (4.187)$$

With these constraints and the substitution

$$dZ = \Phi^{*1/2} dz, \quad (4.188)$$

the integral equation (4.186) defining the image principal ray y_π takes the form

$$\begin{aligned} Y_\pi &= y_\pi \Phi^{*1/2} = \Phi_{-\infty}^{*1/2} - \int_{-\infty}^z \sqrt{\Phi^*} \int_{-\infty}^{\tilde{z}} \frac{Q}{\sqrt{\Phi^*}} Y_\pi d\tilde{z} dz \\ &= \Phi_{-\infty}^{*1/2} - z \sqrt{\Phi^*} \int_{-\infty}^z \frac{Q}{\sqrt{\Phi^*}} Y_\pi d\tilde{z} + \int_{-\infty}^z z Q Y_\pi d\tilde{z} + \frac{1}{2} \int_{-\infty}^z \left\{ \frac{z \gamma_0 \Phi'}{\sqrt{\Phi^*}} \int_{-\infty}^{\tilde{z}} \frac{Q}{\sqrt{\Phi^*}} Y_\pi d\tilde{z} \right\} dz. \end{aligned} \quad (4.189)$$

We solve this equation by employing the method of successive iteration starting with no lens ($Q = 0$). This yields the initial trajectory $y_\pi = Y_\pi \Phi_{-\infty}^{*-1/2} = 1$ as the trivial solution in the absence of the lens. In the first approximation, we set $Y_\pi = \Phi_{-\infty}^{*1/2} = \text{const.}$ in the integrands of the last expression in (4.189). In the field-free image space ($\Phi'_\infty = 0$), the resulting expression for the asymptote of the principal ray differs from that of a round lens:

$$y_{\pi, \text{as}} = \left(\frac{\Phi_{-\infty}^*}{\Phi_\infty^*} \right)^{1/2} \left\{ 1 - z \int_{-\infty}^{\infty} \sqrt{\frac{\Phi_{-\infty}^*}{\Phi^*}} Q dz + \int_{-\infty}^{\infty} z Q dz + \frac{1}{2} \int_{-\infty}^{\infty} \left[\frac{z \gamma_0 \Phi'}{\sqrt{\Phi^*}} \int_{-\infty}^z \frac{Q}{\sqrt{\Phi^*}} d\tilde{z} \right] dz \right\}. \quad (4.190)$$

The prime difference is the last term, which does not exist for the round lens. This term does not affect the focal length $f = z_F - z_P$ but the location z_P of the principal plane. We obtain its location from the defining relation $y_{\pi, \text{as}}(z_P) = 1$ as

$$z_P = \frac{1}{\int_{-\infty}^{\infty} \sqrt{\Phi_{-\infty}^* \Phi^*} Q dz} \left\{ 1 - \sqrt{\frac{\Phi_{-\infty}^*}{\Phi_\infty^*}} + \int_{-\infty}^{\infty} z Q dz + \frac{1}{2} \int_{-\infty}^{\infty} \left[\frac{z \gamma_0 \Phi'}{\Phi^{*1/2}} \int_{-\infty}^z \frac{Q}{\Phi^{*1/2}} d\tilde{z} \right] dz \right\}. \quad (4.191)$$

We derive the location z_F of the focal plane from the condition $y_{\pi, \text{as}}(z_F) = 0$, giving

$$z_F = z_P + \frac{1}{\int_{-\infty}^{\infty} \sqrt{\frac{\Phi_{-\infty}^*}{\Phi^*}} Q dz}. \quad (4.192)$$

By employing the relation (4.180), we obtain for the focusing strength the expression

$$\frac{1}{f} = \frac{1}{z_F - z_P} = \int_{-\infty}^{\infty} \sqrt{\frac{\Phi_{-\infty}^*}{\Phi^*}} Q dz = \sqrt{\Phi_{-\infty}^*} \int_{-\infty}^{\infty} \frac{1 + \gamma_0^2}{4} \frac{\Phi'^2}{\Phi^{*5/2}} dz. \quad (4.193)$$

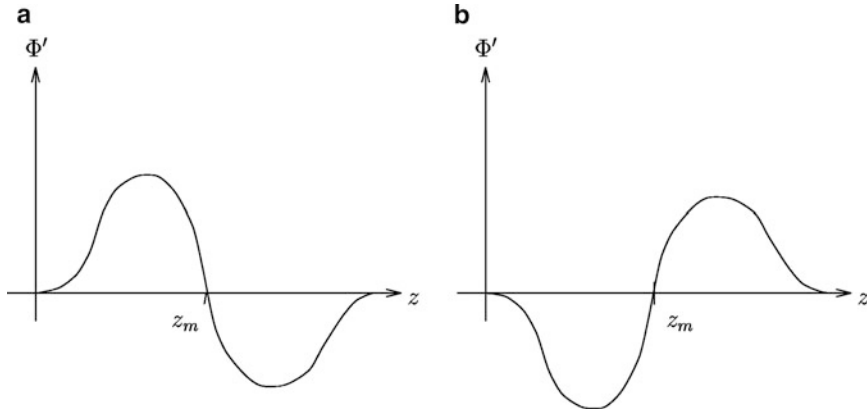


Fig. 4.30 Course of the negative axial electric field strength $\Phi'(z)$ within the region of a cylinder einzel lens in the case of (a) an acceleration lens ($\Phi_m > \Phi_0$) and (b) a deceleration lens ($\Phi_m < \Phi_0$)

We derive the locations of the object cardinal elements from (4.191), (4.192), and (4.193) by replacing in these formulas z by $-z$ and ∞ by $-\infty$. For an einzel cylinder lens ($\Phi_{-\infty} = \Phi_{\infty} = \Phi_0$), the image focal length f and the object focal length $\bar{f}$ coincide. In this case, the separation of the principal planes is

$$\Delta = z_P - z_{\bar{P}} = f \int_{-\infty}^{\infty} \left\{ \frac{z\gamma_0\Phi'}{2\sqrt{\Phi^*}} \left[\int_{-\infty}^z \frac{\mathcal{Q}}{\sqrt{\Phi^*}} d\tilde{z} + \int_z^{\infty} \frac{\mathcal{Q}}{\sqrt{\Phi^*}} d\tilde{z} \right] \right\} dz = \frac{1}{2\sqrt{\Phi_0^*}} \int_{-\infty}^{\infty} \frac{z\gamma_0\Phi'}{\Phi^{*1/2}} dz. \quad (4.194)$$

For a freestanding lens, the integral

$$\int_{-\infty}^{\infty} \frac{\gamma_0\Phi'}{2\sqrt{\Phi^*}} dz = \frac{1}{2} \int_{-\infty}^{\infty} \frac{d\Phi^*}{\sqrt{\Phi^*}} = \sqrt{\Phi_{\infty}^*} - \sqrt{\Phi_{-\infty}^*} = 0 \quad (4.195)$$

vanishes. Therefore, we can substitute $z - z_m$ for z in the integrand of the last integral in (4.194). The plane z_m denotes the location of the maximum or minimum of the axial potential ($\Phi'(z_m) = 0$). The potential adopts a maximum for an acceleration lens and a minimum for a deceleration lens, as shown in Fig. 4.30a and b. These figures demonstrate that the product $(z - z_m)\Phi'$ is everywhere negative within the acceleration lens and positive inside the deceleration lens. Hence, the separation distance Δ of the principal planes is positive for the deceleration lens ($\Phi_m < \Phi_0$) and negative ($\Phi_m > \Phi_0$) for the acceleration lens. It follows from the relation (4.179) that the principal planes are crossed ($\Delta < 0$) for the acceleration lens, but not for the deceleration lens ($\Delta > 0$).

By applying the same considerations to the trajectory displacement (4.195) in the $x - z$ section, we find that trajectory displacement ζ_{∞} is negative for the acceleration

lens and positive for the deceleration lens. Figure 4.28 illustrates this behavior in the case of an acceleration lens whose axial potential $\Phi_m = \Phi(z_m)$ at the center of the middle electrode is larger than the potential Φ_0 applied to the outer electrodes.

4.7 General Systems with Straight Axis

General systems with straight axis are arrangements of round lenses and quadrupoles with arbitrary azimuthal orientation about the straight optic axis. In this case, it is generally not possible to decouple the complex paraxial equation in two real equations, one for each of the two lateral components of the trajectory. The magnetic cylinder lens represents such inseparable system [91]. The light-optical analogue is a medium with an arbitrarily variable index of refraction. Caratheodory [92] investigated mathematically the paraxial imaging properties of such media and classified the systems according to their characteristic paraxial imaging properties.

Electron optical systems with straight optic axis do not have lateral components of the electromagnetic fields along this axis:

$$\Gamma = 0, \quad \Phi_1 = 0, \quad \Psi_1 = 0, \quad D = 0. \quad (4.196)$$

Considering these constraints, the complex path equation (4.29) adopts the form

$$U'' + TU - G\bar{U} = 0, \quad (4.197)$$

with

$$T = \bar{T} = \frac{2 + \gamma_0^2}{16} \frac{\Phi'^2}{\Phi^{*2}} + \frac{eB^2}{8m_e\Phi^*}, \quad G = \left(\gamma_0 \frac{\Phi_2}{\Phi^*} + i \sqrt{\frac{2e}{m_e\Phi^*}} \Psi_2 \right) e^{-2i\chi}. \quad (4.198)$$

We can decouple the complex differential equation (4.196) only if

$$G = \text{const } \bar{G}. \quad (4.199)$$

To determine the azimuthal orientation of the electric and the magnetic quadrupole fields, we write G in the form

$$G = G_r + iG_i = |G| e^{2i(\alpha_2 - \chi)}, \quad \alpha_2 = \frac{1}{2} \arctan \left(\frac{\gamma_0 \Phi_{2s} + \sqrt{2e\Phi^*/m_e} \Psi_{2c}}{\gamma_0 \Phi_{2c} - \sqrt{2e\Phi^*/m_e} \Psi_{2s}} \right). \quad (4.200)$$

Using this representation, we can formulate the separation condition (4.199) as $\chi' = \alpha_2'$ or

$$\sqrt{\frac{e}{2m_e\Phi^*}}B = \frac{d}{dz} \arctan \left(\frac{\gamma_0\Phi_{2s} + \sqrt{2e\Phi^*m_e}\Psi_{2c}}{\gamma_0\Phi_{2c} - \sqrt{2e\Phi^*m_e}\Psi_{2s}} \right). \quad (4.201)$$

The condition (4.199) implies that we can decouple the complex differential equations (4.197) only if the phase $\alpha_2 - \chi$ of G is constant. We obtain this situation in the following cases:

- Purely electrostatic systems with plane principal sections ($\Psi_2 = 0, B = 0, \alpha'_2 = 0$)
- Electric and magnetic quadrupole systems with common plane principal sections ($\alpha_{2e} = \alpha_{2m} \pm \pi/4, \alpha'_{2e} = \alpha'_{2m} = 0$)
- Twisted quadrupole with superposed axial magnetic field such that the twist angle coincides with the angle of Larmor rotation up to a constant angle ($2\alpha_2 = 2\chi + \text{const}$)
- Electric and magnetic quadrupoles with common plane sections and round lenses such that the axial magnetic field does not overlap with the fields of the quadrupoles ($B \cdot \Phi_2 = 0, B \cdot \Psi_2 = 0, \alpha_{2e} = \alpha_{2m} \pm \pi/4, \alpha'_{2e} = \alpha'_{2m} = 0$)

The case (b) is realized in quadrupole–octopole correctors capable of compensating for the unavoidable axial chromatic and the spherical aberrations of round lenses.

4.7.1 Inseparable Systems with Straight Axis

If the system does not fulfill the separation condition (4.201), we can decouple the complex path equation (4.197) by increasing the order of the separated differential equations for the real part U_r and the imaginary part U_i of the complex modified coordinate $U = U_r + iU_i$. The two real components satisfy the set of coupled differential equations:

$$\begin{aligned} U_r'' + (T - G_r)U_r - G_iU_i &= 0, \\ U_i'' + (T + G_r)U_i - G_iU_r &= 0. \end{aligned} \quad (4.202)$$

We eliminate the variable U_i in the second equation by means of the expression for this variable obtained from the first equation. As a result, we derive for the real part U_r of the complex off-axis coordinate the linear fourth-order differential equation

$$G_i \left(\frac{U_r''}{G_i} \right)'' + G_i \left(\frac{T - G_r}{G_i} U_r \right)'' + (T + G_r)U_r'' + (T^2 - G_r^2 - G_i^2)U_r = 0. \quad (4.203)$$

We readily derive the equation for the imaginary coordinate U_i by substituting in (4.203) $-G_r$ for G_r and U_i for U_r .

Let us assume that we have found in some way four linearly independent solutions $U_{r\mu} = U_{r\mu}(z)$, $\mu = 1, 2, 3, 4$, of Eq. (4.203). By inserting each of these

solutions in the first equation of (4.202), we obtain directly the corresponding imaginary part $U_{i\mu} = U_{i\mu}(z)$ of the complex trajectory

$$U_\mu = U_{r\mu}(z) + i U_{i\mu}(z). \quad (4.204)$$

Hence, each pair $U_{r\mu}, U_{i\mu}$ forms the components of one of the four linearly independent paraxial rays. Owing to the linearity of the complex path equation (4.197), we can write the general solution of this equation as the linear combination

$$U(z) = \sum_{\mu=1}^4 a_\mu U_\mu(z), \quad a_\mu = \bar{a}_\mu \quad (4.205)$$

with *real* coefficients a_μ . This requirement is necessary in order that U satisfies Eq. (4.197). More than four arbitrary parameters cannot exist because the two positional coordinates and the two slope components at the initial plane $z = z_0$ determine entirely the course of each trajectory. The phases of the linearly independent rays $U_\mu(z)$ are functions of the z -coordinate. Since these functions differ from each other, it is not possible to find a coordinate system such that two linearly independent rays propagate in orthogonal sections.

Employing the transformations (4.28) and (4.21), we can rewrite the solution (4.205) in the fixed $x - y - z$ coordinate system as

$$w = w(z) = \sum_{\mu=1}^4 w_\mu(z), \quad w_\mu = x_\mu + iy_\mu = e^{i\chi} \left(\frac{\Phi_o^*}{\Phi^*} \right)^{1/4} U_\mu. \quad (4.206)$$

For determining the imaging properties of inseparable systems, we utilize relations existing between the positions and the slopes of any two linearly independent paraxial rays.

4.7.2 Generalized Helmholtz–Lagrange Relations

The canonical momenta of particles emanating from a point source are orthogonal to the surfaces of the associated wave, or the surfaces of constant eikonal, respectively. As a result, fixed relations exist between the positions and the slopes of any two trajectories, even if the particles originate at different source points. Since we have four linearly independent rays, six relations exist in the most general case. We derive them most conveniently by means of the transformed path equation (4.197). The corresponding equations for the linearly independent solutions U_μ and $\bar{U}_\nu$ are

$$\begin{aligned} U_\mu'' + T U_\mu &= G \bar{U}_\mu, \\ \bar{U}_\nu'' + T \bar{U}_\nu &= \bar{G} U_\nu. \end{aligned} \quad (4.207)$$

By multiplying the first equation by $\overline{U}_v$, the second by U_μ , and subtracting the resulting equations from each other, we obtain

$$U_\mu \overline{U}_v'' - U_\mu'' \overline{U}_v = \frac{d}{dz} \left(U_\mu \overline{U}_v' - \overline{U}_v U_\mu' \right) = \overline{G} U_\mu U_v - G \overline{U}_\mu \overline{U}_v = 2i \operatorname{Im} (\overline{G} U_\mu U_v). \quad (4.208)$$

Integration with respect to z yields

$$U_\mu \overline{U}_v' - \overline{U}_v U_\mu' = K_{\mu v} + 2i \operatorname{Im} \int_{z_0}^z \overline{G} U_\mu U_v dz, \quad K_{\mu v} = -\overline{K}_{v\mu} = U_{\mu 0} \overline{U}_{v0}' - \overline{U}_{v0} U_{\mu 0}'. \quad (4.209)$$

The index 0 indicates the value at the starting plane $z = z_0$. By taking the real part of Eq. (4.209), we obtain the generalized Helmholtz–Lagrange relations

$$\operatorname{Re} (U_\mu \overline{U}_v' - \overline{U}_v U_\mu') = \operatorname{Re} (\overline{U}_\mu U_v' - \overline{U}_v U_\mu') = C_{\mu v} = \overline{C}_{\mu v} = -C_{v\mu} = \operatorname{Re} K_{\mu v}, \quad \mu, v = 1, 2, 3, 4. \quad (4.210)$$

These six relations represent the paraxial approximations of the Lagrange brackets of classical mechanics formed by the exact trajectories. The values of their position slope coordinates at the initial plane define the constant $K_{\mu v}$. By employing the transformation (4.206), we can rewrite (4.210) as

$$\sqrt{\Phi^*} \operatorname{Re} (\bar{u}_\mu u_v' - \bar{u}_v u_\mu') = \sqrt{\Phi^*} \operatorname{Re} (\bar{w}_\mu w_v' - \bar{w}_v w_\mu' - 2i \chi' \bar{w}_\mu w_v) = C_{\mu v} \sqrt{\Phi_0^*}. \quad (4.211)$$

We obtain another important relation by taking the imaginary part of (4.209), yielding

$$\operatorname{Im} (U_\mu \overline{U}_v' - \overline{U}_v U_\mu') = \operatorname{Im} K_{\mu v} + 2 \operatorname{Im} \int_{z_0}^z G \overline{U}_\mu \overline{U}_v dz. \quad (4.212)$$

To illustrate the significance of this relation, we choose the axial fundamental rays U_α and U_β as the two linearly independent rays and the object plane z_0 as initial plane z_0 , resulting in $K_{\alpha\beta} = 0$. We assume that both axial rays intersect the optic axis at another plane $z = z_i$. Accordingly, the expression on the left-hand side of Eq. (4.212) is zero at this stigmatic image plane. In the case of undistorted paraxial imaging, we obtain a stigmatic image for any location of the object plane if this plane and the image plane are located in field-free regions. According to this behavior, the integral must vanish at any plane z in the field-free image space:

$$\operatorname{Im} \int_{z_0}^{z_i} G \overline{U}_\alpha \overline{U}_\beta dz = 0. \quad (4.213)$$

Another invariant, formed by four solutions of the complex path equation (4.197) and their derivatives, is the Wronski determinant or Wronskian, respectively

$$D_W = \begin{vmatrix} U_1 & \overline{U}'_1 & \overline{U}_1 & U'_1 \\ U_2 & \overline{U}'_2 & \overline{U}_2 & U'_2 \\ U_3 & \overline{U}'_3 & \overline{U}_3 & U'_3 \\ U_4 & \overline{U}'_4 & \overline{U}_4 & U'_4 \end{vmatrix} = \overline{D}_W. \quad (4.214)$$

This determinant is only nonzero if all four trajectories are linearly independent. The Wronskian is real, as follows by taking the conjugate complex of (4.214) and rearranging the columns of the resulting determinant. In order to evaluate the determinant, we consider that it changes sign if we exchange two columns. Hence, we can write

$$\begin{aligned} D_W &= \sum_p (-)^p U_\kappa \overline{U}'_\lambda \overline{U}_\mu U'_\nu = \frac{1}{2} \sum_p (-)^p (U_\kappa \overline{U}'_\lambda - U_\lambda \overline{U}'_\kappa) \overline{U}_\mu U'_\nu \\ &= \frac{1}{2} \sum_p (-)^p (U_\kappa \overline{U}'_\lambda - \overline{U}_\lambda U'_\kappa + \overline{U}_\kappa U'_\lambda - U_\lambda \overline{U}'_\kappa) \overline{U}_\mu U'_\nu \\ &= \sum_p (-)^p \overline{U}_\mu U'_\nu \operatorname{Re} (U_\kappa \overline{U}'_\lambda - \overline{U}_\lambda U'_\kappa) = \sum_p (-)^p C_{\kappa\lambda} \overline{U}_\mu U'_\nu. \end{aligned} \quad (4.215)$$

To obtain the second row of this expression, we have added two determinants, which are zero because each has two identical columns. We have derived the last expression by employing the Helmholtz–Lagrange relation (4.210). To evaluate this result further, we split it up in two halves and take the conjugate complex of the second half. This procedure does not alter its value because the Wronskian is real. Subsequently, we split up each term once more and exchange the indices in one of each two pairs. We account for the resulting change of sign by multiplying the corresponding terms by -1 giving

$$\begin{aligned} D_W &= \frac{1}{2} \sum_p (-)^p C_{\kappa\lambda} (\overline{U}_\mu U'_\nu + U_\mu \overline{U}'_\nu) \\ &= \frac{1}{4} \sum_p (-)^p C_{\kappa\lambda} (\overline{U}_\mu U'_\nu + U_\mu \overline{U}'_\nu - \overline{U}'_\mu U_\nu - \overline{U}_\nu U'_\mu) \\ &= \frac{1}{2} \sum_p (-)^p C_{\kappa\lambda} \operatorname{Re} (U_\mu \overline{U}'_\nu - \overline{U}_\nu U'_\mu) = \frac{1}{2} \sum_p (-)^p C_{\kappa\lambda} C_{\mu\nu}. \end{aligned} \quad (4.216)$$

The summation runs over all $4! = 24$ permutations p of the four indices κ, λ, μ , and ν . We evaluate the last sum by considering the relation $C_{\mu\nu} = -C_{\nu\mu}$, giving

$$\begin{aligned}
D_W &= \sum_p (-)^p U_\kappa \bar{U}'_\lambda \bar{U}_\mu U'_\nu = \frac{\Phi^*}{\Phi_o^*} \sum_p (-)^p w_\kappa \bar{w}'_\lambda \bar{w}_\mu w'_\nu \\
&= 4(C_{12}C_{34} - C_{13}C_{24} + C_{14}C_{23}). \tag{4.217}
\end{aligned}$$

The determinant (4.214) degenerates into a product of two 2×2 determinants if the path equation (4.197) can be decoupled.

Apart from the Wronskian (4.214), we can form other invariants existing between the rays and their derivatives. We obtain these invariants by forming sums of 4×4 determinants, which have two identical columns. Hence, the resulting invariants are zero. By employing the generalized Helmholtz–Lagrange relations (4.210), we can rewrite the sums in the form

$$\begin{aligned}
\sum_p (-)^p C_{\kappa\lambda} \bar{U}_\mu U_\nu &= \sqrt{\frac{\Phi^*}{\Phi_o^*}} \sum_p (-)^p C_{\kappa\lambda} \bar{w}_\mu w_\nu = 0, \\
\sum_p (-)^p C_{\kappa\lambda} U_\mu U'_\nu &= \sqrt{\frac{\Phi^*}{\Phi_o^*}} \sum_p (-)^p C_{\kappa\lambda} w_\mu w'_\nu = 0, \\
\sum_p (-)^p C_{\kappa\lambda} U'_\mu \bar{U}'_\nu &= \sqrt{\frac{\Phi^*}{\Phi_o^*}} \sum_p (-)^p C_{\kappa\lambda} w'_\mu \bar{w}'_\nu + 2i D_W \chi' = 0. \tag{4.218}
\end{aligned}$$

We shall need these invariants in the context of establishing an algorithm for calculating aberrations.

4.7.3 Imaging Properties

We can derive some information about the imaging properties of inseparable systems with straight axis by means of the general solution (4.216) of the complex path equation (4.197) and the Helmholtz–Lagrange relations (4.211). For reasons of simplicity, we define the four particular solutions $w_\mu = w_\mu(z)$ by specific initial values at the object plane $z = z_o$, given by

$$w_1(z_o) = w_{1o} = 0, \quad w'_{1o} = 1, \quad w_{2o} = 0, \quad w'_{2o} = i, \quad w_{3o} = 1, \quad w'_{3o} = 0, \quad w_{4o} = i, \quad w'_{4o} = 0. \tag{4.219}$$

By using these particular solutions, the trajectory, which intersects the object plane at position with slope $w'_o = x'_o + iy'_o$, has the simple form

$$w(z) = \sum_{\mu=1}^4 a_\mu w_\mu(z) = x'_o w_1 + y'_o w_2 + x_o w_3 + y_o w_4. \tag{4.220}$$

To investigate the conditions for the formation of an astigmatic or stigmatic image, it suffices to consider a homocentric bundle of rays originating at the center $x_o = 0$, $y_o = 0$ of the object plane. In order to achieve stigmatic imaging at plane $z = z_i$, the components of the axial rays must vanish at this plane:

$$x_1(z_i) = y_1(z_i) = x_2(z_i) = y_2(z_i) = 0. \quad (4.221)$$

Hence, at least three free parameters are necessary to achieve a stigmatic image at some image plane.

No additional conditions are necessary for astigmatic imaging. In this case, the system images any point of the object plane in a straight line in the image plane, yielding a line-to-line correlation between a specially oriented line grating at the object plane and its image [93].

An astigmatic image forms at a plane $z = z_i$ where one of the two axial rays is zero or they are located on a straight line intersecting the optic axis. This is the case if the determinant

$$D_{12}(z, z_o) = x_1(z)y_2(z) - x_2(z)y_1(z) = 0. \quad (4.222)$$

The determinant is zero at the object plane and may change its sign several times with increasing distance $z - z_o$ from the object plane depending on the strength and the extension of the field of the round lenses and quadrupoles. The number of astigmatic images is equal to the number of sign changes of the determinant. Because the paraxial rays $w_1 = x_1 + iy_1$ and $w_2 = x_2 + iy_2$ are straight lines in the field-free domain on the far side of the system, they form two line images, each of which is either real or virtual. To prove this conjecture, we assume that we know their lateral positions and slopes at the plane $z = z_o$, yielding

$$x_v = x_{v0} + x'_{v0}(z - z_o), \quad y_v = y_{v0} + y'_{v0}(z - z_o), \quad v = 1, 2. \quad (4.223)$$

By inserting these relations into Eq. (4.222), we derive the quadratic equation

$$(z - z_o)^2(x'_{10}y'_{20} - x'_{20}y'_{10}) + (z - z_o)(x'_{10}y_{20} + x_{10}y'_{20} - x'_{20}y_{10} - x_{20}y'_{10}) + x_{10}y_{20} - x_{20}y_{10} = 0 \quad (4.224)$$

for the locations $z = z_1$ and $z = z_2$ of the two astigmatic images. Employing the Lagrange–Helmholtz relation (4.211) for the axial trajectories w_1 and w_2 , and considering $\chi' = 0$ and $C_{12} = 0$, the solutions of Eq. (4.224) adopt the form

$$z_{1,2} - z_o = -\frac{\text{Im}(w_{20}\bar{w}'_{10} + \bar{w}_{10}w'_{20}) \pm |w_{10}w'_{20} - w'_{10}w_{20}|}{2\text{Im}(\bar{w}'_{10}w'_{20})}. \quad (4.225)$$

Two real line images are formed in the field-free image space if the two expressions on the right-hand side are positive. A negative value corresponds to a virtual line

image. A stigmatic image results in the case $w_{10}w'_{20} = w'_{10}w_{20}$, where the locations of the two astigmatic images coincide:

$$z_1 = z_2 = z_0 - w_{10}/w'_{10}. \quad (4.226)$$

We obtain the constants $C_{\mu\nu}$ of the Lagrange–Helmholtz relations (4.211) by evaluating these expressions at the object plane $z = z_0$ by means of the initial values (4.219) of the fundamental rays w_ν , $\nu = 1, 2, 3, 4$, giving

$$C_{12} = C_{14} = C_{23} = 0, \quad C_{31} = C_{42} = 1, \quad C_{34} = 2\chi'_0. \quad (4.227)$$

Using these values, the first relation in (4.218) adopts the simple form

$$x_3y_1 - x_1y_3 + x_4y_2 - x_2y_4 = 2\chi'_0(x_1y_2 - x_2y_1). \quad (4.228)$$

The expression on the right-hand side vanishes at the astigmatic image plane $z = z_i$ according to the condition (4.222). The image lines at this plane enclose an angle θ_i with the x -direction. The slope of the central image line $\tan \theta_i = y_{1i}/x_{1i} = y_{2i}/x_{2i}$ determines the azimuthal orientation of each line image. The slope angle θ_i depends on the form and strength of the electromagnetic field and on the location z_0 of the object plane. In the case of astigmatic imaging, the system images a line grid, which has a specific azimuthal orientation in the object plane, into a line grid at the image plane. Accordingly, each image line is conjugate to a specific object line, which encloses an angle θ_0 with the x -direction. In order to determine this angle, we consider a trajectory originating at the point $x_0, y_0 = x_0 \tan \theta_0$ of the central object line. Since this line intersects the optic axis, the trajectory must also intersect the central image plane. Hence, the relation

$$\frac{y_i}{x_i} = \frac{x'_0 y_{1i} + y'_0 y_{2i} + x_0(y_{3i} + y_{4i} \tan \theta_0)}{x'_0 x_{1i} + y'_0 x_{2i} + x_0(x_{3i} + x_{4i} \tan \theta_0)} = \tan \theta_i \quad (4.229)$$

must hold for arbitrary values x'_0 and y'_0 of the slope components of an axial ray intersecting the center of the object plane. Considering in addition the relations (4.228) and (4.222), we find the slope of the object lines as

$$\tan \theta_0 = \frac{x_{3i}y_{1i} - y_{3i}x_{1i}}{y_{4i}x_{1i} - x_{4i}y_{1i}} = -\frac{y_{4i}x_{2i} - x_{4i}y_{2i}}{y_{4i}x_{1i} - x_{4i}y_{1i}} = -\frac{x_{2i}}{x_{1i}} = -\frac{y_{2i}}{y_{1i}}. \quad (4.230)$$

The relations (4.229) and (4.230) show that the angles θ_0 and θ_i of the object lines and image lines are completely defined by the components of the axial rays w_1 and w_2 at the astigmatic image plane.

To readily survey the properties of astigmatic imaging, it is advantageous to define intrinsic fundamental trajectories, which are special linear combinations of the rays w_ν , $\nu = 1, 2, 3, 4$. The intrinsic axial ray w_α starts from the center of the object plane and intersects the center of the astigmatic image plane $z_1 = z_\alpha$, while

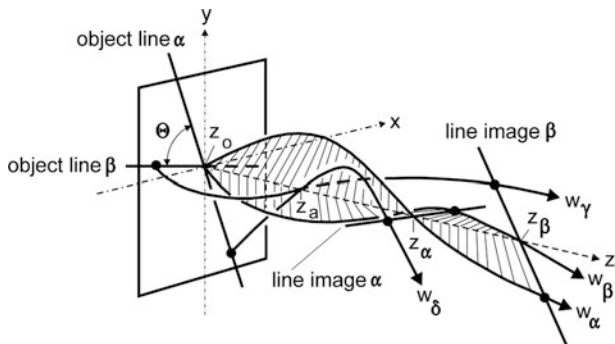


Fig. 4.31 Formation of astigmatic imaging within systems with inseparable Gaussian path equations demonstrated by the course of the intrinsic fundamental rays w_α , w_β , w_γ , and w_δ . Conjugate object and image lines are perpendicular to the (hatched) sections formed by the axial rays w_α and w_β , respectively

the other intrinsic axial ray w_β intersects the center of the second astigmatic image $z_2 = z_\beta$, as illustrated in Fig. 4.31. We choose the two intrinsic field rays such that w_γ originates at distance 1 from the central object line β , whereas w_δ starts at the same distance from the central object line α . Both rays intersect the optic axis at the aperture plane $z = z_a$. The two object lines α and β enclose an angle Θ , which differs from 90° for non-orthogonal systems. Using these definitions for the intrinsic fundamental rays, four of the corresponding Helmholtz constants are zero:

$$C_{\alpha\beta} = C_{\alpha\delta} = C_{\beta\gamma} = C_{\gamma\delta} = 0. \quad (4.231)$$

The lines of the two astigmatic images are orthogonal to each other if the images are located in the field-free region. In this region, the intrinsic fundamental rays are straight lines:

$$w_\alpha = w'_\alpha(z - z_\alpha), \quad w_\beta = w'_\beta(z - z_\beta). \quad (4.232)$$

By inserting these expressions into the Helmholtz–Lagrange relation (4.212) for the two intrinsic axial trajectories, we readily obtain

$$C_{\alpha\beta} = \text{Re}\{\bar{w}'_\alpha w'_\beta(z - z_\alpha) - \bar{w}'_\beta w'_\alpha(z - z_\beta)\} = (z_\beta - z_\alpha)\text{Re}\{\bar{w}'_\alpha w'_\beta\} = 0. \quad (4.233)$$

Since the two astigmatic image planes are spatially separated and the slopes w'_α and w'_β differ from zero, the last expression can only be satisfied if the directions of the slopes are perpendicular to each other.

The entity of transversal lines, each of which intersects the optic axis and a point of an intrinsic axial trajectory, forms an *image section*, which spirals along the optic axis in the region of the electromagnetic field. The corresponding object and image lines are perpendicular to this section. To prove this conjecture, we choose the image section α and consider that w_β is located on the central image

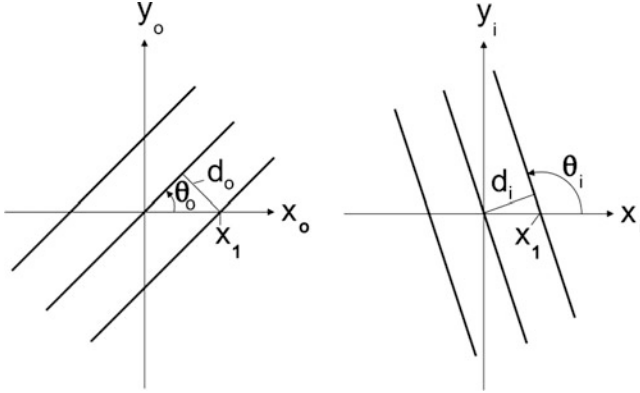


Fig. 4.32 Grid magnification $M_g = d_i/d_o$ in the case of a magnetic cylinder lens

line at the image plane z_α . It readily follows from the Helmholtz–Lagrange relation $\text{Re}\{\bar{w}'_\alpha(z_\alpha)w_\beta(z_\alpha)\} = 0$ at this plane that the central image line is perpendicular with respect to the image section α . Similarly, the Helmholtz–Lagrange relation $\text{Re}\{\bar{w}'_\alpha(z_o)w_\delta(z_o)\} = 0$ for the trajectories w_α and w_δ at the object plane z_o demonstrates that the object line is also perpendicular to the image section α because w_δ intersects this line, as shown in Fig. 4.31.

The constant $C_{\alpha\delta}$ is zero because w_δ also intersects the image line, which is orthogonal to $w'_\alpha(z_\alpha)$. The same behavior holds for the equivalent constant $C_{\beta\gamma}$.

To determine the magnification M_g of the line grid at the image plane z_α , we consider an object line, which intersects the point $w_{\gamma o} = w_\gamma(z_o)$. The vertical distance of this line from the optic axis is given by

$$d_o = \frac{\text{Re}\{w_{\gamma o}\bar{w}'_{\alpha o}\}}{|w'_{\alpha o}|}. \quad (4.234)$$

The corresponding distance at the image plane $z_\alpha = z_i$ is

$$d_i = \frac{\text{Re}\{w_\gamma(z_\alpha)\bar{w}'_\alpha(z_\alpha)\}}{|w'_\alpha(z_\alpha)|} = \frac{\text{Re}\{w_{\gamma o}\bar{w}'_{\alpha o}\}}{|w'_{\alpha o}|} \sqrt{\frac{\Phi_o^*}{\Phi_\alpha^*}}. \quad (4.235)$$

The resulting grid magnification (Fig. 4.32)

$$M_g = \frac{d_i}{d_o} = \left| \frac{w'_\alpha(z_o)}{w'_\alpha(z_\alpha)} \right| \sqrt{\frac{\Phi_o^*}{\Phi_\alpha^*}} \quad (4.236)$$

satisfies the familiar Helmholtz–Lagrange formula (4.58).

4.7.4 Paraxial Pseudo Rays

Any paraxial ray (4.206) of an arbitrary system with a straight optic axis can be written as a linear combination

$$w = a_1 w_1 + a_2 w_2 + a_3 w_3 + a_4 w_4 \quad (4.237)$$

of four linearly independent solutions $w_v = w_v(z) = x_v(z) + i y_v(z)$ of the paraxial path equation. The coefficients $a_v = \bar{a}_v$, $v = 1, 2, 3, 4$ are always real, while the trajectories $w_v(z)$ are generally complex.

In the presence of an axial magnetic field, it is advantageous to employ the rotating coordinate system $u = w e^{-i\chi}$. The trajectory (4.237) consists of a round-lens component and an astigmatic component. To elucidate this behavior, we introduce the paraxial pseudo rays

$$u_\omega = \frac{1}{2}(u_1 - i u_2), \quad u_{\bar{\omega}} = \frac{1}{2}(u_1 + i u_2), \quad u_\rho = \frac{1}{2}(u_3 - i u_4), \quad u_{\bar{\rho}} = \frac{1}{2}(u_3 + i u_4) \quad (4.238)$$

and the complex coefficients

$$\omega = a_1 + i a_2, \quad \rho = a_3 + i a_4. \quad (4.239)$$

Using these relations, we obtain for the paraxial ray in the rotating coordinate system the representation

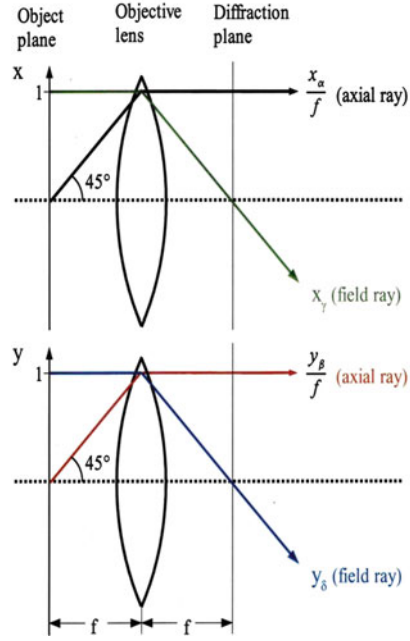
$$u = \omega u_\omega + \eta u_\rho + \bar{\omega} u_{\bar{\omega}} + \bar{\eta} u_{\bar{\rho}}. \quad (4.240)$$

The first and second term on the right-hand side of (4.225) define the round-lens component, while the two other terms describe the astigmatic component of the trajectory. In the case, of rotational symmetry, we have $u_4 = i u_3$, $u_2 = i u_1$, $u_1 = \bar{u}_1$, $u_3 = \bar{u}_3$, giving $u_{\bar{\omega}} = u_{\bar{\rho}} = 0$. The resulting trajectory

$$u = \omega u_1 + \eta u_3 \quad (4.241)$$

contains only the round component, as required by rotational symmetry. The pseudo rays (4.238) are real for orthogonal systems. In the case of plane principal sections ($\chi = 0$), each pseudo ray is a linear combination of a ray located in the x, z -section and a ray embedded in the y, z -section. Each of these rays is composed of the fundamental axial ray x_α or y_β and the field ray x_γ or y_δ , respectively. The course of the fundamental rays is depicted in Fig. 4.33 for the case that the diffraction plane coincides with the back focal plane of the objective lens. By employing these rays as the four linearly independent solutions of the paraxial path equations, we obtain for the pseudo rays the relations

Fig. 4.33 Definition of the fundamental axial rays x_α, y_β and of the field rays x_γ, y_δ for systems with decoupled paraxial path equations; for simplicity the object plane is placed at the front focal plane of the objective lens



$$u_\omega = \frac{1}{2}(x_\alpha + y_\beta), \quad u_{\bar{\omega}} = \frac{1}{2}(x_\alpha - y_\beta), \quad u_\rho = \frac{1}{2}(x_\gamma + y_\delta), \quad u_{\bar{\rho}} = \frac{1}{2}(x_\gamma - y_\delta). \quad (4.242)$$

Since the four fundamental rays are linearly independent in non-rotationally symmetric systems, any ray (4.240) consist of a round-lens component and an astigmatic component. Contrary to the astigmatic component, the round-lens component can never vanish within the entire system. Since the pseudo rays are *complex* linear combinations of the fundamental rays, they represent possible rays only in the degenerate case of rotational symmetry. Moreover, the pseudo rays may be symmetric or antisymmetric with respect to a given plane, although the fundamental rays are not. We can utilize this behavior to cancel out aberrations with twofold symmetry.

4.8 Systems with Curved Axis

Systems with curved axis are dispersive since the curvature of the optic axis depends on the velocity of the electron. Hence, an electron, which initially travels along the optic axis, remains on this axis only if the velocity of the particle coincides with the nominal velocity defined by the axial potential Φ . Most systems with curved axis are distortion-free. In this case, the optic axis is embedded in a plane section,

which is usually the midsection of the system. Typical examples are accelerators, storage rings, spectrometers, and energy filters. Recently, optic axes with torsion are becoming increasingly important for the design of helical wigglers, undulators, wavelength shifters, and for particle motion in helical dipole Siberian Snakes [94]. The paraxial equations for the motion of electrons in systems with arbitrarily curved axis were first introduced by Cotte [95]. However, he restricted most of his investigations to systems in which the design curve is a plane curve. The design curve forms the z -axis of the curvilinear coordinate system [10, 51, 52, 61]. In order that this axis is a possible ray, we must relate the curvature of the axis with the electromagnetic dipole fields in an appropriate way given by the relation (4.14).

4.8.1 General Systems

The paraxial trajectory of a particle with a relative energy deviation $\kappa = \Delta\Phi/\Phi_0$ has the form

$$U(z) = \sum_{\mu=1}^4 a_{\mu} U_{\mu}(z) + \kappa U_{\kappa}(z). \quad (4.243)$$

The sum defines the general solution of the homogeneous part of the inhomogeneous path equation (4.29), while term κU_{κ} denotes the inhomogeneous solution of the equation

$$U'' + TU - G\bar{U} = \kappa \left(\frac{\Phi^*}{\Phi_0^*} \right)^{1/4} D. \quad (4.29)$$

We can obtain this solution by integration if we know four linearly independent solutions $U_{\mu} = U_{\mu}(z)$ of the homogeneous equation $U'' + TU - G\bar{U} = 0$. This equation has the same form as that for systems with a straight axis. Hence, the paraxial properties of systems with curved axis are equivalent to those of systems with straight axis for electrons with nominal energy ($\kappa = 0$). To determine the integral expression for the inhomogeneous solution, we consider an arbitrary inhomogeneous perturbation term

$$P = P(z) = \kappa \left(\frac{\Phi^*}{\Phi_0^*} \right)^{1/4} D \quad (4.244)$$

and employ the method of variation of coefficients. This method assumes that the inhomogeneous solution has the form

$$U_{\text{in}} = \sum_{\mu=1}^4 b_{\mu}(z) U_{\mu}(z), \quad (4.245)$$

where the four functions $b_\mu(z) = \bar{b}_\mu(z)$ are real. We can define two of these functions arbitrarily because the inhomogeneous solution (4.245) must only satisfy the complex equation

$$U_{\text{in}}'' + T U_{\text{in}} - G \bar{U}_{\text{in}} = P, \quad (4.246)$$

which imposes two conditions on the functions $b_\mu = b_\mu(z)$, $\mu = 1, 2, 3, 4$. Therefore, we can impose two additional requirements. The two most appropriate additional conditions are

$$\sum_{\mu=1}^4 b'_\mu U_\mu = 0, \quad \sum_{\mu=1}^4 b'_\mu \bar{U}_\mu = 0. \quad (4.247)$$

By differentiating the expression (4.245) twice with respect to z and considering relations (4.247), we readily obtain

$$U_{\text{in}}'' = \sum_{\mu=1}^4 b'_\mu U'_\mu + \sum_{\mu=1}^4 b_\mu U''_\mu. \quad (4.248)$$

Substituting this result for U_{in}'' and (4.245) for U_{in} in Eq. (4.246) and considering that each U_μ is a solution of the homogeneous equation (4.197), we obtain together with (4.247) the four equations:

$$\sum_{\mu=1}^4 b'_\mu U_\mu = 0, \quad \sum_{\mu=1}^4 b'_\mu \bar{U}_\mu = 0, \quad \sum_{\mu=1}^4 b'_\mu U'_\mu = P, \quad \sum_{\mu=1}^4 b'_\mu \bar{U}'_\mu = \bar{P}. \quad (4.249)$$

This set of linear equations defines unambiguously the derivatives of the four real functions $b_\mu = b_\mu(z)$. They are obtained most conveniently by means of Cramer's rule, giving

$$b'_\mu = \frac{D_\mu}{D_W}, \quad \mu = 1, 2, 3, 4. \quad (4.250)$$

Here D_W denotes the Wronski determinant (4.217) and D_μ the determinant

$$D_\mu = (-1)^{\mu+1} \begin{vmatrix} 0 & U_\nu & U_\sigma & U_\tau \\ 0 & \bar{U}_\nu & \bar{U}_\sigma & \bar{U}_\tau \\ P & U'_\nu & U'_\sigma & U'_\tau \\ \bar{P} & \bar{U}'_\nu & \bar{U}'_\sigma & \bar{U}'_\tau \end{vmatrix} = 2(-1)^{\mu+1} \text{Re} \left\{ P \begin{vmatrix} \bar{U}_\nu & \bar{U}_\sigma & \bar{U}_\tau \\ U'_\nu & U'_\sigma & U'_\tau \\ \bar{U}'_\nu & \bar{U}'_\sigma & \bar{U}'_\tau \end{vmatrix} \right\}. \quad (4.251)$$

The indices differ from each other and have to be chosen such that $\nu < \sigma < \tau$. To evaluate readily the last determinant, we subtract from it the determinant

$$\begin{vmatrix} \bar{U}_\nu & \bar{U}_\sigma & \bar{U}_\tau \\ \bar{U}_\nu & \bar{U}_\sigma & \bar{U}_\tau \\ U'_\nu & U'_\sigma & U'_\tau \end{vmatrix} = 0. \quad (4.252)$$

By employing the Lagrange–Helmholtz relations (4.210), we obtain the result

$$D_\mu = 4(-1)^{\mu+1} \text{Re}\{\overline{P}(\overline{U}_\nu C_{\sigma\tau} - \overline{U}_\sigma C_{\nu\tau} + \overline{U}_\tau C_{\nu\sigma})\}. \quad (4.253)$$

Substituting this expression for D_μ in (4.250) and integrating over z gives

$$b_\mu = (-1)^{\mu+1} \frac{4}{D_W} \sum_{(p)} C_{\sigma\tau} \int_{z_\nu}^z \text{Re}\{P \overline{U}_\nu\} dz, \quad \nu < \sigma < \tau. \quad (4.254)$$

The sum consists of three terms, which we obtain by permutation (p) of the three indices ν, μ, τ . We must choose the lower integration limits z_ν , $\nu = 1, 2, 3, 4$, in such a way that the dispersion is zero at an arbitrary plane $z = z_\nu = z_0$ in front of the deflection fields. By inserting the result (4.254) into the expression (4.245) of the inhomogeneous solution and substituting (4.244) for P , we find the general solution of the inhomogeneous path equation (4.29) as

$$U = U(z) = \sum_{\mu=1}^4 a_\mu U_\mu(z) + \kappa U_\kappa. \quad (4.255)$$

The inhomogeneous solution $\kappa U_\kappa(z)$ defines the modified *dispersion ray*

$$\begin{aligned} U_\kappa &= \frac{-4}{D_W \sqrt[4]{\Phi_o^*}} \sum_{\mu=1}^4 (-1)^\mu U_\mu \sum_{(p)} C_{\sigma\tau} \int_{z_0}^z \text{Re} \left(\sqrt[4]{\Phi^*} D \overline{U}_\nu \right) dz \\ &= -\frac{2}{D_W} \sum_{(P)} (-1)^P C_{\mu\nu} U_\sigma \int_{z_0}^z \sqrt[4]{\frac{\Phi^*}{\Phi_o^*}} \text{Re} (D \overline{U}_\nu) dz, \quad \nu < \sigma < \tau \end{aligned} \quad (4.256)$$

denotes. The summation in the second representation must be taken over all $4! = 24$ permutations (P) of the indices μ, ν, σ and τ . Both expressions simplify considerably if we choose the four linearly independent solutions U_μ of the homogeneous equation such that four of the six coefficients $C_{\mu\nu}$ vanish. We define these solutions to a certain extent by imposing the conditions

$$C_{31} = C_{42} = 1, \quad C_{12} = C_{23} = C_{14} = C_{34} = 0. \quad (4.257)$$

With these specifications, we obtain from (4.217) the Wronski determinant as $D_W = -4$. Using this result together with $u_\mu = U_\mu(\Phi_o^*/\Phi^*)^{1/4}$ and (4.258), the dispersion ray $u_\kappa = U_\kappa(\Phi_o^*/\Phi^*)^{1/4}$ adopts the form

$$\begin{aligned}
u_\kappa = & \frac{u_1}{\sqrt{\Phi_o^*}} \int_{z_0}^z \operatorname{Re} \left(\sqrt{\Phi^*} D \bar{u}_3 \right) dz + \frac{u_2}{\sqrt{\Phi_o^*}} \int_{z_0}^z \operatorname{Re} \left(\sqrt{\Phi^*} D \bar{u}_4 \right) dz \\
& - \frac{u_3}{\sqrt{\Phi_o^*}} \int_{z_0}^z \operatorname{Re} \left(\sqrt{\Phi^*} D \bar{u}_1 \right) dz - \frac{u_4}{\sqrt{\Phi_o^*}} \int_{z_0}^z \operatorname{Re} \left(\sqrt{\Phi^*} D \bar{u}_2 \right) dz. \quad (4.258)
\end{aligned}$$

This expression is valid for arbitrary systems with curved axis. It simplifies considerably for torsion-free systems with midsection symmetry. The first-degree dispersion ray u_κ describes in paraxial approximation the chromatic off-axial displacement of electrons which move initially with relative energy deviation $\kappa = 1$ along the optic axis.

4.8.2 Systems with Midsection Symmetry

Most systems with a curved optic axis embed this axis in a plane section, which is the midsection of the system. We assume that the electric potential φ is symmetric and the scalar magnetic potential ψ antisymmetric with respect to this plane:

$$\varphi(x, -y, z) = \varphi(x, y, z), \quad \psi(x, -y, z) = -\psi(x, y, z). \quad (4.259)$$

As a result, the multipole strengths satisfy the conditions

$$\Phi_\nu = \bar{\Phi}_\nu = \Phi_{\nu c}, \quad \Phi_{\nu s} = 0, \quad \Psi_\nu = -\bar{\Psi}_\nu = i\Psi_{\nu s}, \quad \Psi_{\nu c} = 0, \quad \nu \geq 0. \quad (4.260)$$

In this case, the curvature of the optic axis $\Gamma = \bar{\Gamma}$ is real, which implies that its torsion is zero. In addition, the Larmor rotation vanishes as well ($\chi = 0$) resulting in $u = w = x + iy$. Moreover, the complex path equation decouples (4.34), (4.35), and the term $D(\Phi^*/\Phi_o^*)^{1/2}$ of the integrands in (4.258) takes the form

$$D \sqrt{\frac{\Phi^*}{\Phi_o^*}} = \Lambda_c = \bar{\Lambda}_c = \frac{1}{1 + \gamma_0} \sqrt{\frac{\Phi_o^*}{\Phi^*}} \left[\gamma_0 \sqrt{\frac{e}{2m_e \Phi^*}} \Psi_{1s} - \frac{1 + \gamma_0^2}{2} \frac{\Phi_{1c}}{\Phi^*} \right]. \quad (4.261)$$

For systems with decoupled paraxial equations, we employ the real fundamental rays $x_\alpha, y_\beta, x_\gamma, y_\delta$ shown in Fig. 4.33 which are connected with the paraxial solutions $u_\nu, \nu = 1, 2, 3, 4$, by the relations

$$u_1 = w_1 = x_\alpha, \quad u_2 = iy_\beta, \quad u_3 = x_\gamma, \quad u_4 = iy_\delta. \quad (4.262)$$

And considering that Λ_c is real, the expression (4.259) for dispersion ray reduces to

$$u_k = x_k = x_\alpha \int_{z_0}^z \Lambda_c x_\gamma dz - x_\gamma \int_{z_0}^z \Lambda_c x_\alpha dz. \quad (4.263)$$

Hence, the dispersion ray lies entirely in the midsection, as does the optic axis. A special case is the Wien filter. Its optic axis degenerates to a straight line because the curvature Γ of this axis is zero,

$$\Gamma = \bar{\Gamma} = \frac{\gamma_0}{2} \frac{\Phi_{1c}}{\Phi^*} - \sqrt{\frac{e}{2m_e \Phi^*}} \Psi_{1s} = 0, \quad (4.264)$$

giving

$$\Lambda_c = -\frac{1}{2(1 + \gamma_0)} \frac{\Phi_{1c}}{\Phi^*} \sqrt{\frac{\Phi_o^*}{\Phi^*}}. \quad (4.265)$$

This result demonstrates that the dispersion of the Wien filter decreases rapidly with increasing accelerating voltage Φ within the filter because its electric dipole strength Φ_{1c} cannot be made larger than about 8 kV/mm. This is the reason why one places the monochromator in high-performance electron microscopes near the cathode at potentials of a few kV with respect to that of the cathode ($\Phi = 0$). By considering the relations (4.264) and (4.260), we derive from the relations (4.26) and (4.30) for the round-lens strength T and the quadrupole strength G of the quadrupole dipole (QD) Wien filter the expressions

$$T = \frac{2 + \gamma_0^2}{16} \frac{\Phi'^2}{\Phi^{*2}} + \frac{1}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}}, \quad G = \frac{\gamma_0 \Phi_{2c}}{\Phi^*} - \sqrt{\frac{2e}{m_e \Phi^*}} \Psi_{2s} - \frac{5 - \gamma_0^2}{32} \frac{\Phi_{1c}^2}{\Phi^{*2}}. \quad (4.266)$$

Contrary to the round-lens strength T , we can eliminate the quadrupole effect of the QD Wien filter by choosing the electric and/or the magnetic dipole strength appropriately. In order to satisfy the condition $G = 0$ as precisely as possible, an octopole or dodecapole element should be employed allowing the simultaneous electric and magnetic excitation of each individual electrode or pole piece, respectively. This stigmatic QD Wien filter acts as a round lens combined with a straight view dispersive prism.

In many cases, it is mandatory that the dispersion is zero at any plane on the far side of the system. Such systems are nondispersive as a whole and satisfy the relations

$$\int_{-\infty}^{\infty} \Lambda_c x_\alpha dz = 0, \quad \int_{-\infty}^{\infty} \Lambda_c x_\gamma dz = 0. \quad (4.267)$$

Systems with double symmetry can fulfill these conditions in a very elegant way [90]. To demonstrate this behavior, we consider a system consisting of two identical symmetric subunits. One of the two linearly independent fundamental rays x_α and x_γ is symmetric; the other is antisymmetric with respect to the central plane

z_c of each subunit. In addition, the ray, which is symmetric with respect to the plane z_c , is antisymmetric with respect to the plane z_m located midway between the two subunits and vice versa. As a result, the integrands $\Lambda_c x_\alpha$ and $\Lambda_c x_\gamma$ are antisymmetric either with respect to the central plane of each subunit or the midplane z_m of the entire system. In both cases, the integrals (4.267) are zero. Hence, these systems are nondispersive as a whole. Such nondispersive systems are suitable as monochromators and as beam separators used in mirror correctors for separating the incident beam from the reflected beam. In nondispersive monochromators, one places the energy-selecting slit aperture at the midplane where the dispersion adopts its maximum. These monochromators are especially suitable because they do not affect the size of the effective source and its angular emission characteristic.

4.9 Quadrupole Anastigmat

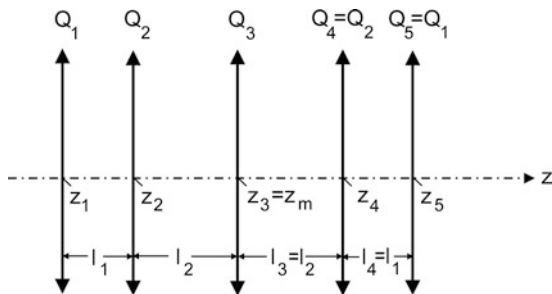
A quadrupole compound lens is a system composed of quadrupoles with mutual plane principal sections. Such orthogonal systems yield generally line images, each of which is orthogonal to one of the two principal sections. Each object point is imaged into two orthogonal lines located at different planes. If we move the object plane along the optic axis, the separation of the two astigmatic image planes either increases or decreases depending on the direction of the object shift. Hence, for a distinct location of the object plane, the locations of the two line images coincide. In this case, the two image lines shrink to a common image point for each point of the object. As a result, the system forms a stigmatic image, which is generally distorted in first order. We obtain a distortion-free stigmatic image for all locations of the object plane only if the cardinal planes of the $x - z$ principal section coincide with the corresponding planes of the $y - z$ section. Since we have two focal planes and two principal planes for each section, the system must satisfy the four conditions

$$z_{Px} = z_{Py}, \quad z_{\bar{P}x} = z_{\bar{P}y}, \quad z_{Fx} = z_{Fy}, \quad z_{\bar{F}x} = z_{\bar{F}y}. \quad (4.268)$$

The subscripts P and F define the image principal plane and the image focal plane of the entire system, whereas $\bar{P}$ and $\bar{F}$ denote the object principal plane and the object focal plane, respectively.

To satisfy the requirements (4.268), we need four free parameters. For a fixed geometry, these parameters are the strengths G_ν of the quadrupoles. Hence, we need at least four quadrupoles, $\nu = 1, 2, 3, 4$, to realize a quadrupole anastigmat, as it is the case for the Russian quadruplet [87]. The focal length of the quadrupole quadruplet depends on the location of the quadrupole elements. Since we cannot vary retrospectively their positions, the focal length of the quadruplet is fixed. In order to vary the focal length such that the system acts like a round lens, we need an additional quadrupole. Hence, we need at least five quadrupoles for achieving a system, which has the properties of a round lens in paraxial approximation [71]. In accordance with the terminology of light optics, we name such a system *quadrupole*

Fig. 4.34 Arrangement of the quadrupoles Q_ν within the symmetric quintuplet forming a pseudo round lens with variable focal length



anastigmat because it does *not* produce astigmatic imaging for any location of the object plane.

To survey the properties of an anastigmat formed by five quadrupoles Q_ν , $\nu = 1, 2, 3, 4, 5$, it suffices to consider thin elements. In this case, the strength $G_\nu = G_\nu(z)$ of each quadrupole is proportional to a delta function:

$$G_\nu = -\frac{1}{f_\nu} \delta(z - z_\nu). \quad (4.269)$$

Here, z_ν defines the location of the ν th quadrupole and f_ν is its focal length, respectively.

For simplicity, we assume a symmetric quadrupole quintuplet, as shown in Fig. 4.34. In this case, we can separate the system in two symmetric halves, each of which has three focal lengths:

$$f_1 = f_5, \quad f_2 = f_4, \quad f_3 = f_3^*/2. \quad (4.270)$$

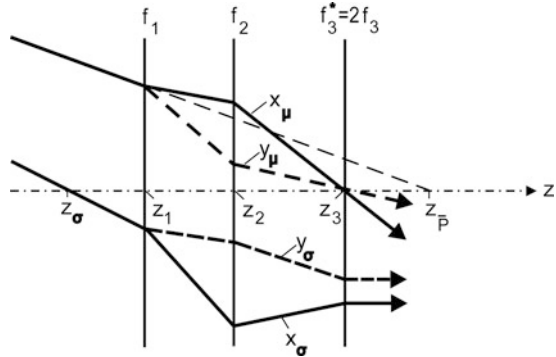
The distances between the thin quadrupoles of the quintuplet satisfy the symmetry relations

$$\begin{aligned} l_1 &= z_2 - z_1 = l_4 = z_5 - z_4, \\ l_2 &= z_3 - z_2 = l_3 = z_4 - z_3. \end{aligned} \quad (4.271)$$

We can consider the focal lengths f_1, f_2, f_3^* and the distances l_1, l_2 as five free parameters. However, we must conceive one of the two distances as a scaling parameter. The focal length f_3^* is twice the focal length f_3 of the third quadrupole. We split up this quadrupole into two elements, each of which belonging to one of the two identical subunits, because the plane $z_3 = z_m$ forms the midplane of the symmetric quadrupole quintuplet.

Owing to the symmetric arrangement and excitation of the quadrupoles, it suffices to determine the image focal length $f = \bar{f}$ and the location $z_p - z_m = z_m - z_{\bar{p}}$ of the image cardinal plane of the anastigmat. Moreover, we need to consider

Fig. 4.35 Course of the symmetric fundamental rays x_σ, y_σ and the antisymmetric rays x_μ, y_μ within the first half of the quadrupole quintuplet



only half of the system because two symmetric fundamental rays x_σ, y_σ and two antisymmetric fundamental rays x_μ, y_μ exist, as shown in Fig. 4.35.

In the case of stigmatic imaging, the two antisymmetric rays are the components of the principal fundamental ray $w_\mu = x_\mu + iy_\mu$, which is rotationally symmetric ($x_\mu(z) = y_\mu(z)$) in the region outside the anastigmat. The intersections of the asymptotes with the optic axis define the locations of the principal planes z_P and $z_{\bar{P}}$, respectively.

We assume a symmetric quadrupole anastigmat, which satisfies the symmetry conditions

$$x'_\sigma(z_m) = y'_\sigma(z_m) = 0, \quad x_\mu(z_m) = y_\mu(z_m) = 0. \quad (4.272)$$

The image principal plane z_P and the object principal plane $z_{\bar{P}}$ of the symmetric anastigmat are located symmetrically about the midplane $z_m = z_3$ at equal distances $z_m - z_P = z_{\bar{P}} - z_m$.

The zeros $z_{\sigma 1} = z_\sigma$ and $z_{\sigma 2}$ of the symmetric ray $w_\sigma = x_\sigma + iy_\sigma$ define the locations of the object plane and the image plane for unit magnification $M = \pm 1$. The distance between the location $z_{\bar{P}}$ of the object principal plane and the intersection z_σ of the symmetric fundamental ray w_σ equals twice the focal length of the anastigmat

$$z_{\bar{P}} - z_\sigma = z_{\sigma 2} - z_P = 2(z_{\bar{P}} - z_{\bar{F}}) = 2(z_F - z_P) = 2\bar{f} = 2f. \quad (4.273)$$

Conjugate object and image cardinal planes are located symmetrically about the midplane $z_m = z_3$ due to the imposed symmetry.

4.9.1 Focal Lengths of the Constituent Quadrupoles of the Anastigmat

For determining the focal lengths of the quadrupoles forming the compound anastigmat, we must know the values of the fundamental rays $x_\sigma, y_\sigma, x_\mu, y_\mu$

and the slopes of these rays at the midplane as functions of the focal lengths f_v , $v = 1, 2, 3$ and the distances l_1 and l_2 . The propagation of the rays through the anastigmat simplifies considerably by employing the thin-lens approximation, which considers the strengths (4.269) of the quadrupoles as delta functions. Without loss of generality, we suppose that at some initial plane $z = z_0$ in front of the anastigmat, the fundamental rays have initial values

$$\begin{aligned} x_\mu(z_0) &= x_{\mu 0} = 1, & y_{\mu 0} &= 1, \\ x_\sigma(z_0) &= x_{\sigma 0} = 1, & y_{\sigma 0} &= 1. \end{aligned} \quad (4.274)$$

We define the initial slopes of the fundamental rays as

$$\begin{aligned} x'_\mu(z_0) &= y'_\mu(z_0) = s_\mu = \frac{1}{z_0 - z_{\bar{p}}}, \\ x'_\sigma(z_0) &= y'_\sigma(z_0) = s_\sigma = \frac{1}{z_0 - z_\sigma}. \end{aligned} \quad (4.275)$$

The locations $z_{\bar{p}}$ and z_σ are given by the zeros of the image principal ray and the symmetric fundamental ray, respectively. Since each principle ray is a linear combination of the symmetric and the antisymmetric ray, we can determine from these rays the locations of the object principal plane $z_{\bar{p}}$ and the object focal plane $z_{\bar{f}} = (z_{\bar{p}} + z_\sigma)/2$.

In accordance with the thin-lens approximation of light optics, we can conceive the central planes z_1, z_2 and z_3 of the quadrupoles as refracting planes. A ray $w = x + iy$ intersecting such a plane only changes its direction but not its distance from the axis. Employing the path equation (4.114) and taking into account the relation (4.269), we obtain the change of the slope components resulting from the refraction at the plane z_v as

$$\begin{aligned} x'(z_v + \varepsilon) - x'(z_v - \varepsilon) &= \int_{z_v - \varepsilon}^{z_v + \varepsilon} Gx \, dz = -\frac{x(z_v)}{f_v}, \\ y'(z_v + \varepsilon) - y'(z_v - \varepsilon) &= -\int_{z_v - \varepsilon}^{z_v + \varepsilon} Gy \, dz = \frac{y(z_v)}{f_v}. \end{aligned} \quad (4.276)$$

Here, ε denotes an infinitely small distance. Since the lateral distance of the ray does not vary when the ray passes the refraction plane, we have

$$x(z_v + \varepsilon) = x(z_v - \varepsilon) = x(z_v), \quad y(z_v + \varepsilon) = y(z_v - \varepsilon) = y(z_v). \quad (4.277)$$

We calculate most conveniently the course of a ray through the system by employing the matrix method. It describes the quadrupole effect of the refracting plane z_v on the x - and y -components of the ray by the refraction matrices

$$\overset{\leftrightarrow}{R}_{xv} = \begin{pmatrix} 1 & 0 \\ -1/f_v & 1 \end{pmatrix}, \quad \overset{\leftrightarrow}{R}_{yv} = \begin{pmatrix} 1 & 0 \\ 1/f_v & 1 \end{pmatrix}. \quad (4.278)$$

Accordingly, we describe the propagation of the ray through the space between the refraction plane $z_v = z_{v-1} + l_v$ and the refraction plane z_{v-1} by the transfer matrices

$$\overset{\leftrightarrow}{T}_{xv} = \overset{\leftrightarrow}{T}_{yv} = \overset{\leftrightarrow}{T}_1 = \begin{pmatrix} 1 & l_v \\ 0 & 1 \end{pmatrix}. \quad (4.279)$$

By placing the initial plane close to the first refraction plane $z_0 = z_1 - \varepsilon$, the propagation of the ray in the x, z -section from this plane to the plane $z_3 + \varepsilon$ is given by the matrix

$$\begin{aligned} \overset{\leftrightarrow}{M}_x &= \begin{pmatrix} a_x & b_x \\ c_x & d_x \end{pmatrix} = \overset{\leftrightarrow}{R}_{x3} \overset{\leftrightarrow}{T}_2 \overset{\leftrightarrow}{R}_{x2} \overset{\leftrightarrow}{T}_1 \overset{\leftrightarrow}{R}_{x1} \\ &= \begin{pmatrix} 1 & 0 \\ -1/f_3^* & 1 \end{pmatrix} \begin{pmatrix} 1 & l_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & l_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix}. \end{aligned} \quad (4.280)$$

We readily obtain from (4.280) the corresponding matrix $\overset{\leftrightarrow}{M}_y$ by replacing each focal length by its negative value ($-f_v \rightarrow f_v$). Using these matrices, we find the slopes $x'_m = x'(z_m)$, y'_m and the lateral positions $x_m = x(z_m)$, y_m of the trajectory components at the midplane $z = z_m = z_3$ as

$$\begin{pmatrix} x_m \\ x'_m \end{pmatrix} = \overset{\leftrightarrow}{M}_x \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix}, \quad \begin{pmatrix} y_m \\ y'_m \end{pmatrix} = \overset{\leftrightarrow}{M}_y \begin{pmatrix} y_1 \\ y'_1 \end{pmatrix}. \quad (4.281)$$

We use these relations for describing position and the slope of both the symmetric fundamental ray $w_\sigma = x_\sigma + iy_\sigma$ and the antisymmetric fundamental ray $w_\mu = x_\mu + iy_\mu$ at the midplane. The initial values of these rays are: $x_{\mu 1} = x_{\sigma 1} = y_{\mu 1} = y_{\sigma 1} = 1$, $x'_{\mu 0} = y'_{\mu 0} = s_\mu$, $x'_{\sigma 0} = y'_{\sigma 0} = s_\sigma$. Employing these initial values and imposing the symmetry requirements on the fundamental rays at the midplane, we obtain from the relations (4.280) and (4.281)

$$\begin{aligned} x'_{\sigma m} = c_{x\sigma} + d_{x\sigma}s_\sigma &= 0, & x_{\mu m} = a_{x\mu} + b_{x\mu}s_\mu &= 0, \\ y'_{\sigma m} = c_{y\sigma} + d_{y\sigma}s_\sigma &= 0, & y_{\mu m} = a_{y\mu} + b_{y\mu}s_\mu &= 0. \end{aligned} \quad (4.282)$$

These conditions can only be satisfied if the matrix coefficients fulfill the relations

$$a_{x\mu}b_{y\mu} - a_{y\mu}b_{x\mu} = 0, \quad c_{x\sigma}d_{y\sigma} - c_{y\sigma}d_{x\sigma} = 0. \quad (4.283)$$

We obtain the matrix coefficients $a_{x\mu}$, $b_{x\mu}$, $c_{x\sigma}$, and $d_{x\sigma}$ by evaluating the expression (4.280) for the x -components of the fundamental rays w_μ and w_σ .

After a lengthy yet straightforward calculation we find

$$a_{x\mu} = 1 - \frac{l_2}{f_2} - \frac{l_1 + l_2}{f_1} + \frac{l_1 l_2}{f_1 f_2}, \quad b_{x\mu} = l_1 + l_2 - \frac{l_1 l_2}{f_2}, \quad (4.284)$$

$$c_{x\sigma} = -\frac{1}{f_1} - \frac{1}{f_2} - \frac{1}{f_3^*} + \frac{l_1}{f_1} \left(\frac{1}{f_2} + \frac{1}{f_3^*} \right) + \frac{l_2}{f_3^*} \left(\frac{1}{f_1} + \frac{1}{f_2} \right) - \frac{l_1 l_2}{f_1 f_2 f_3^*},$$

$$d_{x\sigma} = 1 - \frac{l_1}{f_2} - \frac{l_1 + l_2}{f_3^*} + \frac{l_1 l_2}{f_2 f_3^*}. \quad (4.285)$$

Substitution of f_ν for $-f_\nu$, $\nu = 1, 2, 3$, in (4.284) and (4.285) gives the matrix coefficients for the y -components of the fundamental rays at the midplane. By inserting these results and the expressions (4.284) and (4.285) into (4.283), we eventually derive the expressions

$$\frac{l_1^2 l_2^2}{f_1 f_2^2} = \frac{l_2^2}{f_2} + \frac{(l_1 + l_2)^2}{f_1}, \quad (4.286)$$

$$\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3^*} + \frac{l_1^2 l_2^2}{f_1 f_2^2 f_3^{*2}} = \frac{l_1^2}{f_1} \left(\frac{1}{f_2} + \frac{1}{f_3^*} \right)^2 + \frac{l_2^2}{f_3^{*2}} \left(\frac{1}{f_1} + \frac{1}{f_2} \right) + \frac{2l_1 l_2}{f_1 f_3^{*2}}. \quad (4.287)$$

We must consider the lengths l_1 and l_2 as fixed parameters for a given system. One of these two parameters can serve as a scaling length. However, we can vary the focal lengths because the quadrupole currents and voltages are adjustable. Since the three focal lengths must only satisfy the two conditions (4.286) and (4.287), one focal length serves as the independent variable. The calculations reveal that it is advantageous to choose the focal length f_2 as this variable. Employing the condition (4.286), we obtain the focal length of the first quadrupole as

$$f_1 = \frac{l_1^2}{f_2} - \frac{(l_1 + l_2)^2}{l_2^2} f_2. \quad (4.288)$$

The focal length f_1 changes its sign either if we reverse the sign of the focal length f_2 . For $f_2 > f_{20} = l_1 l_2 / (l_1 + l_2)$, the focal length f_1 is negative, while both focal lengths have the same sign if $f_2 < f_{20}$. In the case $f_2 = f_{20}$, the refraction power $1/f_1$ of the first quadrupole diverges. Hence, we cannot realize this unphysical mode.

To derive the focal length $f_3 = f_3^*/2$ as a function of f_2 , we substitute the expression (4.288) for f_1 in Eq. (4.287). Then all terms cancel out which are quadratic in f_3^* . As a result, the focal length of the third quadrupole adopts the simple form

$$f_3 = \frac{f_3^*}{2} = -\frac{l_1^2 l_2^2 + f_2^2 (l_1 + l_2)^2}{2l_1 f_2 (l_1 + 2l_2)}. \quad (4.289)$$

This result demonstrates that the sign of the focal length of the third quadrupole is always opposite to that of the second quadrupole. In the special case $f_2 = f_{20} = l_1 l_2 / (l_1 + l_2)$, the absolute value of the focal length adopts its minimum

$$|f_3|_{\min} = l_2 \frac{l_1 + l_2}{l_1 + 2l_2}. \quad (4.290)$$

Since the focusing power of the first quadrupole diverges for this value, we can never achieve the minimum (4.290) for a realistic anastigmat.

4.9.2 Cardinal Elements of the Quadrupole Anastigmat

The cardinal elements of a compound lens characterize its optical properties. Since the quadrupole anastigmat as a whole acts like a symmetric round lens, it suffices to calculate the location of one focal plane and one principal plane. Owing to the symmetry of the anastigmat, the locations of conjugate object and image cardinal planes are symmetric with respect to the midplane z_m . For determining the location $z_{\bar{p}}$ of the object principal plane, we employ the relations (4.282), (4.284), and (4.288). As a result, we find

$$z_{\bar{p}} - z_1 = -\frac{1}{s_{\mu}} = \frac{b_{x\mu}}{a_{x\mu}} = \frac{b_{y\mu}}{a_{y\mu}} = \frac{l_1^2 l_2^2 - (l_1 + l_2)^2 f_2^2}{l_1 l_2^2 - (l_1 + l_2) f_2^2}. \quad (4.291)$$

The principal planes collapse in the midplane z_m in the limit of a very weak lens ($f_2 \rightarrow \infty$). Therefore, it is more suitable to define the location of the principal plane by means of its distance $z_{\bar{p}} - z_m = z_{\bar{p}} - z_1 - (z_m - z_1)$ from the midplane. Considering $z_m - z_1 = l_1 + l_2$ and the relation (4.291), we obtain

$$z_{\bar{p}} - z_m = z_m - z_p = \frac{l_1 l_2^3}{(l_1 + l_2) f_2^2 - l_1 l_2^2}. \quad (4.292)$$

Contrary to round lenses, the object and image principal planes may be crossed or not depending on the value of the focal length f_2 . This behavior follows readily from the separation of the principal planes

$$\Delta = z_p - z_{\bar{p}} = -2(z_{\bar{p}} - z_m) = -\frac{2l_1 l_2^3}{(l_1 + l_2) f_2^2 - l_1 l_2^2}. \quad (4.293)$$

This distance is negative for $(l_1 + l_2) f_2^2 > l_1 l_2^2$, as in the case of round lenses, and positive for $(l_1 + l_2) f_2^2 < l_1 l_2^2$, as shown in Fig. 4.36. We will show in the following that the anastigmat acts as a convergent lens in the former case and as a divergent lens in the latter case.

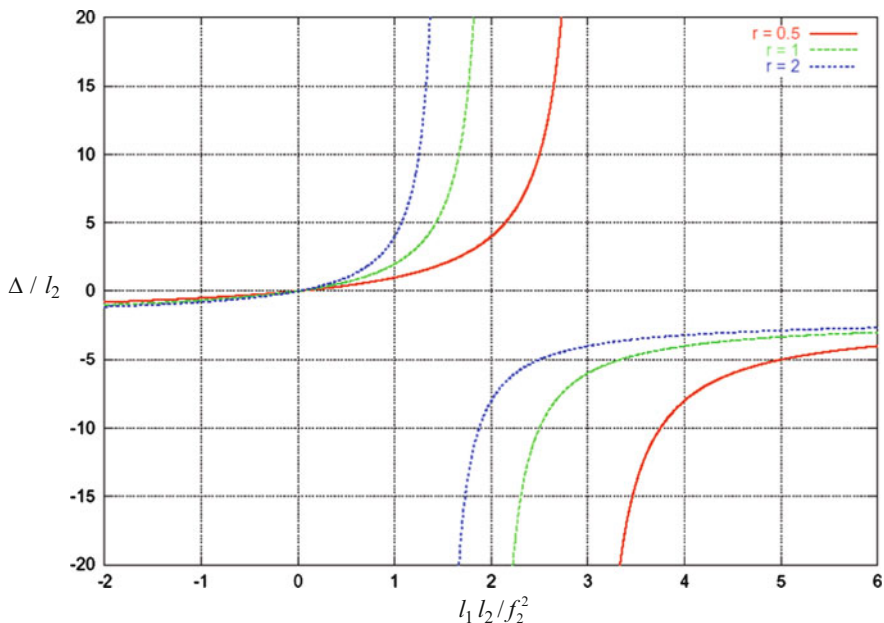


Fig. 4.36 Normalized distance $\Delta/l_2 = (z_P - z_{\bar{P}})/l_2$ between the principal planes as function of the normalized quadratic focal strength of the second quadrupole for different ratios $r = l_2/l_1$

We determine the focal length most conveniently by considering that the anastigmat forms an image with unit magnification $|M| = 1$ if we place the object at the plane $z_o = z_\sigma$. Since the distance $z_{\bar{P}} - z_o = z_i - z_P$ equals twice the focal length (4.273),

$$f = \bar{f} = z_{\bar{P}} - z_{\bar{F}} = \frac{1}{2}(z_{\bar{P}} - z_\sigma) = \frac{1}{2}(z_{\bar{P}} - z_1) - \frac{1}{2}(z_\sigma - z_1). \quad (4.294)$$

We replace the second term on the right-hand side by the relation

$$z_\sigma - z_1 = -\frac{1}{s_\sigma} = \frac{c_{x\sigma}}{d_{x\sigma}} = \frac{c_{y\sigma}}{d_{y\sigma}}. \quad (4.295)$$

The intersection of the object asymptote of the symmetric fundamental ray w_σ with the optic axis defines the plane z_σ . To establish its location, we employ the relations (4.285), (4.288), and (4.289). After a lengthy and involved calculation, we eventually find

$$z_\sigma - z_1 = \frac{1}{l_1 l_2} \frac{[l_1^2 l_2^2 - (l_1 + l_2)^2 f_2^2][l_1^2 l_2 - (l_1 + l_2) f_2^2]}{l_1^2 l_2^2 - \{l_1 l_2 + (l_1 + l_2)^2\} f_2^2}. \quad (4.296)$$

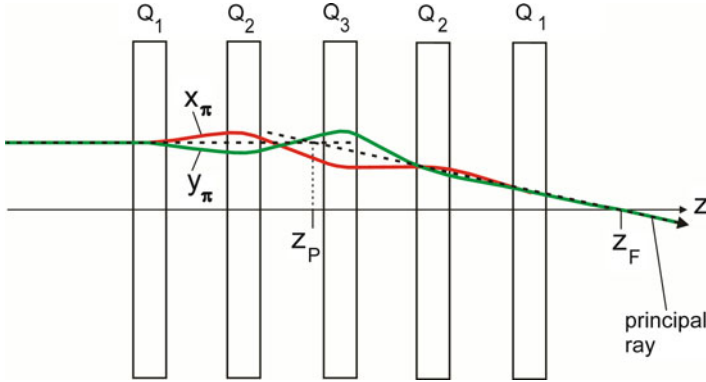


Fig. 4.37 Course of the image principal rays x_π and y_π of the anastigmat; the object and image asymptotes define the image focal plane z_F and the image principal plane z_P of the quadrupole anastigmat

Substituting this relation for $z_\sigma - z_1$ and (4.291) for $z_{\bar{P}} - z_1$ on the right-hand side of (4.294), we obtain

$$f = \bar{f} = \frac{f_2^2}{2l_1l_2} \frac{f_2^4(l_1 + l_2)^4 - l_1^4l_2^4}{[f_2^2(l_1 + l_2) - l_1l_2^2][f_2^2\{l_1l_2 + (l_1 + l_2)^2\} - l_1^2l_2^2]}. \quad (4.297)$$

This expression shows that the focal length of the quadrupole anastigmat can be positive or negative depending on the magnitude of f_2^2 .

We determine the location of the object and image focal planes by means of the expressions (4.294), (4.291), and (4.297), giving

$$z_1 - z_{\bar{F}} = z_F - z_5 = \frac{1}{2l_1} \frac{[f_2^2(l_1 + l_2)^2 - l_1^2l_2^2]^2 [f_2^2 - 2l_1l_2]}{[f_2^2(l_1 + l_2) - l_1l_2^2][f_2^2\{(l_1 + l_2)^2 + l_1l_2\} - l_1^2l_2^2]}. \quad (4.298)$$

The distance between entrance plane z_1 and the object focal plane $z_{\bar{F}}$ can be positive or negative depending on the focal length f_2 . The conjugate focal planes z_F and $z_{\bar{F}}$ are located outside the anastigmat if the distance (4.298) is positive and inside if this distance is negative.

The cardinal elements are defined by the asymptotes of the corresponding cardinal rays. To illustrate the course of these rays within the anastigmat, we have depicted in Fig. 4.37 the courses of the image principal rays x_π and y_π for the horizontal $x-z$ section and the vertical $y-z$ section. The common intersection point of these rays with the optic axis defines the image focal point, and the intersection of their asymptotes determines the location of the image principal plane. Owing to the symmetry of the anastigmat, we obtain the object cardinal elements by mirroring the corresponding image cardinal elements at the midplane of the anastigmat.

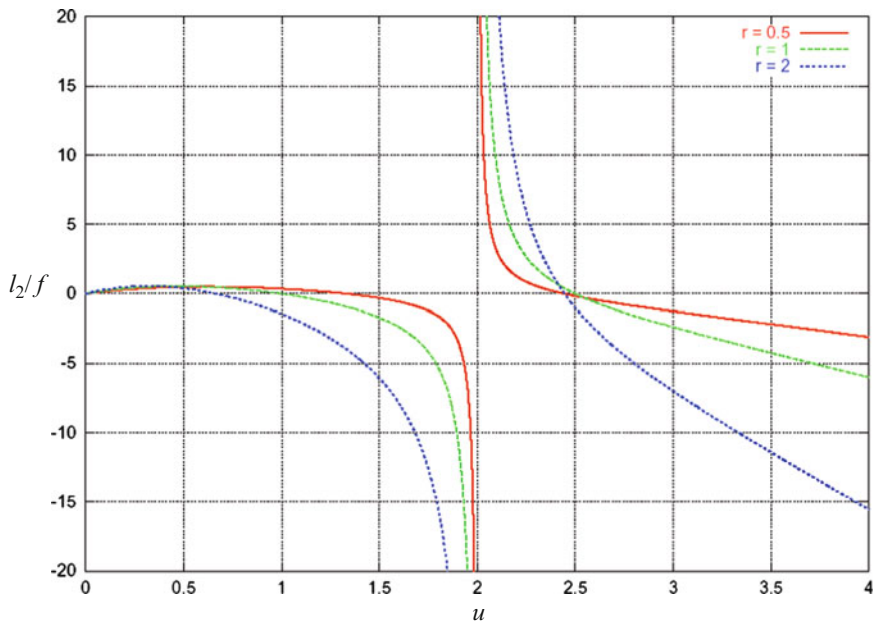


Fig. 4.38 Normalized focal strength l_2/f of the quadrupole anastigmat as function of $u = 2l_1^2 l_2^2 / f_2^2 (l_1 + l_2)^2$ for different ratios $r = l_2/l_1$

It is interesting to note that the cardinal elements of the anastigmat only depend on the square of the focal length f_2 . This behavior becomes evident if we consider that changing the sign of f_2 also inverses the sign of the other focal lengths f_1 and f_3 , as follows readily from (4.288) and (4.289). However, such an inversion merely interchanges the course of the fundamental rays in the x -section with that in the y -section but does not affect the optical properties of the anastigmat as a whole.

The focal length (4.298) of the anastigmat is positive as long as $f_2^2 > l_1 l_2^2 / (l_1 + l_2)$, as illustrated in Fig. 4.38. The focal length and the distances of the focal planes (4.298) from the fixed entrance plane z_1 diverge for $f_2^2 = l_1^2 l_2^2 / [l_1 l_2 + (l_1 + l_2)^2]$ and for $f_2^2 = l_1 l_2^2 / (l_1 + l_2)$. The anastigmat operates in the telescopic mode in these special cases. In the first case, the object and image principal planes are located within the telescopic anastigmat. In the other case, the object principal plane is located at $z = \infty$ and the image principal plane at $z = -\infty$.

For this telescopic mode, the focal lengths of the constituent quadrupoles adopt the values

$$f_1 = -l_2 \sqrt{\frac{l_1 + l_2}{l_1}} = -\frac{l_2^2}{f_2}, \quad f_3^* = 2f_3 = -l_2 \frac{l_2 + 2l_1}{2l_2 + l_1} \sqrt{\frac{l_1 + l_2}{l_1}}. \quad (4.299)$$

The conjugate planes $z_{\sigma 1}$ and $z_{\sigma 2}$ for unit magnification represent the nodal planes z_N and $z_{\bar{N}}$. These planes are located at positions

$$z_1 - z_{\bar{N}} = z_N - z_5 = \frac{l_2^2 - l_1^2}{2l_1 + l_2}. \quad (4.300)$$

The locations of the nodal planes are outside the anastigmat if $l_2 > l_1$. In the opposite case ($l_2 < l_1$), the planes are inside the anastigmat. In this case, we have a virtual image formation with unit magnification.

The telescopic anastigmat forms a distorted stigmatic image of the infinitely distant object plane $z_o = -\infty$ at the midplane z_m . The first-order distortion of this so-called *anamorphic image* is given by the relation

$$\frac{M_x}{M_y} = \frac{x_{\sigma m}}{y_{\sigma m}} = \left(\sqrt{\frac{l_1}{l_2}} + \sqrt{\frac{l_1 + l_2}{l_2}} \right)^2. \quad (4.301)$$

Here, M_x and M_y denote the magnifications in the x -section and in the y -section, respectively. If we further decrease the focal length f_2 , the locations of the object principal planes and the conjugate image principal planes invert and move from opposite directions toward the anastigmat. Then, the focal length $f = \bar{f} = z_{\bar{P}} - z_{\bar{F}}$ is negative, resulting in a diverging lens. In the case $f_2 = l_1 l_2 / (l_1 + l_2)$, the object principal plane catches up with the object focal plane at the entrance plane $z_{\bar{P}} = z_{\bar{F}} = z_1$, resulting in the unrealistic focal length $\bar{f} = f = f_1 = 0$. In the region $l_1^2 l_2^2 / (l_1 + l_2)^2 > f_2^2 > l_1^2 l_2^2 / [(l_1 + l_2)^2 + l_1 l_2]$, the anastigmat is able to act as a strong converging lens because its focal length can be made very small. Such a lens offers a very promising alternative for the projector system of an electron microscope since the quadruple anastigmat reduces considerably the required length of the microscope column. Moreover, a shorter column results in an improved mechanical stability of the microscope.

The anastigmat forms another telescopic mode, such that the asymptotes of the object principal ray and the image principal ray coincide, for $f_2^2 \{(l_1 + l_2)^2 + l_1 l_2\} = l_1^2 l_2^2$. For this mode, the object principal plane is located at position

$$z_{\bar{P}} = z_1 + \frac{l_1^2}{2l_1 + l_2} \quad (4.302)$$

within the region $z_1 < z < z_m$ of the anastigmat.

4.10 Variable-Axis Lens

Electron beams are used among other applications for the fabrication of masks employed in lithography. In order to write a large area of a fixed object, it is advantageous to employ a moving objective lens. Purely electric systems are

advantageous because they allow fast shifts of the electrical field [97]. We can construct a moving electric round lens with arbitrary lateral shift in one direction by superposing a moving quadrupole field on the static field of an electric cylinder lens. The potential of an infinitely extended cylinder lens is two-dimensional. Assuming that the electrodes run parallel to the x -coordinate, the electrostatic potential is given by

$$\varphi_c = \varphi_c(y, z) = \Phi(z) - \frac{1}{2}\Phi''(z)y^2 + \dots \quad (4.303)$$

Because the x -component of the electric field strength is zero, the cylinder lens focuses the electrons only in the y -direction. To obtain stigmatic focusing, we superpose on the potential (4.303) the quadrupole potential

$$\varphi_2 = \Phi_{2c}(z)(x^2 - y^2) + \dots = -\frac{1}{4}\Phi''(z)(x^2 - y^2) + \dots \quad (4.304)$$

We have chosen the quadrupole strength Φ_{2c} in such a way that the refractive power of the quadrupole equals half that of the cylinder lens at any plane along the optic axis. By adding the potentials (4.303) and (4.304), we obtain the potential of a round lens in paraxial approximation:

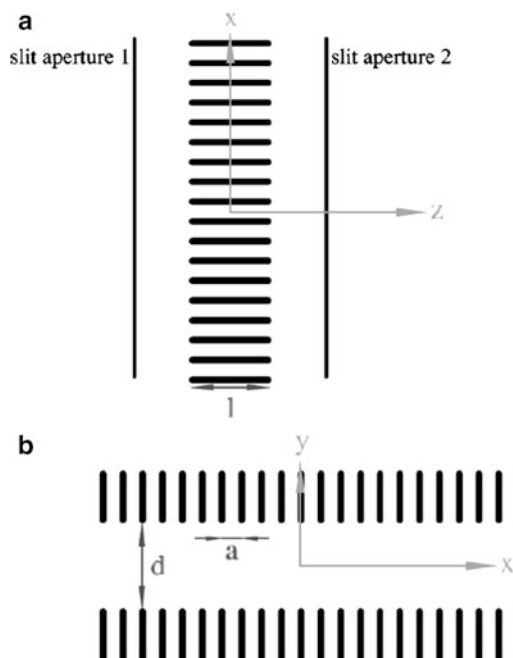
$$\varphi = \varphi_c + \varphi_2 = \Phi(z) - \frac{1}{4}\Phi''(z)(x^2 + y^2) + \dots \quad (4.305)$$

Unfortunately, we cannot realize in practice the required relation $\Phi_{2c} = -\Phi''/4$ by subdividing the slit electrodes in a sequence of equally separated identical stripes forming a comb structure, as illustrated in Fig. 4.39. However, this behavior does not prevent the formation of a movable field, which acts globally as variable-axis anastigmat, as it is the case for the quadrupole anastigmat outlined in Sect. 4.9. Compound systems consisting of at least five non-rotationally symmetric elements can serve as substitutes for round lenses because they are able to operate as anastigmats forming a stigmatic image for any location of the object plane. This behavior is due to the fact that we can vary the focal length of the system without losing stigmaticity. The system depicted in Fig. 4.39 consists of a thick central comb electrode and two slit apertures. This system enables stigmatic imaging only for a single object plane. In order to obtain a moveable round lens, the system must be composed either of three comb electrodes or of a single thick comb electrode and four slit electrodes placed symmetrically about the central comb electrode.

To shift the axis of the quadrupole field along the x -axis by an arbitrary amount, the slit width d between the comb electrodes must be large compared with the distance a between any two adjacent stripes of each comb. About eight pairs of stripes suffice to form a quadrupole field which can be moved continuously along the x -axis, as illustrated in Fig. 4.39. We perform the shift by varying the potentials applied to the individual stripe pairs of the comb electrode.

The most versatile system consists of three twin comb electrodes. The central twin comb is put at an average potential which differs from that of the outer

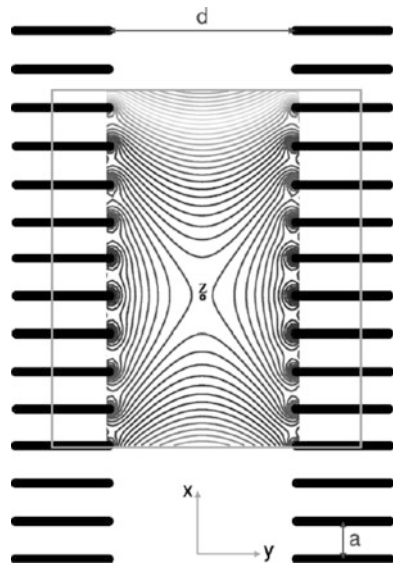
Fig. 4.39 Arrangement of the electrodes of a comb lens: (a) top view and (b) vertical section through the central comb electrode



electrodes. Contrary to systems which incorporate only a single thick twin comb electrode, the elements of the triple comb system can be made thin, thus shortening the extension of the compound lens in the direction of the optic axis. Owing to the symmetric arrangement and excitation of the electrodes, one needs only twice as many voltage supplies as in the case of a single comb lens. The triple comb lens provides stigmatic and distortion-free paraxial imaging because a symmetric system consisting of three quadrupoles and two immersion cylinder lenses can always be adjusted to act like a movable anastigmat. We form the two immersion cylinder lenses by applying an additional voltage between the central comb electrode and the outer comb electrodes. Accordingly, the system forms a cylinder einzel lens if the stripes of each comb structure are at the same potential. The moveable lens of the triple comb system has about the same imaging quality as a rotationally symmetric electrostatic lens. Moreover, the comb system has the advantage to enable an unlimited lateral shift of the optic axis in one direction and to employ simultaneously many spatially separated beams. Owing to these properties, mechanical shifts of the object are not necessary.

To realize a quadruple field, about eight stripe pairs are necessary, as illustrated in Fig. 4.40. Due to the periodicity of the comb structure, we can form many quadrupole fields along the comb. The centers of adjacent quadrupoles must be separated from each other by a distance larger than about $10a$. Because we can mutually shift the resulting lenses in the direction of the comb axis, it is possible to image simultaneously many elements of an extended stripe or to scan the object

Fig. 4.40 Equipotentials of the quadrupole field formed by applying appropriate voltages at 10 stripe pairs of the comb electrode



simultaneously with many beams originating from a linear array of electron sources. The multi-beam comb system avoids a common crossover of the individual beams. Therefore, beam broadening resulting from Coulomb interactions does not depend on the number of beams. Hence, the so-called throughput of the system increases in proportion to the number of beams operating simultaneously.

The system composed of four slit apertures and a thick comb lens has less flexibility than the triple comb lens but has the advantage to need only half as many voltage supplies. Each pair of slit apertures A_i and A_o is arranged symmetrically about the central thick comb structure, as depicted in Fig. 4.41. The two outer apertures A_o are put at the potential Φ_0 of the column. Depending on the potential Φ_i applied to the two inner slit apertures, we differentiate between an accelerating ($\Phi_i > \Phi_0$) and a retarding system ($\Phi_i < \Phi_0$). In the absence of the quadrupole field, the stripes of the comb electrode are at mutual potential Φ_c , which differs from those of the slit apertures. An additional voltage U_v is applied to a given set of sheet pairs $v = 1, 2, \dots, n$ of the comb electrode. The performance of the comb anastigmat is significantly better for the accelerating mode than for the retarding mode.

4.11 Highly Symmetric Telescopic Systems

Highly symmetric telescopic systems are widely applied in light optics. The so-called $4-f$ system, which is composed of two identical round lenses separated from each other by twice the focal length, forms the basic system of coherent optics [98]. The $4-f$ system images an object, placed at the front focal plane of the first lens,

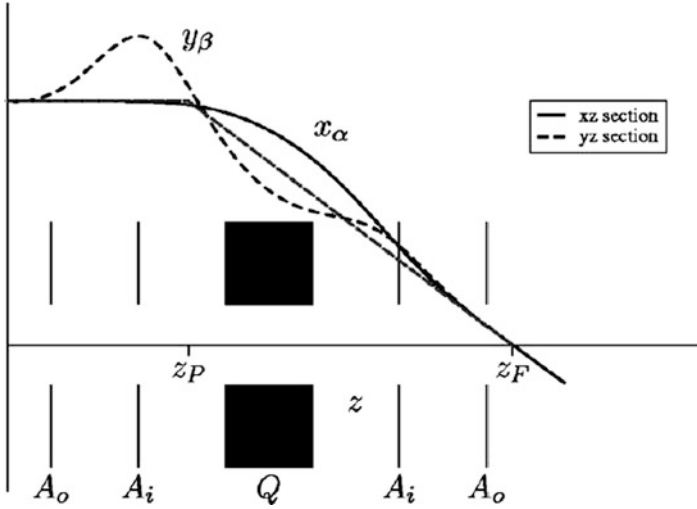


Fig. 4.41 Path of the image principal rays x_α and y_β within the movable anastigmat consisting of four slit apertures and a thick central comb electrode; the *dashed asymptotes* define the location of the image focal plane

with unit negative magnification into the back focal plane of the second lens. In addition, it forms an exact diffraction image of the object transparency at the plane midway between the two lenses. One places a mask or a structured phase plate at this plane in order to manipulate the image of the object in a distinct way. Owing to the importance of this possibility for the image formation in microscopy, Abbe called the diffraction image as *primary image*.

The $4f$ system also serves as largely aberration-free transfer system since it can transfer the asymptotic lateral positions of the rays at a given plane to a plane located at a distance $4f$ with negative unit magnification. If we center identical optical elements at each of the two conjugate focal planes of the $4f$ system, the primary focusing effect of the elements cancels out on the far side of the system. This peculiar property is utilized for compensating the primary second-order deviations introduced by the sextupoles of an electron optical hexapole corrector [22, 99]. The residual axial third-order aberration is rotational symmetric and of opposite sign to that of round electron lenses. Hence, we can adjust the hexapole strength to compensate for the unavoidable spherical aberration of these lenses. Since the hexapoles do not affect the paraxial rays, they cannot eliminate the first-order axial chromatic aberration. In order to compensate for this aberration in systems with a straight optic axis, we must employ electric and magnetic quadrupoles [18, 100–102]. Quadrupole systems also enable the compensation of the third-order geometrical aberrations by incorporating octopole fields. These fields and the quadrupole fields can be excited independently within octopole or dodecapole elements [103].

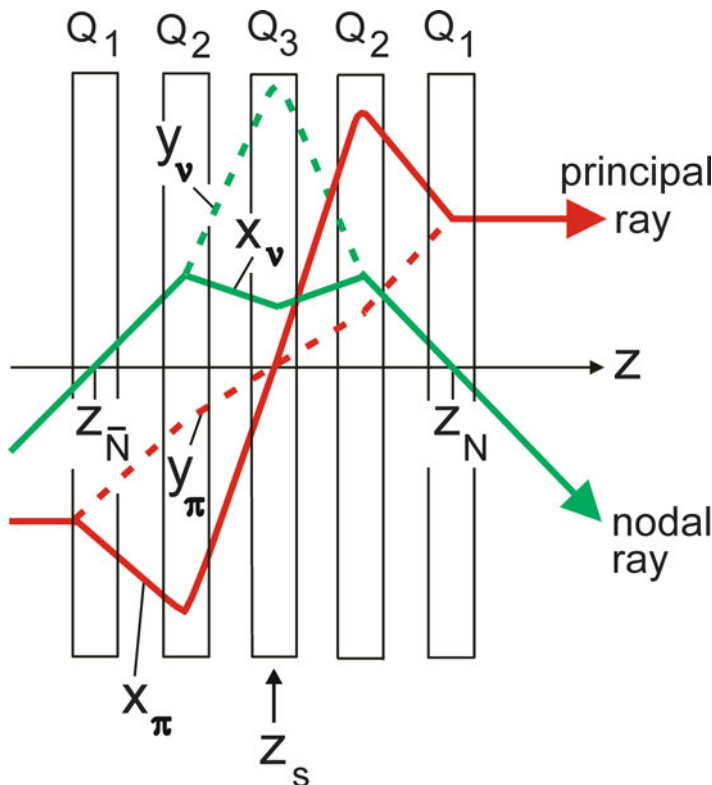


Fig. 4.42 Course of the fundamental rays and location of the nodal planes $z_N, z_{\bar{N}}$ for the telescopic quadrupole anastigmat in the special case $l_1 = l_2 = l$

Telescopic quadrupole systems with a high degree of symmetry of both the arrangement and excitation of the quadrupoles and the internal course of the fundamental trajectories represent an important class of quadrupole compound lenses. These systems are extremely suitable as correctors compensating for the unavoidable aberrations of round lenses, while minimizing the number of additional aberrations introduced by the deviation from rotational symmetry.

As an important example, we consider the quadrupole anastigmat with $l_1 = l_2 = l$ operating in the first telescopic mode, as illustrated in Fig. 4.42. For this system, we find from (4.298) and (4.301) the values $f_1 = -2f_2 = f_3^* = 2f_3 = l\sqrt{2}$ for the focal lengths and the value $M_x/M_y \approx 5.8$ for the first-order distortion of the anamorphic image at the symmetry plane $z_s = z_m$. The nodal ray $w_v = x_v + iy_v$ intersects the optic axis at the entrance plane $z_1 = z_{\bar{N}}$ and at the exit plane $z_5 = z_N$ of the telescopic anastigmat. The lateral distance of the principal ray $w_\pi = x_\pi + iy_\pi$ is opposite at these planes. Therefore, the central quadrupole triplet images the quadrupole Q_1 of the anastigmat with magnification $M_x = M_y = M = -1$ onto the quadrupole $Q_5 = Q_1$.

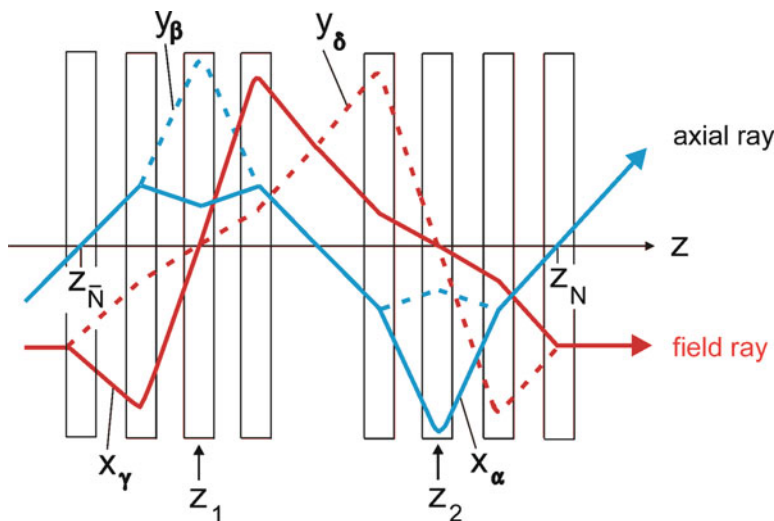


Fig. 4.43 Telescopic quadrupole system forming anamorphic images of the plane $z = -\infty$ at planes z_1 and z_2 ; the system exhibits exchange symmetry with respect to the plane located midway between the two quadruplets

We can utilize the telescopic quadrupole quintuplet shown in Fig. 4.42 to construct a useful system, which has two anamorphic images z_1 and z_2 of the infinitely distant plane with distortions $M_x/M_y = 5.8$ and 0.172 , respectively. The quadrupoles of the second quintuplet are excited with polarity opposite to those of the first quintuplet. By superposing the last quadrupole of the first quintuplet with the first quadrupole of the second quadruplet, these quadrupoles compensate each other and can be omitted. Therefore, the resulting system is composed of two quadrupole quadruplets, as shown in Fig. 4.43. The course of the fundamental rays exhibits exchange symmetry with respect to the midplane of the system. We obtain this symmetry by exchanging in one of the subsystems the path of the fundamental rays in the $x - z$ section with that in the $y - z$ section.

The antisymmetric quadrupole quadruplet is also able to form a telescopic system. We can achieve a system with unit magnification ($M_x = M_y = -1$), by means of two antisymmetric quadrupole doublets separated by the distance $d = 2f$, where $f = f_1 = -f_2$ is the absolute value of the focal length of each quadrupole which are separated by the distance $l = f$. The arrangement of the quadrupoles and the path of the fundamental rays are shown in Fig. 4.44. This system represents the quadrupole equivalent of the $4 - f$ round-lens system because the focal lengths f_{dx} and f_{dy} of each quadrupole doublet for the x -section and the y -section coincide ($f_{dx} = f_{dy} = f$) and are equal to that of the quadrupoles.

We achieve a telescopic quadruplet with different magnifications M_x and M_y most conveniently by replacing the central quadrupole triplet of the symmetric quintuplet by an antisymmetric doublet resulting in an antisymmetrically excited

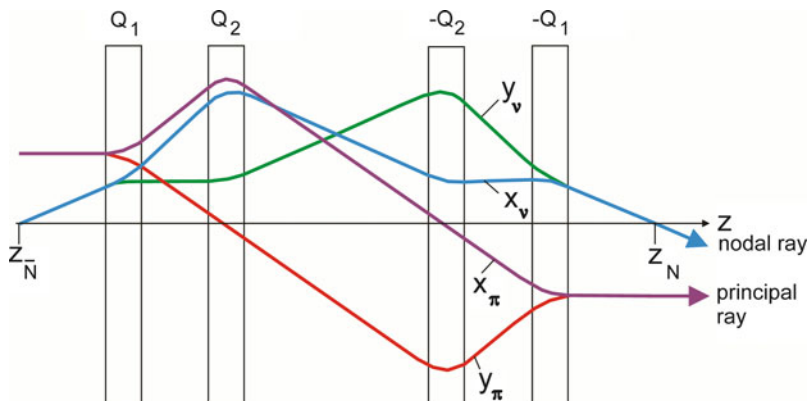


Fig. 4.44 Course of the fundamental rays in the antisymmetric telescopic quadrupole quadruplet forming the equivalent of a telescopic round-lens doublet

quadruplet whose elements are arranged symmetrically about the midplane. Accordingly, the system consists of two identical quadrupole doublets, which are excited antisymmetrically, as it is the case for the system with unit magnification. In order to achieve different magnifications M_x and M_y , we must allow that the focal lengths $f_1 = f_4$ and $f_2 = f_3$ of the quadrupoles and their separation $l_1 = z_2 - z_1 = z_4 - z_3$ differ from each other. For simplicity, we suppose that the field ray $w_f = x_f + iy_f$ intersects the centers $z = z_1$ and $z = z_4$ of the first and fourth quadrupole, respectively. Therefore, only the two inner quadrupoles separated by the distance $2l_2$ affect this ray. Its course is not symmetric with respect to the midplane z_m of the quadruplet because the quadrupoles are excited antisymmetrically ($f_{3x} = -f_{2x}$).

Employing the thin-quadrupole approximation and the matrix method for the propagation of the ray components from the center of the plane $z = z_1$ to the exit plane $z = z_4$ of the quadruplet, we obtain

$$x_f(z_4) = x_{f4} = y_{f4} = 2(l_1 + l_2) - 2\frac{l_1^2 l_2}{f_2^2} = 0, \quad (4.306)$$

resulting in

$$f_2^2 = \frac{l_1^2 l_2}{l_1 + l_2}. \quad (4.307)$$

To obtain different magnifications for the two principal sections, we impose the condition that the image principal ray $w_\pi = x_\pi + iy_\pi$ intersects the center of the midplane located at distance $l_2 = z_m - z_2$ from the second quadrupole. This ray runs parallel to the optic axis in front of the quadruplet. Both components of the image principal ray must vanish at the plane z_4 , thus

$$\begin{aligned}
x_\pi(z_m) &= 1 - \frac{l_1 + l_2}{f_1} + \frac{l_2}{f_2} - \frac{l_1 l_2}{f_1 f_2} = 0, \\
y_\pi(z_m) &= 1 + \frac{l_1 + l_2}{f_1} - \frac{l_2}{f_2} - \frac{l_1 l_2}{f_1 f_2} = 0.
\end{aligned} \tag{4.308}$$

Adding and subtracting these equations gives

$$f_1 f_2 = l_1 l_2, \quad f_1 l_2 = f_2 (l_1 + l_2) \rightarrow f_2^2 = \frac{l_1 l_2^2}{l_1 + l_2}. \tag{4.309}$$

The comparison of the resulting expression for the focal length f_2 with that given by the condition (4.307) shows that we can satisfy both relations only if $l_1 = l_2 = l$. Using this result, we find from (4.309) the focal lengths as

$$f_1 = f_4 = l\sqrt{2}, \quad f_2 = f_3 = \frac{l}{\sqrt{2}}. \tag{4.310}$$

We derive the magnifications M_x and M_y most conveniently from the slopes of the components of the field ray at the plane z_4 of the fourth quadrupole. Employing the Lagrange–Helmholtz relation (4.57), we find

$$M_x = -\frac{x'_y(z_1)}{x'_y(z_4)} = (1 + \sqrt{2})^2 \approx 5.81, \quad M_y = -\frac{y'_\delta(z_1)}{y'_\delta(z_4)} = \frac{1}{M_x} = (\sqrt{2} - 1)^2 \approx 0.168. \tag{4.311}$$

If we place an objective lens in front of this telescopic quadruplet, the resulting system forms an anamorphic (first-order distorted) stigmatic image of the focal plane with magnification aspect ratio $M_x/M_y \approx 33.8$.

For correcting aberrations, we aim for orthomorphic (distortion-free) telescopic systems exhibiting anamorphic images of the diffraction plane in its interior. We construct such a unit by combining two telescopic quadruplets in such a way that they form a symmetric septuplet. Its midplane coincides with the exit plane of the first quadruplet and the entrance plane $z_e = z_4$ of the second quadruplet. The polarities of the quadrupoles of the resulting septuplet or octuplet are symmetric with respect to the midplane $z_m = z_4$.

The components of the image principal ray w_π are symmetric with respect to z_m and coincide with those of the object principal ray w_π , as illustrated in Fig. 4.45. The field ray w_y is antisymmetric with respect to this plane and represents the nodal ray of the septuplet. The intersections $z_1 = z_{\bar{p}}$ and $z_7 = z_p$ of the object and image asymptotes of this ray with the optic axis define the object principal plane and the image principal plane, respectively. The separation of the principal planes is positive and equals the total length of the telescopic quadrupole septuplet:

$$\Delta = z_p - z_{\bar{p}} = z_7 - z_5 = 8l. \tag{4.312}$$

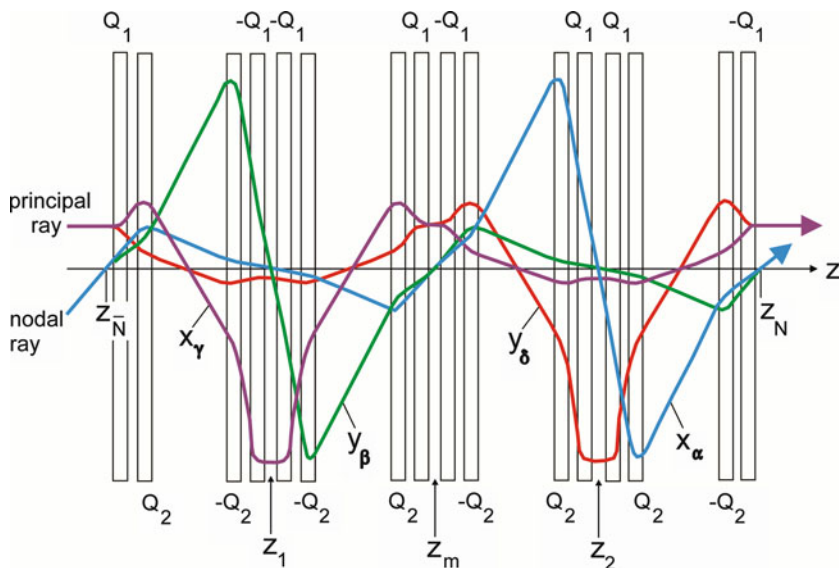


Fig. 4.45 Course of the fundamental rays within the telescopic system formed by two doubly symmetric quadrupole octuplets; strongly anamorphic images of the front nodal plane z_N are formed within the system at the center planes z_1 and z_2 of each subunit. The quadrupoles are excited antisymmetric with respect to the midplane z_m of the system, resulting in an exchange of the path of rays of the $x - z$ section with that of the $y - z$ section

This special device may serve as subunit of the ultracorrector compensating for the primary aberrations of round lenses.

For correcting the spherical aberration, it is necessary to place octipoles at the distorted images of the diffraction plane within the corrector. In order that these fields do overlap the quadrupole fields, it is desirable to place the images between the quadrupole elements. In this case, the excitation of the octipole fields does not affect the quadrupole fields and vice versa. Crosstalk may occur from hysteresis effects, if we excite different magnetic multipole fields within the same element. We avoid this situation by splitting the central quadrupole of the septuplet in two spatially separated quadrupoles yielding a symmetric octuplet. In some cases, it is desirable to form also astigmatic images of the object plane within some of the quadrupole fields to correct for axial chromatic aberration.

We can satisfy both conditions by an octuplet consisting of two antisymmetric quadruplets. Adjacent quadrupoles are separated by the same distance l . Hence, the total length of each quadruplet is $3l$. The focal lengths of the constituent quadrupoles are $f_1 = f_4 = l$, $f_2 = f_3 = 2l/3$, and the separation distance between the two quadruplets is $3l/5$. The first quadruplet forms at the plane midway between the two quadruplets an anamorphic image of the plane $z = z_0$, which is located at a

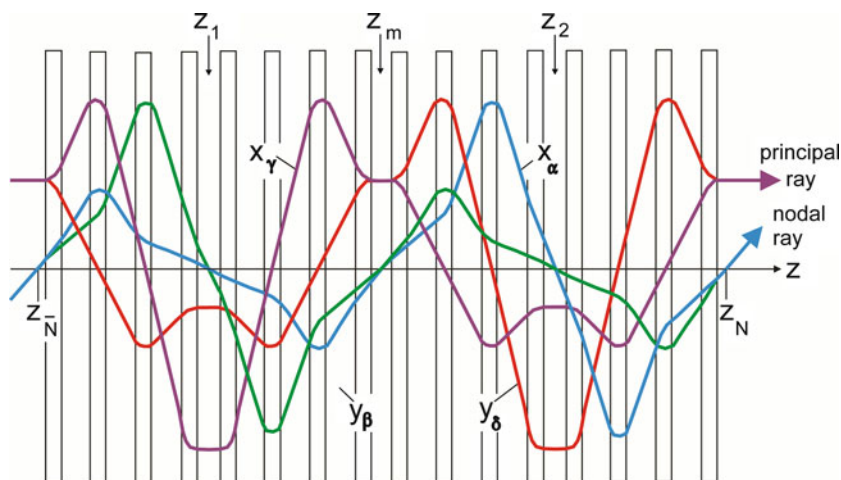


Fig. 4.46 Course of the fundamental rays in a system consisting of two doubly symmetric quadrupole octuplets forming an anamorphic image of the infinitely distant plane at the center of each octuplet

distance $0.3l$ in front of the quadruplet, as shown in Fig. 4.46. The magnification of the anamorphic image is $1/2$ in one section and 2 in the other section. Astigmatic images of the infinitely distant plane are formed in the two inner quadrupoles of each quadruplet. The total length of the quadrupole octuplet is $6.6l$, which is smaller than the corresponding length $8l$ of the septuplet.

While the configurations shown in Figs. 4.45a and 4.46b are superficially similar, there are differences, which affect their application as corrector elements. System B has eight astigmatic images of the infinitely distant plane, four per section. If we place a multipole element at one of these planes, it will have no effect for one component of the principal ray, while deflecting the other component as well as the components of the nodal ray. Suppose one puts identical octupoles at the centers of the second and seventh quadrupoles. Owing to the symmetry of the components of the principal ray and the antisymmetry of those of the nodal ray with respect to the plane z_1 , the effect of the first octupole on coma and distortion is canceled by the effect of the second octupole. If the octupoles are exited with opposite sign, the other aberrations cancel out while coma and distortion add up. Thus, one may apply a very selective correction procedure.

Another difference between the A and B configurations is that in B, the rays do not go as far off-axis as they do in A. Therefore, the aberrations induced by the quadrupoles themselves are smaller. The compromise one makes is that the stigmatic images formed in B between the 4th and 5th and again between the 12th and 13th quadrupoles are not as distorted as in A, so the correction exerted by correcting elements placed at these positions is not as selective as in the A design.

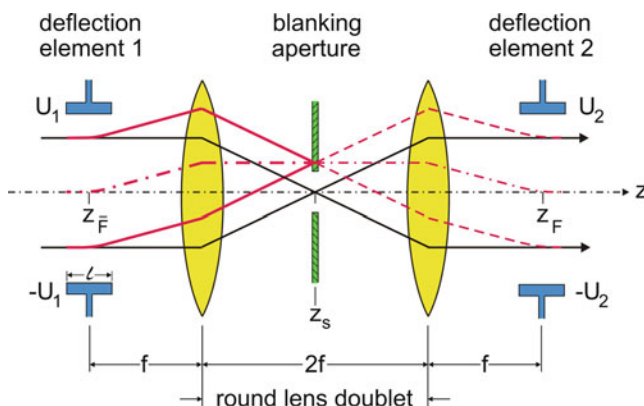


Fig. 4.47 Arrangement of the elements of the streak-free beam chopper and path the rays in the vertical $x - z$ section without deflection (*black*) and for chopping deflection (*red*)

4.11.1 Streak-Free Beam Chopper

Stroboscopic imaging is required for the time-resolved imaging of periodic processes in dynamic transmission electron microscopy (DTEM) and for the inspection of integrated circuits for measuring time-varying voltages applied to specific conducting structures of the chip. Pulsed electron beams are generated by (a) modulation of the Wehnelt voltage, (b) laser-pulsed photoemission cathode, or (c) beam deflection across a chopping aperture or knife edge. Owing to its simplicity, the last approach is widely employed in stroboscopic operating scanning electron microscopes whose electron guns have constant emission current. These instruments are used primarily for functional testing of wafers by voltage-contrast measurements. For this purpose, one needs highly rectangular pulses with very short rise times and with pulse widths as short as about fifty picoseconds. Moreover, the position and the size of the scanning spot at the chip must not be affected by the blanking operation. For visualizing periodic processes in objects by means of dynamic transmission electron microscopy, even shorter pulses are required.

Blanking systems consisting of a single deflection element are not suitable for most applications because they result in streaks preventing the spot from remaining at rest at the specimen plane during the blanking operation. A useful beam chopper must act like a spatially fixed flashlight which is switched alternately on and off. In order to achieve an equivalent device for electrons, we need a system composed of a telescopic round lens doublet and two identical deflection elements, one centered at the front focal plane of the first lens, the other centered at the back focal plane of the second round lens, as depicted in Fig. 4.47. Owing to the symmetric arrangement of the constituent elements of the beam blanker, the deflection, the astigmatism, and the dispersion introduced by the front deflector are compensated for by the second deflector. Because the telescopic lens doublet

images the first deflector with magnification $M = -1$ onto the second deflector, all primary effects of the deflectors cancel out provided that the time-dependent fields of the two deflectors are the same for each electron. In order that the second deflector eliminates completely the path deviation introduced by the first deflector, the voltage $U_2 = \pm U(t - \Delta t)$ applied to the second deflector must be delayed with respect to the voltage $U_1 = \pm U(t)$ by the time of flight Δt of the electrons between the exit of the first deflector and the entrance of the second deflector:

$$\Delta t = (4f - l)/v. \quad (4.313)$$

Here, l , f , and v are the length of each deflector, the focal length of the round lenses, and the electron velocity, respectively. We realize the time delay most conveniently by means of a delay cable of length $c_s(4f - l)/v$, where c_s is the velocity of the signal within the cable.

The time-dependent deflection fields introduce an energy broadening of the beam. We reduce this broadening by exciting each deflection unit antisymmetrically such that one electrode is at voltage U and the opposite electrode at voltage $-U$. Beam blanking is performed by moving the beam over the blanking aperture or knife edge placed at the symmetry plane z_s where the beam is focused. The rise time of the pulse will be the shorter the smaller the spot size is. The electric dipole field introduces an axial astigmatism apart from the deflection. The astigmatism splits up the point focus at the symmetry plane z_s into two line foci, one located in front of the symmetry plane the other located behind this plane. The horizontal line focus in front of the symmetry plane is parallel to the edge of the chopping aperture the other perpendicular to it. To guarantee that the rise time of the pulse will not be affected by the deflection, each deflector must consist of a quadrupole element enabling the excitation of a time-dependent dipole field and a superposed static quadrupole field. The strength of the quadrupole field is adjusted in such a way that the horizontal line focus approaches the symmetry plane when the deflected axis touches the edge of the beam chopper.

Chapter 5

General Principles of Particle Motion

Abstract Hamilton's principle, Liouville's theorem, the Lagrange and Poincare invariants of classical mechanics, and the eikonal procedure are employed to charged-particle optics for describing the properties of ensembles of rays and the deviation of the trajectories from their course in paraxial approximation. In particular, the mixed eikonal is employed for deriving the Abbe sine condition and the Poincare invariant for elucidating the focusing properties of magnetic fields. Based on the eikonal method, integral equations are derived for the deviations of the trajectories from their paraxial course.

Charged-particle optics investigates primarily the properties of bundles of rays in analogy to light optics. Important ensembles are homocentric bundles, whose trajectories originate from a common point, which is usually a point of an object or source. Within the frame of our considerations, we neglect Coulomb interactions between the charged particles of the beam. In this case, we can consider the beam as an ensemble of noninteracting particles, whose trajectories are entirely defined by the external electromagnetic fields.

5.1 Hamiltonian Formulation

When synchrotron radiation is negligibly small, we derive the equation of particle motion in arbitrary electromagnetic fields most conveniently from Hamilton's principle:

$$\delta W = \delta \int L(\vec{r}, \dot{\vec{r}}, t) dt = \delta \int [\vec{p}\dot{\vec{r}} - H] dt = 0. \quad (5.1)$$

The Lagrangian L is given by (2.14) and the canonical momentum is

$$\vec{p} = \text{grad}_{\dot{\vec{r}}} L = \frac{m_e \dot{\vec{r}}}{\sqrt{1 - \dot{\vec{r}}^2/c^2}} - e \vec{A}. \quad (5.2)$$

The Hamilton function H is a constant of motion $H = E = \text{const.}$ if the electromagnetic potentials $\vec{A}$ and φ do not depend on the time t . In this static case, we have $\delta H = 0$, and Hamilton's principle reduces to the *Principle of Maupertuis*, which adopts the form

$$\delta S = \delta(W + Et) = \delta \int_{\vec{r}_0}^{\vec{r}} \vec{p} d\vec{r} = \delta \int_{z_0}^z \tilde{L} dz = 0. \quad (5.3)$$

The optic axis forms the z -coordinate, which may be straight or curved. The reduced Lagrangian is defined as

$$\tilde{L} = \vec{p} \frac{d\vec{r}}{dz} = \tilde{L}(x, y, x', y'; z) = \tilde{L}(w, \bar{w}, w', \bar{w}'; z). \quad (5.4)$$

We must perform the integration along the true path from some initial plane $z = z_0$ to the plane of observation z . Dashes denote derivatives with respect to z , which serves as the independent variable substituting for the time t . We utilize the conservation of the total energy to eliminate the time to reduce the number of dependent variables from three to two. These variables are the lateral components $x = x(z)$ and $y = y(z)$ of the particle trajectory.

Employing complex notation, we find from the relations (3.58), (3.60), and (3.61) the reduced Lagrangian as

$$\tilde{L} = m_e c (\mu_e + \mu_m) = \sqrt{2em_e \phi^*} \sqrt{w' \bar{w}' + g_3^2} - e [g_3 A_z + \text{Re}(\bar{A} w')]. \quad (5.5)$$

The metric coefficient has the form $g_3 = 1 - \text{Re}(\Gamma \bar{w})$. Partial differentiation of the reduced Lagrangian with respect to $\bar{w}'$ gives the complex lateral component of the canonical momentum:

$$\tilde{p} = p = p_x + i p_y = 2 \frac{\partial \tilde{L}}{\partial \bar{w}'} = \sqrt{2em_e \phi^*} \frac{w'}{\sqrt{w' \bar{w}' + g_3^2}} - eA. \quad (5.6)$$

In order to obtain the reduced Hamiltonian $\tilde{H} = \tilde{H}(p, \bar{p}, w, \bar{w}, z)$, we employ the relation

$$\tilde{L} = \text{Re}(p \bar{w}') + p_z = p_x x' + p_y y' - \tilde{H}. \quad (5.7)$$

Employing the relation (5.6) and the conjugate complex expression, we substitute the components of the canonical momentum $p, \bar{p}$ for the slope components $w', \bar{w}'$, giving

$$\begin{aligned} \tilde{H}(p, \bar{p}, w, \bar{w}, z) &= -\tilde{p}_z = -\sqrt{\frac{2em_e \phi^*}{g_3^2 + w' \bar{w}'}} + e g_3 A_z \\ &= -\sqrt{2em_e \phi^* + (p + eA)(\bar{p} + e\bar{A})} + e g_3 A_z. \end{aligned} \quad (5.8)$$

The Lagrangian (5.5) is a simpler function than the Hamiltonian $\tilde{H} = \tilde{H}(x, p_x, y, p_y; z) = \tilde{H}(w, \bar{w}, p, \bar{p}; z)$, which is a function of five canonical coordinates. We can derive the complex path equation either from Hamilton's principle $\delta S = 0$ giving the Lagrange equation

$$\frac{d}{dz} \left(\frac{\partial \tilde{L}}{\partial \bar{w}'} \right) - \frac{\partial \tilde{L}}{\partial \bar{w}} = 0, \quad (5.9)$$

or from the equivalent Hamilton equations of classical mechanics

$$w' = 2 \frac{\partial \tilde{H}}{\partial \bar{p}}, \quad p' = -2 \frac{\partial \tilde{H}}{\partial \bar{w}}. \quad (5.10)$$

The accelerator physics community employs these equations [104], whereas the electron optics community uses primarily the eikonal method with Lagrangian (5.5). This method enables one to determine the trajectory in an elegant iterative way, starting from the linear paraxial approximation [10, 54, 105–107]. The general complex solution of the nonlinear second-order differential equation (5.9)

$$w = w(a_1, a_2, a_3, a_4; z) \quad (5.11)$$

is a function of its position z along the optic axis and of four real parameters $a_\nu = \bar{a}_\nu$, $\nu = 1, 2, 3, 4$. These constants of integration depend on the initial constraints imposed on the trajectory. They are generally the position coordinates at two planes or the position x_o, y_o and the slope components x'_o, y'_o of the ray at the object plane $z = z_o$. Hence, the trajectories as a whole form a four-dimensional manifold. Instead of the slope components, one uses often the components of the lateral canonical momentum $p_o = p(z_o) = p_{x_o} + i p_{y_o}$ at the object plane to define the ray. Then it is advantageous to gauge the vector potential to zero on the axis. The power series representation of the vector potential given in Sect. 3.4 fulfills this requirement. For an electron traveling along the optic axis we have $a_1 = a_2 = a_3 = a_4 = 0$. Since this electron stays on the axis, the relations

$$w(0, 0, 0, 0; z) = w^{(0)}(z) = 0, \quad p(0, 0, 0, 0; z) = p^{(0)}(z) = 0, \quad (5.12)$$

must hold. We obtain most conveniently the lateral component $p = p(a_1, a_2, a_3, a_4; z)$ of the canonical momentum from the relation (5.6) if we know the general solution (5.11) of the trajectory.

5.2 Lagrange Invariants

We can consider the eikonal S as an “optical potential” because the rays are the orthogonal trajectories to the surfaces of constant eikonal in the absence of a magnetic field, as outlined in Sect. 2.1.4. In the presence of a magnetic field, only the canonic momentum of the particle is orthogonal to these surfaces or the wave surfaces, respectively. Due to this behavior, relations exist between the position of the trajectories and their canonical momenta. In accordance with general convention, we call these invariants *Lagrange invariants* or *Lagrange brackets*, respectively.

To specify the Lagrange invariants, we assume that the lateral position and momentum of an arbitrary trajectory

$$w = w(a_1, a_2, a_3, a_4; z), \quad p = p(a_1, a_2, a_3, a_4; z) \quad (5.13)$$

are known functions of the z -coordinate and the ray parameters a_ν , $\nu = 1, 2, 3, 4$. The optic axis may be straight or curved. We consider an adjacent trajectory, whose ray parameters differ by small deviations δa_ν from the parameters a_ν of the chosen reference trajectory (5.13). In this case, we find the lateral separation and the difference of the lateral canonical momenta of the two trajectories as

$$\delta w = \sum_{\nu=1}^4 \frac{\partial w}{\partial a_\nu} \delta a_\nu, \quad \delta p = \sum_{\nu=1}^4 \frac{\partial p}{\partial a_\nu} \delta a_\nu. \quad (5.14)$$

If we choose the optic axis as the reference trajectory, the partial derivatives

$$\frac{\partial w}{\partial a_\nu} \big|_{a_\nu=0} = w_\nu(z), \quad \nu = 1, 2, 3, 4 \quad (5.15)$$

are identical with the paraxial rays $w_\nu = w_\nu(z)$ of the paraxial trajectory

$$w^{(1)}(z) = \sum_{\nu=1}^4 a_\nu w_\nu(z) \quad (5.16)$$

with nominal energy ($\kappa = \Delta E / E_0 = 0$) and lateral canonical momentum

$$p^{(1)} = \sum_{\nu=1}^4 a_\nu p_\nu(z), \quad p_\nu = p_\nu(z) = 2m_e c \frac{\partial \mu^{(2)}}{\partial \tilde{w}'} \big|_{w=w_\nu}. \quad (5.17)$$

Here $m_e c \mu^{(2)} = \tilde{L}^{(2)}$ represents the paraxial approximation of the reduced Lagrange function (5.5). By employing the relations (4.3) and (4.10), we obtain the relation

$$p_v = \frac{\partial p^{(1)}}{\partial a_v} = \sqrt{2m_e e \Phi^*} w'_v - \frac{i}{2} e B w_v = \sqrt{2m_e e \Phi^*} e^{i\chi} \frac{d(w_v e^{-i\chi})}{dz} = e^{i\chi} \sqrt{2m_e e \Phi^*} u'_v \quad (5.18)$$

existing between the canonical momentum p_v , the position w_v , and the slope w'_v of the fundamental ray. The angle χ defined by (4.24) is half the angle of the axial Larmor rotation. We can conceive

$$p_v e^{-i\chi} = \sqrt{2m_e e \Phi^*} u'_v \quad (5.19)$$

as the lateral canonical momentum of the fundamental ray $u_v = w_v e^{-i\chi}$ referred to the rotating (u, z) -coordinate system;

$$\sqrt{2m_e e \Phi^*} = p_z^{(0)} = q \quad (5.20)$$

is the z -component of the canonical momentum taken along the optic axis. This component coincides with that of the corresponding kinetic momentum because the magnetic vector potential vanishes along this axis owing to the chosen gauge.

By changing over from the reference trajectory w to the neighboring trajectory $w + \delta w$, the eikonal $S = S(a_1, a_2, a_3, a_4; z_0, z)$ changes by the amount

$$\delta S = \text{Re}(\bar{p} \delta w) - \text{Re}(\bar{p}_0 \delta w_0) \quad (5.21)$$

if the locations z_0 and z of the initial plane and the plane of observation, respectively, are kept fixed ($\delta z_0 = 0, \delta z = 0$). Varying the trajectory by changing the parameter a_μ by a small amount δa_μ gives

$$\delta w = \frac{\partial w}{\partial a_\mu} \delta a_\mu, \quad \delta w_0 = \frac{\partial w_0}{\partial a_\mu} \delta a_\mu. \quad (5.22)$$

By substituting the expressions (5.22) for δw and δw_0 in (5.21), we obtain

$$\frac{\partial S}{\partial a_\mu} = \text{Re} \left\{ \bar{p} \frac{\partial w}{\partial a_\mu} - \bar{p}_0 \frac{\partial w_0}{\partial a_\mu} \right\}. \quad (5.23)$$

Derivation of this equation with respect to another parameter a_v yields

$$\frac{\partial^2 S}{\partial a_v \partial a_\mu} = \text{Re} \left\{ \frac{\partial w}{\partial a_\mu} \frac{\partial \bar{p}}{\partial a_v} - \frac{\partial w_0}{\partial a_\mu} \frac{\partial \bar{p}_0}{\partial a_v} + \bar{p} \frac{\partial^2 w}{\partial a_v \partial a_\mu} - \bar{p}_0 \frac{\partial^2 w_0}{\partial a_v \partial a_\mu} \right\}. \quad (5.24)$$

We interchange the indices μ and v and subtract the resulting expression from (5.24) giving

$$[a_\mu, a_v] = \text{Re} \left\{ \frac{\partial w}{\partial a_\mu} \frac{\partial \bar{p}}{\partial a_v} - \frac{\partial w}{\partial a_v} \frac{\partial \bar{p}}{\partial a_\mu} \right\} = \text{Re} \left\{ \frac{\partial w_0}{\partial a_\mu} \frac{\partial \bar{p}_0}{\partial a_v} - \frac{\partial w_0}{\partial a_v} \frac{\partial \bar{p}_0}{\partial a_\mu} \right\} = I_{\mu v} = -I_{v\mu}. \quad (5.25)$$

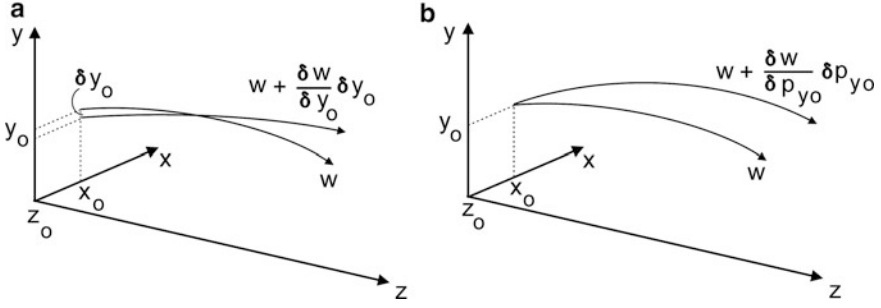


Fig. 5.1 Variation of the trajectory resulting from (a) an infinitesimal change δy_0 of the position of the initial point and (b) from a small change δp_{y0} of the initial canonical momentum

The third expression of this equation depends only on the location of the initial plane $z = z_0$. Therefore, the Lagrange bracket $[a_\mu, a_\nu] = I_{\mu\nu}$ is an invariant of motion for each pair of parameters. The six invariants $I_{\mu\nu} = -I_{\nu\mu}$ adopt most simple values if we fix the trajectory by its lateral position and momentum components

$$\begin{aligned} a_1 &= p_x(z_0) = p_{x0}, & a_3 &= x(z_0) = x_0, \\ a_2 &= p_y(z_0) = p_{y0}, & a_4 &= y(z_0) = y_0. \end{aligned} \quad (5.26)$$

Using these trajectory parameters, we obtain readily from the third expression in formula (5.25) the six invariants

$$I_{31} = I_{42} = 1, \quad I_{21} = I_{32} = I_{41} = I_{43} = 0. \quad (5.27)$$

To visualize the effect of an infinitesimal variation δa_μ of the ray parameters, we have plotted schematically in Fig. 5.1 the change of the trajectory for the variations $\delta a_\mu = \delta a_4 = \delta y_0$ and $\delta a_\nu = \delta a_2 = \delta p_{y0}$. The Lagrange brackets correlate the differential quotients

$$\frac{\partial w}{\partial a_\mu}, \quad \frac{\partial w}{\partial a_\nu}, \quad \frac{\partial p}{\partial a_\mu}, \quad \frac{\partial p}{\partial a_\nu} \quad (5.28)$$

of the lateral position and momentum of a trajectory with each other. Since all trajectories form a quadruple manifold of rays, six Lagrange relations exist. If we choose the optic axis as the reference trajectory, the Lagrange brackets degenerate to the Helmholtz–Lagrange relations. To prove this behavior, we insert the relations

$$\frac{\partial w}{\partial a_\mu} = \frac{\partial w}{\partial a_\mu} \Big|_{a_\nu=0} = \frac{\partial w^{(1)}}{\partial a_\mu} = w_\mu, \quad \frac{\partial \bar{p}}{\partial a_\nu} = \frac{\partial \bar{p}}{\partial a_\nu} \Big|_{a_\mu=0} = \frac{\partial \bar{p}^{(1)}}{\partial a_\nu} = \bar{p}_\nu \quad (5.29)$$

into (5.25). Considering further expressions (5.18) and (5.19), we obtain

$$\begin{aligned}
[a_\mu, a_\nu] &= \text{Re}\{w_\mu \bar{p}_\nu - w_\nu \bar{p}_\mu\} \\
&= q_0 \text{Re}\{u_\mu \bar{u}'_\nu - u_\nu \bar{u}'_\mu\} = I_{\mu\nu} \sqrt{2em_e \Phi_0^*} C_{\mu\nu},
\end{aligned} \tag{5.30}$$

which represents the Helmholtz–Lagrange relation (4.197) of Gaussian optics.

5.3 Liouville's Theorem

The trajectory of each particle is unambiguously defined by four real ray parameters. We can conceive these parameters as four degrees of freedom, which form the coordinates of a four-dimensional parameter space. A special case of this space is the so-called *phase space* spanned by the x - and y -coordinates and the lateral components p_x, p_y of the canonical momentum. This space is well established in statistical mechanics [40]. A point in this space defines entirely the position and the direction of a particle for a given value $z = z_0$ of the independent variable z . By changing this variable, we obtain a curve in phase space. We can conceive the manifold of all lines as the streamlines of an incompressible four-dimensional fluid. Such a fluid has the property that each volume element is an invariant of motion. The shape of the fluid element may change arbitrarily along its path, yet the enclosed volume stays constant.

Contrary to the streamlines of a real fluid, the trajectories in phase space can intersect each other. To avoid this difficulty, one extends the four-dimensional phase space to the five-dimensional *state space* by considering the z -variable as the fifth coordinate. In this space, each trajectory connects uniquely a given point of the four-dimensional initial “plane” z_0 with the plane of observation z , as illustrated in Fig. 5.2. The projections of these trajectories onto the initial four-dimensional plane $z = z_0$ represent the streamlines in phase space. The volume element in this space equals the four-dimensional surface element

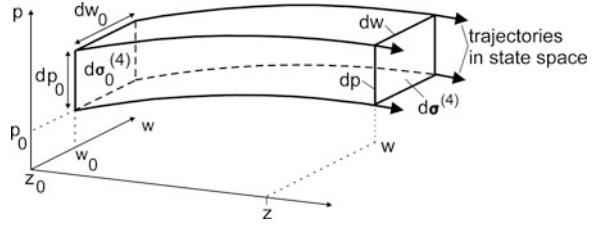
$$d\sigma_0^{(4)} = dx_0 dy_0 dp_{x0} dp_{y0} = -\frac{1}{4} dw_0 d\bar{w}_0 dp_0 d\bar{p}_0 \tag{5.31}$$

of the state space. The intersection points of sixteen adjacent trajectories with the four-dimensional plane z_0 form the corners of this surface element. The points of intersection of these trajectories at any other plane define unambiguously the conjugate element

$$d\sigma^{(4)} = dx dy dp_x dp_y = -\frac{1}{4} dw d\bar{w} dp d\bar{p}. \tag{5.32}$$

We can conceive the correlation between $d\sigma^{(4)}$ and $d\sigma_0^{(4)}$ as an imaging in state space. This so-called symplectic mapping describes how the phase-space coordinates of a particle at a plane z of the state space relate with those at the initial plane z_0 . The surface element (5.32) in phase space is related to the volume element

Fig. 5.2 Symplectic mapping in the five-dimensional phase space



$dV_a = da_1 da_2 da_3 da_4$ of the four-dimensional parameter space because the trajectories are functions of the ray parameters a_v . Therefore, the relation

$$d\sigma^{(4)} = -\frac{1}{4} D_a dV_a \quad (5.33)$$

holds, where

$$D_a = \sum_{(p)} (-1)^p \frac{\partial w}{\partial a_\mu} \frac{\partial \bar{p}}{\partial a_v} \frac{\partial \bar{w}}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} = \begin{vmatrix} \frac{\partial w}{\partial a_1} & \frac{\partial w}{\partial a_2} & \frac{\partial w}{\partial a_3} & \frac{\partial w}{\partial a_4} \\ \frac{\partial \bar{p}}{\partial a_1} & \frac{\partial \bar{p}}{\partial a_2} & \frac{\partial \bar{p}}{\partial a_3} & \frac{\partial \bar{p}}{\partial a_4} \\ \frac{\partial \bar{w}}{\partial a_1} & \frac{\partial \bar{w}}{\partial a_2} & \frac{\partial \bar{w}}{\partial a_3} & \frac{\partial \bar{w}}{\partial a_4} \\ \frac{\partial p}{\partial a_1} & \frac{\partial p}{\partial a_2} & \frac{\partial p}{\partial a_3} & \frac{\partial p}{\partial a_4} \end{vmatrix} \quad (5.34)$$

is the corresponding Jacobi determinant. We must perform the summation (p) with respect to all 24 permutations of the indices μ, v, σ , and τ , each of which has values 1, 2, 3, and 4. We evaluate the Jacobi determinant (5.34) as follows:

$$\begin{aligned} D_a &= \overline{D}_a = \frac{1}{2} \sum_{(p)} (-1)^p \left[\frac{\partial w}{\partial a_\mu} \frac{\partial \bar{p}}{\partial a_v} - \frac{\partial w}{\partial a_v} \frac{\partial \bar{p}}{\partial a_\mu} \right] \frac{\partial \bar{w}}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} \\ &= \sum_{(p)} (-1)^p \text{Re} \left[\frac{\partial w}{\partial a_\mu} \frac{\partial \bar{p}}{\partial a_v} - \frac{\partial w}{\partial a_v} \frac{\partial \bar{p}}{\partial a_\mu} \right] \frac{\partial \bar{w}}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} = \sum_{(P)} (-1)^P I_{\mu\nu} \frac{\partial \bar{w}}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} \\ &= \frac{1}{4} \sum_{(p)} (-1)^p I_{\mu\nu} \left[\frac{\partial \bar{w}}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} + \frac{\partial w}{\partial a_\sigma} \frac{\partial \bar{p}}{\partial a_\tau} - \frac{\partial \bar{w}}{\partial a_\tau} \frac{\partial p}{\partial a_\sigma} - \frac{\partial w}{\partial a_\tau} \frac{\partial \bar{p}}{\partial a_\sigma} \right] \\ &= \frac{1}{2} \sum_{(p)} (-1)^p I_{\mu\nu} I_{\sigma\tau} = 4 (I_{12} I_{34} - I_{13} I_{24} + I_{14} I_{23}). \end{aligned} \quad (5.35)$$

In the first step, we have split the sum (5.34) in two halves and in the second sum exchanged the indices μ and v . The minus sign considers the fact that this permutation changes the sign of the determinant. By adding two determinants with

two identical columns and substituting the Lagrange invariants (5.25) for the first factor of the second sum, we derive the expression on the right-hand side of the second row. Subsequently, we repeat this procedure. We split the result up in two sums and take the conjugate complex of one of them without changing its value because each sum is real. Since the second factor of the resulting expression in the third row is twice the Lagrange invariant $I_{\sigma\tau}$, we readily derive the last result in the fourth row.

If we fix the trajectory by its canonical initial values w_0 and p_0 , the ray parameters take the values (5.26) and the Lagrange invariants the values (5.27) resulting in

$$D_a = D_{\text{can}} = -4. \quad (5.36)$$

Substituting this expression for D_a in (5.33), we obtain

$$d\sigma^{(4)} = -\frac{1}{4}D_a da_1 da_2 da_3 da_4 = dx_0 dy_0 dp_{x0} dp_{y0} = d\sigma_0^{(4)}. \quad (5.37)$$

This result proves Liouville's theorem, which states that the differential volume element of the phase space is an invariant.

5.3.1 Paraxial Approximation

We obtain the paraxial approximation of the Jacobi determinant (5.34) by substituting the expressions (5.15) and (5.18) for $\partial w/\partial a_\mu$ and $\partial p/\partial a_\mu$, respectively:

$$D_a^{(1)} = \begin{vmatrix} w_1 & w_2 & w_3 & w_4 \\ \bar{p}_1 & \bar{p}_2 & \bar{p}_3 & \bar{p}_4 \\ \bar{w}_1 & \bar{w}_2 & \bar{w}_3 & \bar{w}_4 \\ p_1 & p_2 & p_3 & p_4 \end{vmatrix} = 2em_e \Phi^* \begin{vmatrix} w_1 & w_2 & w_3 & w_4 \\ \bar{w}'_1 & \bar{w}'_2 & \bar{w}'_3 & \bar{w}'_4 \\ \bar{w}_1 & \bar{w}_2 & \bar{w}_3 & \bar{w}_4 \\ w'_1 & w'_2 & w'_3 & w'_4 \end{vmatrix} = 2em_e \Phi_0^* D_W. \quad (5.38)$$

Hence, the Jacobi determinant degenerates to the Wronski determinant (4.202) in paraxial approximation apart from the constant factor $q_0^2 = 2em_e \Phi_0^*$.

5.3.2 Abbe Sine Condition

Ideal optical instruments image stigmatically points in the object plane $z = z_0$ into conjugate points in the image plane $z = z_i$. In this case, all trajectories originating from a point in the object plane intersect each other at the conjugate point in the image plane irrespective of their ray gradients. At first, we assume an instrument, which only images the center of the object plane $w_0 = 0$ ideal in the center $w_i = 0$ of the image plane. We achieve this by eliminating the spherical aberration of the

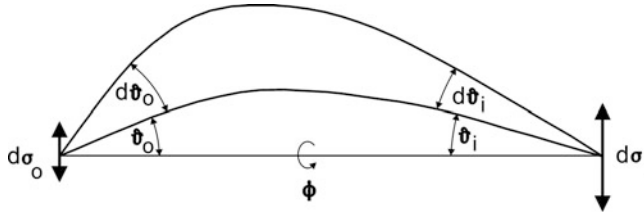


Fig. 5.3 Diagram for illustrating the derivation of the Abbe sine condition

instrument. Next, we want to know which additional condition must hold in order that a small region $d\sigma_o = dx_o dy_o$ around the object center is imaged ideally. We can answer this question most conveniently by utilizing the invariance of the phase-space element. Since the conjugate surface elements $d\sigma$ and $d\sigma_o$ are centered on the axis, we have

$$dp_{x_i} dp_{y_i} = 2em_e \Phi_i^* \cos \vartheta_i d\Omega_i, \quad dp_{x_o} dp_{y_o} = 2em_e \Phi_o^* \cos \vartheta_o d\Omega_o. \quad (5.39)$$

Here we have considered that the vector potential vanishes along the optic axis, so that the canonical momentum coincides with the kinetic momentum of the particle. The central axial trajectory of an infinitesimal bundle of rays has starting angle ϑ_o and intersects the center of the image plane with slope angle ϑ_i , as shown in Fig. 5.3.

The differential solid angles

$$d\Omega_o = 2\pi \sin \vartheta_o d\vartheta_o, \quad d\Omega_i = 2\pi \sin \vartheta_i d\vartheta_i \quad (5.40)$$

are hollow cones, which confine the homocentric bundle of rays in the vicinity of the image center and object center, respectively. Due to the invariance of the phase-space element, we have

$$\Phi_o^* d\sigma_o \sin \vartheta_o \cos \vartheta_o d\vartheta_o = \Phi_i^* d\sigma_i \sin \vartheta_i \cos \vartheta_i d\vartheta_i. \quad (5.41)$$

The zonal magnification

$$M = M(\vartheta_o, \vartheta_i) = \sqrt{\frac{d\sigma_i}{d\sigma_o}} \quad (5.42)$$

of the surface element depends only on the slope angles of the central axial ray at the object and image plane. This surprising fact is a consequence of the eikonal or Liouville's theorem, respectively. In order to obtain an ideal image, the zonal magnification must be the same, regardless of the ray gradients of the corresponding central ray:

$$M = M_0 = \text{const.} \quad (5.43)$$

Assuming the instrument satisfies this requirement, we can perform the integrations in (5.41) over the angles ϑ_o and ϑ_i giving

$$M_0^2 \Phi_i^* \sin^2 \vartheta_i = \Phi_o^* \sin^2 \vartheta_o. \quad (5.44)$$

By taking the square root of this expression and employing the electron optical index of refraction $n_o \propto \Phi^{*1/2}$ on the optic axis, we derive the Abbe sine condition of light optics

$$\frac{n_{o0} \sin \vartheta_o}{n_{oi} \sin \vartheta_i} = M_0 = \text{const.} \quad (5.45)$$

When this condition is satisfied irrespectively of the ray angles, the optical system images sharply a small area around the center of the object plane onto the image plane. This behavior implies that spherical aberration and off-axial coma must vanish in all orders. Systems, which are free of these aberrations, are called *aplanats* in light optics. One has adopted this terminology in electron optics for systems, which are corrected for spherical aberration and off-axial coma in third order. Hence, the zonal magnification (5.45) of these systems is not constant for zones governed by the higher-order aberrations. The angles ϑ_o and ϑ_i are small if we confine the trajectories to the paraxial regime. Then we can approximate with a sufficient degree of accuracy the sine functions in (5.45) by their arguments resulting in the Helmholtz–Lagrange relation (4.58). Since this relation is valid only in the Gaussian regime, we cannot apply it in the case of large ray gradients. Contrary to light-optical aplanats, any electron aplanat must contain non-rotationally symmetric elements. This requirement is a consequence of the *Scherzer theorem*, which states that axial chromatic and spherical aberrations of static round electron lenses are unavoidable.

5.4 Generalized Symplectic Matrices

We can view the propagation of charged particles as a so-called symplectic mapping in phase space. Such a mapping describes how the canonical variables w and p at some plane of observation z relate to their initial value w_0 and p_0 . By combining w and p to a complex canonical vector in phase space $\vec{R}^t = (w, p)$ and employing the antisymmetric matrix

$$\overleftrightarrow{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (5.46)$$

we can write Hamilton equations (5.10) as the following single equation in matrix form:

$$\vec{R}' = \begin{pmatrix} w' \\ p' \end{pmatrix} = 2 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \partial \tilde{H} / \partial \bar{w} \\ \partial \tilde{H} / \partial \bar{p} \end{pmatrix} = 2 \overleftrightarrow{j} \partial_{\vec{R}} \tilde{H}, \quad \partial_{\vec{R}} = \begin{pmatrix} \partial / \partial \bar{w} \\ \partial / \partial \bar{p} \end{pmatrix}. \quad (5.47)$$

We can consider the matrix (5.46) as a symplectic 2×2 matrix. If we use real variables, a vector in phase space has *four* components, and the fundamental symplectic matrix $\overleftrightarrow{J}$ is a 4×4 matrix. It has the properties

$$\overleftrightarrow{J}^2 = -\overleftrightarrow{I}, \quad \overleftrightarrow{J}^{-1} = \overleftrightarrow{J}^t = -\overleftrightarrow{J}, \quad \det \overleftrightarrow{J} = 1. \quad (5.48)$$

Here $\overleftrightarrow{I}$ is the unit matrix and $\overleftrightarrow{J}^t$ the transposed matrix. The definitions (5.48) are not unique because several representations exist for $\overleftrightarrow{J}$, which satisfy these conditions. In deriving relatively simple expressions for the higher-order deviations of the trajectory from its paraxial approximation, it is advantageous to introduce the matrix

$$\overleftrightarrow{J}_C = -\frac{4}{D_W} \begin{pmatrix} 0 & C_{34} & C_{42} & C_{23} \\ C_{43} & 0 & C_{14} & C_{31} \\ C_{24} & C_{41} & 0 & C_{12} \\ C_{32} & C_{13} & C_{21} & 0 \end{pmatrix}. \quad (5.49)$$

The elements of this matrix are the Helmholtz–Lagrange invariants (4.211) and (5.30).

We obtain the elements of row μ and column ν by cyclic permutation of the indices ν, σ, τ with $\nu < \sigma < \tau$. These indices differ from each other, and each of them takes successively one of the values 1, 2, 3, and 4. The matrix is antisymmetric due to the relation $C_{\mu\nu} = -C_{\nu\mu}$. Accordingly, the transposed matrix is $\overleftrightarrow{J}_C^t = -\overleftrightarrow{J}_C$. Using (4.217) for the Wronski determinant D_W , we find the determinant of the matrix $\overleftrightarrow{J}_C$ as

$$\det \overleftrightarrow{J}_C = \left(\frac{4}{D_W} \right)^4 (C_{12}C_{34} - C_{13}C_{24} + C_{14}C_{23})^2 = (C_{13}C_{24} - C_{12}C_{34} + C_{14}C_{23})^{-2}. \quad (5.50)$$

We cannot provide the six elements of the matrix (5.49) arbitrarily, because the four linearly independent solutions of the paraxial path equation define these invariants.

However, we can put a constraint on these solutions by requiring that the matrix (5.49) is symplectic. This is, for example, the case if we put four elements of the matrix equal to zero and the two nonvanishing equal to 1. However, these constants must be chosen in such a way that $\det \overleftrightarrow{J}_C = 1$. The choice of the constants determines the structure of the fundamental symplectic matrix $\overleftrightarrow{J}_C = \overleftrightarrow{J}$. Three different representations exist, because the determinant (5.50) contains three terms, two of them can be put zero, while the remaining term must be 1. For example, the choice

$$C_{12} = -C_{34} = 1, \quad C_{13} = C_{14} = C_{23} = C_{24} = 0 \quad (5.51)$$

yields the representation

$$\overleftrightarrow{J} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \quad (5.52)$$

The choice

$$C_{31} = C_{42} = 1, \quad C_{12} = C_{14} = C_{23} = C_{34} = 0 \quad (5.53)$$

gives

$$\overleftrightarrow{J} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad (5.54)$$

while the last choice

$$C_{14} = -C_{23} = C_{32} = 1, \quad C_{12} = C_{13} = C_{24} = C_{34} \quad (5.55)$$

results in the representation

$$\overleftrightarrow{J} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (5.56)$$

We derive the transposed matrices if we exchange the indices of the coefficients in (5.51), (5.53), and (5.55). The choice of the representation of the fundamental symplectic matrix largely determines the initial conditions of the four linearly independent solutions w_1, w_2, w_3 , and w_4 of the paraxial path equation. The matrix (5.49) is very suitable for writing the inhomogeneous solution of the path equation (4.29) in a concise form. For this purpose, we introduce the complex four-dimensional *paraxial* trajectory vector

$$\vec{U}^{(1)} = \vec{U}^{(1)}(z) = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix}, \quad \vec{U}^{(1)\dagger} = (U_1 \ U_2 \ U_3 \ U_4) \quad (5.57)$$

and the corresponding vectors

$$\vec{w}^{(1)} = \vec{U}^{(1)} \left(\frac{\Phi_0^*}{\Phi^*} \right)^{1/4} e^{i\chi} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}, \quad \vec{w}^{(1)\dagger} = (w_1 \ w_2 \ w_3 \ w_4). \quad (5.58)$$

The components $w_\mu = w_\mu(z)$, $u_\mu = u_\mu(z) = w_\mu e^{-i\chi}$, $U_\mu = U_\mu(z) = u_\mu (\Phi^* / \Phi_0^*)^{1/4}$ are four linearly independent solutions of the homogeneous part of the path equation

(4.29). The column vectors are the transposed vectors. By considering an arbitrary complex perturbation

$$P = P(U, \bar{U}, U', \bar{U}'; z) = P(w, \bar{w}, w', \bar{w}'; z), \quad (5.59)$$

we can transform the nonlinear differential equation

$$U'' + TU - G\bar{U} = P \quad (5.60)$$

into the inhomogeneous integral equation

$$U = U^{(1)} + \bar{U}^{\dagger} \overset{\leftrightarrow}{J}_C \int_{z_0}^z \text{Re} \left(\bar{P} \bar{U} \right) dz. \quad (5.61)$$

Here we have assumed that the trajectory is defined by some initial conditions at the object plane. The term $U^{(1)} = U^{(1)}(z)$ is the solution of the linear part $U'' + TU - G\bar{U} = 0$ of the complex nonlinear path equation (5.60). The transformation of a differential equation into an integral equation allows one to incorporate the initial conditions defining a distinct trajectory. Moreover, if the nonlinear perturbation P is sufficiently weak, we can solve the integral equation (5.61) iteratively by employing the method of successive approximation, which starts with the paraxial solution $U = U^{(1)}(z)$.

If the perturbation has the form $P = P(z)$, as it is the case for the dispersion (4.244), equation (5.61) represents the solution of (5.60) for a distinct trajectory. In this case, formula (5.61) is the matrix representation of the dispersion ray (4.256).

5.5 Poincaré's Invariant

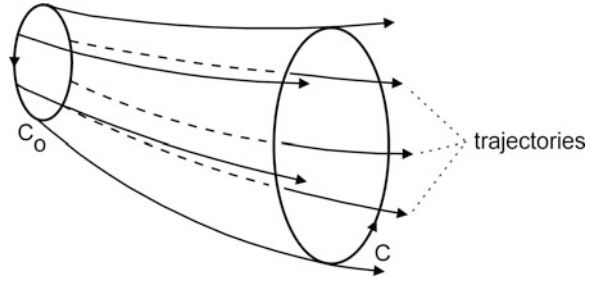
Poincaré's integral invariant is closely connected with the Lagrange invariant. To derive the Poincaré invariant, we consider a tube of nonintersecting trajectories, as illustrated in Fig. 5.4. On the mantle surface, we choose two closed loops C_0 and C so that the mantle trajectories intersect both contours. By going from a trajectory to the neighboring trajectory, the optical path length or eikonal, taken along the initial trajectory between the two loops, changes by

$$dS = \vec{p} d\vec{r} - \vec{p}_0 d\vec{r}_0. \quad (5.62)$$

Here $\vec{p} = p_x \vec{e}_x + p_y \vec{e}_y + p_z \vec{e}_z$ is the three-dimensional canonical momentum vector; $d\vec{r}$ and $d\vec{r}_0$ are the infinitesimal displacements along the contour C and the initial contour C_0 , respectively. Because dS is a total differential, its loop integral must vanish:

$$\oint_C dS = \oint_C \vec{p} d\vec{r} - \oint_{C_0} \vec{p}_0 d\vec{r}_0 = 0. \quad (5.63)$$

Fig. 5.4 Integration loops C and C_0 on the mantle of a bundle of trajectories employed for obtaining Poincaré's invariant



Since we can choose the location of the loops arbitrarily, the expression

$$I_P = \oint_C \vec{p} d\vec{r} = \oint_C m \dot{\vec{r}} d\vec{r} - e \oint_C \vec{A} d\vec{r} = \oint_C m \dot{\vec{r}} d\vec{r} - e \Phi_m = \oint_{C_0} \vec{p}_0 d\vec{r}_0 \quad (5.64)$$

must be an invariant, which is called *Poincaré's invariant*.

By employing Stoke's theorem, we find that the loop integral

$$\Phi_m = \oint_C \vec{A} d\vec{r} = \iint_{\sigma} \text{curl } \vec{A} d\vec{\sigma} = \iint_{\sigma} \vec{B} d\vec{\sigma} \quad (5.65)$$

is the magnetic flux through the surface σ embraced by the loop C . The Poincaré invariant is zero if the trajectories of the tube originate from a common source forming a homocentric bundle of rays. Since we can move the loop C_0 along the surface of the tube without changing the invariant (5.64), the loop shrinks to a point at the origin. Hence, the Poincaré invariant is nonzero only if the trajectories on the tube do not emanate from a common source point. In the following, we demonstrate the usefulness of the Poincaré invariant by means of two illustrating examples.

First, we consider a homocentric bundle of rays, which propagate through a magnetic field. Since the rays form a normal congruence, the Poincaré invariant vanishes ($I_P = 0$). Let us assume that the common origin of the trajectories is located in a field-free region. Then, the paths of the charged particles are perpendicular to the surfaces of constant eikonal due to the condition $\vec{p} = m\dot{\vec{r}} = \text{grad } S$. Once these particles enter the magnetic field, their canonical momenta $\vec{p} = m\dot{\vec{r}} - e\vec{A}$ remain normal to the wave surfaces while their trajectories form a skew (non-normal) congruence. To demonstrate this behavior, we surround the mantle of the rays by a curve, which intersects these trajectories at right angles, as shown in Fig. 5.5. Within the field-free region, this curve is a closed loop located on a surface of constant eikonal. However, in the region of the magnetic field, the curve forms a spiral. It reaches the trajectory through the starting point A at a point B some distance away from A. We close the curve by the section $BA = l$ on the starting trajectory. Since we have assumed a purely magnetic field, the velocity $v = \dot{r}$ of the particles along their trajectories is constant. Since $I_P = 0$, we readily find from (5.64) that the skewness

Fig. 5.5 Path of integration for determining the twist of a homocentric bundle of skew rays

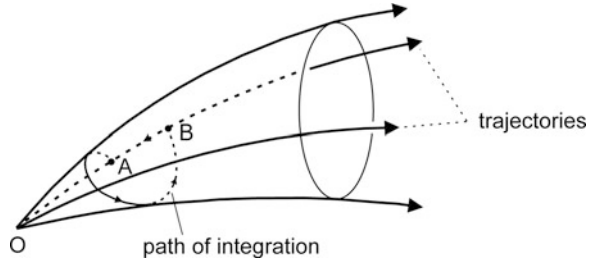
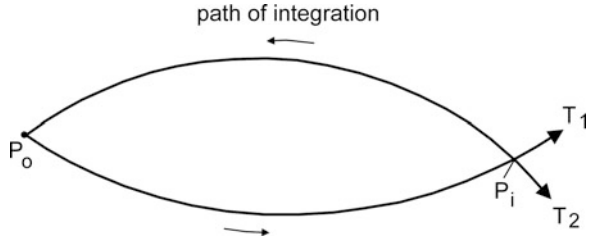


Fig. 5.6 Path of integration for determining the difference between the path lengths of the trajectories T_1 and T_2 connecting the conjugate points P_0 and P_i



$$l = \frac{e}{mv} \Phi_m \quad (5.66)$$

of the bundle is proportional to the magnetic flux enclosed by the chosen loop.

Next, we consider two trajectories T_1 and T_2 of an ensemble of rays intersecting the conjugate the points P_0 and P_i , as depicted in Fig. 5.6. We form a closed path by traveling along the trajectory T_1 from P_0 to P_i and back along the trajectory T_2 . Since the velocity is constant in a magnetic field, we can readily perform the integration in (5.64) along the closed contour to obtain

$$l_1 - l_2 = \frac{e}{mv} \Phi_m, \quad (5.67)$$

where l_1 and l_2 are the arc lengths of the trajectories T_1 and T_2 , respectively, between the points P_0 and P_i .

Although Eqs. (5.66) and (5.67) are formally identical, their physical consequences are different. The latter formula demonstrates that a magnetic field can form an image, whereas formula (5.66) states that this field generally twists the rays, as it happens in rotationally symmetric systems. Such a twist does not arise if the magnetic field is perpendicular to the motion of the particle. In this two-dimensional case, the trajectories lie on a common plane. We assume a magnetic field, which is homogeneous in some sections and zero else such that there is a sharp cut-off at the field boundaries. Although such an assumption is unrealistic, it allows one to demonstrate that it is possible to focus all particles, emanating from point in a plane perpendicular to the magnetic field, into a common conjugate image point. We can realize approximately a homogeneous magnetic field with sharp cut-off fringing field by two plane-parallel iron plates with small gap width illustrated in Fig. 5.7.

Fig. 5.7 Formation of a homogeneous magnetic field $\vec{B}$ within the gap between two plane-parallel iron plates

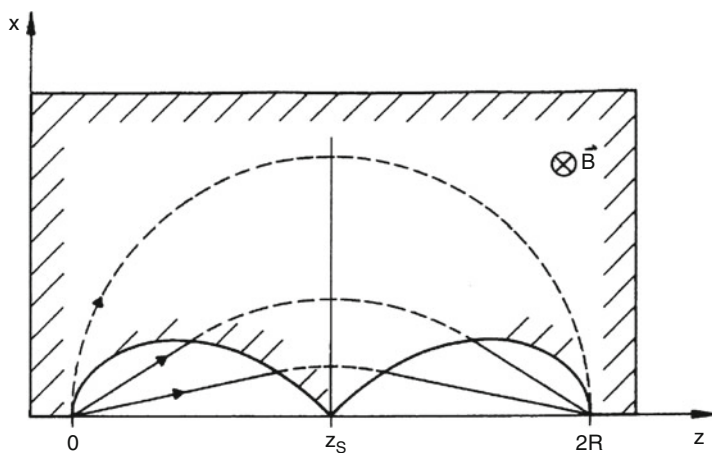
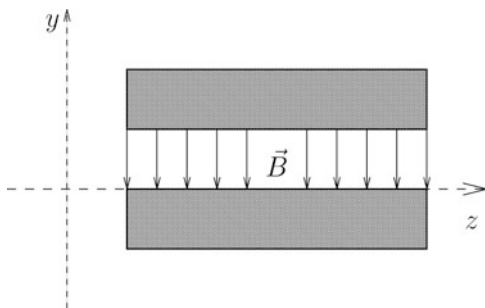


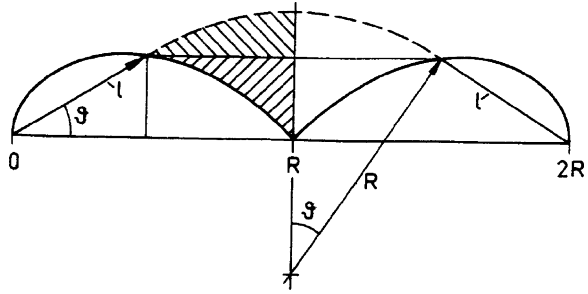
Fig. 5.8 Form of the cutouts of the iron plates providing perfect imaging in the midplane $y = 0$ between the two plane-parallel plates, as illustrated by different trajectories

To obtain field-free sections, we cut the corresponding areas out of the plates. However, we must know the precise shape of these areas in order to achieve ideal two-dimensional imaging. The trajectories perpendicular to the homogeneous magnetic field are circles with radius

$$R = \frac{mv}{eB}. \quad (5.68)$$

Adjacent trajectories starting from the origin ($x = 0, y = 0, z = 0$) intersect each other at points whose locations depend on the initial direction of the trajectories. The locus of these points forms a caustic. In order that the caustic degenerates to a point, we place two field-free sections between the two conjugate points. The sections are symmetric with respect to the y, z -plane. Figure 5.8 shows the upper halves of these sections together with three trajectories. The conjugate image point

Fig. 5.9 Area of the magnetic field of the homogeneous magnet shown in Fig. 5.8 enclosed by the optic axis and a trajectory connecting the conjugate points 0 and $2R$; the *hatched region* is the left half of the enclosed area



is located on the optic axis at a distance $2R$ from the source point. The trajectories are symmetric with respect to the plane $z = z_s = R$. It follows from Fig. 5.9 that the rays are composed of two straight lines with length

$$l = l(\vartheta) = R \frac{1 - \sin \vartheta}{\cos \vartheta} \quad (5.69)$$

and a circular arc with length $2R\vartheta$. The angle ϑ is the starting angle of the trajectory with respect to the z -axis. Equation (5.69) follows from the condition $l \cos \vartheta + R \sin \vartheta = R$, which guarantees that all trajectories run parallel to the optic axis at the symmetry plane z_s . We can also derive the curve (5.69) for the field boundary by means of (5.67) derived from the Poincaré invariant (5.74). The length of the trajectory is

$$l_1 = 2(l + R\vartheta). \quad (5.70)$$

We choose the optic axis as the other trajectory giving $l_2 = 2R$. The magnetic flux enclosed by this loop is

$$\Phi_m = \sigma B = \frac{mv}{eR} \sigma, \quad (5.71)$$

where the surface $\sigma = \sigma_u + \sigma_l$ is twice the hatched area shown in Fig. 5.9. We readily find the upper part of this area as

$$\frac{\sigma_u}{2} = \frac{1}{2} R^2 \vartheta - \frac{R^2}{2} \sin \vartheta \cos \vartheta. \quad (5.72)$$

The lower part of the hatched area is more difficult to obtain because it requires integration over the angle ϑ to yield

$$\begin{aligned} \frac{\sigma_l}{2} &= \frac{l^2}{2} \sin \vartheta \cos \vartheta + l \sin \vartheta (R - l \cos \vartheta) - \frac{1}{2} \int_0^{\vartheta} l^2(\vartheta) d\vartheta \\ &= \frac{Rl}{2} \sin \vartheta (1 + \sin \vartheta) - \frac{R^2}{2} \int_0^{\vartheta} \frac{(1 - \sin \vartheta)^2}{\cos^2 \vartheta} d\vartheta. \end{aligned} \quad (5.73)$$

Integration by part gives

$$\frac{1}{2} \int_0^{\vartheta} \frac{(1 - \sin \vartheta)^2}{\cos^2 \vartheta} d\vartheta = -\frac{1 - \sin \vartheta}{\cos \vartheta} + 1 - \frac{\vartheta}{2}. \quad (5.74)$$

By inserting this expression in (5.73) and substituting subsequently (5.69) for l , we obtain

$$\sigma_l = R^2 \left[\sin \vartheta \cos \vartheta + \vartheta - 2 + 2 \frac{1 - \sin \vartheta}{\cos \vartheta} \right]. \quad (5.75)$$

Substituting this expression for σ_l and (5.72) for σ_u in the relation

$$l_1 - l_2 = \frac{e}{mv} \Phi_m = \frac{\sigma_u + \sigma_l}{R}, \quad (5.76)$$

we prove that the result

$$l_1 - l_2 = 2R \left[\vartheta - 1 + \frac{1 - \sin \vartheta}{\cos \vartheta} \right], \quad R = \frac{mv}{eB}, \quad (5.77)$$

coincides with that obtained by substituting the paths length (5.70) for l_1 and $2R$ for l_2 on the left side of (5.77). A small deviation ΔE of the energy from its nominal value E_0 shifts the location of the image plane by

$$\Delta l_2 = 2\Delta R = \frac{2m\Delta v}{eB} = \frac{mv_0}{eB} \frac{\Delta E}{E_0} = R_0 \kappa. \quad (5.78)$$

One has utilized this chromatic shift in the so-called *orange spectrometer*. We can conceive this spectrometer as an ensemble of two-dimensional spectrometers centered about a common axis such that the azimuth angle is the same between any two adjacent segments, as it is the case for the slices of an orange. Instead of iron plates, one incorporates coils with proper shape of their windings. Our result for the two-dimensional case is very suitable as a first approximation for the optimum shape of the coils.

5.6 Eikonals

In 1895, the German mathematician Bruns introduced a characteristic function for calculating light-optical problems and named it *eikonal* derived from the Greek word icon meaning image [108]. This image function depends on four variables, because any trajectory is defined entirely by four parameters. Hence, the manifold of all trajectories is four-dimensional. In charged-particle optics, the so-called point eikonal is identical with the reduced action

$$S = \int_{\vec{r}_0}^{\vec{r}} \vec{p} d\vec{r} = q_0 \int_{z_0}^z \mu dz = S(x_0, y_0; x, y) = S(w_0, \bar{w}_0; w, \bar{w}). \quad (5.79)$$

The point eikonal (5.79) is a function of the lateral coordinates $w_0 = x_0 + iy_0$ and $w = x + iy$ of the intersection points of a ray with the fixed terminal planes z_0 and z , respectively. This restriction on the planes is of no importance because we can choose their locations arbitrarily. Since the z -coordinate takes over the role of the time or path length, respectively, we must vary the variational function $\mu = \mu(w, \bar{w}; w', \bar{w}'; z)$ with respect to the position and slope variables giving

$$\delta S = \text{Re} \left\{ p \delta \bar{w} - p_0 \delta \bar{w}_0 + \int_{z_0}^z \delta \bar{w} \left(2q_0 \frac{\partial \mu}{\partial \bar{w}} - \frac{dp}{dz} \right) dz \right\} = \text{Re} \{ p \delta \bar{w} - p_0 \delta \bar{w}_0 \}, \quad (5.80)$$

where

$$p = 2q_0 \frac{\partial \mu}{\partial \bar{w}'} \quad (5.81)$$

is the complex lateral component of the canonical momentum. The expression in the parenthesis of the integrand vanishes according to Hamilton's principle $\delta S = 0$ for fixed terminal points ($\delta \bar{w} = 0, \delta \bar{w}_0 = 0$). We derive the complex lateral components of the canonical momentum at the terminal planes from the variation (5.80) of S with respect to the lateral position of the ray-defining points as

$$p = 2 \frac{\partial S}{\partial \bar{w}}, \quad p_0 = -2 \frac{\partial S}{\partial \bar{w}_0}. \quad (5.82)$$

The function (5.79) is called *point eikonal* because it depends on the coordinates of the terminal points. Usually one defines the ray by its lateral position and momentum components at the starting plane $z = z_0$. To find the position w of the ray at the end-plane z , we must solve the implicit second equation of (5.82) with respect to w . Due to this difficulty and because we can solve the Hamilton–Jacobi equation analytically only in rather trivial cases, one has often argued that the eikonal method is unsuitable for practical calculations. However, this conjecture does not hold true if we apply well-established perturbation techniques for determining the eikonal iteratively.

Apart from the point eikonal, other eikonals exist. However, they all have the property that two of the four ray parameters belong to the initial plane and two to the end plane, which is usually the image plane.

We obtain the different eikonals by considering that $\delta S = \text{Re}\{p d\bar{w} - p_0 d\bar{w}_0\}$ is a total differential. Therefore, we can construct other eikonal by adding on both sides of this relation a total differential. This procedure corresponds to a *Legendre transformation*, which replaces one set of variables by another [29, 40]. In this way, we may construct the *mixed* eikonals

$$\begin{aligned} V &= V(p, \bar{p}; w_0, \bar{w}_0) = S - \text{Re}\{p\bar{w}\}, \\ \hat{V} &= \hat{V}(p_0, \bar{p}_0; w, \bar{w}) = S + \text{Re}\{p_0\bar{w}_0\}. \end{aligned} \quad (5.83)$$

The terminology *mixed* indicates the use of different kinds of variables. In light optics, one uses predominantly the *angle* eikonal, which corresponds to the *momentum* eikonal

$$M = M(p, \bar{p}; p_0, \bar{p}_0) = S + \text{Re}\{p\bar{w} - p_0\bar{w}_0\} \quad (5.84)$$

in charged-particle optics. The eikonals S , V , $\hat{V}$, and M are related to each other by Legendre transformations in the same way as the thermodynamic potentials. Accordingly, we can conceive the eikonals as optical potentials. The eikonals normalized with respect to the momentum $m_e c$ represent optical path lengths. For example, the mixed eikonal V is the optical path length of the ray between the starting point w_0, z_0 , and the foot point of the perpendicular dropped upon the ray at the origin $w = 0$ of the final plane z . If this plane is not located in a field-free region, we must drop the perpendicular upon the tangent of the ray taken at the intersection point w, z .

We obtain the lateral components w and p_0 of the trajectory from the variation

$$\delta V = \delta S - \text{Re}\{p\delta\bar{w} + \bar{w}\delta p\} = -\text{Re}\{\bar{w}\delta p + p_0\delta\bar{w}_0\} \quad (5.85)$$

of the mixed eikonal V as

$$w = -2\frac{\partial V}{\partial \bar{p}}, \quad p_0 = -2\frac{\partial V}{\partial \bar{w}_0}, \quad (5.86)$$

because the infinitesimal variations $\delta\bar{w}_0$ and $\delta\bar{p}$ can be chosen arbitrarily. Accordingly, we find from the variation $\delta\hat{V} = \text{Re}\{p\delta\bar{w} + \bar{w}_0\delta p_0\}$ of the mixed eikonal and the variation $\delta M = \text{Re}\{w_0\delta\bar{p}_0 - w\delta\bar{p}\}$ of the momentum eikonal the relations

$$p = 2\frac{\partial \hat{V}}{\partial \bar{w}}, \quad w_0 = 2\frac{\partial \hat{V}}{\partial \bar{p}_0}, \quad w = -2\frac{\partial M}{\partial \bar{p}}, \quad w_0 = 2\frac{\partial M}{\partial \bar{p}_0}. \quad (5.87)$$

The expressions (5.82), (5.86), and (5.87) demonstrate that we derive the lateral components of the trajectory at the terminal planes by partial differentiation with respect to their conjugate momentum variables and vice versa. We can also use these expressions for obtaining relations between partial derivatives of the ray variables at one terminal plane taken with respect to those of the other plane. Crosswise differentiation of each row in (5.82), (5.86), and (5.87) yields the relations

$$\frac{\partial p}{\partial \bar{w}_0} = -\frac{\partial p_0}{\partial \bar{w}}, \quad \frac{\partial w}{\partial \bar{w}_0} = \frac{\partial p_0}{\partial \bar{p}}, \quad \frac{\partial p}{\partial \bar{p}_0} = \frac{\partial w_0}{\partial \bar{w}}, \quad \frac{\partial w_0}{\partial \bar{p}} = -\frac{\partial w}{\partial \bar{p}_0}. \quad (5.88)$$

The complex conjugates of these expressions form another set of relations.

5.6.1 Mixed Eikonal and Sine Condition

As demonstrated by Abbe [40], the sine condition must be fulfilled in order that a small area centered about the optic axis at object plane $z_0 = z_o$ is imaged perfectly by the optical system into the conjugate image plane $z = z_i$. Optical systems which satisfy the sine condition are free of spherical aberration and off-axial coma in any order. An optical arrangement that fulfills these requirements is called an *aplanatic* system or *aplanat*, respectively. We derive the Abbe sine condition most conveniently by considering the mixed eikonal $V_i = V_i(w_o, \bar{w}_o; p_i, \bar{p}_i)$ taken at the image plane [107]. By expanding this eikonal in a power series with respect to the off-axial object coordinates w_o and $\bar{w}_o$, we obtain

$$V_i = V_{i0}^{(0)} + \text{Re}[V_{i0}^{(1)}\bar{w}_o] + V_{i0}^{(1,1)}w_o\bar{w}_o + \text{Re}[V_{i0}^{(2)}\bar{w}_o^2] + \dots \quad (5.89)$$

The expansion coefficients

$$V_{i0}^{(\mu,\nu)} = V_{i0}^{(\mu,\nu)}(p_i, \bar{p}_i) = \frac{1}{\mu!\nu!} \left. \frac{\partial^\mu \partial^\nu V(p_i, \bar{p}_i, w_o, \bar{w}_o)}{\partial \bar{w}_o^\mu \partial w_o^\nu} \right|_{w_o=0, \bar{w}_o=0}, \quad V_{i0}^{(\mu,0)} = V_{i0}^{(\mu)} \quad (5.90)$$

are functions of the complex lateral component p_i of the canonical momentum at the image plane. The coefficients are real for $\mu = \nu$, $\mu, \nu = 0, 1, 2, \dots$ and may be complex else.

Neglecting the nonlinear terms in the expansion (5.89), we find from (5.86) the relations

$$\begin{aligned} w_i &= -2 \frac{\partial V_{i0}^{(0)}}{\partial \bar{p}_i} - w_o \frac{\partial \bar{V}_{i0}^{(1)}}{\partial \bar{p}_i} - \bar{w}_o \frac{\partial V_{i0}^{(1)}}{\partial \bar{p}_i}, \\ p_{o0} &= p_o(w_o = 0, \bar{w}_o = 0) = -V_{i0}^{(1)}(p_i, \bar{p}_i). \end{aligned} \quad (5.91)$$

Here p_{o0} is the object lateral component of the canonical momentum of the axial trajectory, which starts from the center $w_o = \bar{w}_o = 0$ of the object plane and has lateral canonical momentum p_i at the image plane. It follows from the first relation in (5.91) that the axial trajectory intersects the center of the image plane irrespectively of its slope only if

$$\frac{\partial V_{i0}^{(0)}}{\partial \bar{p}_i} = 0. \quad (5.92)$$

In this case, the axial aberration is eliminated to any order of the power series expansion of $V_i^{(0)} = V_i^{(0)}(p_i, \bar{p}_i)$ with respect to p_i and $\bar{p}_i$. To guarantee that all points of a small area of the image center are also imaged perfectly without distortion, the magnification

$$M = \frac{w_i}{w_o} \quad (5.93)$$

must be a constant $M = M_0$. It follows from the first relation in (5.91) that we can achieve this requirement only if

$$\frac{\partial V_{i0}^{(1)}}{\partial \bar{p}_i} = 0, \quad \frac{\partial \bar{V}_{i0}^{(1)}}{\partial \bar{p}_i} = -M = -M_0. \quad (5.94)$$

However, we obtain ideal aplanatic imaging also in the case

$$\frac{\partial \bar{V}_{i0}^{(1)}}{\partial \bar{p}_i} = 0, \quad \frac{\partial V_{i0}^{(1)}}{\partial \bar{p}_i} = -\frac{w_i}{\bar{w}_o} = -\hat{M} = -\hat{M}_0 \quad (5.95)$$

yielding a mirror image. Hence, the eikonal coefficient $V_i^{(1)} = V_i^{(1)}(p_i, \bar{p}_i)$ of an aplanatic system must have one of the two forms:

$$V_{i0}^{(1)} = -M_0 p_i, \quad V_{i0}^{(1)} = -\hat{M}_0 \bar{p}_i. \quad (5.96)$$

The magnifications M_0 and $\hat{M}_0$ may be complex indicating a rotation of the image with respect to the object. By inserting the first expression into the second relation in (5.91), we find the conditions for aplanatism as

$$V_{i0}^{(0)} = V_0 = \text{const}, \quad \frac{p_{o0}}{p_i} = M_0. \quad (5.97)$$

Conditions (5.96) imply that we must eliminate the off-axial coma to any order. We can transform the second expression in (5.97) into a more familiar form by taking its absolute value and considering the relations

$$|p_{o0}| = \sqrt{2em_e\Phi_o^*} \sin \vartheta_o, \quad |p_i| = \sqrt{2em_e\Phi_i^*} \sin \vartheta_i \quad (5.98)$$

yielding the sine condition

$$\sqrt{\frac{\Phi_o^*}{\Phi_i^*}} \frac{\sin \vartheta_o}{\sin \vartheta_i} = |M_0|. \quad (5.99)$$

The angles ϑ_o and ϑ_i are the slope angles of the axial ray taken at the center of the object and image plane, respectively.

In order that all points of an extended object are imaged perfectly into the image plane, it does not suffice to fulfill the sine condition. In addition, the second- and higher-order off-axial terms in the expansion (5.89) must also be eliminated or sufficiently suppressed. The second-order terms in this expansion account for image

curvature and field astigmatism. It follows from these considerations that the mixed eikonal of a perfect optical system must have the simple form:

$$V_i = V_0 - \text{Re}(M_0 \bar{w}_0 p_i) \quad (5.100)$$

at the image plane $z = z_i$. To my knowledge, this simple result has not yet been stated in the literature.

5.6.2 Perturbation Eikonal

In most systems, one confines the particle trajectories to the vicinity of the axis, which may be either straight or curved. Then the path deviation $\Delta w = w - w^{(1)}$ of the exact ray $w = w(z)$ from its paraxial approximation $w^{(1)} = w^{(1)}(z)$ will generally be small. The path deviation at the image plane $\Delta w_i = \Delta w(z_i)$ determines the aberration. Therefore, we differentiate between path deviations and aberrations. In light optics, one classifies the geometrical aberrations according to their *Seidel* order n . We obtain the aberrations formally by expanding the path deviation in a power series with respect to the four ray parameters:

$$\Delta w(z) = w(z) - w^{(1)}(z) = \sum_{n=2}^{\infty} w^{(n)}(z). \quad (5.101)$$

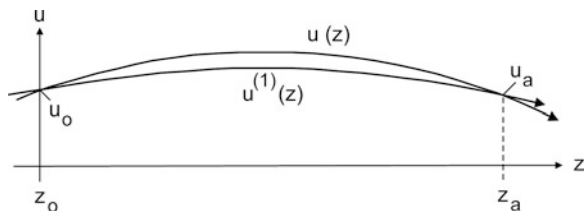
Each path deviation $w^{(n)}(z)$ of order n is a polynomial of degree n in the four ray parameters a_1, a_2, a_3 , and a_4 . The coefficients of the constituent monomials are generally complex and functions of the z -coordinate. Their values at the image plane define the aberration coefficients. Interrelations exist between various coefficients due to the existence of the eikonal. These connections become rather involved with increasing order of the aberrations. The interrelations are very simple if we define the trajectory by its position $w_0 = a_3 + i a_4$ at the object plane $z_0 = z_0$ and its lateral canonical momentum $p_i = a_1 + i a_2$ at the image plane $z = z_i$. In this case, it is advantageous to employ the mixed eikonal, which we expand in a power series expansion of the mixed eikonal at this plane:

$$V_i = V(z_i) = V_i^{(2)} + \Delta V_i = -\text{Re}(M_0 w_0 \bar{p}_i) + \sum_{m=3}^{\infty} V_i^{(m)}. \quad (5.102)$$

Here $V_i^{(m)}$ is the polynomial of degree m in the four complex ray parameters $w_0, \bar{w}_0, p_i$, and $\bar{p}_i$. By applying relation (5.86) to the image plane, substituting the expansion (5.102) for $V(z_i)$ and (5.101) for $w(z)$, we find

$$w(z_i) = w^{(1)}(z_i) + \sum_{n=2}^{\infty} w^{(n)}(z_i) = -2 \frac{\partial V_i}{\partial \bar{p}_i} = M_0 w_0 - 2 \sum_{m=3}^{\infty} \frac{\partial V_i^{(m)}}{\partial \bar{p}_i}. \quad (5.103)$$

Fig. 5.10 Fixing of the real trajectory u and its paraxial approximation $u^{(1)}$ by their common lateral distances $u(z_0) = u^{(1)}(z_0) = u_0$ and $u(z_a) = u^{(1)}(z_a) = u_a$ at the object plane z_0 and the aperture plane z_a , respectively



Since this equation must be valid for arbitrary values of the ray parameters, we find the relations

$$w^{(1)}(z_i) = w_i^{(1)} = M_0 w_0, \quad w^{(n)}(z_i) = w_i^{(n)} = -2 \frac{\partial V_i^{(n+1)}}{\partial \bar{p}_i}. \quad (5.104)$$

The first relation describes the Gaussian approximation, while the second relation reveals that the expansion polynomial of order $n + 1$ of the mixed eikonal at the image plane determines unambiguously the total n th-order aberration.

Because the eikonal polynomials are real, relations must exist between the aberration coefficients. These relations may become rather involved if we fix the ray by its lateral position and momentum components at the object plane or by its intersection points with two given planes; one is usually the object plane and the other the aperture plane $z = z_a$, as illustrated in Fig. 5.10. In the latter case, two eikonals are required to determine the ray data at the plane of observation. Owing to this difficulty, one has often argued that the eikonal method would be of little use for determining higher-order aberrations in the case of arbitrary ray-defining parameters [109]. However, the eikonal method enables one to construct a systematic iteration algorithm for the calculation of the path deviations according to their order. The algorithm gives integral expressions for the aberration coefficients yielding information on their structure. This insight provides elegant procedures for compensating the deleterious aberrations at the image plane. The accelerator community and others [110] favor the matrix method claiming that the matrix formalism gives the same insight with relative little effort. However, so far this method has not given any novel design for high-performance electron optical elements.

To minimize the calculation expenditure and to find optimum means for eliminating performance-limiting aberrations, it is, therefore, very desirable to find the representation for the eikonal terms that involves path deviations with lowest Seidel order. Owing to this possibility, the eikonal approach offers an elegant and straightforward procedure for calculating the path and momentum deviations $w^{(n)}$, $p^{(n)}$ and elucidating their internal structure [52, 54, 106]. Such an insight yields invaluable hints for the optimum design of correctors eliminating aberrations.

Ideal imaging is achieved if all path deviations with $n > 2$ vanish at the image plane. In this case, the optical path length, or eikonal, is the same for all trajectories connecting two conjugate points. Accordingly, the difference

$$\Delta S_i = S_i - S_i^{(0)} - S_i^{(2)} = q_0 \int_{z_0}^{z_i} \mu dz - q_0 \int_{z_0}^{z_i} (\mu^{(0)}(z) + \mu^{(2)}(w^{(1)}, p^{(1)}, z)) dz \quad (5.105)$$

between the eikonal S_i of the true ray and the eikonal $S_i^{(0)} + S_i^{(2)}$ of its paraxial approximation must vanish at the image plane. The first-order term $S^{(1)}$ is zero because we have imposed the condition that the optic axis is a trajectory; $S_i^{(0)}/q_0$ is the optical path length of the optic axis between the object and the image plane. This length does not depend on the ray parameters.

To provide an efficient and lucid iteration algorithm, it is advantageous to introduce another eikonal, termed perturbation eikonal E . We obtain this eikonal by adding an appropriate total differential to the variational function $\mu - \mu^{(0)} - \mu_1^{(2)}$ of the eikonal (5.105), where $\mu_1^{(2)} = \mu^{(2)}(w^{(1)}, \bar{w}^{(1)}, p^{(1)}, \bar{p}^{(1)}; z)$. This addition does not affect the path equation because the integral of the added term depends only on the fixed ray components at the terminal planes. Considering this fact, we define the *perturbation eikonal* as

$$E_v = \int_{z_v}^z (\mu - \mu^{(0)} - \mu_1^{(2)}) dz - \frac{1}{q_0} \text{Re} [p^{(1)} (\bar{w} - \bar{w}^{(1)})]_{z_v}^z = \int_{z_v}^z \mu_E dz. \quad (5.106)$$

We can readily interpret this formula in geometrical terms, as illustrated in Fig. 5.11. The perturbation eikonal represents the difference between two optical lengths, one connecting the point P_0 at the initial plane $z_v = z_0$ with the point P at the final plane z , the other connecting P_0 with the point Q . The former length is the optical path of the true ray between P_0 and the intersection P of the ray with the plane z . The other optical path length must be taken along the paraxial ray from P_0 to the foot Q of the perpendicular dropped from P upon the tangent of the paraxial momentum $\bar{p}^{(1)}$ at the intersection point P' of the paraxial ray with the plane z .

The initial plane z_v does not have to be the object plane because we need different eikonals if we define the ray at planes, which differ from the terminal planes. For convenience, we have normalized the lateral component p of the canonical momentum and its paraxial approximation $p^{(1)}$ such that

$$p = 2q_0 \frac{\partial \mu}{\partial \bar{w}},$$

$$p^{(1)} = 2q_0 \frac{\partial \mu_1^{(2)}}{\partial \bar{w}^{(1)'}} = \sqrt{2m_e e \Phi^*} w^{(1)'} - \frac{i}{2} e B w^{(1)} = \sqrt{2m_e e \Phi^*} e^{i\chi} u^{(1)'}. \quad (5.107)$$

The subtraction of the term, which only depends on the coordinates of the ray at the terminal planes, is equivalent to a subtraction of

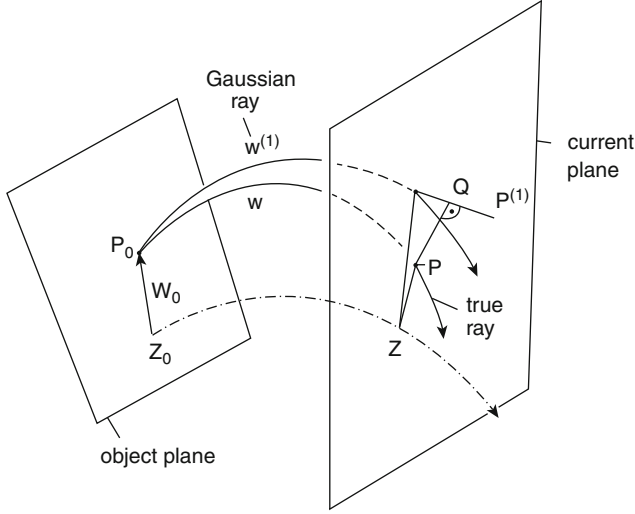


Fig. 5.11 Geometrical illustration of perturbation eikonal in the case $z_v = z_0$

$$\begin{aligned} & \frac{1}{q_0} \operatorname{Re} \int_{z_v}^z \left\{ \frac{d}{dz} \operatorname{Re} [\bar{p}^{(1)}(w - w^{(1)})] dz \right\} \\ &= 2 \int_{z_v}^z \operatorname{Re} \left[(w - w^{(1)}) \frac{\partial \mu_1^{(2)}}{\partial w^{(1)}} + (w' - w^{(1)'}) \frac{\partial \mu_1^{(2)}}{\partial w^{(1)'}} \right] dz. \end{aligned} \quad (5.108)$$

For deriving the last expression, we have made use of the relation

$$\frac{1}{q_0} \frac{d\bar{p}}{dz} = 2 \frac{d}{dz} \left(\frac{\partial \mu}{\partial w'} \right) = 2 \frac{\partial \mu}{\partial w}. \quad (5.109)$$

Employing the result (5.198), we obtain for the integrand μ_E of the perturbation eikonal

$$E_v = \int_{z_v}^z \mu_E dz = \frac{1}{q_0} \operatorname{Re} \int_{z_v}^z (\bar{p} dw - \bar{p}^{(1)} dw^{(1)}) - \frac{1}{q_0} \operatorname{Re} [\bar{p}^{(1)}(w - w^{(1)})]_{z_v}^z, \quad (5.110)$$

the relation

$$\mu_E = \mu - \mu_1^{(2)} - 2 \operatorname{Re} \left[(w - w^{(1)}) \frac{\partial \mu_1^{(2)}}{\partial w^{(1)}} + (w' - w^{(1)'}) \frac{\partial \mu_1^{(2)}}{\partial w^{(1)'}} \right]. \quad (5.111)$$

Equations (5.110) and (5.111) are most suitable for deriving an efficient algorithm, which yields iteratively the path deviations according to their order. The method

of successive approximation works most effectively for inhomogeneous integral equations. To establish such an approach, we transform the complex differential path equation (5.109) into a set of appropriate integral equations. This transformation has the additional advantage that it includes the boundary constraints imposed on a distinct ray.

We derive the most suitable form of the integral equations by varying the perturbation eikonal (5.110) with respect to the lateral position and momentum coordinates at the terminal planes z and z_v yielding

$$\begin{aligned} q_0 \delta E_v &= \text{Re} \left[\bar{p} \delta w - w \delta \bar{p}^{(1)} - \bar{p}^{(1)} \delta w + w^{(1)} \delta \bar{p}^{(1)} \right]_{z_v}^z \\ &= \text{Re} \left[(p - p^{(1)}) \delta \bar{w}^{(1)} - (w - w^{(1)}) \delta \bar{p}^{(1)} + (\bar{p} - \bar{p}^{(1)}) \delta (w - w^{(1)}) \right]_{z_v}^z. \end{aligned} \quad (5.112)$$

The perturbation eikonal E_v , the lateral canonical momentum p , $p^{(1)}$, and the position w , $w^{(1)}$ of the particle are functions of the ray-defining parameters a_v , $v = 1, 2, 3, 4$. Therefore, we can also vary the perturbation eikonal (5.112) with respect to each of these parameters separately. By considering in addition the relations

$$w_v = \frac{\partial w^{(1)}}{\partial a_v}, \quad p_v = \frac{\partial p^{(1)}}{\partial a_v}, \quad (5.113)$$

we obtain the following set of four integral equations

$$\begin{aligned} \text{Re} \left[(w - w^{(1)}) \bar{p}_v - (p - p^{(1)}) \bar{w}_v \right]_{z_v}^z &= -\frac{\partial E_v}{\partial a_v} + \text{Re} \left[(\bar{p} - \bar{p}^{(1)}) \frac{\partial (w - w^{(1)})}{\partial a_v} \right]_{z_v}^z, \\ v &= 1, 2, 3, 4. \end{aligned} \quad (5.114)$$

These equations show that we have attributed a distinct eikonal or optical path length E_v to each ray parameter a_v . This parameter relates to a distinct ray coordinate at the plane z_v , which defines the lower integration limit in the integral expression (5.110) of the perturbation eikonal E_v . In the most general case, the number of required eikonals E_v is identical with the number of planes at which one defines the ray. The two terms in the bracket on the left-hand side linearly related to the difference between the off-axis position and the lateral canonical momentum, respectively, of the true ray and its paraxial approximation. The terms in the bracket on the right-hand side are bilinear in these deviations and hence do not contribute to the primary aberrations obtained in the first iteration step. The left-hand side of (5.114) is a linear combination of the deviations $w - w^{(1)}$, $\bar{w} - \bar{w}^{(1)}$, $p - p^{(1)}$, and $\bar{p} - \bar{p}^{(1)}$. Since we have four equations, we can solve these equations with respect to the path and momentum deviations $w - w^{(1)}$ and $p - p^{(1)}$, respectively.

In most cases, one defines the ray by its lateral position and/or momentum coordinates at distinct planes. Then the contribution of the lower limit on the right-hand side in (5.114) vanishes. In order that the corresponding contribution also

vanishes on the left-hand side, we impose the condition

$$\operatorname{Re} [(w - w^{(1)})\bar{p}_v - (\bar{p} - \bar{p}^{(1)})w_v]_{z=z_v} = 0, \quad v = 1, 2, 3, 4. \quad (5.115)$$

Using the abbreviation

$$\begin{aligned} b_v &= \frac{1}{q_0} \operatorname{Re} [\bar{p}_v(z_v)w(z_v) - w_v(z_v)\bar{p}(z_v)] \\ &= \operatorname{Re} \left[\sqrt{\frac{\Phi_v^*}{\Phi_0^*}} u(z_v) \bar{u}'_v(z_v) - \bar{u}_v(z_v) \frac{p(z_v)e^{-i\chi_v}}{p_{z0}} \right] \end{aligned} \quad (5.116)$$

and the representations

$$w^{(1)} = \sum_{\mu=1}^4 a_\mu w_\mu, \quad p^{(1)} = \sum_{\mu=1}^4 a_\mu p_\mu \quad (5.117)$$

for the paraxial position $w^{(1)}$ and the lateral canonical momentum $p^{(1)}$ of the particle, we can rewrite (5.115) as the following set of four linear equations in the ray parameters a_μ :

$$\frac{1}{q_0} \sum_{\mu=1}^4 a_\mu \operatorname{Re} (w_\mu \bar{p}_v - w_v \bar{p}_\mu) = \sum_{\mu=1}^4 a_\mu C_{\mu v} = b_v, \quad v = 1, 2, 3, 4. \quad (5.118)$$

We derived the second expression in this equation by employing the Lagrange–Helmholtz relation (5.30). In order to attribute only one of the four ray components to each fundamental ray w_μ or momentum p_μ , we must choose them in such a way that most of the constants $C_{\mu v}$ are zero. By choosing

$$C_{31} = C_{42} = 1, \quad C_{12} = C_{14} = C_{23} = C_{34} = 0, \quad (5.119)$$

the sum in (5.118) degenerates to a single term $a_\mu C_{\mu v} = b_v$, $\mu, v = 1, 2, 3, 4$, giving

$$a_1 = -b_3, \quad a_2 = -b_4, \quad a_3 = b_1, \quad a_4 = b_2. \quad (5.120)$$

The requirements (5.119) do not suffice to fix the fundamental rays entirely. The proper fixing of the ray w_v depends on the boundary condition, which one imposes on the true ray at the plane z_v . If we define the ray by its lateral position at this plane, we must impose the condition

$$w_v(z_v) = 0. \quad (5.121)$$

On the other hand, we must require

$$p_v(z_v) = 0 \quad (5.122)$$

if we fix the ray by the lateral component of the canonical momentum. If we fix the ray by its position and momentum coordinates at the object plane $z_1 = z_2 = z_3 = z_4 = z_0$, we must impose condition (5.121) on the rays w_1 and w_2 and condition (5.122) on the rays w_3 and w_4 . Considering these constraints, we obtain from the relations (5.116) and (5.120) the ray parameters

$$a_1 = \frac{1}{q_0} \text{Re}(\bar{p}_0 w_3(z_0)), \quad a_2 = \frac{1}{q_0} \text{Re}(\bar{p}_0 w_4(z_0)), \quad (5.123)$$

$$a_3 = \frac{1}{q_0} \text{Re}(w_0 \bar{p}_1(z_0)), \quad a_4 = \frac{1}{q_0} \text{Re}(w_0 \bar{p}_2(z_0)). \quad (5.124)$$

These relations simplify further if we consider that $\Phi_0^* = \Phi_0^*$, employ relation (5.19) with $\chi(z_0) = 0$ and specify the initial values of the fundamental rays at the object plane as

$$w_3(z_0) = 1, \quad w_4(z_0) = i, \quad (5.125)$$

$$p_1(z_0) = q_0 u'_1(z_0) = q_0, \quad p_2(z_0) = q_0 u'_2(z_0) = i q_0 \quad (5.126)$$

resulting in

$$a_1 = \frac{p_{0x}}{q_0}, \quad a_2 = \frac{p_{0y}}{q_0}, \quad a_3 = x_0, \quad a_4 = y_0. \quad (5.127)$$

If we define the ray by its lateral positions at the object plane $z_1 = z_2 = z_0$ and the aperture plane $z_3 = z_4 = z_a$, the canonical momentum of the true ray differs from that of its paraxial approximation at the boundary planes, as illustrated in Fig. 5.12. As a result, we must choose the fundamental rays in such a way that they satisfy the constraint (5.121) so that

$$w_1(z_0) = w_2(z_0) = 0, \quad w_3(z_a) = w_4(z_a) = 0. \quad (5.128)$$

We fix the fundamental rays further by imposing the initial constraints (5.126) on the lateral canonical momentum of the rays w_1 and w_2 , and the corresponding conditions

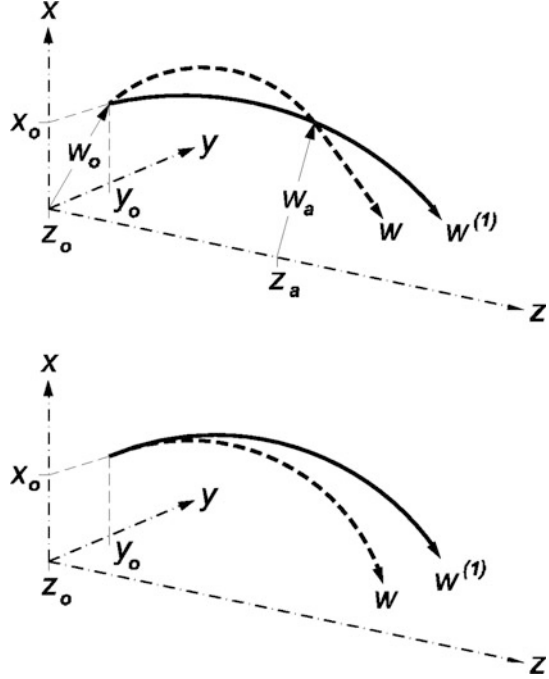
$$p_3(z_a) = -q_a e^{i\chi_a} u'_3(z_a), \quad p_4(z_a) = -q_a e^{i\chi_a} u'_4(z_a) \quad (5.129)$$

on the rays w_3 and w_4 , respectively. This specification of the fundamental rays gives parameters

$$\begin{aligned} a_1 &= -\sqrt{\frac{\Phi_a^*}{\Phi_0^*}} \text{Re}\{u_a \bar{u}'_{3a}\}, \quad a_2 = -\sqrt{\frac{\Phi_a^*}{\Phi_0^*}} \text{Re}\{u_a \bar{u}'_{4a}\}, \quad u_a = u(z_a), \\ a_3 &= \text{Re}\{u_0 \bar{u}'_{10}\}, \quad a_4 = \text{Re}\{u_0 \bar{u}'_{20}\}, \quad u_0 = u(z_0). \end{aligned} \quad (5.130)$$

With these parameters, the paraxial trajectory defined by its intersection coordinates at the object and aperture planes has the form

Fig. 5.12 Influence of the initial constraints on the deviation $w(z) - w^{(1)}(z)$ of the paraxial ray $w^{(1)} = w^{(1)}(z)$ from the exact trajectory $w = w(z)$



$$w^{(1)} = -\sqrt{\frac{\Phi_a^*}{\Phi_0^*}} [w_1 \text{Re}\{u_a \bar{u}'_{3a}\} + w_2 \text{Re}\{u_a \bar{u}'_{4a}\}] + w_3 \text{Re}\{u_0 \bar{u}'_{10}\} + w_4 \text{Re}\{u_0 \bar{u}'_{20}\}. \quad (5.131)$$

To check the correctness of this expression, we take its value at the object plane. Considering $\chi(z_0) = 0$, $w_{30} = u_{30}$, $w_{40} = u_{40}$, and $u_{10} = u_{20} = 0$, we obtain

$$\begin{aligned} w^{(1)}(z_0) &= \frac{u_0}{2} [u_{30} \bar{u}'_{10} + u_{40} \bar{u}'_{20}] + \frac{\bar{u}_0}{2} [u_{30} u'_{10} + u_{40} u'_{20}] \\ &= \frac{u_0}{2} \text{Re}[u_{30} \bar{u}'_{10} - u'_{30} \bar{u}_{10} + u_{40} \bar{u}'_{20} - u'_{40} \bar{u}_{20}] = u_0 = w_0. \end{aligned} \quad (5.132)$$

The expression in the second bracket vanishes as follows from the second relation in (4.218). The expression in the first bracket is real because we have chosen the fundamental rays appropriately. Therefore, we can apply the Lagrange–Helmholtz relations for evaluating the first bracket. The equivalent relation

$$\bar{u}'_{3a} u_{1a} = u'_{3a} \bar{u}_{1a}, \quad \bar{u}'_{4a} u_{2a} = u'_{4a} \bar{u}_{2a}, \quad \bar{u}'_{4a} u_{2a} = u'_{4a} \bar{u}_{2a} \quad (5.133)$$

holds true for the fundamental rays at the aperture plane.

By taking into account the Lagrange–Helmholtz relation

$$\text{Re}[\sqrt{\Phi_a^*} (u_{1a} \bar{u}'_{3a} - u'_{1a} u_{3a})] = \sqrt{\Phi_a^*} u_{1a} \bar{u}'_{3a} = C_{13} \sqrt{\Phi_0^*} = -\sqrt{\Phi_0^*} \quad (5.134)$$

and the equivalent relation for the rays u_2 and u_4 , we find that the ray parameter a_3 and a_4 adopt the simple form

$$a_3 = \operatorname{Re} \left(\frac{u_a}{u_{1a}} \right) = \operatorname{Re} \left(\frac{w_a}{w_{1a}} \right), \quad a_4 = \operatorname{Re} \left(\frac{u_a}{u_{2a}} \right) = \operatorname{Re} \left(\frac{w_a}{w_{2a}} \right). \quad (5.135)$$

5.6.3 Integral Equations of the Path and Momentum Deviations

By imposing the condition (5.115), the equations (5.114) for the path and momentum deviations $w - w^{(1)}$ and $p - p^{(1)}$ adopt the form

$$(w - w^{(1)})\bar{p}_v + (\bar{w} - \bar{w}^{(1)})p_v - (p - p^{(1)})\bar{w}_v - (\bar{p} - \bar{p}^{(1)})w_v = 2q_0 Q_v, \\ Q_v = -\frac{\partial E_v}{\partial a_v} + \frac{1}{q_0} \operatorname{Re} \left[(\bar{p} - \bar{p}^{(1)}) \frac{\partial (w - w^{(1)})}{\partial a_v} \right], \quad v = 1, 2, 3, 4. \quad (5.136)$$

We can solve this set of equations with respect to the path and momentum deviations most conveniently by multiplying this equation with factors $(-1)^p C_{\kappa\lambda} w_\mu$ and $(-1)^p C_{\kappa\lambda} p_\mu$, respectively, and subsequently sum over all 24 permutations p of the four indices κ, λ, μ , and v . By employing the relations

$$\sum_{(p)} (-1)^p C_{\kappa\lambda} w_\mu p_v = 0, \quad \sum_{(p)} (-1)^p C_{\kappa\lambda} p_\mu p_v = 0, \\ \sum_{(p)} (-1)^p C_{\kappa\lambda} w_\mu \bar{w}_v = 0, \quad \sum_{(p)} (-1)^p C_{\kappa\lambda} p_\mu \bar{p}_v = 0, \\ \sum_{(p)} (-1)^p C_{\kappa\lambda} w_\mu w_v = 0, \quad \sum_{(p)} (-1)^p C_{\kappa\lambda} w_\mu \bar{p}_v = q_0 D_W, \quad (5.137)$$

we eventually find

$$w - w^{(1)} = \frac{2}{D_W} \sum_{(p)} (-1)^p C_{\kappa\lambda} w_\mu Q_v, \\ p - p^{(1)} = \frac{2}{D_W} \sum_{(p)} (-1)^p C_{\kappa\lambda} p_\mu Q_v. \quad (5.138)$$

The sums reduce considerably, if we choose the fundamental rays such that the constants $C_{\kappa\lambda}$ of the corresponding Lagrange–Helmholtz relations adopt the values listed in (5.119) resulting in $D_W = -4$ for the Wronski determinant (4.217). With this value and the values (5.119) for $C_{\kappa\lambda}$, the equations (5.138) for the deviations take the simple form

$$w - w^{(1)} = w_1 Q_3 - w_3 Q_1 + w_2 Q_4 - w_4 Q_2, \quad (5.139)$$

$$p - p^{(1)} = p_1 Q_3 - p_3 Q_1 + p_2 Q_4 - p_4 Q_2. \quad (5.140)$$

The integrand (5.111) of the perturbation eikonal (5.110) is a function of the z -coordinate and the position $w = w(z)$ and slope w' of the true trajectory. Therefore, it is obvious that equation (5.139) represents an inhomogeneous complex integral equation for the lateral position w of the true ray. Equation (5.140) is the equivalent integral equation for the lateral momentum of this ray. These integral equations are most suitable for determining the path deviations with respect to their order by employing the method of successive approximation. We will develop the iteration algorithm in Chap. 7.

5.7 Poisson Brackets

We define the Poisson bracket $\{F, G\}$ of any two complex functions $F = F(w, \bar{w}, p, \bar{p})$ and $G = G(w, \bar{w}, p, \bar{p})$ of the complex canonical ray coordinates $w = w(a_1, a_2, a_3, a_4; z)$ and $p = p(a_1, a_2, a_3, a_4; z)$ by the equation

$$\begin{aligned} \{F, G\} &= \frac{\partial F}{\partial a_1} \frac{\partial G}{\partial a_3} - \frac{\partial F}{\partial a_3} \frac{\partial G}{\partial a_1} + \frac{\partial F}{\partial a_2} \frac{\partial G}{\partial a_4} - \frac{\partial F}{\partial a_4} \frac{\partial G}{\partial a_2} \\ &= 2 \left\{ \frac{\partial F}{\partial \omega} \frac{\partial G}{\partial \bar{\rho}} + \frac{\partial F}{\partial \bar{\omega}} \frac{\partial G}{\partial \rho} - \frac{\partial F}{\partial \rho} \frac{\partial G}{\partial \bar{\omega}} - \frac{\partial F}{\partial \bar{\rho}} \frac{\partial G}{\partial \omega} \right\}, \end{aligned} \quad (5.141)$$

with complex ray parameters $\omega = a_1 + i a_2$ and $\rho = a_3 + i a_4$. In the special case of canonical boundary conditions we have $\omega = p_0/q_0$ and $\rho = w_0$. Note that our definition of the Poisson bracket differs by the factor $1/2$ from the usual definition. We have normalized the Poisson bracket such that the *fundamental* Poisson brackets will be unity or zero, respectively. The Poisson bracket has the remarkable property that it is preserved under canonical transformations, which replace the initial canonical ray components w_0 and p_0 by the corresponding components w and p at any other plane z . For our purpose, it suffices to prove this behavior for the so-called fundamental Poisson brackets, which are obtained by putting F and G equal to one of the ray components $w, \bar{w}, p$, and $\bar{p}$. We can readily evaluate these brackets by the relations

$$\sum_{(p)} (-1)^p I_{\mu\nu} \frac{\partial w}{\partial a_\sigma} \frac{\partial p}{\partial a_\tau} = 0, \quad \sum_{(p)} (-1)^p I_{\mu\nu} \frac{\partial w}{\partial a_\sigma} \frac{\partial \bar{w}}{\partial a_\tau} = 0, \quad \sum_{(p)} (-1)^p I_{\mu\nu} \frac{\partial p}{\partial a_\sigma} \frac{\partial \bar{p}}{\partial a_\tau} = 0, \quad (5.142)$$

$$\sum_{(p)} (-1)^p I_{\mu\nu} \frac{\partial w}{\partial a_\sigma} \frac{\partial \bar{p}}{\partial a_\tau} = 4 (I_{12} I_{34} - I_{13} I_{24} + I_{14} I_{23}). \quad (5.143)$$

The formula (5.125) restates the expression (5.35) for the Jacobian (5.34). We obtain the relations (5.142) by equalizing two rows in this determinant in three different ways. If we fix the rays by their initial canonical values $a_1 + i a_2 = p_0$, $a_3 + i a_4 = \rho = w_0 = x_0 + i y_0$, we have $I_{31} = I_{42} = 1$, $I_{12} = I_{14} = I_{23} = I_{34} = 0$. Then, (5.143) represents the fundamental Poisson bracket, which may be written as

$$\{w, \bar{p}\} = 2 \left\{ \frac{\partial w}{\partial w_0} \frac{\partial \bar{p}}{\partial \bar{p}_0} + \frac{\partial w}{\partial \bar{w}_0} \frac{\partial \bar{p}}{\partial p_0} - \frac{\partial w}{\partial \bar{p}_0} \frac{\partial \bar{p}}{\partial w_0} - \frac{\partial w}{\partial p_0} \frac{\partial \bar{p}}{\partial \bar{w}_0} \right\} = 1 = \{\bar{w}, p\}. \quad (5.144)$$

Using the same procedure for the remaining combinations (5.142) of the four ray components, we eventually obtain

$$\{w, w\} = \{w, \bar{w}\} = \{w, p\} = \{p, p\} = \{p, \bar{p}\} = 0. \quad (5.145)$$

The Poisson brackets are conjugate to the Lagrange brackets (5.25), because their validity is a consequence of the properties of the eikonal. The invariance of these brackets is the reason that we can view the propagation of charged particles as symplectic mapping in phase space.

Chapter 6

Beam Properties

Abstract The properties of charged-particle beams, representing ensembles of trajectories, are investigated. It is shown that the beam properties depend strongly on the emission characteristics of the source such as the brightness or emittance, respectively. The emittance and the acceptance of beam guiding elements are discussed in detail, and it is shown that the emittance domain of an optical element must be embedded entirely in the so-called acceptance domain of the subsequent element in order to achieve maximum throughput.

So far, we have considered exclusively the propagation of single particles in external electromagnetic fields. However, in many cases, one is also interested in the behavior of the charged-particle beam, which represents an ensemble of trajectories. In most cases, one characterizes the beam by its current and mean energy E_0 . However, these quantities do not suffice to describe its focusing properties, which strongly depend on the emission characteristics of the source.

Interaction of particles within the beam occurs predominantly at high current densities realized even at low currents when the particles pass through a focus or crossover. The statistical interactions between the particles within this regime produce an energy broadening called (energetic) Boersch [111, 112] effect and an angular spread (angular Boersch effect) [113, 114]. The stochastic ray deflections reduce the resolution of micro-lithographic systems because these deflections broaden the image spot. As a result, the usable current is limited.

In the following, we restrict our considerations to low-density beams assuming that the current density or trajectory density, respectively, stays sufficiently small along the entire course of the beam. In this case, we can neglect the effect of the space charge forces and conceive the beam as a bundle of independent rays. Then we can represent the properties of each particle at any point along its trajectory by a point in the six-dimensional phase space with coordinates x, p_x, y, p_y, z, E . Instead of the energy, one uses generally the energy deviation $\Delta E = E - E_0$ or the relative energy deviation $\kappa = \Delta E/E_0$, respectively. If we set $\Delta E = 0$, we can represent the properties of each particle by a point in the five-dimensional state

space [115]. At a given plane z in this space, the beam intersects a certain area, which is known as *hyperemittance*. We can project this four-dimensional area onto the two-dimensional phase planes x, p_x and y, p_y . The sum of these projections forms the *total transverse emittance*. In orthogonal systems, the motion of the particle in the vertical principal section decouples from that in the horizontal section. In the absence of coupling between these degrees of freedom, we may split the total transverse emittance into two independent two-dimensional *emittances*, one for the x -section, the other for the y -section. For a real beam, these emittances are defined by the extension of the source and/or apertures, which limit the maximum width and the maximum lateral momentum of the beam along the optic axis.

The *brightness* is another important beam quantity. It corresponds to that used in light optics where it describes the photon density in phase space. In accordance with this concept, we define the so-called *reduced brightness* of a charged-particle beam as the current density in four-dimensional phase space. The concepts of brightness and emittance are closely related with each other in the sense that low emittance corresponds to high brightness and vice versa.

6.1 Brightness

A beam consisting of N particles represents a system with $3N$ degrees of freedom. Due to the Coulomb force, charged particles interact with each other. The strength of this interaction depends on the current density. The interaction forces are the larger the higher the particle density, as it is the case in the region of caustics whose tips form the Gaussian focal points. The exact description of the motion of N interacting particles necessitates the introduction of a $6N$ -dimensional phase space. Since one can tackle this task only numerically for a limited number of particles, we restrict our considerations to the propagation of noninteracting particles whose initial positions and lateral canonical momenta are given at the plane $z = z_0$ by the distribution function

$$f_0 = f(x(z_0), y(z_0), p_x(z_0), p_y(z_0), \kappa). \quad (6.1)$$

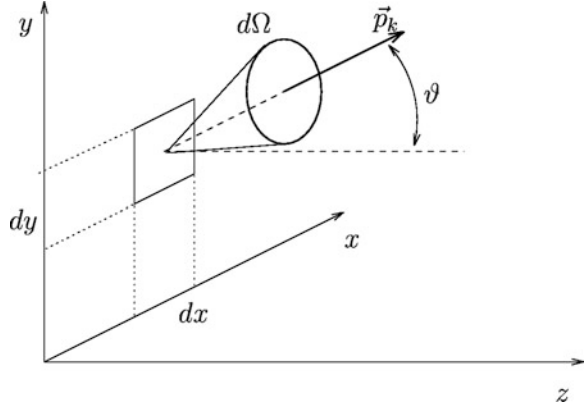
The distribution function accounts for the probability that a particle of the beam occupies a distinct trajectory. In the five-dimensional state space, a particle covers the distance

$$dz = v_z dt = v \cos \vartheta dt \quad (6.2)$$

in the direction of the optic axis during the infinitesimal time interval dt . The number of particles passing through the surface element $dx dy$ into the differential solid angle $d\Omega$ at given plane z of the state space during this time is

$$\frac{d^6 N}{dt} = f(x, y, p_x, p_y, \kappa) v \cos \vartheta dp_x dp_y dp_z dx dy = f \vec{p}^2 \vec{v} d\vec{p} \cos \vartheta d\Omega dx dy. \quad (6.3)$$

Fig. 6.1 Definition of the differential solid $d\Omega$ angle of the momentum volume element $dV_p = dp_x dp_y dp_z = \vec{p}^2 d|\vec{p}| d\Omega$



Considering the conservation of energy

$$\vec{v} d\vec{p} = dH = dE = E_0 d\kappa, \quad (6.4)$$

we obtain the corresponding differential current as

$$d^5 J = e \frac{d^6 N}{dt} = f \vec{p}^2 dE dx dy \cos \vartheta d\Omega. \quad (6.5)$$

The brightness function $B = B(x, y, \vartheta, \phi; z)$ is defined as the differential current density per differential solid angle $d\Omega = \sin \vartheta d\vartheta d\phi$, given by

$$B(x, y, \vartheta, \phi; z) = \frac{d^3 J}{dx dy \cos \vartheta d\Omega} = \frac{dj}{d\Omega} = e E_0 \int_0^\infty f(x, y, \vartheta, \phi, \kappa) \vec{p}^2 d\kappa. \quad (6.6)$$

Here, ϕ is the azimuth angle about the direction of the surface element $dx dy$, which points in the direction of the z -axis, as illustrated in Fig. 6.1. The variables x, y, ϑ, ϕ are functions of the z -coordinate, since they define the position and the direction of a distinct trajectory whose initial values are fixed at the starting plane $z = z_0$. In the absence of a magnetic field, the relation

$$\vec{p}^2 = 2m_e(e\varphi^* + \Delta E) \quad (6.7)$$

holds, which shows that the brightness depends on the accelerating potential $\varphi = \varphi(x, y, z)$. Since we can vary the potential arbitrarily, we aim for a measure of the emission characteristic of the source that does not depend on φ . The kinetic energy of the nominal electron ($\Delta E = 0$) is generally much larger than the maximum

energy width of the beam. In this case, it is advantageous to introduce the reduced brightness

$$\beta(x, y, \vartheta, \phi) = \frac{2em_e B}{\vec{p}_0^2} = \frac{2em_e d^4 J}{dx dy dp_x dp_y} \quad (6.8)$$

where $\vec{p}_0 = \vec{p}(\kappa = 0)$ is the canonical momentum vector of a particle with nominal energy.

In most cases, one characterizes the source by the axial brightness or the reduced axial brightness

$$\beta_0 = \beta(x = 0, y = 0, \vartheta = 0, \phi = 0) = 2e^2 m_e \int_0^\infty f(\Delta E) d\Delta E. \quad (6.9)$$

The reduced brightness is an invariant of the beam as long as we can neglect the effect of particle collisions. This behavior follows directly from the last relation in (6.8) by considering Liouville's theorem. The distribution function $f = f(x, y, z, p_x, p_y; \Delta E)$ relates closely with the emission characteristic of the source at the plane z_s . If we know this distribution and the trajectories (w, p) , we can determine the distribution function for noninteracting particles at any other plane $z > z_s$. We can approximate the emission characteristic of most sources with a sufficient degree of accuracy by means of a Maxwell distribution for the emission energy ΔE and a Gaussian distribution for both the angular and the local emission:

$$f_s = f(z_s, x_s, y_s, \vartheta_s, \phi_s; \Delta E) \approx A_s e^{-\frac{\Delta E}{E_s}} e^{-\frac{x_s^2 + y_s^2}{\rho_s^2}} e^{-\frac{\vartheta_s^2}{\vartheta_s^2}}. \quad (6.10)$$

The radius ρ_s of the source defines the mean-emitting area of the source, which we assume to be rotationally symmetric as well as the angular emission characterized by the mean emission angle ϑ_s ; E_s is the mean energy width of the source. The angular distribution is sufficiently accurate as long as $\sin \vartheta_s \approx \vartheta_s$. If this approximation does not hold, we must substitute $\sin \vartheta$ for ϑ in (6.10). We determine the constant A_s by assuming that we know the reduced axial brightness of the source. Inserting (6.10) into the integral (6.9), we obtain

$$\beta_0 = 2e^2 m_e A_s \int_0^\infty e^{-\frac{\Delta E}{E_s}} d\Delta E = 2e^2 m_e E_s A_s, \quad (6.11)$$

which gives

$$A_s = \frac{\beta_0}{2e^2 m_e E_s}. \quad (6.12)$$

The reduced axial brightness and the mean energy width of the source are characteristic parameters, which we must determine from the experiment.

6.2 Emittance

The particles in a beam occupy a certain domain in phase space. We can calculate in principle the trajectory of each of these particles when we know its lateral position $w_0 = x_0 + iy_0$ and canonical momentum $p = p_{x0} + ip_{y0}$ at the initial plane $z = z_0$. To survey the propagation of a beam confined to the region near the optic axis, it is more appropriate to describe the beam as a whole. In many cases of practical importance, the systems exhibit two orthogonal principle sections. Particles, which initially propagate in one of these sections, will stay in this section throughout their entire paths. In the case of a curved axis, it is common to name the $x - z$ section, which embeds the optic axis, *horizontal section* and the $y - z$ section *vertical section*. In the absence of coupling between these sections, they form principal sections. One generally assumes that this situation is valid and describes the transverse properties of the beam by the two-dimensional emittances

$$\varepsilon_x = \frac{1}{\pi q_0} \iint_{a_x} dp_x dx, \quad \varepsilon_y = \frac{1}{\pi q_0} \iint_{a_y} dp_y dy. \quad (6.13)$$

The areas a_x and a_y define the projections of the occupied domain of the four-dimensional phase space onto the two-dimensional x, p_x and y, p_y subspaces at the plane z . When we confine the beam to the paraxial regime, it is customary to choose ellipses for the areas a_x and a_y . Hence, one surrounds all particles in each of the two subspaces by an elliptical contour.

We do not presuppose the absence of coupling between the horizontal and vertical sections. Instead, we generalize the two-dimensional emittances in such a way that they are valid for arbitrary systems and degenerate into (6.13) in the absence of coupling. In this case, the trajectories in the five-dimensional state space are twisted about the z -axis. Owing to this twist, a two-dimensional element of the phase space rotates along this axis. We start from the element

$$da_{x0} = dx_0 dp_{x0} \quad (6.14)$$

located initially in the x, p_x sheet of the phase space at the plane z_0 . At some other plane $z > z_0$, this element has the form

$$da_1 = dx dp_x + dy dp_y = \text{Re}\{d\bar{p} dw\}. \quad (6.15)$$

Since $p = p(x_0, y_0 = 0, p_{x0}, p_{y0} = 0; z)$ and $w = w(x_0, y_0 = 0, p_{x0}, p_{y0} = 0; z)$ depend only on the initial beam parameters x_0 and p_{x0} , we can express da_x in terms of the initial differentials dx_0 and dp_{x0} by means of the corresponding Jacobi determinant resulting in

$$da_1 = \text{Re} \left(\frac{\partial w}{\partial x_0} \frac{\partial \bar{p}}{\partial p_{x0}} - \frac{\partial w}{\partial p_{x0}} \frac{\partial \bar{p}}{\partial x_0} \right) dx_0 dp_{x0} = dx_0 dp_{x0}. \quad (6.16)$$

Here, we have made use of the fact that the Jacobi determinant coincides with one of the fundamental Lagrange brackets. The result reveals that the size of the two-dimensional phase-space element is preserved regardless of any coupling. The same behavior holds also true for the element da_y , which lies entirely in the y, p_y sheet at the initial plane z_0 . Hence, we may define the generalized two-dimensional emittances as

$$\begin{aligned}\varepsilon_1 &= \frac{1}{\pi q_0} \text{Re} \iint_{a_1} dw d\bar{p} = \frac{1}{\pi q_0} \iint_{a_{x0}} dx_0 dp_{x0} = \frac{a_{x0}}{\pi q_0}, \\ \varepsilon_2 &= \frac{1}{\pi q_0} \text{Re} \iint_{a_2} dw d\bar{p} = \frac{1}{\pi q_0} \iint_{a_{y0}} dy_0 dp_{y0} = \frac{a_{y0}}{\pi q_0}.\end{aligned}\quad (6.17)$$

In the absence of coupling, the generalized emittances (6.17) degenerate into the emittances (6.13) such that $\varepsilon_1 = \varepsilon_x$ and $\varepsilon_2 = \varepsilon_y$. We have normalized the two-dimensional phase space by q_0 in order that the emittance has the conventional dimension of a length.

Unfortunately, the definition of emittance is not standardized. In many cases, one uses the slope of the ray instead of the canonical momentum in the integrals (6.17). As a result, the emittance is not an invariant but decreases with increasing acceleration of the electron beam. In order to avoid this behavior, we choose the definition (6.17).

6.2.1 Paraxial Approximation

We can describe conveniently the characteristics of a beam at any plane z in the state space if we confine the beam to the Gaussian regime. In paraxial approximation, we can describe the contours of the domains a_{x0} and a_{y0} at the starting plane z_0 with a sufficient degree of accuracy by ellipses. It is customary in accelerator physics to choose the slope components of the trajectories instead of the components of the lateral canonical momentum. This choice is of no concern in the paraxial approximation since these quantities are then linearly related with each other (5.18). The relations simplify further if we describe the trajectories in the rotating coordinate system. A ray, which starts in the x, p_x subspace of the state space at the plane z_0 , has the form

$$u = u^{(1)}(z) = a_1 u_1(z) + x_0 u_3(z), \quad \bar{u}' = a_1 \bar{u}'_1 + x_0 \bar{u}'_3, \quad a_1 = p_{x0}/q_0. \quad (6.18)$$

In the absence of an axial magnetic field at the starting plane z_0 , we have $a_1 \approx x'_0$. Multiplying the first equation by $\bar{u}'_3 = u'_{3r} - i u'_{3i}$, the second by $u_3 = u_3(z) = u_{3r}(z) + i u_{3i}(z)$, and subsequently subtracting the resulting equations from each other gives

$$a_1(u_3\bar{u}'_1 - u_1\bar{u}'_3) = \bar{u}'u_3 - u\bar{u}'_3. \quad (6.19)$$

By taking the real part of this equation and considering the Lagrange–Helmholtz relation, we obtain

$$a_1 = \sqrt{\frac{\Phi^*}{\Phi_0^*}} \operatorname{Re}(u'\bar{u}_3 - u\bar{u}'_3) = \sqrt{\frac{\Phi^*}{\Phi_0^*}} (u'_ru_{3r} + u'_iu_{3i} - u_ru'_{3r} - u_iu'_{3i}). \quad (6.20)$$

Employing the same procedure with the fundamental ray $u_1 = u_1(z) = u_{1r}(z) + iu_{1i}(z)$ yields

$$x_0 = \sqrt{\frac{\Phi^*}{\Phi_0^*}} \operatorname{Re}(u\bar{u}'_1 - u'\bar{u}_1) = \sqrt{\frac{\Phi^*}{\Phi_0^*}} (u_ru'_{1r} + u_iu'_{1i} - u'_ru_{1r} - u'_iu_{1i}). \quad (6.21)$$

Let the equation of the horizontal phase ellipse at the starting point be

$$\gamma_{10}x_0^2 + 2\alpha_{10}x_0a_1 + \beta_{10}a_1^2 = \varepsilon_1. \quad (6.22)$$

The ellipse parameters $\alpha_{10} = \alpha_1(z_0)$, $\beta_{10} = \beta_1(z_0)$, $\gamma_{10} = \gamma_1(z_0)$ are called *Twiss parameters*. We should not mix up these parameters with trajectory angles or the relativistic factor, respectively. The ellipse (6.22) is tilted by the angle θ given by the formula

$$\tan 2\theta = \frac{2\alpha_{10}}{\gamma_{10} - \beta_{10}}. \quad (6.23)$$

We find the semiaxes b_1 , b_2 of the ellipse as

$$b_{1,2}^2 = \frac{2\varepsilon_1}{\gamma_{10} + \beta_{10} \pm \sqrt{\gamma_{10}\beta_{10} + 4\alpha_{10}^2}}. \quad (6.24)$$

From this relation and (6.17), we obtain for the horizontal emittance the relation

$$\varepsilon_1 = b_1b_2 = \frac{\varepsilon_1}{\sqrt{\beta_{10}\gamma_{10} - \alpha_{10}^2}}. \quad (6.25)$$

Since the area $\pi\varepsilon_1$ of the ellipse is an invariant of motion, it follows from (6.25) that the Twiss parameters must satisfy the condition

$$\beta_{10}\gamma_{10} - \alpha_{10}^2 = 1. \quad (6.26)$$

By substituting expression (6.20) for a_1 and (6.21) for x_0 in Eq. (6.22), we obtain the equation

$$\begin{aligned}
& \gamma_{10} (u_r u'_{1r} + u_i u'_{1i} - u'_r u_{1r} - u'_i u_{1i})^2 \\
& + 2\alpha_{10} (u_r u'_{1r} + u_i u'_{1i} - u'_r u_{1r} - u'_i u_{1i}) (u'_r u_{3r} + u'_i u_{3i} - u_r u'_{3r} - u_i u'_{3i}) \\
& + \beta_{10} (u_r u'_{3r} + u_i u'_{3i} - u'_r u_{3r} - u'_i u_{3i})^2 = \frac{\Phi_0^*}{\Phi^*} \varepsilon_1,
\end{aligned} \tag{6.27}$$

which represents an ellipse in the four-dimensional subspace u_r, u_i, u'_r, u'_i . These coordinates are the real parts and the imaginary parts of the complex lateral position $u = u_r + i u_i$ and the complex slope $u' = u'_r + i u'_i$ of a ray. The shape and the angular orientation of the ellipse in the four-dimensional subspace are functions of the location z of this hyperplane in the five-dimensional state space. The ellipse is centered on the z -axis.

We employ the same considerations for the ray, which starts with components y, p_y at the plane z_0 . As a result, we find for the paraxial emittance ε_2 another ellipse, which we obtain from the expression (6.27) by substituting the index 2 for the index 1 and the index 4 for the index 3. In the case of rotational symmetry, we have

$$u_{1i} = u_{3i} = 0, \quad u_{1r} = u_1 = \bar{u}_1, \quad u_{3r} = u_3 = \bar{u}_3, \quad u = \bar{u} = u_r, \quad u' = \bar{u}' = u'_r. \tag{6.28}$$

Moreover, the emittances $\varepsilon_1 = \varepsilon_2 = \varepsilon$ coincide, and the two ellipses adopt the common form

$$\gamma u^2 + 2\alpha u \sqrt{\frac{\Phi^*}{\Phi_0^*}} u' + \beta \frac{\Phi^*}{\Phi_0^*} u'^2 = \varepsilon. \tag{6.29}$$

The Twiss parameters are given by

$$\gamma = \gamma(z) = \frac{\Phi^*}{\Phi_0^*} (\gamma_0 u_1'^2 - 2\alpha_0 u_1' u_3' + \beta_0 u_3'^2), \tag{6.30}$$

$$\beta = \beta(z) = \gamma_0 u_1^2 - 2\alpha_0 u_1 u_3 + \beta_0 u_3^2, \tag{6.31}$$

$$\alpha = \alpha(z) = -\sqrt{\frac{\Phi^*}{\Phi_0^*}} [\gamma_0 u_1 u_1' - \alpha_0 (u_1 u_3' + u_3 u_1') + \beta_0 u_3 u_3']. \tag{6.32}$$

Equation (6.29) reveals that the area of the ellipse is invariant only if we choose the normalized lateral momentum $u' \sqrt{\Phi^* / \Phi_0^*}$ instead of the slope u' as one of the coordinates of the ellipse.

Choosing the slope as coordinate results in a modified emittance

$$\hat{\varepsilon} = \varepsilon \sqrt{\frac{\Phi_0^*}{\Phi^*}}, \tag{6.33}$$

which decreases with increasing acceleration of the beam parallelizing the rays. The Twiss parameters (6.30), (6.31), and (6.32) are functions of the location z of the ellipse along the beam axis. Since these parameters satisfy the relation

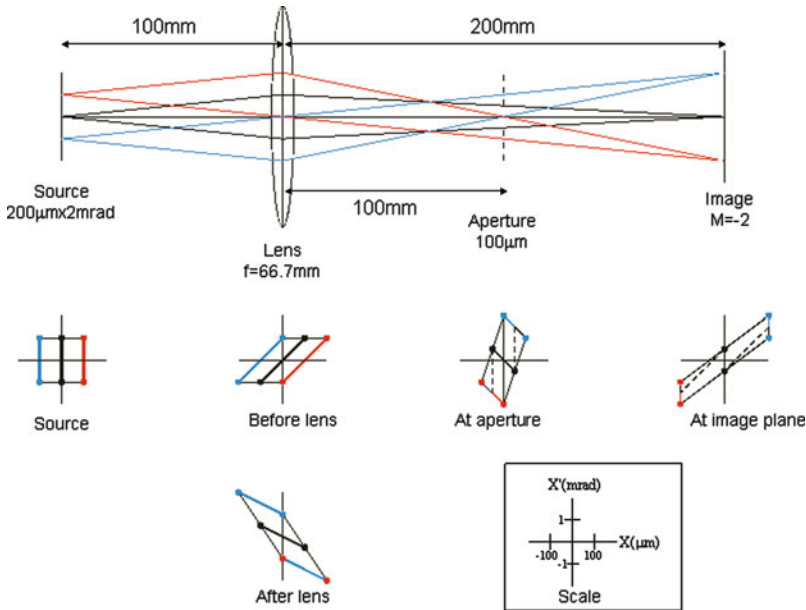


Fig. 6.2 Change of the phase-space element and path of rays along an optical system consisting of a quadratic source, a round lens and a beam-limiting aperture. The dashed lines in the rightmost two phase-space plots illustrate the effect of the aperture, thus showing that it vignettes the off-axis regions of the image. In fact, only the bundle of rays originating at the center of the source remains unaffected by the aperture

$$\beta\gamma - \alpha^2 = \beta_0\gamma_0 - \alpha_0^2 = 1, \quad (6.34)$$

the area of the ellipse remains constant along the optic axis, although its shape may vary considerably. We readily confirm the validity of (6.34) by substituting expressions (6.30), (6.31), and (6.32) for γ , β and α , respectively, and by considering the Lagrange–Helmholtz relation for the fundamental rays u_1 and u_3 . In the presence of an axial magnetic field, the ellipse rotates about the axis in the x, x', y, y' hyperplane. Our considerations have shown that always two ellipses exist for the beam regardless if the system decouples or not. The two ellipses coincide in the case of rotational symmetry. In decoupled systems with two-section symmetry, the ellipses do not rotate. We then have an (x, p_x) ellipse for the horizontal motion and an (y, p_y) ellipse for the vertical motion.

Phase-space diagrams can sometimes be useful for following beams through systems to see how they propagate from one element to another in the paraxial approximation. For instance, Fig. 6.2 shows a simple optical system consisting of a lens and an aperture illuminated by a beam described by a rectangle in phase space. Such a rectangle could be considered an approximation of an ellipse or the illumination provided by an extended source. In these diagrams, drift space produces

a horizontal shear, while lens action shears the figure in the vertical direction. At the image plane, one can see that the phase-space figure is elongated along a diagonal. This shows that the image is non-telecentric, in that the rays at the edges are off-normal, on average. We have illustrated the effect of the aperture by the dotted lines on the right side. One can easily see how big the aperture must be to avoid vignetting. While it is true that the same information can be gained by tracing the rays directly, as in the top half of the figure, the phase-space method is often more direct, once one gets used to it. These diagrams are useful in designing systems so that the beam from one part may be fully used by the succeeding elements. We discuss this topic in somewhat more detail in the next subsection.

6.2.2 *Matching*

Apart from the task to design systems with precise image formation by considering individual trajectories, one is also interested in the transmission and conservation of the beam as a whole. These quantities are of importance in beam-transport systems, such as accelerators and storage rings, and in high-performance analytical electron microscopes equipped with an imaging energy filter. An appropriate measure is the four-dimensional phase-space area, which describes the overall properties of the beam as it propagates through the optical system. The maximum size of the transferable phase-space area (transmissivity) depends on the geometry and the focusing properties of the constituent elements of the system.

Most systems, such as storage rings or electron microscopes, consist of numerous elements. In order that all particles entering the system can pass through it without being lost, it is necessary to embed the emittance domain entirely in the so-called *acceptance* domain throughout the entire system. The inner faces of the electrodes and/or pole pieces or apertures limit the acceptance domain. Since these boundaries are located in most cases outside the paraxial regime, the shape of the acceptance domain is generally not an ellipse due to the nonlinear forces in the non-Gaussian region. In order to transmit all particles of the beam through the system, the emittance of the beam behind any subunit must be located entirely in the acceptance domain of the following unit. We can meet this requirement the best if the emittance

diagram only fills the paraxial region of the acceptance domain, which we can describe by an ellipse. Optimum conditions occur if the emittance ellipse matches the acceptance ellipse in each plane of reference. In order to enlarge the Gaussian regime, it is necessary to reduce the effect of the nonlinear forces. Although it is not possible to eliminate these effects everywhere, we can compensate for the resulting aberrations at distinct planes, thus preventing an uncontrolled expansion of the beam. The problem of optimum matching an imaging energy filter with the lenses of an analytical electron microscope has been investigated by Uhlemann and Rose [116].

6.3 Propagation of Beam Intensity

Hitherto, we have been concerned with the course of individual electrons without considering the trajectory density of electrons emanating from a distinct point of the source. In the case of a fixed-beam transmission electron microscope (TEM), we conceive the current density at the exit plane $z_e = z_o$ of the object as the “source.” Its emission characteristic is determined by scattering of the incident electrons with the atoms of the object. We define this radiation characteristic by the object brightness function

$$B_o = B(x_o, y_o, p_{x_o}, p_{y_o}, z_o) \quad (6.35)$$

In the case of the scanning electron microscope or the scanning transmission electron microscope (STEM), the initial emission characteristic is given by the brightness function $B_s = B(x_s, y_s, p_{x_s}, p_{y_s}, z_s)$ of the source. Employing the relation (6.8), we find that the differential current originating at the emission plane from the surface element $dx_o dy_o$ within the momentum element $dp_{x_o} dp_{y_o}$ is given by

$$d^4 J_o = \frac{B_o}{\bar{p}_o^2} dx_o dy_o dp_{x_o} dp_{y_o}. \quad (6.36)$$

We assume that all trajectories originating within the four-dimensional differential element intersect the recording plane z within the surface element $dx dy$. Because the number of particles is conserved, we obtain for the current density at the field-free recording plane the relation

$$d^2 j_z = \frac{d^4 J}{dx dy} = \frac{B_o}{m_e^2 v_o^2} dx_o dy_o D_o. \quad (6.37)$$

Employing the relation (5.82) and the point eikonal $S = S(\vec{r}_o, \vec{r})$, we write the Jacobi determinant $D_o = D(\vec{r}_o, \vec{r})$ in the form

$$D_o = \frac{\partial p_{x_o}}{\partial x} \frac{\partial p_{y_o}}{\partial y} - \frac{\partial p_{x_o}}{\partial y} \frac{\partial p_{y_o}}{\partial x} = \frac{\partial^2 S}{\partial x_o \partial x} \frac{\partial^2 S}{\partial y_o \partial y} - \frac{\partial^2 S}{\partial x_o \partial y} \frac{\partial^2 S}{\partial y_o \partial x}. \quad (6.38)$$

We cannot employ the relation (6.37) if the point of observation (x, y, z) is located on the caustic where an infinite number of trajectories, emanating from a distinct source point, intersect each other on the caustic. This focal surface degenerates to a point the electrons propagate through the field of a perfect lens providing ideal imaging. In this case, each point of the object plane is associated with a conjugate point in the image plane. The breakdown of the relation (6.37) arises because the determinant (6.38) becomes infinite on the caustic. We can conceive this divergence from the relation (6.8), which defines the initial slope of the ray connecting the source point with the point of observation. If these points are conjugate points, any ray originating from the source point intersects the conjugate point regardless of its initial slope. In reality, this singularity does not arise because of the wave nature of

the electron. Due to this property, we must attribute a partial wave to each electron. As result, strong interference effects show up in the vicinity of the caustic where partial waves with different direction of propagation overlap and interfere, as it is the case in holography. Integrating the differential current density (6.37) over the entire emission (object) plane, we obtain for the current density at the recording plane the expression

$$j_z(x, y, z) = \iint \frac{B_o}{m_e^2 v_o^2} D_o(x, y, x_o, y_o) dx_o dy_o. \quad (6.39)$$

Although the geometrical current density (6.39) is infinite along the curve formed by the intersection of the caustic with the recording plane $z = \text{const}$, the total current

$$J = \iint j_z(x, y) dx dy \quad (6.40)$$

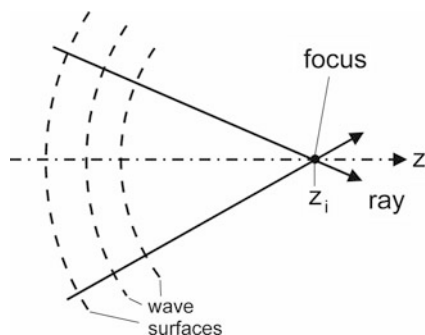
is finite because the number of particles must be conserved [9].

6.4 Caustics

Caustics are important quantities of geometrical optics because they can be made visible, whereas wave fronts and rays are not realizable. Caustics dominate the wave fields when the wavelength is short compared with the characteristic distance $d_r = n / |\text{grad } n|$ of the refracting medium with index of refraction $n = n(x, y, z)$. This condition is satisfied for the macroscopic electromagnetic fields of charged-particle optics. Most electron optical elements have the properties to confine or deflect the beam which we conceive as a bundle of trajectories. We consider a small set of neighboring rays of a homocentric bundle of rays. The current density of this differential pencil of rays is inversely proportional to the cross-sectional area of the pencil. If the rays are converging, they are somehow focused, so that the cross section of the pencil may have zero area on a surface, a line, or even a point. According to geometrical optics, the intensity is infinite on these caustics. However, this is not true in reality due to the particle and wave properties of elementary particles. Wave mechanics attributes a partial wave to each trajectory. These partial waves are coherent with each other if the trajectories originate from the same object point. If this point is located in field-free space, the superposition of all partial waves forms an outgoing spherical wave. The attributed trajectories are then straight lines which originate from the center of the sphere.

The electromagnetic field of any electron optical device steadily deforms the surfaces of constant phase when the electron wave propagates through the electromagnetic field. In the case of ideal imaging, the outgoing spherical wave is transformed into an incident spherical wave in the field-free image space. The center of this wave forms the conjugate image point. Within the frame of

Fig. 6.3 Generation of a point-like caustic by an incident spherical wave



geometrical optics, the initial homocentric bundle of outgoing rays is transformed in a homocentric bundle of incident rays, as depicted in Fig. 6.3. However, we can never achieve this ideal situation in reality because of aberrations. In the case of rotationally symmetric fields, only the paraxial trajectories form a conical pencil behind the lens. Neighboring non-paraxial rays intersect each other at other points. The loci of all these points produce the caustic which generally consists of several sheets representing focal surfaces. The paraxial image point forms the tip of one of the surfaces of a rotationally symmetric caustic named *cusp*. The curves obtained by the intersection of the caustic with the recording plane represent the caustic curves or focal curves, respectively. These curves exhibit very high intensity, as demonstrated convincingly in light optics. We can utilize the shape of the caustic to judge the optical performance of an imaging system, in particular the effect of the aberrations. For example, this possibility is exploited to align the scanning spot in the aberration-corrected STEM by utilizing so-called Ronchigrams produced by the strong interference effects in the vicinity of the caustic.

6.4.1 Determination of the Caustic

The simplest caustic is produced by the intersection points of rays emanating from a common source point. More involved caustic phenomenon show up for an extended source due to the superposition of numerous caustics each of which belonging to a distinct point of the source. In order to elucidate the basic properties of the caustic, we restrict our considerations to a single point source fixed at position x_0, y_0, z_0 . We consider two neighboring trajectories which intersect the aperture plane z_a at positions

$$w_a = x_a + iy_a, \quad w_a + dw_a = x_a + iy_a + dx_a + diy_a. \quad (6.41)$$

Using these boundary conditions, the corresponding trajectories are given by

$$w_1 = w(x_a, y_a, z),$$

$$w_2 = w(x_a + dx_a, y_a + dy_a, z) = w_1 + \frac{\partial w}{\partial x_a} dx_a + \frac{\partial w}{\partial y_a} dy_a. \quad (6.42)$$

The two rays intersect each other at the plane $z_c > z_a$ on the far side of the aperture plane if

$$w_1(z_c) = w_2(z_c). \quad (6.43)$$

By employing relation (6.42), we readily derive the condition

$$\begin{aligned} \frac{\partial x(z_c)}{\partial x_a} dx_a + \frac{\partial x(z_c)}{\partial y_a} dy_a &= 0, \\ \frac{\partial y(z_c)}{\partial x_a} dx_a + \frac{\partial y(z_c)}{\partial y_a} dy_a &= 0. \end{aligned} \quad (6.44)$$

This set of two equations has a nontrivial solution only if the determinant of the coefficients is zero:

$$D_c = D_c(z_c) = \frac{\partial x(z_c)}{\partial x_a} \frac{\partial y(z_c)}{\partial y_a} - \frac{\partial x(z_c)}{\partial y_a} \frac{\partial y(z_c)}{\partial x_a} = 0. \quad (6.45)$$

The equation may have several roots

$$z_c = z_{cv}(x_a, y_a), \quad v = 1 \dots n. \quad (6.46)$$

By substituting this relation for z_c in $w(x_a, y_a, z_c)$, we obtain the parametric representation of the n different sheets of the caustic as

$$z_c = z_{cv}(x_a, y_a), \quad w_c = w_{cv}(x_a, y_a), \quad v = 1 \dots n. \quad (6.47)$$

The sheets need not to be separated and may intersect each other. At most two sheets are formed if the space behind the aperture plane is field-free. In the following, we assume this situation because it allows us to investigate the properties of the caustic in greater detail. The intersection of the caustic with the fixed recording plane z_c yields the caustic curves which are closed loops in most cases. We describe the curves by a single-parameter representation obtained by putting $z_c = z_r = \text{const}$ in (6.47).

6.4.2 Formation of the Caustic in Free Space

For reasons of simplicity, we use in the following the optical path length $L = S/m_e v_0$ instead of the eikonal $S = S(\vec{r}, \vec{r}_o, v_0)$ where v_0 denotes the velocity of the electron in the field-free space behind the focusing elements and $\vec{r}_o$ is the origin of the optical path. Using L for S , the eikonal equation (2.21) adopts in field-free space the simple form

$$\left(\frac{\partial L}{\partial x}\right)^2 + \left(\frac{\partial L}{\partial y}\right)^2 + \left(\frac{\partial L}{\partial z}\right)^2 = 1. \quad (6.48)$$

A special solution of this equation is the distance of the straight line connecting the point x_a, y_a in the aperture plane z_a with the point x, y in the current plane z :

$$L(\vec{r}, \vec{r}_a) = s = \sqrt{(x - x_a)^2 + (y - y_a)^2 + (z - z_a)^2}. \quad (6.49)$$

In the following, we assume that the origin $\vec{r}_o$ is located either at the object plane z_o in front of or within the electromagnetic fields of the lenses, whereas the plane of observation and the aperture plane z_a positioned in the field-free space on the far side of the optical elements. We assume that the trajectory originating from $\vec{r}_o$ intersects the point $\vec{r}_a$ and the point of observation $(x, y, z > z_a)$. The optical path length between the points $\vec{r}_o$ and $\vec{r}$ is

$$L = L_a(\vec{r}_a, \vec{r}_o) + s. \quad (6.50)$$

Here, $L_a = L_a(\vec{r}_a, \vec{r}_o)$ is the optical path length between the points $\vec{r}_o$ and $\vec{r}_a$. Because the trajectory is entirely defined by its lateral positions at the object plane and the plane of observation, it must intersect the aperture plane z_a at a distinct position x_a, y_a . On the other hand, if the ray is fixed by the points $\vec{r}_o$ and $\vec{r}_a$, the connecting line (6.49) between the point $\vec{r}_a$ and the point of observation $\vec{r}$ must be part of the trajectory which intersects the points $\vec{r}_o$ and $\vec{r}_a$. Because the optical path length is an extremum, it must satisfy the continuity conditions

$$\frac{\partial L}{\partial x_a} = 0, \quad \frac{\partial L}{\partial y_a} = 0, \quad \frac{\partial L}{\partial z_a} = 0. \quad (6.51)$$

By employing the relations (6.49) and (6.50) and the conditions (6.51), we obtain for the direction cosines of the trajectory in field-free space the expressions

$$\frac{x - x_a}{s} = \frac{\partial L_a}{\partial x_a} = \partial L_{ax}, \quad \frac{y - y_a}{s} = \frac{\partial L_a}{\partial y_a} = \partial L_{ay}, \quad \frac{z - z_a}{s} = \frac{\partial L_a}{\partial z_a} = \partial L_{az}. \quad (6.52)$$

For our considerations, we fix the trajectory by its intersection points $\vec{r}_o$ and $\vec{r}_a$. Using the relations (6.52) and complex notation, we write the trajectory in field-free space in the form

$$w(z) = x(z) + iy(z) = w(z, \vec{r}_a, \vec{r}_o) = w(z_a) + 2 \frac{z - z_a}{\partial L_{az}} \frac{\partial L_a(w_a, \bar{w}_a)}{\partial \bar{w}_a}. \quad (6.53)$$

In order to obtain an explicit expression for the caustic, we assume that we know the optical path length $L_a = L_a(w_a, \bar{w}_o)$. Moreover, for reasons of mathematical simplicity, we denote partial differentiations of L_a with respect to the coordinates x_a and y_a by indices, as we have done in (6.52). Thus, we write the second derivatives as

$$\begin{aligned}
\frac{\partial^2 L_a}{\partial x_a^2} &= \partial^2 L_{axx}, & \frac{\partial^2 L_a}{\partial y_a^2} &= \partial^2 L_{ayy}, & \frac{\partial^2 L_a}{\partial x_a \partial y_a} &= \partial^2 L_{axy}, \\
\frac{\partial^2 L_a}{\partial x_a \partial z_a} &= \partial^2 L_{axz}, & \frac{\partial^2 L_a}{\partial y_a \partial z_a} &= \partial^2 L_{ayz}.
\end{aligned} \tag{6.54}$$

To solve the caustic condition (6.45), we need the partial derivatives of the lateral coordinates of the trajectory (6.53) with respect to the coordinates x_a and y_a . The resulting expressions are

$$\begin{aligned}
\frac{\partial x}{\partial x_a} &= 1 + (z - z_a) \frac{\partial^2 L_{axx} \partial L_{az} - \partial L_{ax} \partial^2 L_{axz}}{(\partial L_{az})^2}, \\
\frac{\partial y}{\partial y_a} &= 1 + (z - z_a) \frac{\partial^2 L_{ayy} \partial L_{az} - \partial L_{ay} \partial^2 L_{ayz}}{(\partial L_{az})^2}, \\
\frac{\partial x}{\partial y_a} &= (z - z_a) \frac{\partial^2 L_{axy} \partial L_{az} - \partial L_{ax} \partial^2 L_{ayz}}{(\partial L_{az})^2}, \\
\frac{\partial y}{\partial x_a} &= (z - z_a) \frac{\partial^2 L_{axy} \partial L_{az} - \partial L_{ay} \partial^2 L_{axz}}{(\partial L_{az})^2}.
\end{aligned} \tag{6.55}$$

We can eliminate the mixed derivatives L_{axz} and L_{ayz} by means of the eikonal equation

$$(\partial L_{ax})^2 + (\partial L_{ay})^2 + (\partial L_{az})^2 = 1. \tag{6.56}$$

Partial differentiation of this equation with respect to x_a and y_a , respectively, gives

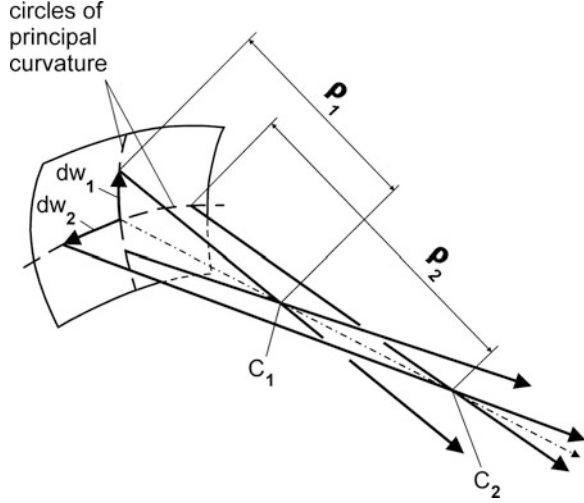
$$\begin{aligned}
\partial L_{ax} \partial^2 L_{axx} + \partial L_{ay} \partial^2 L_{ayx} + \partial L_{az} \partial^2 L_{axz} &= 0, \\
\partial L_{ax} \partial^2 L_{axy} + \partial L_{ay} \partial^2 L_{ayx} + \partial L_{az} \partial^2 L_{ayz} &= 0, \\
\partial L_{ax} \partial^2 L_{axz} + \partial L_{ay} \partial^2 L_{ayz} + \partial L_{az} \partial^2 L_{azz} &= 0.
\end{aligned} \tag{6.57}$$

Introducing the relations (6.55), (6.56), and (6.57) and the distance $d_c = z_c - z_a$ into the caustic condition (6.45), we eventually obtain the caustic equation

$$\begin{aligned}
d_c^2 D_a + d_c (\partial L_{az})^3 \Delta L_a + (\partial L_{az})^4 &= 0, \\
\Delta_a L_a &= \frac{\partial^2 L_a}{\partial x_a^2} + \frac{\partial^2 L_a}{\partial y_a^2} + \frac{\partial^2 L_a}{\partial z_a^2} = 4 \frac{\partial^2 L_a}{\partial w_a \partial \bar{w}_a} + \frac{\partial^2 L_a}{\partial z_a^2}, \\
D_a &= \frac{\partial^2 L_a}{\partial x_a^2} \frac{\partial^2 L_a}{\partial y_a^2} - \left(\frac{\partial^2 L_a}{\partial x_a \partial y_a} \right)^2 = 4 \left\{ \left(\frac{\partial^2 L_a}{\partial w_a \partial \bar{w}_a} \right)^2 - \frac{\partial^2 L_a}{\partial w_a^2} \frac{\partial^2 L_a}{\partial \bar{w}_a^2} \right\}.
\end{aligned} \tag{6.58}$$

The determinant D_a is related with the Gaussian curvature of the wave surface $L(\vec{r}, \vec{r}_o)$ at the point $\vec{r} = \vec{r}_a$; Δ_a denotes the Laplace operator with respect to the

Fig. 6.4 Formation of the caustic points C_1 and C_2 representing the centers of the circles of principal curvature of a surface element of the wave surface $L(\vec{r}, \vec{r}_0) = \text{const}$



coordinates x_a, y_a, z_a . We can interpret $\Delta_a L_a$ as the mean curvature of the wave surface at the position x_a, y_a, z_a .

The quadratic equation in (6.58) has the two solutions

$$d_{c1,2} = (z_{c1,2} - z_a) = -\frac{(\partial L_{az})^2}{2D_a} \left(\partial L_{az} \Delta L_a \pm \sqrt{(\partial L_{az} \Delta L_a)^2 - 4D_a} \right). \quad (6.59)$$

The distances d_{c1} and d_{c2} satisfy the relation

$$d_{c1} d_{c2} = (\partial L_{az})^4 / D_a. \quad (6.60)$$

We illustrate the formation of the caustic by considering the surface element of the eikonal $L_a(w_a, \bar{w}_a)$, as illustrated in Fig. 6.4. The radii of the circles of principle curvature of the surface element are given by

$$\rho_{1,2} = \frac{d_{c1,2}}{\partial L_{az}}. \quad (6.61)$$

The centers C_1 and C_2 of the two circles are located on different sheets of the caustic. The centers are defined by neighboring trajectories intersecting the surface element at positions $w_{a1}, w_{a1} + dw_{a1}$ and $w_{a2}, w_{a1} + dw_{a2}$, respectively. The radii are both either positive or negative if $D_a > 0$. Negative radii show up for $\partial L_{ax} \Delta L_a > 0$. In this case, the trajectories do not form a caustic in the field-free space behind the lens system. However, the asymptotes of the trajectories intersect each other in the left half space $z < z_a$ where they form a virtual caustic. The distances d_{c1} and d_{c2} are both positive for $D_a > 0$ and $\partial L_{ax} \Delta L_a < 0$. In this case, a caustic

of two sheets is formed in the right half space $z > z_a$. If the Gaussian curvature is negative, $D_a < 0$, one distance is positive and the other negative regardless of the sign of $\partial L_{ax} \Delta L_a$. Accordingly, the caustic consists in this case of a single real sheet located in the right half space and a single virtual sheet positioned in the left half space. By inserting the relation (6.59) into the expression (6.53), we readily obtain the parametric representation of the caustic:

$$\begin{aligned} z_{c1,2} - z_a &= -\frac{2\partial L_{az}}{\Delta L_a \pm \sqrt{(\Delta L_a)^2 - 4D_a/(\partial L_{az})^2}}, \\ w(z_{c1,2}) &= w_a - \frac{4\partial L_{az}}{\Delta L_a \pm \sqrt{(\Delta L_a)^2 - 4D_a/(\partial L_{az})^2}} \frac{\partial L_a(w_a, \bar{w}_a)}{\partial \bar{w}_a}. \end{aligned} \quad (6.62)$$

We utilize the relations (6.59) and (6.62) for determining the mean curvature κ_m and the Gaussian curvature κ_G of the wave surface at the point in free space. The resulting expressions have the simple form:

$$\begin{aligned} \kappa_m &= \frac{1}{2} \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right) = \frac{1}{2} \Delta L_a, \\ \kappa_G &= \frac{1}{\rho_1 \rho_2} = \frac{D_a}{(\partial L_{az})^2}. \end{aligned} \quad (6.63)$$

In the special case $\kappa_G = \kappa_m^2$, the caustic degenerates to a single sheet because then, the radicand of the square root in the formulas (6.62) is zero. The Gaussian curvature κ_G vanishes if the eikonal $L_a(w_a, \bar{w}_a) = L_a(x_a, y_a)$ depends only on a single variable. For example, this situation is realized for an electric cylinder lens whose field is two-dimensional. In this case, the caustic consists of a single sheet. Without loss of generality, we can choose the coordinate system such that $L_a = L(z_a, y_a)$. Using this choice, we obtain from (6.58) and (6.53) for the caustic the single-parameter representation

$$\begin{aligned} z_c &= z_a - \frac{\partial L_{az}}{\partial^2 L_{ayy} + \partial^2 L_{azz}}, \\ y_c &= y_a - \frac{\partial L_{ay}}{\partial^2 L_{ayy} + \partial^2 L_{azz}}. \end{aligned} \quad (6.64)$$

To demonstrate the validity of our consideration, we consider the caustic formed by an ideal lens. The point source located on the optic axis is imaged by the lens into the conjugate image point $x_i = 0$, $y_i = 0$, $z_i > z_a$. With these assumptions, the eikonal at the aperture plane is given by

$$L_a = L_0 - s_a, \quad s_a = \sqrt{x_a^2 + y_a^2 + (z_a - z_i)^2}. \quad (6.65)$$

The optical path length L_0 is an arbitrary constant. Substituting the expression (6.65) for the eikonal in the relations (6.58), we readily obtain

$$\begin{aligned} 4D_a &= (\partial L_{az} \Delta L_a)^2 = 4(z_a - z_i)^2/s_a^4, \\ L_{az} &= (z_i - z_a)/s_a. \end{aligned} \quad (6.66)$$

By introducing these relations into (6.59), (6.62), and (6.63), we find the trivial result

$$\begin{aligned} z_{c1} = z_{c2} = z_c = z_i, \quad \rho_1 = \rho_2 = s_a, \\ w(z_c) = w(z_i) = 0, \quad \kappa_G = \kappa_m^2 = 1/s_a^2. \end{aligned} \quad (6.67)$$

In this singular case, the caustic collapses to a point which represents the ideal image of the point source. Unfortunately, the ideal form (6.64) cannot be realized in the space behind rotationally symmetric electron lenses due to the validity of the Scherzer theorem. The unavoidable spherical aberration of these lenses prevents the outgoing spherical wave to be transformed into an incident spherical wave. However, the spherical form of the eikonal (6.65) is not mandatory in order that the caustic collapses to a point. To prove this conjecture, we consider the eikonal

$$L_a = z_a - z_0 - \frac{x_a^2 + y_a^2}{2b} = z_a - z_0 - \frac{w_a \bar{w}_a}{2b}. \quad (6.68)$$

Here, z_0 and b are arbitrary constant parameters. For this eikonal, we obtain

$$\partial L_{az} = \frac{\partial L_a}{\partial z_a} = 1, \quad \partial^2 L_{azz} = \frac{\partial^2 L_a}{\partial z_a^2} = 0, \quad \Delta L_a = -\frac{1}{b}, \quad D_a = \frac{1}{b^2}. \quad (6.69)$$

Inserting these values into the expressions into (6.62) and (6.63), we readily find

$$\begin{aligned} z_{c1} = z_{c2} = z_c = b + z_a, \quad \rho_1 = \rho_2 = b, \\ w(z_c) = 0, \quad \kappa_G = \kappa_m^2 = 1/b^2. \end{aligned} \quad (6.70)$$

These relations closely resemble those derived from the eikonal (6.65). However, the radii and the quantities (6.69) are constants, whereas the corresponding quantities (6.66) and (6.67), derived from the eikonal (6.65), are functions of the coordinates x_a, y_a, z_a .

6.4.3 Classification of the Fundamental Caustic Forms

The caustics observed in nature possess generally a very involved structure. In order to elucidate the formation of intricate caustics, it is advantageous to investigate the

so-called *elementary* caustics. Caustics are also an important subject of catastrophe theory where the elementary caustics are called *elementary catastrophes* [117]. Unfortunately, we cannot simply add the elementary caustics for obtaining more complicated caustics, although their generating functions are linear combinations of the generating functions of the elementary caustics. Stigmatic imaging corresponds to the highest singularity a caustic can have. In this case, the caustic collapses to a point. Within the context of catastrophe theory, this singularity represents the largest possible catastrophe. We can observe the elementary catastrophes in electron optics by employing distinct compound lenses consisting of round lenses and specific multipoles. For reasons of simplicity, we suppose that the aperture plane is located in field-free image space. Moreover, we choose our coordinate system in such a way that the z -axis coincides with the symmetry axis of the caustic. This implies that the source point is always located on the z -axis, which is generally tilted with respect to the optic axis of the optical system. If the point source is not located on the optic axis, the off-axial coma of a lens, centered on the optic axis, shows up as axial coma in the tilted system. By introducing this coordinate system, the linear terms in x_a and y_a of the power series expansion of the eikonal L_a vanish. For obtaining the elementary caustics, we expand the eikonal $L_a = L(w_a, \bar{w}_a, z_a, \vec{r}_o)$ in a series of polynomials of increasing order n in the complex lateral coordinates w_a and $\bar{w}_a$, giving

$$L_a = z_a + \sum_{n=2}^{\infty} \sum_{k=0}^{[n/2]} \frac{1}{n} \operatorname{Re} \left\{ L_{n-2k}^{(n)} \left(\frac{\bar{w}_a}{b} \right)^{n-k} \left(\frac{w_a}{b} \right)^k \right\}. \quad (6.71)$$

The upper summation limit $[n/2]$ is the integer value of $n/2$. The eikonal coefficients $L_{n-2k}^{(n)} = L_{n-2k}^{(n)}(w_o, \bar{w}_o, z_o; v_o)$ are real for multiplicity $m = n - 2k = 0$; they are generally complex else. The series (6.71) converges sufficiently fast because the beams within most systems of charged-particle optics are confined by apertures to the axial regime in order to avoid large slopes w_a/b of the trajectories. If the point source is centered on the optic axis ($w_o = \bar{w}_o = 0$) of a rotationally symmetric system, all coefficients with multiplicity $m = n - 2k \neq 0$ vanish. The expansion adopts the special form (6.68) if $L_0^{(2)} = -b$, $L_2^{(2)} = 0$ and all eikonal coefficients of order $n > 2$ vanish. In this case, the caustic produced by the eikonal (6.71) collapses to a point. One tries to achieve this form approximately by means of multipole correctors, which nullify the higher-order terms up to $n = 5$ inclusively. A point-like caustic is rather instable because already small disturbances of the optical medium split the point singularity in two sheets. To avoid an unduly large splitting in an aberration-corrected electron microscope, the electromagnetic fields of the lenses and correctors must be precisely adjusted and kept stable with extreme accuracy.

The elementary caustics are produced by the generating functions

$$L_{gm}^{(n)} = z_a + \frac{1}{nb^n} \operatorname{Re} \left\{ L_{n-2k}^{(n)} \bar{w}_a^{n-k} w_a^k \right\} - \frac{w_a \bar{w}_a}{2b}, \quad m = n - 2k, \quad k = 0 \dots [n/2]. \quad (6.72)$$

We assume that the coefficients $L_{n-2k}^{(n)}$ do not depend appreciably on the position z_a of the aperture plane so that

$$\partial L_{gmz}^{(n)} = \frac{\partial L_{gm}^{(n)}}{\partial z_a} \approx 1, \quad \frac{\partial^2 L_{gm}^{(n)}}{\partial z^2} \approx 0. \quad (6.73)$$

By applying this approximation to the relations (6.58), the caustic equations (6.62) adopt the form

$$z_{c1,2} - z_a = -\frac{1}{2} \left(\frac{\partial^2 L_{gm}^{(n)}}{\partial w_a \partial \bar{w}_a} \pm \left| \frac{\partial^2 L_{gm}^{(n)}}{\partial \bar{w}_a^2} \right| \right)^{-1},$$

$$w(z_{c1,2}) = w_a + 2(z_{c1,2} - z_a) \frac{\partial L_{gm}^{(n)}(w_a, \bar{w}_a)}{\partial \bar{w}_a}. \quad (6.74)$$

The expressions demonstrate that the use of complex coordinates simplifies considerably the parametric representation of the caustic.

Each elementary caustic is related with a characteristic aberration produced by electron lenses and/or multipoles. For example, the two sheets of the elementary line-foci caustic are degenerated in two spatially separated orthogonal straight lines, which are produced in the electron microscope by the axial astigmatism of the objective lens. In the following, we suppose without loss of generality that the eikonal coefficients are real:

$$L_m^{(n)} = \bar{L}_m^{(n)} = b^2 / l_m^{(n)}. \quad (6.75)$$

6.4.4 Third-Order Elementary Caustics

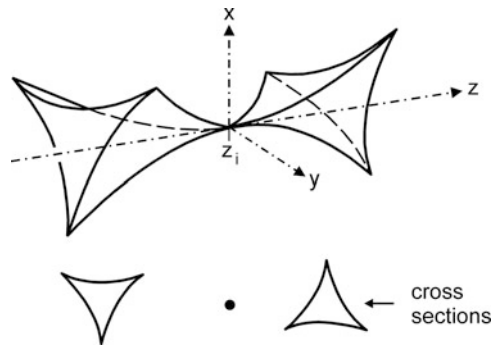
The elementary caustics of third order consist of the *threefold umbilic* and the *hyperbolic fold*. The threefold umbilic originates from a threefold deformation of the wave surface at the aperture plane. As a result, the non-paraxial rays miss the Gaussian image point and produce the threefold astigmatism. The associated caustic is named threefold umbilic (umbilicus) due to its funnel-shaped form. By employing the generating eikonal $L_g^{(3)}$ listed in Table 6.1 and the parameters $l_3^{(3)}$ and $\rho_a = \sqrt{x_a^2 + y_a^2}$, we readily obtain the parametric representation of the caustic from the equations (6.74) in the form

$$z_{c1,2} - z_a = \frac{b l_3^{(3)}}{l_3^{(3)} \pm 2\rho_a},$$

$$w(z_{c1,2}) = \frac{\bar{w}_a^2 \pm 2w_a \rho_a}{l_3^{(3)} \pm 2\rho_a}. \quad (6.76)$$

Table 6.1 Classification of the elementary caustics and their generating functions up to order $n = 4$ inclusively

Order n	Multiplicity m	Name	Generating function $L_{gm}^{(n)}$
2	0	Point focus	$z_a - w_a \bar{w}_a / 2b$
	2	Line foci	$z_a + \operatorname{Re} \left\{ L_2^{(2)} \bar{w}_a^2 \right\} / 2b^2 - w_a \bar{w}_a / 2b$
3	3	Threefold umbilic	$z_a + \operatorname{Re} \left\{ L_3^{(3)} \bar{w}_a^3 \right\} / 3b^3 - w_a \bar{w}_a / 2b$
	1	Double fold	$z_a + \operatorname{Re} \left\{ L_1^{(3)} \bar{w}_a^2 w_a \right\} / 3b^3 - w_a \bar{w}_a / 2b$
4	0	Rotational cusp	$z_a + L_0^{(4)} \bar{w}_a^2 w_a^2 / 4b^4 - w_a \bar{w}_a / 2b$
	2	Plane double cusp	$z_a + \operatorname{Re} \left\{ L_2^{(4)} \bar{w}_a^3 w_a \right\} / 4b^4 - w_a \bar{w}_a / 2b$
	4	Fourfold umbilic	$z_a + \operatorname{Re} \left\{ L_4^{(4)} \bar{w}_a^4 \right\} / 4b^4 - w_a \bar{w}_a / 2b$

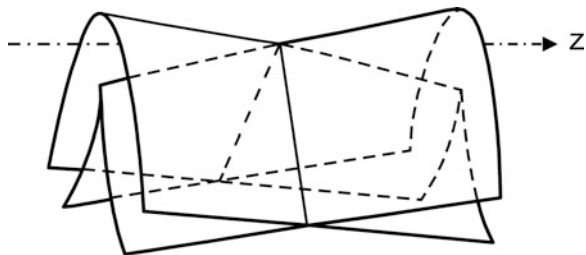
Fig. 6.5 Caustic and cross sections of the threefold umbilic produced in the electron microscope by the threefold astigmatism

This caustic and the caustic curves are shown for several recording planes in Fig. 6.5. The two sheets of the threefold umbilic are located on opposite sides of their common tip located at the center of the plane $z = z_a + b$. Therefore, we obtain for each location of the recording plane only a single threefold caustic curve which represents a deltoid.

The hyperbolic fold consists of two hyperbola-like sheets located in the same half space $x > x_0$. Hence, we obtain two caustic curves for each recording plane. The generating function $L_{g1}^{(3)}$ listed in Table 6.1 differs somewhat from that of the equivalent hyperbolic umbilic investigated in catastrophe theory. To avoid confusion, we have characterized our choice by a new name based on the shape of the caustic. The corresponding generating function has the advantage to connect directly the double fold with the axial coma of microscopes.

We obtain the equations for the caustic of the *double fold* by introducing $L_{g1}^{(3)}$ for $L_{gm}^{(n)}$ into the relations (6.74), giving

Fig. 6.6 Caustic of the double fold produced by the axial coma in the electron microscope



$$\begin{aligned}
 z_{c1,2} - z_a &= \frac{3bl_1^{(3)}}{3l_1^{(3)} - 4x_a \mp 2\rho_a}, \\
 w_{c1} &= -\frac{w_a^2 + 2w_a\rho_a}{3l_1^{(3)} - 4x_a - 2\rho_a}, \\
 w_{c2} &= -\frac{w_a^2 - 2w_a\rho_a}{3l_1^{(3)} - 4x_a + 2\rho_a}
 \end{aligned} \tag{6.77}$$

Each of the two sheets of the caustic has the shape of a fold, as illustrated schematically in Fig. 6.6. The sheets intersect each other at the Gaussian image plane $z_i = z_{c1} = z_{c2} = z_a + b$, where the caustic curves form two straight lines given by

$$y(z_i) = \pm\sqrt{3}x(z_i), \quad x_i \geq 0. \tag{6.78}$$

The *bifurcation lines* originate from the center $x(z_i) = y(z_i) = 0$ and enclose an angle of 120° . This angle coincides with that of the conventional coma figure formed by round lenses.

If we move the recording plane z_r away from the Gaussian image plane, the straight lines split in two curves. Both curves are symmetric with respect to the x -axis, as shown in Fig. 6.7. One curve intersects the y -axis, the other not. Its sharp cusp is located on the x -axis at position

$$x_{c2} = \frac{3l_1^{(3)}}{4b} \frac{(b - z_r + z_a)^2}{z_r - z_a} > 0. \tag{6.79}$$

The other caustic curve resembles a hyperbola which intersects the y -axis twice. Its apex is positioned on the negative x -axis at the point

$$x_{c1} = -\frac{l_1^{(3)}}{4b} \frac{(b - z_r + z_a)^2}{z_r - z_a} > 0. \tag{6.80}$$

It readily follows from the relations (6.79) and (6.80) that the cusp of the inner curve is three times farther away from the z -axis than the apex of the other curve.

Fig. 6.7 Cross section of the double fold shown in Fig. 6.6 with the recording plane $z_r \neq z_i$

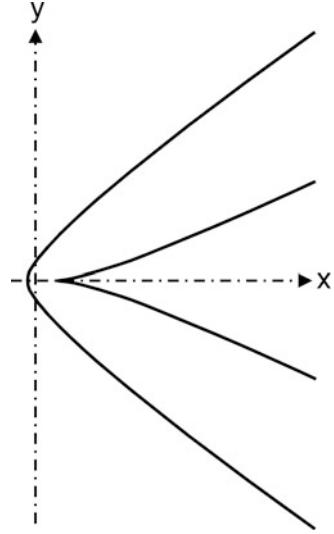
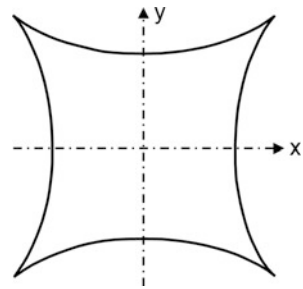


Fig. 6.8 Cross section of the fourfold umbilic forming an astroid

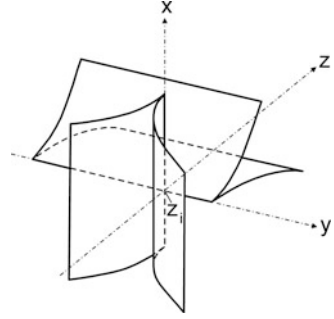


6.4.5 Fourth-Order Elementary Caustics

The elementary caustics of fourth order consist of the *rotational cusp*, the *plane double cusp*, and the *fourfold umbilic*. We readily obtain the caustic of the fourfold umbilic from that of the threefold umbilic by replacing its funnel-shaped threefold caustic by an equivalent fourfold caustic. Hence, we do not need to treat this caustic in detail. The intersection of the fourfold funnel-shaped caustic with the recording plane has the form of a star, as shown in Fig. 6.8. The fourfold umbilic originates in electron optical systems from the fourfold astigmatism.

We derive the equations of the plane double cusp by introducing the eikonal term $L_{g2}^{(4)}$ listed in table 6.1 for $L_{gm}^{(2)}$ into (6.74). The resulting parametric equations for the two sheets of the caustic are found as

Fig. 6.9 Caustic of the plane double cusp formed by the star aberration



$$\begin{aligned}
 z_{c1} - z_a &= \frac{b^2 l_2^{(4)}}{b l_2^{(4)} - 3x_a^2}, & z_{c2} - z_a &= \frac{b^2 l_2^{(4)}}{b l_2^{(4)} + 3y_a^2}, \\
 x_{c1} &= -\frac{2x_a^3}{b l_2^{(4)} - 3x_a^2}, & x_{c2} &= \frac{x_a^3 + 3x_a y_a^2}{b l_2^{(4)} + 3y_a^2}, \\
 y_{c1} &= -\frac{y_a^3 + 3x_a^2 y_a}{b l_2^{(4)} - 3x_a^2}, & y_{c2} &= \frac{2y_a^3}{b l_2^{(4)} + 3y_a^2}.
 \end{aligned} \tag{6.81}$$

The intersection of the two sheets of the caustic with the fixed recording plane $z_r = z_{c1} = z_{c2}$ yields the caustic curves. In order to obtain a single-parameter representation of these curves, we replace for the first curve the parameter x_a and for the second curve the parameter y_a by the distance $d_r = z_c - z_a = z - z_a$. These substitutions give

$$\begin{aligned}
 x_{c1} &= -\frac{2}{3} \sqrt{\frac{l_2^{(4)}}{3b} \frac{(z - z_i)^3}{z - z_a}}, & x_{c2} &= \frac{x_a}{b} \left(z_i - z + \frac{z - z_a}{b l_2^{(4)}} x_a^2 \right), \\
 y_{c1} &= -\frac{y_a}{b} \left(z - z_i + \frac{z - z_a}{b l_2^{(4)}} y_a^2 \right), & y_{c2} &= \frac{2}{3} \sqrt{\frac{l_2^{(4)}}{3b} \frac{(z_i - z)^3}{z - z_a}}, \\
 l_2^{(4)}(z - z_i) &\geq 0, & l_2^{(4)}(z_i - z) &\geq 0.
 \end{aligned} \tag{6.82}$$

These relations demonstrate that the caustic consists of two line-shaped cusps which are perpendicular to each other, as shown schematically in Fig. 6.9. One cusp is located in front of and the other behind the Gaussian image plane $z = z_i$. They touch each other at the center $x_i = y_i = 0$ of this plane.

The fourth-order eikonal $L_{g0}^{(4)}$ produces the rotational cusp whose sharp tip is located at the center of the Gaussian image plane. The coefficient $L_0^{(4)}$ of the generating function is always negative for round electron lenses because their coefficient of the spherical aberration $C_3 \propto -L_0^{(4)} > 0$ is always positive. Therefore,

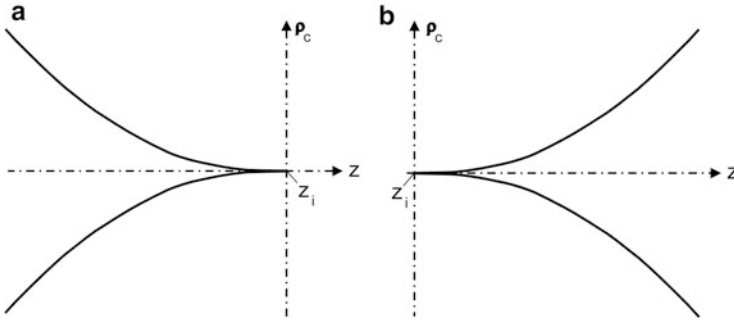


Fig. 6.10 Axial sections $\rho_c = \rho_2(z) = 3\sqrt{3}\rho_1(z)$ of the rotational cusp in the case (a) $L_0^{(4)} = -C_3 < 0$ and (b) $L_0^{(4)} = -C_3 > 0$

these lenses can never provide aberration-free imaging. Substituting $L_{g0}^{(4)}$ from Table 6.1 for $L_g^{(n)}$ in the equations (6.74) and considering the relation (6.75), we readily find

$$\begin{aligned} d_{c1,2} = z_{c1,2} - z_a &= \frac{b^2 l_0^{(4)}}{b l_0^{(4)} - 2\rho_a^2(2 \mp 1)}, \\ w_{c1,2} &= -\frac{w_a \rho_a^2}{b l_0^{(4)} - 2\rho_a^2(2 \mp 1)}. \end{aligned} \quad (6.83)$$

The first equation shows that the caustic is located between the aperture plane and the Gaussian image plane if $L_0^{(4)} < 0$, as illustrated in Fig. 6.10a. This situation exists always for rotationally symmetric electron lenses whose caustic, caused by spherical aberration, is confined to the region $z_a \leq z \leq z_i$. We can realize the case $L_0^{(4)} > 0$ in electron optics only by means of multipole correctors or electron mirrors, which are able to overcompensate for the spherical aberration of round lenses. In scanning electron microscopes, the tip of the rotational cusp is located at the back focal plane $z_i = z_F$ of the objective lens. In this case, the aperture plane coincides with the principal plane located at $z_a = z_P = z_F - f$ where $f = b$ is the focal length of the objective lens. Using these relations together with $L_0^{(4)} = -C_3$, we obtain the curves $\rho_{1,2} = \rho_{1,2}(z)$ of the axial section of the associated caustic by substituting $d_{c1,2} = z - z_F + f$ from the first equation of (6.83) for $\rho_a = |w_a|$ in the second equation, giving

$$\rho_1(z) = |w_1(z_c = z)| = \frac{\rho_2(z)}{3\sqrt{3}} = \frac{1}{6} \sqrt{\frac{f(z_F - z)^3}{6C_3(z + f - z_F)}}. \quad (6.84)$$

This result demonstrates that the two caustic curves $\rho_1(z)$ and $\rho_2(z)$ differ from each other only by the scaling factor $3\sqrt{3}$. The form of the caustic observed in realistic systems differs considerably from that of the elementary caustics because its generating function is the sum of the generating functions of several elementary caustics. In electron microscopy, the precise determination of the eikonal coefficients $L_m^{(n)}$ is an indispensable prerequisite for aligning aberration correctors with the required accuracy. These correctors nullify or reduce sufficiently the eikonal coefficients up to a distinct order n , which is defined by the required resolution of the microscope.

Chapter 7

Path Deviations

Abstract The inhomogeneous integral equations for the path deviations derived in Chap. 5 are solved by successive iteration starting from the paraxial trajectories which are supposed to be known. The iteration algorithm yields the deviations with increasing rank which is the sum of the exponents of the chromatic parameter and of the four geometrical ray parameters which define the trajectory.

We can solve most conveniently the set (5.124) of integral equations for the lateral position $w = w(z)$ and canonical momentum $p = p(z)$ of the particle at a given plane z by iteration. This method is well known from the theory of inhomogeneous integral equations as Neumann iteration procedure. The resulting Neumann series converges the faster the smaller is the kernel of the integral equation. With respect to our task, this behavior implies that the paraxial approximations is the better the closer the beam is confined to the optic axis because the kernel (5.121) of the integral equations (5.124) increases the larger the slope and the off-axial distance of the trajectory. We presuppose that the paraxial approximations $w^{(1)}(z)$ and $p^{(1)}(z)$ are known functions and that the geometrical ray parameters a_μ and the chromatic parameter κ are small quantities. The set (5.124) of integral equations is best suited for obtaining successively the power series expansions of the lateral position and the lateral component of the canonical momentum of the ray with respect to the ray parameters. In order to obtain recurrence formulas for the momentum and path deviations, it is advantageous to introduce an expansion parameter ε , which we will put equal to unity after the expansion. Using this parameter, we expand the complex ray coordinate in the form

$$w \Rightarrow w(\varepsilon a_\mu, \varepsilon \kappa; z) = \sum_{r=1}^{\infty} \varepsilon^r w^{(r)}(z) \Rightarrow \sum_r w^{(r)}(z). \quad (7.1)$$

Here $w^{(r)}(z)$ is a polynomial of rank r in the geometrical ray parameters a_1, a_2, a_3, a_4 and the chromatic parameter κ . The coefficients of the constituent

monomials of each polynomial are functions of the z -coordinate. The introduction of the fictitious sorting parameter ε will prove useful for separating the path deviations according to their rank. We will put this parameter equal to unity at the end of our calculations. The rank $r = n + l$ is composed of the exponent l of the chromatic parameter and of the so-called *Seidel order* n , which is the sum of the exponents of the geometrical ray parameters. We define the exponent l as the *degree* of the deviation. Accordingly, the rank of the deviation is the sum of its order and degree. For example, the primary chromatic aberration is of first order and first degree and, therefore, an aberration of second rank. The rank is a measure for the magnitude of the deviation. Since the ray parameters are small, the deviation $w^{(r)}(z)$ of the ray from its paraxial course $w^{(1)}(z)$ is the smaller the higher the rank of the deviation is.

The path deviation of rank r has the form

$$w^{(r)}(z) = \sum_{n_v, l} w_{n_1 n_2 n_3 n_4 l}^{(r)}(z) a_1^{n_1} a_2^{n_2} a_3^{n_3} a_4^{n_4} \kappa^l, \quad v = 1, 2, 3, 4, \quad (7.2)$$

with the constraint

$$r = n + l = n_1 + n_2 + n_3 + n_4 + l. \quad (7.3)$$

We denote the coefficient $w_{n_1 n_2 n_3 n_4 \kappa}^{(r)}(z)$ of each monomial as the r th-rank fundamental ray of suborders n_1, n_2, n_3, n_4 and degree l . Each coefficient is a function of the *current* plane z and the planes z_v at which we define the ray by the components of its lateral position and its slope, respectively. The fundamental rays of rank r determine the course of the r th-rank path deviation (7.2) along the optic axis. In many cases, it is advantageous to replace the real parameters a_v by the complex ray parameters

$$\omega = a_1 + i a_2, \quad \rho = a_3 + i a_4 \quad (7.4)$$

and their conjugate complex values. Using these parameters, the r th-rank path deviation has the form

$$w^{(r)}(z) = \sum_{n_\mu} w_{n_\omega n_{\bar{\omega}} n_\rho n_{\bar{\rho}} l}^{(r)}(z) \omega^{n_\omega} \bar{\omega}^{n_{\bar{\omega}}} \rho^{n_\rho} \bar{\rho}^{n_{\bar{\rho}}} \kappa^l, \quad \mu = \omega, \bar{\omega}, \rho, \bar{\rho} \quad (7.5)$$

with the constraint

$$r = n + l = n_\omega + n_{\bar{\omega}} + n_\rho + n_{\bar{\rho}} + l. \quad (7.6)$$

The appropriate choice of representation (7.2) or (7.6) for the path deviation depends on the symmetry of the system.

7.1 Iteration Algorithm

We aim for a recurrence formula, which yields the path deviations successively with increasing rank. The iteration starts with the lateral position and canonical momentum components of the Gaussian ray

$$w^{(1)}(z) = \sum_{\mu=1}^4 a_{\mu} w_{\mu}(z) + \kappa w_{\kappa}(z), \quad (7.7)$$

$$p^{(1)}(z) = \sum_{\mu=1}^4 a_{\mu} p_{\mu}(z) + \kappa p_{\kappa}(z), \quad p_{\kappa} = q w'_{\kappa}(z). \quad (7.8)$$

The paraxial approximations (7.8) are the inhomogeneous terms of the integral equations (5.124). In order to obtain the recurrence formulas, we expand the variational function (3.59) of the eikonal S in a power series with respect to the sorting parameter ε to give

$$\mu = \mu^{(0)} + \sum_{k=2}^{\infty} \mu^{(k)}(w, \bar{w}, w', \bar{w}', \kappa; z) = \mu^{(0)} + \sum_{r=1}^{\infty} \varepsilon^{r+1} m^{(r+1)}(z). \quad (7.9)$$

Here $m^{(r+1)} = m^{(r+1)}(w^{(1)}, w^{(2)}, \dots, w^{(r)}, z)$ denotes the variational polynomial of degree $r + 1$ in the ray parameters a_v and κ . Each polynomial originates from contributions of the polynomials $\mu^{(k)}$ with $k \leq r + 1$. We can express the structure of $m^{(r+1)}$ in terms of the polynomials $\mu^{(k)}$ and the path deviations $w^{(\lambda)}, \lambda \leq r$ in the concise operator form

$$m^{(r+1)} = \frac{1}{(r+1)!} \left\{ \sum_{k=2}^{r+1} \frac{\partial^{r+1}}{\partial \varepsilon^{r+1}} \left[\varepsilon^k \exp \left(\sum_{j=2}^{\infty} \varepsilon^{j-1} D^{(j)} \right) \mu_1^{(k)} \right] \right\}_{\varepsilon=0}. \quad (7.10)$$

The differential operator

$$D^{(j)} = w^{(j)} \frac{\partial}{\partial w^{(1)}} + \bar{w}^{(j)} \frac{\partial}{\partial \bar{w}^{(1)}} + w^{(j)'} \frac{\partial}{\partial w^{(1)'}} + \bar{w}^{(j)'} \frac{\partial}{\partial \bar{w}^{(1)'}} \quad (7.11)$$

replaces one of each of the four paraxial ray components $w^{(1)}, \bar{w}^{(1)}, w^{(1)'}$, and $\bar{w}^{(1)'}$ in the polynomials

$$\mu_1^{(k)} = \mu^{(k)}(w^{(1)}, \bar{w}^{(1)}, w^{(1)'}, \bar{w}^{(1)'}, \kappa; z) \quad (7.12)$$

by the corresponding components of the j th-rank path deviation [54, 106].

For evaluating the operator expression (7.10), we must expand the exponential function in a Taylor series in order to perform the differentiations. Only terms with

the same factor ε^{r+1} contribute to the polynomial $m^{(r+1)}$. We can write the result of the rather lengthy calculation as

$$\begin{aligned} m^{(r+1)} = & \mu_1^{(r+1)} + \sum_{k=0}^{r-2} D^{(k+2)} \mu_1^{(r-k)} + \frac{1}{2!} \sum_{k=1}^{r-2} \sum_{h=1}^{r-k} D^{(k+2-h)} D^{(h+1)} \mu_1^{(r-k)} \\ & + \frac{1}{3!} \sum_{k=3}^{r-3} \sum_{h=1}^{k-1} \sum_{j=1}^{k-h} D^{(k+2-h-j)} D^{(h+1)} D^{(j+1)} \mu_1^{(r-k)} + \dots \end{aligned} \quad (7.13)$$

In this formula, we must put all terms zero if the upper summation index of a sum is smaller than the lower index. Considering this requirement, we obtain the following expressions:

$$\begin{aligned} m^{(2)} &= \mu_1^{(2)}, \\ m^{(3)} &= \mu_1^{(3)} + D^{(2)} \mu_1^{(2)} = \mu_1^{(3)} + D^{(2)} m^{(2)}, \\ m^{(4)} &= \mu_1^{(4)} + D^{(2)} \mu_1^{(3)} + D^{(3)} \mu_1^{(2)} + \frac{1}{2} D^{(2)^2} \mu_1^{(2)} \\ &= \mu_1^{(4)} + \frac{1}{2} D^{(2)} \mu_1^{(3)} + \frac{1}{2} D^{(2)} m^{(3)} + D^{(3)} m^{(2)}, \\ m^{(5)} &= \mu_1^{(5)} + D^{(2)} \mu_1^{(4)} + D^{(3)} \mu_1^{(3)} + D^{(4)} \mu_1^{(2)} + D^{(3)} D^{(2)} \mu_1^{(2)} + \frac{1}{2} D^{(2)^2} \mu_1^{(3)} \\ &= \mu_1^{(5)} + D^{(2)} \mu_1^{(4)} + \frac{1}{2} D^{(2)^2} \mu_1^{(3)} + D^{(3)} m^{(3)} + D^{(4)} m^{(2)}. \end{aligned} \quad (7.14)$$

We apply the different representations of the variational polynomials for obtaining most suitable expressions for the path deviations. To find these representations, we utilize the linear relation (7.1) between w and $w^{(1)}$, giving

$$\frac{1}{\varepsilon} \frac{\partial}{\partial \bar{w}^{(1)}} = \frac{1}{\varepsilon} \frac{\partial}{\partial \bar{w}} \frac{\partial \bar{w}}{\partial \bar{w}^{(1)}} = \frac{\partial}{\partial \bar{w}}. \quad (7.15)$$

Employing this result together with the equivalent relation for the derivative $\bar{w}'$ and the expansion (7.9) for the variational function μ , we can rewrite the Euler-Lagrange equation (4.12) as

$$\frac{\partial \mu}{\partial \bar{w}} - \frac{d}{dz} \frac{\partial \mu}{\partial \bar{w}'} = \sum_{r=1}^{\infty} \varepsilon^r \left[\frac{\partial m^{(r+1)}}{\partial \bar{w}^{(1)}} - \frac{d}{dz} \frac{\partial m^{(r+1)}}{\partial \bar{w}^{(1)'}} \right] = 0. \quad (7.16)$$

The last relation must be satisfied for arbitrary values of the sorting parameter ε . We can meet this requirement only if the expression in the bracket vanishes. The result

$$\frac{\partial m^{(r+1)}}{\partial \bar{w}^{(1)}} = \frac{d}{dz} \frac{\partial m^{(r+1)}}{\partial \bar{w}^{(1)'}} = \frac{1}{2q_0} \frac{dp^{(r)}}{dz} \quad (7.17)$$

reveals that the r th-rank deviation of the lateral canonical momentum is related with the $(r + 1)$ th-rank variational polynomial via

$$p^{(r)} = 2q_0 \frac{\partial m^{(r+1)}}{\partial \bar{w}^{(1)'}}. \quad (7.18)$$

The relation (7.17) enables one to integrate terms of the form $D^{(j)}m^{(k+1)}$ because they are total differentials. We prove this behavior by partial integration as follows:

$$\begin{aligned} \int D^{(j)}m^{(k+1)}dz &= 2\text{Re} \int \left(w^{(j)} \frac{\partial m^{(k+1)}}{\partial w^{(1)}} + \frac{dw^{(j)}}{dz} \frac{\partial m^{(k+1)}}{\partial w^{(1)'}} \right) dz \\ &= 2\text{Re} \int w^{(j)} \left(\frac{\partial m^{(k+1)}}{\partial w^{(1)}} - \frac{d}{dz} \frac{\partial m^{(k+1)}}{\partial w^{(1)'}} \right) dz + 2\text{Re} \left(w^{(j)} \frac{\partial m^{(k+1)}}{\partial w^{(1)'}} \right) \\ &= \frac{1}{q_0} \text{Re} \left(w^{(j)} \bar{p}^{(k)} \right). \end{aligned} \quad (7.19)$$

We shall use this result for integrating terms in the integrand of the polynomials

$$E_v^{(r+1)} = m_{ec} \int_{z_v}^z m_E^{(r+1)} dz, \quad m_E^{(r+1)} = m^{(r+1)} - D^{(r)}\mu_1^{(2)}, \quad (7.20)$$

of the perturbation eikonal (5.120).

For obtaining the path deviation of rank r , we substitute the power series expansion (7.1) for w in the left-hand side of the integral equation (5.149) and in the expression (5.146) for Q_v together with the corresponding expansion of the lateral canonical momentum

$$p(z) = \sum_{r=1}^{\infty} \varepsilon^r p^{(r)}(z). \quad (7.21)$$

In addition, we substitute in this formula the series

$$\frac{\partial E_v}{\partial a_v} = \sum_{r=2}^{\infty} \varepsilon^r \frac{\partial E_v^{(r+1)}}{\partial a_v} \quad (7.22)$$

for the derivative of the perturbation eikonal. The resulting series representation of (5.149) must be satisfied for arbitrary values of the sorting parameter ε . We can fulfill this requirement only if all terms with a fixed arbitrary rank satisfy the resulting equation separately. By performing the same procedure for Eq. (5.150), we eventually obtain

$$w^{(r)} = w_1 Q_3^{(r)} - w_3 Q_1^{(r)} + w_2 Q_4^{(r)} - w_4 Q_2^{(r)}, \quad (7.23)$$

$$p^{(r)} = p_1 Q_3^{(r)} - p_3 Q_1^{(r)} + p_2 Q_4^{(r)} - p_4 Q_2^{(r)}. \quad (7.24)$$

The functions $Q_v^{(r)}$ have the form

$$\begin{aligned} Q_v^{(r)} &= \frac{\partial E_v^{(r+1)}}{\partial a_v} - \frac{1}{q_0} \sum_{j=2}^{r-1} \operatorname{Re} \left[\bar{p}^{(j)} \frac{\partial w^{(r+1-j)}}{\partial a_v} \right] \\ &= \frac{\partial E_v^{(r+1)}}{\partial a_v} - \frac{1}{q_0} \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j, r+1}} \operatorname{Re} \left[\bar{p}^{(j)} \frac{\partial w^{(r+1-j)}}{\partial a_v} + \bar{p}^{(r+1-j)} \frac{\partial w^{(j)}}{\partial a_v} \right]. \end{aligned} \quad (7.25)$$

The lower summation limit $j_0 = [(r+1)/2]$ in the second sum denotes the integer value of the expression in the bracket; $\delta_{2j, r+1}$ is the Kronecker symbol, which is unity for $2j = r+1$ and zero else. We obtain the second relation in (7.25) by splitting the first sum in half and by exchanging the upper indices in one of the resulting sums. This exchange does not alter the sum. The Kronecker symbol arises because the first sum may consist of an odd number of terms.

7.2 Canonical Representation

We aim for a representation of $Q_v^{(r)}$ which requires path deviations of the lowest possible rank. In order to derive this representation, we utilize the relation (7.19), which allows us to integrate certain terms of the integrand of the eikonal polynomials $E_v^{(r+1)}$. We can conceive such integrations as a gauge transformation of the eikonal. In order to find the appropriate transformations, we try to establish *canonical expressions*, which are either symmetric or antisymmetric with respect to the momentum and position coordinates. Different transformations exist, which yield canonical representations for the (7.25). For our purpose, we must transform the polynomials of the perturbation eikonal as follows:

$$\begin{aligned} E_v^{(r+1)} &= \int_{z_v}^z \left[m_E^{(r+1)} - \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j, r+1}} D^{(j)} m^{(r+2-j)} \right] dz \\ &\quad + \frac{1}{q_0} \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j, r+1}} \operatorname{Re} (w^{(j)} \bar{p}^{(r+1-j)}) \\ &= \hat{E}_v^{(r+1)} + \frac{1}{q_0} \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j, r+1}} \operatorname{Re} (w^{(j)} \bar{p}^{(r+1-j)}). \end{aligned} \quad (7.26)$$

We prove the validity of this result by integrating the terms of the sum in the integrand by means of (7.19) and considering that either $w^{(j)}$ or $\bar{p}^{(r+1-j)}$ vanish at the ray-defining plane $z = z_v$. As a result, the two sums in the first relation cancel out each other. The polynomials of the modified perturbation eikonal have the form

$$\hat{E}_v^{(r+1)} = \int_{z_v}^z m_{\hat{E}}^{(r+1)} dz. \quad (7.27)$$

Employing the expression (7.20) for $m_E^{(r+1)}$, we obtain for the integrand of the modified eikonal polynomials the relation

$$m_{\hat{E}}^{(r+1)} = m_E^{(r+1)} - \sum_{j=j_0}^r \frac{1}{1 + \delta_{2j,r+1}} D^{(j)} m^{(r+2-j)}. \quad (7.28)$$

Using the relations (7.13) and (7.14), respectively, we derive the following explicit expressions up to the rank $r + 1 = 6$ for these integrands:

$$\begin{aligned} m_{\hat{E}}^{(3)} &= m_E^{(3)} = \mu_1^{(3)}, \\ m_{\hat{E}}^{(4)} &= m_E^{(4)} = \mu_1^{(4)} + \frac{1}{2} D^{(2)} \mu_1^{(3)}, \\ m_{\hat{E}}^{(5)} &= \mu_1^{(5)} + D^{(2)} \mu_1^{(4)} + \frac{1}{2} D^{(2)^2} \mu_1^{(3)}, \\ m_{\hat{E}}^{(6)} &= \mu_1^{(6)} + D^{(2)} \mu_1^{(5)} + \frac{1}{2} D^{(3)} \mu_1^{(4)} + \frac{1}{2} D^{(2)} D^{(3)} \mu_1^{(3)} + \frac{1}{2} D^{(2)^2} \mu_1^{(4)} + \frac{1}{6} D^{(2)^3} \mu_1^{(3)}. \end{aligned} \quad (7.29)$$

The corresponding polynomials of the perturbation eikonal E_v have operators $D^{(j)}$ up to the rank $j = r - 1$. Hence, in this case, we need to know all path deviations up to the rank $r - 1$ in order to calculate the path deviation of the next higher-rank r . On the other hand, the expressions (7.29) show that we need only path deviations up to the rank $j = j_0 = [(r + 1)/2]$ inclusively in order to determine the polynomial $\hat{E}_v^{(r+1)}$ of the modified perturbation eikonal. This result demonstrates convincingly that the modified perturbation eikonal is best suited for efficiently calculating the higher-rank path deviations.

By substituting relation (7.26) for the eikonal polynomial $E_v^{(r+1)}$ in (7.25), we obtain the function $Q_v^{(r)}(z)$ in the canonical form

$$Q_v^{(r)} = \frac{\partial \hat{E}_v^{(r+1)}}{\partial a_v} + \frac{1}{q_0} \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j,r+1}} \operatorname{Re} \left[w^{(j)} \frac{\partial \bar{p}^{(r+1-j)}}{\partial a_v} - \bar{p}^{(j)} \frac{\partial w^{(r+1-j)}}{\partial a_v} \right]. \quad (7.30)$$

We must put the sum equal to zero for the lowest rank $r = 2$.

7.2.1 Recurrence Formula

The representation (7.30) enables us to find a recurrence formula for the deviation functions $Q_v^{(r)}$, $v = 1, 2, 3, 4$. To derive this formula, we substitute (7.23) and (7.24) for the path and momentum deviations in the sum of (7.30). Using the abbreviations

$$A_\mu = Q_\mu^{(j)}, \quad B_\mu = \frac{\partial Q_\mu^{(r+1-j)}}{\partial a_v}, \quad (7.31)$$

we write

$$\begin{aligned} P &= \text{Re} \left[w^{(j)} \frac{\partial \bar{p}^{(r+1-j)}}{\partial a_v} - \bar{p}^{(j)} \frac{\partial w^{(r+1-j)}}{\partial a_v} \right] \\ &= \text{Re} \left[\{w_1 A_3 - w_3 A_1 + w_2 A_4 - w_4 A_2\} \{ \bar{p}_1 B_3 - \bar{p}_3 B_1 + \bar{p}_2 B_4 - \bar{p}_4 B_2 \} \right. \\ &\quad \left. - \{w_1 B_3 - w_3 B_1 + w_2 B_4 - w_4 B_2\} \{ \bar{p}_1 A_3 - \bar{p}_3 A_1 + \bar{p}_2 A_4 - \bar{p}_4 A_2 \} \right]. \end{aligned} \quad (7.32)$$

Since the coefficients (7.31) are real, the factors of the terms $A_\mu B_v - A_v B_\mu$ represent the Helmholtz–Lagrange relations (5.128), which are invariants. Considering the values (5.129) of the corresponding coefficients $C_{\lambda\kappa}$, we obtain

$$\begin{aligned} P &= q_0 [A_3 B_1 - A_1 B_3 + A_4 B_2 - A_2 B_4] \\ &= q_0 \left[Q_3^{(j)} \frac{\partial Q_1^{(r+1-j)}}{\partial a_v} - Q_1^{(j)} \frac{\partial Q_3^{(r+1-j)}}{\partial a_v} + Q_4^{(j)} \frac{\partial Q_2^{(r+1-j)}}{\partial a_v} - Q_2^{(j)} \frac{\partial Q_4^{(r+1-j)}}{\partial a_v} \right]. \end{aligned} \quad (7.33)$$

Introducing this expression into (7.30), we derive the recurrence relation

$$\begin{aligned} Q_v^{(r)} &= \frac{\partial \hat{E}_v^{(r+1)}}{\partial a_v} + \sum_{j=j_0}^{r-1} \frac{1}{1 + \delta_{2j, r+1}} \left[Q_3^{(j)} \frac{\partial Q_1^{(r+1-j)}}{\partial a_v} - Q_1^{(j)} \frac{\partial Q_3^{(r+1-j)}}{\partial a_v} \right. \\ &\quad \left. + Q_4^{(j)} \frac{\partial Q_2^{(r+1-j)}}{\partial a_v} - Q_2^{(j)} \frac{\partial Q_4^{(r+1-j)}}{\partial a_v} \right]. \end{aligned} \quad (7.34)$$

Owing to the nonlinearity of this equation, each term $Q_v^{(r)}$ is composed of products of derivatives of all eikonal polynomials $\hat{E}_\mu^{(m)}$, $\mu = 1, 2, 3, 4$, with $m \leq r$ apart from the linear term with rank $r + 1$. The number of products grows rapidly with increasing rank. We derive the expressions for the components $Q_v^{(r)}$ iteratively by starting with the lowest rank $r = 2$, giving

$$Q_v^{(2)} = \frac{\partial \hat{E}_v^{(3)}}{\partial a_v}. \quad (7.35)$$

The next step of the iteration yields the third-rank components as

$$Q_v^{(3)} = \frac{\partial \hat{E}_v^{(4)}}{\partial a_v} + \frac{1}{2} \left[\frac{\partial^2 \hat{E}_1^{(3)}}{\partial a_v \partial a_1} \frac{\partial \hat{E}_3^{(3)}}{\partial a_3} - \frac{\partial^2 \hat{E}_3^{(3)}}{\partial a_v \partial a_3} \frac{\partial \hat{E}_1^{(3)}}{\partial a_1} + \frac{\partial^2 \hat{E}_2^{(3)}}{\partial a_v \partial a_2} \frac{\partial \hat{E}_4^{(3)}}{\partial a_4} - \frac{\partial^2 \hat{E}_4^{(3)}}{\partial a_v \partial a_4} \frac{\partial \hat{E}_2^{(3)}}{\partial a_2} \right]. \quad (7.36)$$

The geometrical part of the polynomials $\hat{E}_v^{(r)}$ with odd Seidel order $n = r = 2m + 1, m = 1, 2, \dots$ is zero for systems whose electromagnetic potentials have only multipole components with even multiplicity. This is the case for rotationally symmetric systems and for orthogonal systems with plane section symmetry. Equations (7.35) and (7.36) show that for these systems the primary geometrical path deviations are of third order and given by the partial derivatives of the fourth-order polynomials of the perturbation eikonal with respect to the ray parameters. The primary chromatic path deviations of these systems, however, are of second rank.

7.2.2 Canonical Representation of the Path Deviations

If we define the ray by its lateral position and slope or momentum components at the object plane $z = z_o$, all eikonal polynomials $\hat{E}_v^{(r)}$ of the same rank coincide:

$$\hat{E}_v^{(r)} = \hat{E}_o^{(r)}, \quad v = 1, 2, 3, 4. \quad (7.37)$$

Employing our definition (5.151) of the Poisson bracket, we can then rewrite (7.36) in the concise Lie-algebraic form [118]:

$$Q_v^{(3)} = \frac{\partial \hat{E}_o^{(4)}}{\partial a_v} + \frac{1}{2} \left\{ \frac{\partial \hat{E}_o^{(3)}}{\partial a_v}, \hat{E}_o^{(3)} \right\}. \quad (7.38)$$

We can also use the Poisson bracket for obtaining simple expressions for the path deviations. In order to derive this representation for the second-rank path deviation, we introduce the relation (7.35) into the formula (7.23) and write the fundamental rays w_μ as

$$w_\mu = \frac{\partial w^{(1)}}{\partial a_\mu}, \quad \mu = 1, 2, 3, 4. \quad (7.39)$$

Substituting this expression for the fundamental rays in (7.23) gives

$$w^{(2)} = \left\{ w^{(1)}, \hat{E}_o^{(3)} \right\}. \quad (7.40)$$

We find the canonical representation for the third-rank path deviation by substituting (7.38) for $Q_v^{(3)}$ in formula (7.23). By considering that the second derivatives of $w^{(1)}$ with respect to the ray parameters vanish, we eventually obtain for third-rank path deviation the Lie-algebraic expression

$$w^{(3)} = \left\{ w^{(1)}, \hat{E}_o^{(4)} \right\} + \frac{1}{2} \left\{ \hat{E}_o^{(3)}, \left\{ \hat{E}_o^{(3)}, w^{(1)} \right\} \right\}. \quad (7.41)$$

The integrand

$$m_{\hat{E}}^{(4)} = \mu_1^{(4)} + D^{(2)} \mu_1^{(3)} / 2 \quad (7.42)$$

of the eikonal polynomial

$$\hat{E}_o^{(4)} = \int_{z_0}^z m_{\hat{E}}^{(4)} dz \quad (7.43)$$

depends on the Gaussian ray $w^{(1)}$ and on the second-rank path deviation $w^{(2)}$. By utilizing the relation (7.40), we can recast this integrand into a form which does not contain $w^{(2)}$ and its derivative $w^{(2)'}$ explicitly. To remove these quantities, we take the derivative of the expression (7.40), giving

$$\begin{aligned} w^{(2)'} &= w_1' \frac{\partial \hat{E}_o^{(3)}}{\partial a_3} - w_3' \frac{\partial \hat{E}_o^{(3)}}{\partial a_1} + w_2' \frac{\partial \hat{E}_o^{(3)}}{\partial a_4} - w_4' \frac{\partial \hat{E}_o^{(3)}}{\partial a_2} \\ &+ w_1 \frac{\partial \mu_1^{(3)}}{\partial a_3} - w_3 \frac{\partial \mu_1^{(3)}}{\partial a_1} + w_2 \frac{\partial \mu_1^{(3)}}{\partial a_4} - w_4 \frac{\partial \mu_1^{(3)}}{\partial a_2}. \end{aligned} \quad (7.44)$$

Replacing each partial derivative by

$$\frac{\partial \mu_1^{(3)}}{\partial a_v} = \frac{\partial \mu_1^{(3)}}{\partial w^{(1)}} w_v + \frac{\partial \mu_1^{(3)}}{\partial \bar{w}^{(1)}} \bar{w}_v + \frac{\partial \mu_1^{(3)}}{\partial w^{(1)'}} w_v' + \frac{\partial \mu_1^{(3)}}{\partial \bar{w}^{(1)'}} \bar{w}_v' \quad (7.45)$$

and employing the Lagrange–Helmholtz relations, we derive

$$\begin{aligned} w^{(2)'} &= \left\{ w^{(1)'}, \hat{E}_o^{(3)} \right\} + \frac{\partial \mu_1^{(3)}}{\partial \bar{w}^{(1)'}} [w_1 \bar{w}_3' - w_3 \bar{w}_1' + w_2 \bar{w}_4' - w_4 \bar{w}_2'] \\ &= \left\{ w^{(1)'}, \hat{E}_o^{(3)} \right\} - \frac{q_0}{q} \frac{\partial \mu_1^{(3)}}{\partial \bar{w}^{(1)'}} = \left\{ w^{(1)'}, \hat{E}_o^{(3)} \right\} - \sqrt{\frac{\Phi_0^*}{\Phi^*}} \frac{\partial \mu_1^{(3)}}{\partial \bar{w}^{(1)'}}, \end{aligned} \quad (7.46)$$

where $q = \sqrt{2m_e e \Phi^*(z)}$ and q_0 are the momentum along the optic axis (5.20) and the normalization momentum (2.43), respectively. Considering this relation together with (7.45) and (7.40), we eventually find

$$\begin{aligned}
D^{(2)}\mu_1^{(3)} &= w^{(2)}\frac{\partial\mu_1^{(3)}}{\partial w^{(1)}} + \bar{w}^{(2)}\frac{\partial\mu_1^{(3)}}{\partial\bar{w}^{(1)}} + w^{(2)'}\frac{\partial\mu_1^{(3)}}{\partial w^{(1)'}} + \bar{w}^{(2)'}\frac{\partial\mu_1^{(3)}}{\partial\bar{w}^{(1)'}} \\
&= \frac{\partial\mu_1^{(3)}}{\partial a_1}\frac{\partial\hat{E}_o^{(3)}}{\partial a_3} - \frac{\partial\mu_1^{(3)}}{\partial a_3}\frac{\partial\hat{E}_o^{(3)}}{\partial a_1} + \frac{\partial\mu_1^{(3)}}{\partial a_2}\frac{\partial\hat{E}_o^{(3)}}{\partial a_4} - \frac{\partial\mu_1^{(3)}}{\partial a_4}\frac{\partial\hat{E}_o^{(3)}}{\partial a_2} - \frac{4q_0}{q}\left|\frac{\partial\mu_1^{(3)}}{\partial w^{(1)'}}\right|^2 \\
&= \left\{\mu_1^{(3)}, \hat{E}_o^{(3)}\right\} - \sqrt{\frac{\Phi_0^*}{\Phi^*}}\left|\frac{\partial\mu_1^{(3)}}{\partial w^{(1)'}}\right|^2. \tag{7.47}
\end{aligned}$$

Substituting this expression for $D^{(2)}\mu_1^{(3)}$ in the integrand (7.42), we obtain the fourth-rank eikonal polynomial (7.43) of the modified perturbation eikonal in the canonical form

$$\hat{E}_o^{(4)} = \int_{z_0}^z \left[\mu_1^{(4)} - 2\sqrt{\frac{\Phi_0^*}{\Phi^*}}\left|\frac{\partial\mu_1^{(3)}}{\partial w^{(1)'}}\right|^2 + \frac{1}{2}\left\{\mu_1^{(3)}, \hat{E}_o^{(3)}\right\} \right] dz. \tag{7.48}$$

The first two terms of the integrand furnish the contributions to the local third-rank deviation, whereas the Poisson bracket produces the combination deviations, which originate from the concatenation of second-rank deviations at separate planes. The combination deviations grow rapidly with increasing distance between adjacent focusing elements. The geometrical eikonal polynomials $\hat{E}_o^{(6)}$ of rotationally symmetric systems and of systems with two orthogonal symmetry sections have the same structure as (7.47). Since all eikonal polynomials of odd Seidel order are zero for such systems, the second iteration step yields the fifth-order path deviation and the sixth-order eikonal polynomial, respectively.

We readily obtain this polynomial by replacing in formula (7.47) the upper index 4 by 6 and the index 3 by 4. The representation (7.47) does not involve the second-rank path deviation. However, the form (7.42) of the eikonal integrand has the advantage to elucidate much better the symmetry properties of the integrand. We shall utilize this behavior for illustrating the properties of the hexapole corrector, which compensates for the unavoidable third-order spherical aberration of round lenses.

7.3 Expansion Polynomials of the Variational Function

To obtain analytical expressions for the integrands of the eikonal polynomials, we must determine the polynomials $\mu^{(r)} = \mu_g^{(r)} + \mu_c^{(r)}$ of rank $r \geq 3$ of the variational function (3.59). We assume the most general case of a system with curved axis with complex curvature Γ defined by (3.6). To derive the geometrical part $\mu_g^{(r)}$ of the polynomials of the third and the fourth rank, it suffices to insert the power series expansion (3.54) for the electric potential φ and the corresponding expansions

(3.92) and (3.93) for the components of the magnetic vector potential into the terms (3.60) and (3.61) of the variational function. We obtain the chromatic terms $\mu_c^{(r)}$ by expanding the expression (4.6) for the electric part of the variational function in a power series with respect to $w, \bar{w}, w', \bar{w}'$ and the chromatic parameter κ (4.2). Since the higher-rank polynomials become excessively large for arbitrary systems with curved axis, we only list the geometric and chromatic terms of the polynomials $\mu^{(3)}$ and $\mu^{(4)}$ without restating their very laborious yet straightforward derivation [52]. The third-rank geometric and chromatic terms have the form

$$\begin{aligned} \mu_g^{(3)} = & \sqrt{\frac{\Phi^*}{\Phi_0^*}} \text{Re} \left\{ \left[\frac{\gamma_0}{4} \frac{\Phi_1}{\Phi^*} + \frac{\Gamma}{2} \right] w' \bar{w}' \bar{w} + \frac{ie}{8q} [\Psi_1' - B\Gamma] (2\bar{w}' w \bar{w} - w' \bar{w}^2) \right. \\ & + \left[\frac{\gamma_0}{2} \frac{\Phi_3}{\Phi^*} - \frac{\Phi_1 \Phi_2}{8\Phi^{*2}} - \frac{\gamma_0}{4} \frac{\Phi_2}{\Phi^*} \Gamma + \frac{\gamma_0}{64} \frac{\Phi_1^3}{\Phi^{*3}} + \frac{\Gamma}{32} \frac{\Phi_1^2}{\Phi^{*2}} + \frac{ie}{q} \left(\Psi_3 - \frac{\Gamma}{3} \Psi_2 \right) \right] \bar{w}^3 \\ & - \left[\frac{\Phi_2 \bar{\Phi}_1}{8\Phi^{*2}} + \frac{\gamma_0}{8} \frac{\Phi_2}{\Phi^*} \Gamma + \frac{\gamma_0}{16} \frac{\Phi''_1 + \Phi' \Gamma'}{\Phi^*} - \frac{\Phi_1 \Phi''}{16\Phi^{*2}} + \frac{\gamma_0}{32} \frac{\Phi''}{\Phi^*} \Gamma - \frac{3\gamma_0}{64} \frac{\Phi_1^2 \bar{\Phi}_1}{\Phi^{*3}} \right. \\ & \left. \left. - \frac{\Gamma}{32} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} + \frac{\gamma_0 \Gamma}{32} \text{Re} \left(\frac{\Phi_1}{\Phi^*} \Gamma \right) + \frac{ie}{q} \left(\frac{\bar{\Gamma}}{4} \Psi_2 - \frac{\Gamma}{16} \text{Re}(\Psi_1 \bar{\Gamma}) - \frac{1}{16} B \bar{\Gamma} \right) \right] w \bar{w}^2 \right\}, \end{aligned} \quad (7.49)$$

$$\begin{aligned} \mu_c^{(3)} = & \frac{\Phi_o}{\sqrt{\Phi^* \Phi_0^*}} \kappa \text{Re} \left\{ \frac{\gamma_0}{4} w' \bar{w}' - \left[\frac{1}{4} \frac{\Phi_2}{\Phi^*} - \frac{\Gamma}{8} \frac{\Phi_1}{\Phi^*} - \frac{3\gamma_0}{32} \frac{\Phi_1^2}{\Phi^{*2}} \right] \bar{w}^2 \right. \\ & + \left[\frac{3\gamma_0}{32} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} + \frac{1}{16} \frac{\Phi_1}{\Phi^*} \bar{\Gamma} + \frac{1}{16} \frac{\Phi''}{\Phi^*} \right] w \bar{w} \\ & \left. + \frac{\Phi_o}{\Phi^*} \left[\frac{3\gamma_0}{16} \frac{\Phi_1}{\Phi^*} + \frac{1}{8} \Gamma \right] w \kappa \right\}. \end{aligned} \quad (7.50)$$

In the presence of dipole fields, the paraxial ray

$$w^{(1)} = w_g^{(1)} + w_c^{(1)}, \quad w_g^{(1)} = \sum_{\mu=1}^4 a_\mu w_\mu(z), \quad w_c^{(1)} = \kappa w_\kappa \quad (7.51)$$

is composed of the geometrical component $w_g^{(1)}$ and the chromatic component (dispersion) $w_c^{(1)}$, which is proportional to the chromatic parameter κ . If we substitute $w^{(1)}$ for w in the expression (7.49), we also obtain chromatic terms contributing to the chromatic deviations. These chromatic effects occur because particles with different energies follow different trajectories. Particles, which travel initially along the optic axis, are deflected from this axis if their energy differs from the nominal energy E_0 . Owing to this displacement, these particles will experience a force if they pass through subsequent quadrupole and/or hexapole elements centered about the optic axis. Hence, dipole fields produce a coupling between the geometric action of these multipole elements and the dispersion. In this case, the optic axis is

curved apart from the Wien filter (4.15). Within this filter, the deflection resulting from the electric dipole field counterbalances the deflection of the magnetic dipole field:

$$\Gamma = \frac{\gamma_0}{2} \frac{\Phi_1}{\Phi^*} + i \sqrt{\frac{e}{2m_e \Phi^*}} \Psi_1 = 0. \quad (7.52)$$

The geometric third-rank polynomial (7.49) vanishes in the absence of dipole and hexapole fields $\Phi_1 = \Psi_1 = 0$, $\Phi_3 = \Psi_3 = 0$. This is not the case for the chromatic term (7.50), which reduces to

$$\mu_c^{(3)} = \frac{q}{q_0} \frac{\Phi_o}{\Phi^*} \frac{\kappa}{4} \text{Re} \left\{ \gamma_0 w' \bar{w}' - \frac{\Phi_2}{\Phi^*} \bar{w}^2 + \frac{1}{4} \frac{\Phi''}{\Phi^*} w \bar{w} \right\}. \quad (7.53)$$

It readily follows from this expression that $\mu_c^{(3)} > 0$ for magnetic systems ($\Phi_2 = \Phi'' = 0$) with a straight optic axis. According to this behavior, the chromatic aberration is unavoidable at the image plane of any magnetic system with a straight optic axis. We will demonstrate this behavior in the context of the Scherzer theorem.

In order to obtain the third-rank path deviations, we need the third- and the fourth-rank variational polynomials. We find the geometric part of the fourth-rank variational polynomial as

$$\begin{aligned} \mu_g^{(4)} = & \frac{q}{q_0} \text{Re} \left\{ -\frac{1}{8} w'^2 \bar{w}'^2 + \left[\frac{\gamma_0}{4} \frac{\Phi_2}{\Phi^*} - \frac{1}{32} \frac{\Phi_1^2}{\Phi^{*2}} + \frac{\gamma_0}{8} \frac{\Phi_1}{\Phi^*} \Gamma + \frac{1}{4} \Gamma^2 \right] w' \bar{w}' \bar{w}^2 \right. \\ & - \left[\frac{1}{2} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} - \frac{3\gamma_0}{16} \frac{\Phi_1}{\Phi^*} \Gamma - \frac{1}{4} \Gamma \bar{\Gamma} + \frac{1}{16} \frac{\Phi''}{\Phi^*} \right] w' \bar{w}' w \bar{w} + \frac{ie}{4q} [\Psi_2' + \Psi_1' \Gamma - B \Gamma^2] \bar{w}' w \bar{w}^2 \\ & - \frac{ie}{12q} \Psi_2' w' \bar{w}^3 + \frac{ie}{16q} \left[\text{Re} \left(3 \bar{\Gamma} \Psi_1' + \Psi_1' \bar{\Gamma}' \right) - 2B \Gamma \bar{\Gamma} + B'' \right] \bar{w}' w^2 \bar{w} + \left[\frac{\gamma_0}{2} \frac{\Phi_4}{\Phi^*} - \frac{\Phi_3 \Phi_1}{8 \Phi^{*2}} - \frac{\gamma_0}{4} \frac{\Phi_3}{\Phi^*} \Gamma \right. \\ & - \frac{1}{16} \frac{\Phi_2^2}{\Phi^{*2}} + \frac{3\gamma_0}{64} \frac{\Phi_2 \Phi_1^2}{\Phi^{*3}} + \frac{1}{16} \frac{\Phi_2 \Phi_1}{\Phi^{*2}} \Gamma - \frac{1+4\gamma_0^2}{1024} \frac{\Phi_1^4}{\Phi^{*4}} - \frac{\gamma_0}{128} \frac{\Phi_1^3}{\Phi^{*3}} \Gamma + \frac{ie}{q} \left(\Psi_4 - \frac{3}{8} \Psi_3 \Gamma \right) \left. \right] \bar{w}^4 \\ & + \left[\frac{1}{16} \frac{\Phi_2 \Phi''}{\Phi^{*2}} - \frac{1}{8} \frac{\Phi_3 \bar{\Phi}_1}{\Phi^{*2}} - \frac{\gamma_0}{8} \frac{\Phi_3}{\Phi^*} \bar{\Gamma} - \frac{\gamma_0}{24} \frac{\Phi_2''}{\Phi^*} + \frac{3\gamma_0}{32} \frac{\Phi_2 \Phi_1 \bar{\Phi}_1}{\Phi^{*3}} + \frac{1}{32} \frac{\Phi_2 \bar{\Phi}_1}{\Phi^{*2}} \Gamma - \frac{\gamma_0}{96} \frac{\Phi_2}{\Phi^*} \Gamma \bar{\Gamma} \right. \\ & + \frac{1}{64} \frac{\Phi_1' \Phi_1}{\Phi^{*2}} - \frac{\gamma_0}{64} \frac{\Phi_1'}{\Phi^*} \Gamma - \frac{\gamma_0}{48} \frac{\Phi_1'}{\Phi^*} \Gamma' - \frac{3\gamma_0}{128} \frac{\Phi_1^2 \Phi''}{\Phi^{*3}} + \frac{1}{128} \frac{\Phi_1 \Phi''}{\Phi^{*2}} \Gamma - \frac{\gamma_0}{128} \frac{\Phi''}{\Phi^*} \Gamma^2 + \frac{1}{64} \frac{\Phi_1 \Phi'}{\Phi^{*2}} \Gamma' \\ & - \frac{7\gamma_0}{192} \frac{\Phi'}{\Phi^*} \Gamma' \Gamma - \frac{1+4\gamma_0^2}{256} \frac{\Phi_1^2 \bar{\Phi}_1}{\Phi^{*4}} + \frac{\gamma_0}{256} \frac{\Phi_1^3}{\Phi^{*3}} \bar{\Gamma} - \frac{3\gamma_0}{256} \frac{\Phi_1^2 \bar{\Phi}_1}{\Phi^{*3}} \Gamma + \frac{1}{128} \left(\frac{\Phi_1}{\Phi^*} - \gamma_0 \Gamma \right) \Gamma \text{Re} \left(\frac{\Phi_1}{\Phi^*} \bar{\Gamma} \right) \\ & - \frac{ie}{q} \left(\frac{\bar{\Gamma}}{4} \Psi_3 - \frac{1}{48} \Psi_2 \Gamma \bar{\Gamma} - \frac{1}{32} \Gamma \Psi_1'' - \frac{1}{24} \Gamma' \Psi_1' + \frac{7}{96} B \Gamma' \Gamma + \frac{1}{64} B' \Gamma^2 - \frac{1}{64} \Gamma^2 \text{Re}(\Psi_1 \bar{\Gamma}) \right) \left. \right] w \bar{w}^3 \\ & + \left[\frac{3\gamma_0}{64} \frac{\Phi_2 \bar{\Phi}_1^2}{\Phi^{*3}} - \frac{1}{16} \frac{\Phi_2 \bar{\Phi}_2}{\Phi^{*2}} + \frac{1}{32} \frac{\Phi_2 \bar{\Phi}_1}{\Phi^{*2}} \bar{\Gamma} - \frac{\gamma_0}{64} \frac{\Phi_2}{\Phi^*} \bar{\Gamma}^2 + \frac{1}{64} \frac{\Phi_1' \bar{\Phi}_1}{\Phi^{*2}} - \frac{\gamma_0}{64} \frac{\Phi_1'}{\Phi^*} \bar{\Gamma} - \frac{\gamma_0}{32} \frac{\Phi_1'}{\Phi^*} \Gamma' \right. \\ & - \frac{\gamma_0}{128} \frac{\Phi_1}{\Phi^*} \bar{\Gamma}'' + \frac{1}{64} \frac{\Phi_1 \Phi'}{\Phi^{*2}} \Gamma' - \frac{3\gamma_0}{128} \frac{\Phi'}{\Phi^*} \Gamma' \bar{\Gamma} + \frac{3}{128} \frac{\Phi_1 \Phi''}{\Phi^{*2}} \bar{\Gamma} - \frac{3\gamma_0}{128} \frac{\Phi_1 \bar{\Phi}_1 \Phi''}{\Phi^{*3}} \\ & - \frac{3(1+4\gamma_0^2)}{1024} \frac{\Phi_1^2 \bar{\Phi}_1^2}{\Phi^{*4}} + \frac{\gamma_0}{256} \frac{\Phi''}{\Phi^*} \Gamma \bar{\Gamma} - \frac{3\gamma_0}{256} \frac{\Phi_1}{\Phi^*} \Gamma \bar{\Gamma}^2 + \frac{\gamma_0}{128} \frac{\Phi'''}{\Phi^*} - \frac{1}{128} \frac{\Phi'^2}{\Phi^{*2}} \\ & \left. - \frac{ie}{q} \left(\frac{1}{32} \Psi_2 \bar{\Gamma}^2 - \frac{1}{64} \bar{\Gamma} \Psi_1' + \frac{1}{64} B \Gamma' \bar{\Gamma} \right) \right] w^2 \bar{w}^2 \left. \right\}. \quad (7.54) \end{aligned}$$

This expression is valid regardless if the optic axis represents a possible ray or not. In most cases, however, one requires that the optic axis is forming a ray. Then, we must substitute expression (4.14) for the complex curvature Γ in the formulas (7.49), (7.50), and (7.54). The lengthy formula (7.54) reduces considerably for systems with a straight optic axis such as round lenses and orthogonal systems composed of multipoles with even multiplicity $2m$ and mutual plane principal sections ($\Phi_{2m} = \bar{\Phi}_{2m} = \Phi_{2m,c}$, $\Psi_{2m} = -\bar{\Psi}_{2m} = i\Psi_{2m,s}$).

The chromatic part of the fourth-rank variational polynomial $\mu^{(4)}$ has the form

$$\begin{aligned} \mu_c^{(4)} = \frac{q}{q_0} \kappa \text{Re} \left\{ \left[\frac{\gamma_0}{4} \Gamma - \frac{1}{8} \frac{\Phi_1}{\Phi^*} \right] w' \bar{w}' \bar{w} - \left[\frac{1}{4} \frac{\Phi_3}{\Phi^*} - \frac{3\gamma_0}{16} \frac{\Phi_2 \Phi_1}{\Phi^{*2}} - \frac{1}{8} \frac{\Phi_2}{\Phi^*} \Gamma \right. \right. \\ + \frac{1 + 4\gamma_0^2}{128} \frac{\Phi_1^3}{\Phi^{*3}} + \frac{3\gamma_0}{64} \frac{\Phi_1^2}{\Phi^{*2}} \Gamma \left. \right] \bar{w}^3 + \left[\frac{3\gamma_0}{16} \frac{\Phi_2 \bar{\Phi}_1}{\Phi^{*2}} + \frac{1}{16} \frac{\Phi_2}{\Phi^*} \bar{\Gamma} + \frac{1}{32} \frac{\Phi_1''}{\Phi^*} + \frac{1}{32} \frac{\Phi'}{\Phi^*} \Gamma' \right. \\ - \frac{3\gamma_0}{32} \frac{\Phi_1 \Phi''}{\Phi^{*2}} + \frac{1}{64} \frac{\Phi''}{\Phi^*} \Gamma - \frac{3(1 + 4\gamma_0^2)}{128} \frac{\Phi_1^2 \bar{\Phi}_1}{\Phi^{*3}} - \frac{3\gamma_0}{64} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} \Gamma + \frac{\Gamma}{64} \text{Re} \left(\frac{\Phi_1}{\Phi^*} \bar{\Gamma} \right) \left. \right] w \bar{w}^2 \\ - \frac{1}{16} \kappa w' \bar{w}' + \left[\frac{3\gamma_0}{16} \frac{\Phi_2}{\Phi^*} - \frac{3(1 + 4\gamma_0^2)}{128} \frac{\Phi_1^2}{\Phi^{*2}} - \frac{3\gamma_0}{32} \frac{\Phi_1}{\Phi^*} \Gamma \right] \kappa \bar{w}^2 \\ \left. - \left[\frac{3(1 + 4\gamma_0^2)}{128} \frac{\Phi_1 \bar{\Phi}_1}{\Phi^{*2}} + \frac{3\gamma_0}{64} \frac{\Phi_1}{\Phi^*} \bar{\Gamma} + \frac{3\gamma_0}{64} \frac{\Phi''}{\Phi^*} \right] \kappa w \bar{w} \right\}. \end{aligned} \quad (7.55)$$

This term is of importance only at planes where the second-rank chromatic deviation is zero such as the image plane of an electron microscope corrected for primary chromatic and geometric aberrations.

7.4 Path Equation Approach

We can also obtain the path deviations by means of an alternative iteration procedure [52, 119]. This method starts from the inhomogeneous path equation (4.231). To obtain the perturbation function P , we start from the Euler–Lagrange equation (4.12) and write the expansion of the variational function (4.1) as

$$\mu = \mu^{(0)} + \mu^{(1)} + \mu^{(2)} + \Delta\mu, \quad \Delta\mu = \sum_{r=2}^{\infty} \mu^{(r+1)}. \quad (7.56)$$

Using this separation and introducing transformed coordinates (4.28) in the expressions (4.3), (4.9), and (4.10), we may write Eq. (4.12) in the form

$$U'' + T U - G \bar{U} - \kappa \left(\frac{\Phi^*}{\Phi_o^*} \right)^{1/4} D = P, \quad (7.57)$$

where the perturbation function is given by

$$P = 2 \frac{q_0}{q_o} \left(\frac{\Phi_o^*}{\Phi^*} \right)^{1/4} \left(\frac{\partial \Delta \mu}{\partial \bar{w}} - \frac{d}{dz} \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) e^{-i\chi}. \quad (7.58)$$

To simplify the calculations, we put in the following the normalization potential Φ_0 equal to the axial potential at the object plane, $\Phi_0 = \Phi(z_0) = \Phi_o$, resulting in $q_0 = q_o$. We transform the inhomogeneous equation (7.57) in an integral equation by means of the procedure outlined in Sect. 4.4.1, giving

$$U = \hat{U}^{(1)} + U_P, \quad (7.59)$$

$$\hat{U}^{(1)} = \sum_{\mu=1}^4 \hat{a}_\mu U_\mu + \kappa U_\kappa, \quad (7.60)$$

$$U_P = -\frac{4}{D_W} \sum_{\mu=1}^4 (-1)^\mu U_\mu \sum_{(p)} C_{\sigma\tau} \operatorname{Re} \int_{z_v}^z P \bar{U}_v dz, \quad v < \sigma < \tau. \quad (7.61)$$

We must perform the summation (p) in the second sum over the three permutations of the indices v, σ, τ , which differ from the index μ . The ray parameters $\hat{a}_\mu$ differ from the parameters a_μ listed in (5.123) and (5.127) if we fix the ray by its position and lateral canonical momentum at the object plane. In order to demonstrate this behavior, we rewrite the inhomogeneous part (7.61) of the integral equation (7.59) containing all nonlinear terms. Using the expression (7.58), employing (4.191), and replacing the transformed fundamental rays U_μ by the fundamental rays w_μ , we obtain by partial integration

$$\begin{aligned} \int_{z_v}^z P \bar{U}_v dz &= 2 \int_{z_v}^z \left(\frac{\partial \Delta \mu}{\partial \bar{w}} - \frac{d}{dz} \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) \bar{w}_v dz \\ &= -2 \bar{w}_v \frac{\partial \Delta \mu}{\partial \bar{w}'} \Big|_{z_v}^z + 2 \int_{z_v}^z \left(\frac{\partial \Delta \mu}{\partial \bar{w}} \bar{w}_v + \frac{\partial \Delta \mu}{\partial \bar{w}'} \bar{w}_v' \right) dz. \end{aligned} \quad (7.62)$$

We substitute (7.62) for the integral in (7.61). By considering the third and fifth relation in (5.137), we find that the terms derived from partial integration contribute only at their lower limits $z = z_v$ to the sums. The sums taken with the terms at the upper limit cancel out. Expression (7.61) simplifies considerably by choosing the fundamental rays in such a way that the constants $C_{\sigma\tau}$ of the Helmholtz–Lagrange relations adopt the form (5.119) resulting in

$$U_P = U_P^{(1)} + \Delta U, \quad (7.63)$$

$$\begin{aligned}
U_P^{(1)} &= \frac{1}{q_o} [U_1 \text{Re}(\bar{w}_3 \Delta p)_{z_3} - U_3 \text{Re}(\bar{w}_1 \Delta p)_{z_1} + U_2 \text{Re}(\bar{w}_4 \Delta p)_{z_4} - U_4 \text{Re}(\bar{w}_2 \Delta p)_{z_2}], \\
\Delta U &= 2 \left[U_1 \text{Re} \int_{z_3}^z \left(\bar{w}_3 \frac{\partial \Delta \mu}{\partial \bar{w}} + \bar{w}'_3 \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) dz - U_3 \text{Re} \int_{z_1}^z \left(\bar{w}_1 \frac{\partial \Delta \mu}{\partial \bar{w}} + \bar{w}'_1 \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) dz \right] \\
&\quad + 2 \left[U_2 \text{Re} \int_{z_4}^z \left(\bar{w}_4 \frac{\partial \Delta \mu}{\partial \bar{w}} + \bar{w}'_4 \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) dz - U_4 \text{Re} \int_{z_2}^z \left(\bar{w}_2 \frac{\partial \Delta \mu}{\partial \bar{w}} + \bar{w}'_2 \frac{\partial \Delta \mu}{\partial \bar{w}'} \right) dz \right].
\end{aligned} \tag{7.64}$$

The quantity

$$\Delta p = p - p^{(1)} = 2q_o \frac{\partial \Delta \mu}{\partial \bar{w}'} \tag{7.65}$$

represents the difference between the true lateral canonical momentum and its paraxial approximation. The part $U_P^{(1)}$ is linear in the fundamental rays, and hence, we can add it to the paraxial approximation (7.60). If we fix the ray by its lateral positions at the object plane $z_1 = z_2 = z_o$ and the aperture plane $z_3 = z_4 = z_a$, the path deviation (7.63) must vanish at these planes. We satisfy this requirement by imposing the additional condition (5.121) on the fundamental rays. In most cases, one defines the ray by its position w_o and lateral canonical momentum p_o at the object plane so that $z_1 = z_2 = z_3 = z_4 = z_o = z_a$. In this case, $U_{in}^{(1)}$ does not vanish and must be added to (7.60), resulting in

$$U^{(1)} = \hat{U}^{(1)} + U_P^{(1)} = \sum_{\mu=1}^4 a_\mu U_\mu + \kappa U_\kappa, \tag{7.66}$$

where the coefficients a_μ coincide with the ray parameters (5.123) and (5.124) obtained from the variation of the eikonal outlined in Sect. 5.6.2.

7.4.1 Primary Deviations

We derive the *primary* deviation ΔU_1 by inserting the paraxial approximation (7.66) into the integrands of expression (7.64) for the path deviation. This procedure gives contributions to all deviations of rank $r \geq 2$ and neglects deviations resulting from the combination of lower-rank deviations originating at separate planes along the optic axis. Within the frame of validity of this approximation, we can considerably simplify the integrands in (7.64) by utilizing the differential relation

$$w_\mu \frac{\partial}{\partial w^{(1)}} + \bar{w}_\mu \frac{\partial}{\partial \bar{w}^{(1)}} + w'_\mu \frac{\partial}{\partial w^{(1)'}} + \bar{w}'_\mu \frac{\partial}{\partial \bar{w}^{(1)'}} = 2\text{Re} \left(\bar{w}_\mu \frac{\partial}{\partial \bar{w}^{(1)}} + \bar{w}'_\mu \frac{\partial}{\partial \bar{w}^{(1)'}} \right) = \frac{\partial}{\partial a_\mu}. \tag{7.67}$$

Employing this relation and substituting $\Delta\mu_1 = \Delta\mu(w^{(1)}, \bar{w}^{(1)}, w^{(1)'}, \bar{w}^{(1)'}; z)$ for $\Delta\mu$ in (7.64), we readily derive the primary deviation as

$$\Delta U_1 = U_1 \int_{z_3}^z \frac{\partial \Delta\mu_1}{\partial a_3} dz - U_3 \int_{z_1}^z \frac{\partial \Delta\mu_1}{\partial a_1} dz + U_2 \int_{z_4}^z \frac{\partial \Delta\mu_1}{\partial a_4} dz - U_4 \int_{z_2}^z \frac{\partial \Delta\mu_1}{\partial a_2} dz. \quad (7.68)$$

This formula yields correctly the deviations $U^{(2)} = U^{(2)}(z)$ of lowest rank $r = 2$ if we ignore in the expression (7.56) for $\Delta\mu$ the terms $\mu^{(r+1)}$ with $r > 2$. By substituting $\mu^{(3)}$ for $\Delta\mu_1$ in (7.68) and replacing the transformed fundamental rays U_v by w_v (4.191), we obtain the second-rank path deviation in the eikonal representation

$$w^{(2)} = w_1 \frac{\partial E_3^{(3)}}{\partial a_3} - w_3 \frac{\partial E_1^{(3)}}{\partial a_1} + w_2 \frac{\partial E_4^{(3)}}{\partial a_4} - w_4 \frac{\partial E_2^{(3)}}{\partial a_2}. \quad (7.69)$$

Considering the relation $E_v^{(3)} = \hat{E}_v^{(3)}$ and fixing the ray by its initial position and lateral canonical momentum at the object plane, $E_v^{(3)} = \hat{E}_o^{(3)}$, the second-rank path deviation adopts the canonical form (7.40).

7.5 Second-Rank Path Deviations of Systems with Midsection Symmetry

To reduce the number of terms contributing to the eikonal polynomials, one imposes symmetry conditions on the electromagnetic potentials. The higher the degree of symmetry, the lower is the number of constituent monomials of each eikonal polynomial regardless of its rank. By introducing a plane section of symmetry embedding the curved optic axis, we cut the number of monomials in half. Most systems with a curved axis exhibit midsection symmetry regarding the arrangement of the pole pieces and/or the electrodes. In order that the optic axis lies on the horizontal midsection ($y = 0$), the electromagnetic potentials must satisfy the conditions $\varphi(x, -y, z) = \varphi(x, y, z)$, $\psi(x, -y, z) = -\psi(x, y, z)$, as outlined in Sect. 4.7.2. In order to satisfy these conditions, we must arrange all multipoles in such a way that their skew components vanish:

$$\Phi_{\nu s} = \Psi_{\nu c} = 0, \quad B = 0 \quad \rightarrow \quad \Phi_\nu = \Phi_{\nu c}, \quad \Psi_\nu = i\Psi_{\nu s}, \quad \nu \geq 1. \quad (7.70)$$

Most accelerators and beam-guiding systems fulfill these requirements as well as imaging energy filters and monochromators employed in analytical electron microscopes.

7.5.1 Wien Filter

The dipole components of the electromagnetic field produce dispersion and a curvature of the optic axis with the exception of the Wien filter. For this filter, the curvature produced by the electric dipole field cancels out the opposite curvature resulting from the magnetic dipole field for electrons with nominal energy. The Wien filter is a very versatile electron optical element. Special types have been proposed which act as mass separator [120], spectrometer [121, 122], monochromator [123–125], imaging energy filter [126], and as a corrector compensating for the chromatic aberration of round lenses [77]. In the ideal case, the Wien condition (4.15) is fulfilled at any plane along the optic axis. By choosing $\Phi_1 = \Phi_{1c}$, $\Psi_1 = i\Psi_{1s}$, we rewrite this condition in the form

$$\Phi_1 + i\nu_0\Psi_1 = \Phi_{1c} - \frac{q}{\gamma_0 m_e}\Psi_{1s} = 0. \quad (7.71)$$

Moreover, we allow for a superposition of regular electric and magnetic multipoles as defined by (7.71), with the exception of round-lens fields ($\Phi' = 0$, $B = 0$). Since the higher-order multipoles produce an inhomogeneous field, we define the resulting filter as an *inhomogeneous Wien filter*. Considering the conditions (7.70) and (7.71), we derive the paraxial path equation for this filter from the general equation (4.13) as

$$w'' + \frac{1}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} w + \left(\frac{1}{8} \frac{\Phi_{1c}^2}{\Phi^{*2}} - \gamma_0 \frac{\Phi_{2c}}{\Phi^*} + \frac{q}{m_e \Phi^*} \Psi_{2s} \right) \bar{w} = -\frac{\kappa}{4} \frac{\Phi_o \Phi_{1c}}{\Phi^{*2}}. \quad (7.72)$$

This equation reveals that the homogeneous Wien filter ($\Phi_{2c} = \Psi_{2s} = 0$) has the combined action of a cylindrical lens focusing in the horizontal section plus a straight-vision prism. The prism deflects electrons in the horizontal direction whose velocities differ from the nominal velocity (4.5). The horizontal focusing action of the cylinder lens refracts the dispersion ray toward the optic axis, thus preventing a large angular dispersion. We can avoid this effect by superposing electric and/or quadrupole fields with strengths

$$\gamma_0 \Phi_{2c} - \frac{q}{m_e} \Psi_{2s} = \frac{1}{4} \frac{\Phi_{1c}^2}{\Phi^*} \quad (7.73)$$

onto the dipole fields so that the resulting inhomogeneous Wien filter focuses in the vertical y -direction. The superposition (7.73) only changes the orientation of the cylinder lens but not its strength. The dispersion stays in the horizontal section. We readily obtain the dispersion ray by twofold integration of the equation

$$x'' = -\frac{\kappa}{4} \frac{\Phi_o}{\Phi^{*2}} \Phi_{1c}(z), \quad (7.74)$$

giving

$$x_K = -\frac{\Phi_o}{4\Phi^{*2}} \int_{z_o}^z \left\{ \int_{z_o}^z \Phi_{1c}(\zeta) d\zeta \right\} dz = -\frac{\Phi_o}{4\Phi^{*2}} \left(z \int_{z_o}^z \Phi_{1c} dz - \int_{z_o}^z z \Phi_{1c} dz \right). \quad (7.75)$$

Assuming a homogeneous dipole field within the Wien filter, it readily follows from the last relation that the dispersion increases quadratic with the distance z inside the filter. The axial electric potential Φ within the Wien filter may be different from the potential Φ_o at the object plane z_o located in front of the filter. Since the dispersion of the Wien filter is rather small for voltages above about 20 kV, one reduces the velocity within the filter by placing it between a retarding and an accelerating electrostatic immersion lens.

The cylindrical lens action of the standard Wien filter prevents its direct use in analytical electron microscopes as stigmatic imaging energy filter or monochromator, respectively. For this purpose, we must superimpose quadrupole fields, which transform the cylinder lens action into that of a round lens. This happens if the third term on the left of (7.72) vanishes. In this case, the quadrupole strengths Φ_{2c} and Ψ_{2s} satisfy the anastigmatism condition (4.18), which we rewrite as

$$\gamma_o \Phi_{2c} - \frac{q}{m_e} \Psi_{2s} = \gamma_o \Phi_{2c} - \frac{2e}{q} \Phi^* \Psi_{2s} = \frac{1}{8} \frac{\Phi_{1c}^2}{\Phi^*}. \quad (7.76)$$

The conditions (7.71), (7.73), and (7.76) are most difficult to fulfill in the fringing field regions. We obtain the best fit by using an electric and magnetic dodecapole element, allowing independent excitation of the dipole, quadrupole, and hexapole components of the electromagnetic field.

To derive an explicit analytical expression for the second-rank path deviation (7.69) produced by the Wien filter, it suffices to determine the eikonal polynomial

$$E_o^{(3)} = \int_{z_o}^z \left[\mu_{g1}^{(3)}(z) + \mu_{c1}^{(3)}(z) \right] dz. \quad (7.77)$$

We obtain the analytical expression for the integrand from the relation (7.70) by substituting $\Phi_1 = \Phi_{1c}$ from (7.71) for $\Psi_1 = i\Psi_{1s}$ in (7.49) and (7.50) and putting $\Gamma = 0$. The geometrical part $\mu_{g1}^{(3)}$ of the integrand has two terms, which contain derivatives of the electric and magnetic dipole strengths. In order to easily evaluate the integrals, we eliminate the derivatives Ψ'_1 and Φ'_1 of the dipole strengths by partial integrations and suppose that the dipole fields vanish at the object plane ($\Psi_{1s}(z_o) = 0, \Phi_{1c}(z_o) = \Phi'_{1c}(z_o) = 0$). The partial integrations create terms with $w^{(1)''}$ and $\bar{w}^{(1)''}$, which we remove by means of the paraxial path equation (7.72). We employ the same procedure for transforming the first term in the expression (7.53) for the chromatic part $\mu_{c1}^{(3)}$. The straightforward calculations yield

$$E_o^{(3)} = \sqrt{\frac{\Phi^*}{\Phi_o^*}} \frac{\gamma_o}{8\Phi} \text{Re} \left(\left[\frac{\Phi_{1c}}{\Phi^*} w^{(1)'} - \frac{\Phi'_{1c}}{2\Phi^*} w^{(1)} \right] \bar{w}^{(1)2} \right) + \frac{\gamma_o}{4} \sqrt{\frac{\Phi^*}{\Phi_o^*}} \frac{\Phi_o}{\Phi^*} \kappa \text{Re} \left(w^{(1)'} \bar{w}^{(1)} \right) \Big|_{z_o}^z + L_o^{(3)},$$

$$\begin{aligned}
L_o^{(3)}(z) = \int_{z_0}^z \sqrt{\frac{\Phi_o}{\Phi_o^*}} \text{Re} \left[\left(\frac{\gamma_0}{2} \frac{\Phi_{3c}}{\Phi_o^*} - \frac{e}{q} \Psi_{3s} - \frac{1 + \gamma_0^2}{8} \frac{\Phi_{2c} \Phi_{1c}}{\Phi_o^{*2}} + \frac{e \gamma_0}{4q} \frac{\Phi_{1c}}{\Phi_o^*} \Psi_{2s} + \frac{\gamma_0}{32} \frac{\Phi_{1c}^3}{\Phi_o^{*3}} \right) \bar{w}^{(1)3} \right. \\
- \left(\frac{\Phi_{2c} \Phi_{1c}}{8 \Phi_o^{*2}} - \frac{\gamma_0}{16} \frac{\Phi_{1c}^3}{\Phi_o^{*3}} \right) w^{(1)} \bar{w}^{(1)2} + \frac{\Phi_o}{\Phi_o^*} \left(\frac{\gamma_0 e}{2q} \Psi_{2s} - \frac{1 + \gamma_0^2}{4} \frac{\Phi_{2c}}{\Phi_o^*} + \frac{5 \gamma_0}{32} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} \right) \bar{w}^{(1)2} \kappa \\
\left. + \frac{\gamma_0 \Phi_o}{8 \Phi_o^*} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} w^{(1)} \bar{w}^{(1)} \kappa + \frac{\gamma_0}{4} \frac{\Phi_o^2}{\Phi_o^{*2}} \frac{\Phi_{1c}}{\Phi_o^*} x^{(1)} \kappa^2 \right] dz. \quad (7.78)
\end{aligned}$$

The first term on the right hand represents a contribution to the second-rank path deviation, which vanishes in the region outside the dipole field. The second term does not explicitly depend on the dipole fields. This chromatic deviation is solely a function of the position and the slope of the paraxial ray and, hence, results from all elements of the beam line, which affect the course of the paraxial ray $w^{(1)}$.

The integrand of the third-rank eikonal $L_o^{(3)}$ consists of five terms whose coefficients are functions of only the multipole strengths. Therefore, the contribution of $L_o^{(3)}(z)$ to the second-rank path deviation does not depend on the slope of the multipole strengths. However, since their derivatives strongly affect the higher-rank deviations, one should avoid large slopes of the multipole strengths. If we can adjust their course along the z -axis, it is possible to nullify at least two terms of the integrand without affecting the paraxial path. By considering the quadrupole strengths Φ_{1z} , Ψ_{2s} and one hexapole strength, Φ_{3c} or Ψ_{3s} , as three free parameters, we can eliminate the first three terms of the integrand of $L_o^{(3)}(z)$ and simultaneously satisfy the anastigmatism condition (7.76). We achieve this situation by choosing the multipole strengths as

$$\Phi_{2c} = \frac{\gamma_0}{2} \frac{\Phi_{1c}^2}{\Phi_o^*}, \quad \frac{e}{q} \Psi_{2s} = \frac{4\gamma_0^2 - 1}{16} \frac{\Phi_{1c}^2}{\Phi_o^*}, \quad \Phi_{3c} - v \Psi_{3s} = \frac{3}{32} \frac{\Phi_{1c}^3}{\Phi_o^{*2}}. \quad (7.79)$$

Here $v = q/\gamma_0 m_e$ is the nominal velocity of a particle moving along the optic axis. A Wien filter, which satisfies the conditions (7.79), yields stigmatic imaging up to the second order inclusively. However, the rotationally symmetric chromatic deviation of such a filter is unavoidable because the corresponding aberration coefficient is positive definite. We obtain this coefficient from the fourth term of the integrand of $L_o^{(3)}(z)$.

We shall show in the context of aberrations that we can compensate for the chromatic aberrations of the Wien filter by lifting one of the constraints (7.79). Although we cannot meet these conditions exactly in the region of the fringing fields, the residual second-order deviations will be small for a dodecapole Wien filter because the spacings between opposite electrodes and/or magnetic pole pieces are identical in this case. Accordingly, the electric and magnetic strengths of a given multipole component have the same shape along the optic axis.

7.5.2 Magnetic Systems

Magnetic systems with midsection symmetry are widely employed in practice. Accelerators, storage rings, and spectrometers used in high-energy physics and energy filters employed in analytical electron microscopes [79, 127–131] are important examples of these systems. A particle moving in the midsection will not experience a force normal to this section, which embeds the curved optic axis. Owing to this property, we choose the midsection as the horizontal x - z section of the curvilinear coordinate system. Since the field is entirely magnetic, all electric multipole components are zero apart from the constant axial potential $\Phi = \Phi_0 = \Phi_o$:

$$\Phi_v = 0, \quad \Psi_v = i\Psi_{vs}, \quad v > 0, \quad \Phi' = 0. \quad (7.80)$$

Moreover, we require that the optic axis is representing a possible ray with curvature given by (4.15) as

$$\Gamma = \bar{\Gamma} = -\frac{e}{q}\Psi_{1s}. \quad (7.81)$$

Because the axial electric potential is constant, this holds also true for the axial kinetic momentum $q = q_0 = q_o$. We derive the Gaussian path equation of magnetic systems with midsection symmetry from Eq. (4.13) by employing (7.80) and (7.81). The resulting complex equation decouples into the simple real equations (4.36), which we rewrite in the form

$$x'' + \left(\frac{e^2}{q^2}\Psi_{1s}^2 + \frac{2e}{q}\Psi_{2s} \right) x = \kappa^* \frac{e}{2q}\Psi_{1s}, \quad (7.82)$$

$$y'' - \frac{2e}{q}\Psi_{2s}y = 0. \quad (7.83)$$

Because the axial electric potential is constant for magnetic systems, it is advantageous to introduce the relativistic modified chromatic parameter

$$\kappa^* = \frac{1 + e\Phi/m_e c^2}{1 + e\Phi/2m_e c^2} \kappa = \frac{2\gamma_0}{1 + \gamma_0} \kappa. \quad (7.84)$$

The choice κ^* instead of κ (4.2) simplifies the expressions for the variational polynomials considerably. The third-rank variational polynomials (7.49) and (7.50) reduce to

$$\begin{aligned} \mu^{(3)} &= \mu_g^{(3)} + \mu_c^{(3)} \\ &= -\text{Re} \left[\frac{e}{2q} \Psi_{1s} w' \bar{w}' \bar{w} + \frac{e}{8q} \Psi'_{1s} (2w \bar{w} \bar{w}' - w' \bar{w}^2) + \frac{e}{q} \left(\Psi_{3s} + \frac{e}{3q} \Psi_{2s} \Psi_{1s} \right) \bar{w}^3 \right. \\ &\quad \left. + \frac{e^2}{4q^2} \Psi_{2s} \Psi_{1s} \bar{w}^2 w \right] + \frac{1}{4} w' \bar{w}' \kappa^* + \frac{e}{8\gamma_0^2 q} \Psi_{1s} x \kappa^{*2}. \end{aligned} \quad (7.85)$$

We substitute the paraxial ray $w^{(1)} = w_g^{(1)} + \kappa^* x_{\kappa^*}$ for w in this expression and separate the resulting integrand of the third-rank polynomial of the perturbation eikonal in a geometric part $\mu_{1g}^{(3)}$ and a chromatic part $\mu_{1c}^{(3)}$, comprising all terms that are linear and quadratic in the chromatic parameter κ . Replacing w by $w_g^{(1)}$, we readily obtain the geometric part $\mu_{1g}^{(3)}$ as

$$\begin{aligned} \mu_{1g}^{(3)} = & -\text{Re} \left[\frac{e}{2q} \Psi_{1s} w_g^{(1)'} \bar{w}_g^{(1)'} \bar{w}_g^{(1)} + \frac{e}{8q} \Psi_{1s}' \left(2w_g^{(1)} \bar{w}_g^{(1)} \bar{w}_g^{(1)'} - w_g^{(1)'} \bar{w}_g^{(1)2} \right) \right. \\ & \left. + \frac{e^2}{4q^2} \Psi_{2s} \Psi_{1s} \bar{w}_g^{(1)2} w_g^{(1)} + \frac{e}{q} \left(\Psi_{3s} + \frac{e}{3q} \Psi_{2s} \Psi_{1s} \right) \bar{w}_g^{(1)3} \right]. \end{aligned} \quad (7.86)$$

We derive the chromatic part by substituting expression (7.51) for w in (7.85) and retaining the terms, which are linear and quadratic in the relativistic modified chromatic parameter κ^* . Considering that the cubic chromatic terms do not contribute to the path deviation, we obtain

$$\begin{aligned} \mu_{c1}^{(3)} = & -\kappa^* \left[\frac{e}{2q} \Psi_{1s} \left\{ x_{\kappa^*} (x_g^{(1)2} + y_g^{(1)2}) + 2x_{\kappa^*}' x_g^{(1)} x_g^{(1)'} \right\} - \frac{1}{4} (x_g^{(1)2} + y_g^{(1)2}) \right. \\ & + \frac{e}{8q} \Psi_{1s}' \left\{ 2x_{\kappa^*} (x_g^{(1)} x_g^{(1)'} - y_g^{(1)} y_g^{(1)'}) + x_{\kappa^*}' (x_g^{(1)2} + 3y_g^{(1)2}) \right\} + \frac{e^2}{4q^2} \Psi_{2s} \Psi_{1s} x_{\kappa^*} (3x_g^{(1)2} + y_g^{(1)2}) \\ & + \frac{e}{q} \left(3\Psi_{3s} + \frac{e}{q} \Psi_{2s} \Psi_{1s} \right) x_{\kappa^*} (x_g^{(1)2} - y_g^{(1)2}) \left. \right] - \kappa^{*2} \left[\frac{e}{2q} \Psi_{1s} x_{\kappa^*}' (2x_{\kappa^*} x_g^{(1)'} + x_{\kappa^*}' x_g^{(1)}) - \frac{1}{2} x_{\kappa^*}' x_g^{(1)'} \right. \\ & \left. + \frac{e}{4q} \Psi_{1s}' x_{\kappa^*} (2x_{\kappa^*}' x_g^{(1)} + x_{\kappa^*} x_g^{(1)'}) + \frac{e}{q} \left(3\Psi_{3s} + \frac{7e}{4q} \Psi_{2s} \Psi_{1s} \right) x_{\kappa^*}^2 x_g^{(1)} - \frac{e}{8\gamma_0^2 q} \Psi_{1s} x_g^{(1)} \right]. \end{aligned} \quad (7.87)$$

The terms in the second bracket produce the second-degree dispersion. This deviation is proportional to the square of the chromatic parameter. We obtain the chromatic deviation of first order and first degree from the terms linear in κ^* . Relation (7.87) shows that the primary chromatic deviation results from the dipole, quadrupole, and hexapole fields. However, the latter do not contribute if we place the sextupole elements at regions that are free of dispersion ($x_{\kappa^*} = 0$). Since the hexapole fields do not affect the paraxial rays, we can utilize them for correcting the chromatic deviations without influencing the Gaussian beam. One uses this possibility to compensate for the chromaticity of accelerators by placing sextupole elements in the lattice at positions where the dispersion is large. In order to correct for the vertical and horizontal components, it is necessary that the ratio formed by any two fundamental paraxial rays within a sextupole is different from the corresponding ratio within all other sextupoles. Since each monomial of second order in (7.87) is quadratic or bilinear in the geometric ray parameters a_1, a_3 and a_2, a_4 , respectively, six monomials of first degree and second order exist. Hence, at least six sextupoles are generally necessary to compensate for the chromatic deviation of first order and first degree. In order to demonstrate this requirement, we consider that the complex paraxial path equation decouples into the two real equations (7.82) and (7.83) for the x -component and the y -component, respectively.

Accordingly, the fundamental rays have the form

$$w_1(z) = x_1(z), \quad w_3(z) = x_3(z), \quad w_2(z) = iy_2(z), \quad w_4(z) = iy_4(z), \quad (7.88)$$

giving

$$\begin{aligned} x_g^{(1)} &= a_1 x_1 + a_3 x_3, \\ y_g^{(1)} &= y^{(1)} = a_2 y_2 + a_4 y_4. \end{aligned} \quad (7.89)$$

For reasons of simplicity, we define the ray at the object plane in such a way that we only need to consider the chromatic part $E_{co}^{(3)}$ of the single third-rank eikonal $E_o^{(3)}$. To elucidate the chromatic effect of the sextupoles, we analyze their contribution

$$\begin{aligned} E_{co,s}^{(3)}(z_r) &= -\frac{3e}{q_0} \kappa^* \int_{z_o}^{z_r} \Psi_{3s} \left(\begin{array}{c} a_1^2 x_1^2 + 2a_1 a_3 x_1 x_3 + a_3^2 x_3^2 \\ -a_2^2 y_2^2 - 2a_2 a_4 y_2 y_4 - a_4^2 y_4^2 \end{array} \right) x_{\kappa^*} dz \\ &\quad - \frac{3e}{q_0} \kappa^{*2} \int_{z_o}^{z_r} \Psi_{3s} (a_1 x_1 + a_3 x_3) x_{\kappa^*}^2 dz \end{aligned} \quad (7.90)$$

to the polynomial $E_{co}^{(3)}(z)$ taken at the recording plane $z = z_r$. By imposing symmetry conditions on the multipole fields and the course of the fundamental rays, the integrands of several monomials of $E_{co}^{(3)}(z_r)$ become antisymmetric functions with respect to the plane z_m located midway between the terminal planes z_o and z_r . Using this procedure, we can eliminate all geometric monomials and the chromatic monomials, which are linear or bilinear in the geometric ray parameters a_v . Unfortunately, this is not possible for the chromatic monomials, which are quadratic in one of the four geometric ray parameters because the integrands of the monomial coefficients contain positive definite terms. To compensate for these chromatic defects, we must incorporate sextupole elements. Owing to the imposed symmetry, eight sextupoles are necessary to compensate for the chromatic second-rank deviations without introducing geometric deviations at the recording plane. One defines systems corrected for all second-rank deviations as second-order achromats. High-performance accelerators and storage rings are composed of such cells.

7.6 Second-Rank Path Deviations of Systems with Straight Axis

Systems with a straight optic axis do not contain deflection elements. Hence, the electromagnetic potentials of these systems do not possess dipole components, which curve the optic axis and introduce dispersion. Considering $\Phi_1 = 0, \Psi_1 = 0$

and $\Gamma = 0$, the expressions (7.49) and (7.50) for the geometric and the chromatic components of the third-rank variational polynomial reduce to

$$\mu_g^{(3)} = \frac{q}{q_0} \operatorname{Re} \left(\frac{\gamma_0}{2} \frac{\Phi_3}{\Phi^*} + i \frac{e}{q} \Psi_3 \right) \bar{w}^3, \quad (7.91)$$

$$\mu_c^{(3)} = \frac{q}{q_0} \frac{\Phi_o}{\Phi^*} \kappa \operatorname{Re} \left[\frac{\gamma_0}{4} w' \bar{w}' + \frac{1}{16} \frac{\Phi''}{\Phi^*} w \bar{w} - \frac{1}{4} \frac{\Phi_2}{\Phi^*} \bar{w}^2 \right]. \quad (7.92)$$

Relations (7.91) and (7.92) reveal that the second-order (geometrical) path deviations result exclusively from the hexapole fields, whereas the chromatic deviations depend on the axial and quadrupole fields. Sextupoles are important elements of aberration correctors because their secondary third-order aberrations are equivalent to those of axially symmetric lenses and the spherical aberration has opposite sign to that of round lenses [132]. At first glance, the dominant second-order aberrations seem to rule out the use of sextupoles for correcting the much smaller third-order spherical aberration of a good objective lens [133]. However, by employing symmetry conditions, one can eliminate the deleterious second-order aberrations for special systems composed of sextupoles and round lenses [22, 99].

We fix the trajectory by its positions at the planes $z_1 = z_2 = z_o$ and $z_3 = z_4 = z_a$ so that the fundamental rays meet the constraints (5.131) and (5.138), respectively. Substituting the paraxial approximation $w^{(1)}$ for w , we obtain the integrands of the geometric and chromatic parts of the third-rank eikonal polynomial $E_v^{(3)} = E_{gv}^{(3)} + E_{cv}^{(3)}$ as

$$E_{gv}^{(3)} = \operatorname{Re} \int_{z_v}^z \sqrt{\frac{\Phi^*}{\Phi_o^*}} \left(\frac{\gamma_0}{2} \frac{\Phi_3}{\Phi^*} + i \frac{e}{q} \Psi_3 \right) \bar{w}^{(1)3} dz, \quad v = 1, 2, 3, 4, \quad (7.93)$$

$$\begin{aligned} \frac{q_0}{q_o} E_{cv}^{(3)} &= \frac{\kappa}{16} \frac{\Phi_o}{\sqrt{\Phi_o^*}} \frac{\Phi'}{\Phi^{*3/2}} w^{(1)} \bar{w}^{(1)} \Big|_{z_v}^z \\ &+ \frac{\kappa}{4} \frac{\Phi_o}{\Phi_o^*} \operatorname{Re} \int_{z_v}^z \sqrt{\frac{\Phi_o^*}{\Phi^*}} \left(\gamma_0 w^{(1)'} \bar{w}^{(1)'} + \frac{3\gamma_0}{8} \frac{\Phi'^2}{\Phi^{*2}} w^{(1)} \bar{w}^{(1)} - \frac{1}{2} \frac{\Phi'}{\Phi^*} \bar{w}^{(1)} w^{(1)'} - \frac{\Phi_2}{\Phi^*} \bar{w}^{(1)2} \right) dz. \end{aligned} \quad (7.94)$$

We obtained expression (7.94) by removing the second derivative of the axial potential by means of partial integration. The integrated part does not contribute to the second-rank path deviation (7.69) because the terms taken at the upper integration limit z cancel out each other. This behavior follows from the relations (5.147) and (5.155), respectively. The terms taken at the lower limit $z = z_v$ do not contribute too because each derivative $\partial E_v^{(3)} / \partial a_v$ vanishes since we have $\partial w^{(1)}(z_v) / \partial a_v = w_v(z_v) = 0$. Therefore, we can neglect the first term on the right-hand side of formula (7.94).

To survey the properties of the chromatic deviation, we rewrite the integrand I in (7.94) as

$$I = \sqrt{\frac{\Phi_o^*}{\Phi^*}} \left[\gamma_0 \left| w^{(1)'} - \frac{1}{4\gamma_0} \frac{\Phi'}{\Phi^*} w^{(1)} \right|^2 + \frac{6\gamma_0^2 - 1}{16\gamma_0} \frac{\Phi'^2}{\Phi^{*2}} |w^{(1)}|^2 - \frac{\Phi_2}{\Phi^*} \bar{w}^{(1)2} \right]. \quad (7.95)$$

Since the relativistic factor γ_0 is larger than unity, the integrand (7.95) is positive definite in the absence of an electric quadrupole field ($\Phi_2 = 0$). Hence, the chromatic path deviation is unavoidable for round lenses and magnetic systems with straight optic axis except at planes where $w_c^{(2)}(z) = 0$. However, these planes differ from the image plane or images of the diffraction plane.

7.6.1 Second-Order Path Deviation

Hexapoles are the basic elements of correctors compensating for the third-order spherical aberration of round lenses. The arrangement of the electrodes of an electrostatic hexapole and the force acting on electrons traveling in the direction of the optic axis are depicted schematically in Fig. 7.1.

The primary effect of hexapole or sextupole elements is of second order. The resulting path deviation depends on the location of the sextupole elements and on the distance $w^{(1)}(z) = u^{(1)}(z)e^{-i\chi}$ of the paraxial ray within the hexapole field. We define the strength of the electric and the magnetic hexapoles by the hexapole function

$$H = e^{-3i\chi} \left(\frac{\gamma_0}{2} \sqrt{\frac{\Phi^*}{\Phi_o^*}} \frac{\Phi_3}{\Phi^*} + i \frac{e}{q_o} \Psi_3 \right). \quad (7.96)$$

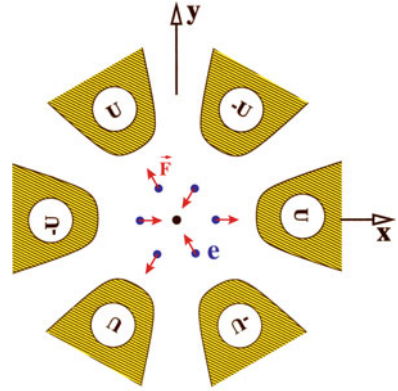
By employing this abbreviation, we obtain for the second-order perturbation eikonal resulting from the sextupoles the expression

$$\begin{aligned} L_{so}^{(3)} &= E_s^{(3)} = \operatorname{Re} \int_{z_o}^z H \bar{u}^{(1)3} dz \\ &= \operatorname{Re} \int_{z_o}^z H \left[\sum_{\mu=1}^4 a_\mu^3 \bar{u}_\mu^3 + 3 \sum_{\mu \neq \nu}^4 a_\mu^2 a_\nu \bar{u}_\mu^2 \bar{u}_\nu + 6 \sum_{\lambda=3}^4 \sum_{\mu=2}^{\lambda-1} \sum_{\nu=1}^{\mu-1} a_\lambda a_\mu a_\nu \bar{u}_\lambda \bar{u}_\mu \bar{u}_\nu \right] dz. \end{aligned} \quad (7.97)$$

This polynomial consists of 20 monomials. In order that the second-order path deviation

$$u^{(2)}(z) = \{u^{(1)}, L_{so}^{(3)}\} \quad (7.98)$$

Fig. 7.1 Arrangement of the electrodes of an electrostatic hexapole element and action of its field on electrons propagating parallel to the optic axis



vanishes on the far side $z \geq z_s$ of the sextupoles, we must eliminate all 20 monomials. The number of monomials cuts in half if $H/\bar{H} = \exp(6i\vartheta_H) = \text{const.}$ and the complex paraxial path equation decouples so that two of the fundamental rays are real ($u_1 = \bar{u}_1, u_3 = \bar{u}_3$) and two are imaginary ($\bar{u}_2 = -u_2, \bar{u}_3 = -u_3$). The number of monomials reduces to four if the path of the Gaussian rays is rotationally symmetric within the region of the hexapole fields. In this case, we find with (7.4) and $u_2 = iu_1 = iu_\alpha, u_4 = iu_3 = iu_\gamma$ the paraxial ray as

$$u^{(1)} = (a_1 + ia_2)u_\alpha + (a_3 + ia_4)u_\gamma = \omega u_\alpha + \rho u_\gamma. \quad (7.99)$$

Since the axial ray $u_\alpha = u_\omega$ and the field ray $u_\gamma = u_\rho$ are real, the eikonal coefficients

$$F_m^{(3,3)}(z) = \int_{z_0}^z H(z) u_\alpha^{3-m}(z) u_\gamma^m(z) dz, \quad m = 0, 1, 2, 3, \quad (7.100)$$

possess the same phase angle ϑ_H as H . This angle accounts for the azimuthal orientation of the sextupole with respect to the x -axis of the rotating coordinate system. The two upper indices indicate the order and the multiplicity of the monomials

$$L_{sm}^{(3,3)} = \text{Re} (F_m^{(3,3)} \bar{\omega}^{3-m} \bar{\rho}^m) \quad (7.101)$$

of the normalized eikonal

$$L_s^{(3)} = \sum_{m=0}^3 L_m^{(3,3)} = \sum_{m=0}^3 \frac{3!}{m!(3-m)!} F_m^{(3,3)} \text{Re} (\bar{\omega}^{3-m} \bar{\rho}^m). \quad (7.102)$$

We readily derive the second-order path deviation from (7.98) by employing the complex form (5.153) of the Lagrange bracket, giving

$$u^{(2)} = 2 \left(u_\alpha \frac{\partial L_s^{(3)}}{\partial \bar{\rho}} - u_\gamma \frac{\partial L_s^{(3)}}{\partial \bar{\omega}} \right) = \bar{\omega}^2 u_{11} + \bar{\omega} \bar{\rho} u_{12} + \bar{\rho}^2 u_{22}. \quad (7.103)$$

The second-order fundamental rays u_{11} , $u_{12} = u_{21}$ and u_{22} have the form

$$u_{\mu\nu} = 3(2 - \delta_{\mu\nu}) \{ u_\alpha F_{\mu+\nu-1}^{(3,3)} - u_\gamma F_{\mu+\nu-2}^{(3,3)} \}, \quad \mu, \nu = 1, 2. \quad (7.104)$$

The Kronecker symbol is defined as $\delta_{\mu\nu} = 1$ for $\mu = \nu$ and zero else. In order that the second-order fundamental rays (7.104) vanish in the entire region $z > z_s$ behind the sextupoles, we must nullify the four coefficients (7.100) at the plane $z = z_s$:

$$F_m^{(3,3)}(z_s) = \int_{z_0}^{z_s} H u_\alpha^{3-m} u_\gamma^m dz = 0, \quad m = 0, 1, 2, 3. \quad (7.105)$$

We can satisfy these requirements most conveniently by imposing symmetry conditions on the paraxial fundamental rays and on the total hexapole strength H such that the integrand of the integrals (7.105) becomes an antisymmetric function either with respect to the midplane of the sextupole arrangement or with respect to the central planes of each half of the system. Since the fundamental rays u_α and u_γ are linearly independent, they cannot have the same symmetry about a given plane. Therefore, it is not possible to eliminate all second-order path deviations by a single symmetry condition. For this purpose, we must introduce a double symmetry such that one of the fundamental rays is symmetric and the other antisymmetric with respect to the midplane and the central planes of each half of the sextupole arrangement. If in addition the hexapole field is symmetric with respect to the three symmetry planes, all second-order fundamental rays vanish outside the system. We can also apply this procedure for eliminating the second-order deviations of systems with curved axis and midsection symmetry by imposing the same conditions on the quadrupole and dipole fields and by requiring that each of the two axial rays $w_1 = x_\alpha$ and $w_2 = iy_\beta$ and each of the two field rays $w_3 = x_\gamma$ and $w_4 = iy_\delta$, respectively, satisfies the same symmetry conditions. One has utilized this procedure for eliminating second-order path deviations of imaging energy filters and monochromators with curved axis.

Figure 7.2 shows the simplest system satisfying the conditions (7.105). The arrangement consists of a telescopic round-lens doublet and two identical sextupoles centered about the outer focal planes of the round lenses. These planes also represent the two nodal planes N_1 and N_2 of the doublet. The plane midway between the round lenses is the midplane of the entire system, while the nodal planes represent the central planes of each half of the sextupole system. Since these planes are also the symmetry planes of the fundamental rays, the system fulfills the requirement necessary for complete elimination of the second-order deviations outside the system, as demonstrated in Fig. 7.3. Because this system introduces a negative third-order spherical aberration, one uses it as a corrector for eliminating the unavoidable

Fig. 7.2 Arrangement of the elements of a spherical-aberration corrector, which does not introduce any second-order deviations outside the system

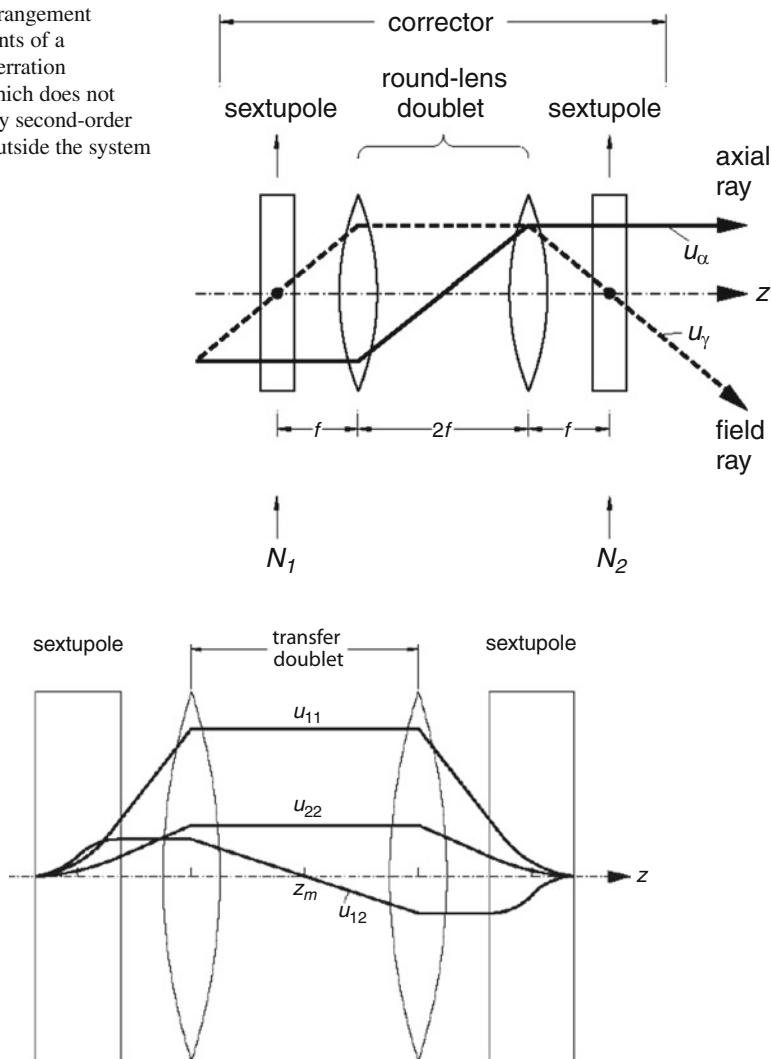


Fig. 7.3 Course of the secondary fundamental rays u_{11} , u_{12} , u_{22} within the hexapole corrector

spherical aberration of the objective lens in electron microscopes. We shall discuss this property in the context of aberrations and aberration correction.

To avoid a rotation of the image of the first sextupole, the coils of the magnetic round lenses must be connected in series opposition so that the excitations of the two identical lenses are equal and opposite, whatever is the strength of the current. In this case, the doublet images the front sextupole with magnification $M = -1$ exactly onto the second sextupole so that their primary effect on the course of the

particles cancels out. This holds also true if we substitute the telescopic quadrupole quadruplet shown in Fig. 4.33 for the round-lens doublet. However, the third-order path deviations of this system will differ from rotational symmetry.

The method of canceling path deviations by imposing symmetry conditions is not limited to second order. For example, if we replace the sextupoles in the arrangement shown in Fig. 7.2 by octupoles, we introduce primary path deviations of the third order and secondary path deviations of fifth order. In order that the second octupole compensates for the primary deviations, its excitation must be excited opposite to that of the first octupole. In the absence of sextupoles, systems with a straight optic axis do not introduce second-order path deviations. Hence, the primary geometrical path deviations of these systems are of third order and produced by the secondary effects of the round lenses and quadrupoles and by the primary action of the octupoles. A multipole with multiplicity m produces an m -fold deformation of order $n = m$ in the wave surface and of order $n = m - 1$ with respect to the beam trajectories, if we place this element in the stigmatic region of the paraxial rays. This multipole also introduces path deviations with lower multiplicities, if we place it within the astigmatic regime of the paraxial rays.

To derive in analytical form the course of the secondary fundamental rays within the hexapole elements shown in Fig. 7.3, we approximate the hexapole function (7.96) by a box function and assume that the fields of the round lenses do not overlap with the hexapole fields. In this case, the paraxial fundamental rays form straight lines,

$$u_1 = u_\alpha = f_0, \quad u_3 = u_\gamma = (z - z_{N_1})/f_0, \quad (7.106)$$

inside the region of the two sextupoles, whose approximated strengths is given by

$$H(z) = \sum_{\mu=1}^2 H_0 \Theta_\mu(z), \quad (7.107)$$

where the step function $\Theta_\mu(z)$ is defined as

$$\Theta_\mu(z) = \begin{cases} 1 & \text{for } z_\mu \leq z \leq z_\mu + l, \\ 0 & \text{else.} \end{cases} \quad (7.108)$$

Here f_0 denotes the objective focal length, and l is the length of each of the two box-shaped hexapole fields, one centered at the front nodal plane $z_{N_1} = z_F = z_1 - l/2$ of the transfer doublet, the other at the back nodal plane $z_{N_2} = z_F = z_2 - l/2$, as illustrated in Fig. 7.2. Owing to the symmetric arrangement and the symmetry of the fundamental rays, it suffices to determine the course of the secondary fundamental rays within the first sextupole element. Employing the approximation (7.107) and the expressions (7.106) for the primary fundamental rays, we can readily perform the integration in the formula (7.100), giving

$$\begin{aligned}
F_m^{(3,3)}(z) &= \int_{z_{N_1}-l/2}^z H u_\alpha^{3-m} u_\gamma^m dz = H_0 f_o^{3-2m} \int_{z_{N_1}-l/2}^z (z - z_{N_1})^m dz \\
&= \frac{H_0}{m+1} f_o^{3-2m} \left(\frac{l}{2}\right)^{m+1} \left[\left(2 \frac{z - z_{N_1}}{l}\right)^{m+1} + (-1)^m \right], \quad m = 0, 1, 2, 3.
\end{aligned} \tag{7.109}$$

The terms with odd index m vanish at the exit plane $z = z_e = z_{N_1} + l/2$ of the first sextupole:

$$F_1^{(3,3)}(z_e) = F_3^{(3,3)}(z_e) = 0. \tag{7.110}$$

The terms

$$F_0^{(3,3)}(z_e) = H_0 l f_o^3, \quad F_2^{(3,3)}(z_e) = \frac{1}{12} H_0 l^3 / f_o \tag{7.111}$$

with even index ($m = 0, 2$) do not vanish at this plane. They are opposite to those of the second sextupole because the fundamental rays change sign after passing the telescopic round-lens doublet. Accordingly, all secondary fundamental rays (7.104) vanish in the region behind the second sextupole. Considering the result (7.110), it follows from the expression (7.104) that the asymptotes of the secondary fundamental rays u_{11} and u_{22} intersect the nodal points N_1 and N_2 . Because these points coincide with the outer focal points of the round lenses, the rays u_{11} and u_{22} are symmetric with respect to the midplane of the system. The mixed secondary fundamental ray $u_{12} = u_{21}$ is antisymmetric with respect to this plane since it follows from (7.104) that this ray is proportional to u_α in the region between the two sextupoles.

Chapter 8

Aberrations

Abstract The path deviations at the image plane represent the aberrations of an electron optical system. Each aberration referred back to the object plane is the product of a monomial of the ray parameters times the aberration coefficient. Integral expressions are obtained for the aberration coefficients of the primary aberrations of systems with straight and curved axes. The chromatic aberrations and the geometrical aberrations of rotationally symmetric systems and of quadrupole–octopole systems are discussed in detail. In particular, the Scherzer theorem is proven by showing that the coefficients of the axial chromatic aberration and the spherical aberration of round lenses can never change sign.

The main task of electron optics concerns the design of systems, which possess distinct imaging or beam-guiding properties. Therefore, we must solve an inverse problem by finding the geometry of the electrodes and pole pieces and the strengths of the currents and voltages, which produce the electromagnetic fields required for refracting the electrons appropriately. Owing to this difficulty, entire numerical methods are not suited for finding optimum systems composed of numerous different elements, such as solenoids and multipoles. However, numerical methods are indispensable for the final design of the system after its outlay has been roughly determined by means of the analytical calculations and symmetry considerations employing the paraxial approximation for the trajectories and aberration integrals. Computer programs are nowadays available for calculating numerically field distributions throughout a given system very accurately by means of high-order finite-element or finite-difference procedures. Computing the particle trajectories by direct ray tracing [134, 135] yields directly the overall aberrations. The main disadvantages of this method are that the individual Seidel order aberration terms cannot be determined with reliable accuracy and that it does not provide information on how to suppress or eliminate appropriately the performance-limiting aberrations. In order to find such means, we must calculate analytically the integral expressions for the aberration integrals and investigate the structure of these integrals giving

information how to nullify them. To determine the performance of systems corrected for the primary aberrations, one must calculate the next higher aberrations. Unfortunately, the number of aberrations and the complexity of the aberration coefficients increase drastically with increasing order. In order to avoid errors in the time-consuming analytical calculations, special algebraic computer programs have been developed for automatically deriving expressions for the aberration coefficients. These programs are particularly useful for calculating the higher-order aberration coefficients of multielement systems, such as the SMART microscope [136].

Aberrations are path deviations at particular planes within the system. In an electron microscope, these planes are image planes of the object or of the diffraction plane. At each of these planes, either the two axial fundamental rays $w_1 = w_\alpha, w_2 = w_\beta$ or the two field rays $w_3 = w_\gamma, w_4 = w_\delta$ vanish. As a result, the path deviations adopt a rather simple form at these planes. We classify the aberrations as geometrical aberrations and chromatic aberrations in accordance with the definition of the path deviations. The eikonal method has the inherent advantage to reveal automatically any interrelations between the various aberration coefficients.

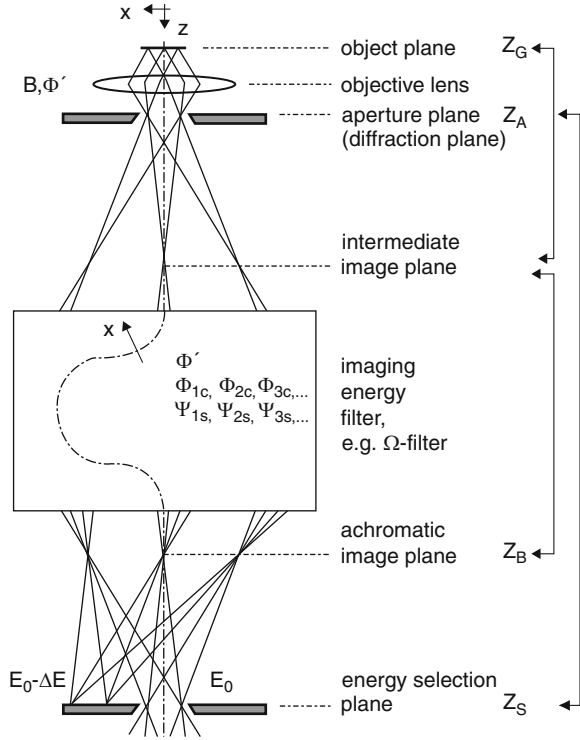
The interrelations increase with rising rank of the aberrations. Moreover, a distinct eikonal coefficient can be proportional to the coefficients of two different types of aberrations, one observed at the image plane and the other at an image of the diffraction plane if these planes are located in the field-free region behind the imaging system. In order to demonstrate this behavior, we consider the second-order aberrations introduced by an uncorrected imaging energy filter at the image plane z_i and the energy-selection plane z_e located behind the filter.

The filter images the polychromatic demagnified diffraction pattern located in front of the filter with unit magnification into a series of laterally displaced monochromatic spots in the energy-selection plane [127–131]. Simultaneously, the filter transfers the intermediate image of the object stigmatically into the achromatic (dispersion-free) image plane, as illustrated in Fig. 8.1. One places a slit aperture at the center of the energy-selection plane. The slit width defines the range of the energy window. By changing the acceleration voltage, we move the energy-loss spectrum in the direction perpendicular to the slit enabling energy selection. Most filters are symmetric with respect to their midplane. In this case, the filter images with unit magnification the intermediate images of the object plane and the diffraction plane located in front of the filter into the conjugate images behind the filter. Moreover, by introducing the symmetry, we eliminate half of the second-order geometric aberrations at these planes. Since two of the four fundamental rays vanish at each of these planes, the second-rank deviations (7.69), giving the aberrations, adopt a rather simple form.

Assuming distortion-free Gaussian images, we have

$$\begin{aligned} w_2(z_e) = i w_1(z_e) = i w_{1e} = i w_{\alpha e}, \quad w_3(z_e) = w_4(z_e) = 0, \\ w_4(z_i) = i w_3(z_i) = i w_{3i} = i w_{\gamma i}, \quad w_1(z_i) = w_2(z_i) = 0. \end{aligned} \quad (8.1)$$

Fig. 8.1 Action of an in-column imaging energy filter illustrating energy selection and the formation of the series of laterally displaced monochromatic demagnified diffraction patterns at the energy-selection plane $z_e = z_s$; the filter may be composed of electric and/or magnetic multipole fields of the type indicated in the box [137]. For simplicity, we have omitted the round lenses in front and behind the filter



The planes z_v at which we define the trajectories are located in front of the filter. Accordingly, the filter contributes the same amount,

$$E_f^{(3)} = E_{f1}^{(3)} = E_{f2}^{(3)} = E_{f3}^{(3)} = E_{f4}^{(3)}, \quad (8.2)$$

to the modified third-rank eikonals $E_v^{(3)}$ of the total system. Because the energy-selection plane z_e and the image plane z_i are located behind the filter, the eikonal $E_f^{(3)} = E_f^{(3)}(z_e) = E_f^{(3)}(z_i)$ is the same for both planes. Considering this fact and the relations (8.1) and (8.20), we obtain from the formula (7.69) the second-rank aberrations at these planes as

$$w_e^{(2)} = 2w_{\alpha e} \frac{\partial E_f^{(3)}}{\partial \rho}, \quad w_i^{(2)} = -2w_{\gamma i} \frac{\partial E_f^{(3)}}{\partial \bar{\omega}}. \quad (8.3)$$

The third-rank eikonal is a polynomial of third degree in the real ray parameters a_v or the complex ray parameters $\omega = a_1 + ia_2$, $\bar{\omega} = a_1 - ia_2$ and $\rho = a_3 + ia_4$, $\bar{\rho} = a_3 - ia_4$. Hence, relations (8.3) demonstrate that we obtain different kinds of aberrations from the same eikonal at the planes z_e and z_i . On going from the image plane to the energy-selection plane, we must exchange the meaning of the geometric

ray parameters, because the axial rays produce aberrations at the image plane and distortions at the energy-selection plane [79]. At this plane, the complex parameter ρ forms the initial axial slope because the field ray w_γ is the axial ray for the diffraction plane. Accordingly, the parameter ω determines the field of view of the diffraction plane since the ray w_α adopts the role of the field ray. The quality of the image of the diffraction pattern formed at the back focal plane of the objective lens does not depend on the optics in front of this plane. This is the reason why we generally need different eikonals for describing image formation of different initial planes. These considerations convincingly illustrate the advantage of the eikonal method even in the rather simple case of the primary aberrations. The virtue of the eikonal method becomes invaluable for designing high-performance aberration-corrected systems. For this purpose, the precise knowledge of the structure of the aberration coefficients is necessary in order to find appropriate means for the simultaneous correction of all disturbing aberrations because the induced higher-order aberrations often increase dramatically, thereby preventing an appreciable improvement of the optical performance.

Electron optical elements with curved axis suffer from both geometric and chromatic second-rank aberrations, whereas systems with straight axis and double-section symmetry ($\Phi_3 = \Psi_3 = 0$) only introduce a chromatic second-rank aberration, which is of first order and first degree. The primary geometric aberrations of these systems are of third order. We derive these aberrations from the perturbation eikonal polynomial $E_v^{(4)}$ obtained by substituting the paraxial ray for the true ray w in the expansion term (7.54) of the variational function.

8.1 Second-Rank Aberrations

We derive completely the second-rank aberrations (7.69) at the image plane $z = z_i$ by means of the third-rank perturbation eikonal:

$$E_{oi}^{(3)} = E_{gi}^{(3)} + E_{ci}^{(3)}. \quad (8.4)$$

This eikonal is composed of a geometric term $E_{gi}^{(3)}$ and a chromatic term $E_{ci}^{(3)}$ given by the relations

$$E_{gi}^{(3)} = L_{gi}^{(3)} = \int_{z_0}^{z_i} \mu_{g1}^{(3)} dz, \quad E_{ci}^{(3)} = L_{ci}^{(3)} = \int_{z_0}^{z_i} \mu_{c1}^{(3)} dz. \quad (8.5)$$

The geometric term $E_{gi}^{(3)} = L_{gi}^{(3)}$ yields the second-order aberrations, and the chromatic term $E_{ci}^{(3)} = L_{ci}^{(3)}$ yields the chromatic second-rank aberrations. The term $L_{gi}^{(3)}$ comprises all monomials of third order in the geometric ray parameters. We derive

the analytical form of the integrand $\mu_{\text{g}1}^{(3)}$ from the expression (7.47) by replacing the true ray by its Gaussian approximation $w^{(1)} = w_{\text{g}}^{(1)} + w_{\text{c}}^{(1)}$ and by retaining only the geometrical part given by

$$w_{\text{g}}^{(1)} = \omega w_{\omega} + \bar{\omega} w_{\bar{\omega}} + \rho w_{\rho} + \bar{\rho} w_{\bar{\rho}} = \sum_{\mu=1}^4 a_{\mu} w_{\mu}. \quad (8.6)$$

The proper choice of the fundamental paraxial rays depends on the symmetry properties of the system.

The number of monomials of a polynomial of degree l in m variables is at most:

$$N_{lm} = \frac{(l + m - 1)!}{l!(m - 1)!}. \quad (8.7)$$

Therefore, the modified third-rank perturbation eikonal (8.4) consists in the most general case of 35 real monomials. Twenty of these monomials are of geometric nature and 15 of chromatic nature. The monomial, which is of third degree in the chromatic parameter, does not contribute to the aberrations (8.3). Hence, we need to consider only 34 real eikonal terms.

8.1.1 Systems with Midsection Symmetry

By introducing midsection symmetry, we cut the number of geometrical monomials in half and reduce the number of chromatic terms from 14 to 8. In this case, the geometric component of the paraxial ray adopts the form (7.89), and the chromatic component is given by $w_{\text{c}}^{(1)} = \kappa x_{\kappa}$. Fixing the ray at the object plane, we obtain the aberrations at the image z_{i} of the object plane and at the image of the diffraction plane (8.3) from the single third-rank eikonal:

$$\begin{aligned} E_{\text{o}}^{(3)}(z) &= E_{\text{go}}^{(3)}(z) + E_{\text{co}}^{(3)}(z) = A_{111}a_1^3/3 + A_{113}a_1^2a_3 + A_{133}a_1a_3^2 + A_{333}a_3^3/3 \\ &\quad + B_{122}a_1a_2^2 + 2B_{124}a_1a_2a_4 + B_{144}a_1a_4^2 + B_{223}a_2^2a_3 + 2B_{234}a_2a_3a_4 \\ &\quad + B_{344}a_3a_4^2 \\ &\quad + \kappa (C_{11\kappa}a_1^2 + 2C_{13\kappa}a_1a_3 + C_{22\kappa}a_2^2 + 2C_{24\kappa}a_2a_4 + C_{33\kappa}a_3^2 + C_{44\kappa}a_4^2) / 2 \\ &\quad + D_{1\kappa\kappa}a_1\kappa^2 + D_{3\kappa\kappa}a_3\kappa^2. \end{aligned} \quad (8.8)$$

If the space between these images is field-free, we derive the aberrations (8.3) from the same eikonal $E_{\text{o}}^{(3)}(z_{\text{i}}) = E_{\text{oi}}^{(3)}$. Inserting (8.8) into the expression (7.40), the second-rank path deviation adopts the perspicuous form

$$x^{(2)} = a_1^2 x_{11} + a_2^2 x_{22} + a_1 a_2 x_{12} + a_2 a_4 x_{24} + a_3^2 x_{33} \\ + a_4^2 x_{44} + a_1 \kappa x_{1\kappa} + a_3 \kappa x_{3\kappa} + \kappa^2 x_{\kappa\kappa}, \quad (8.9)$$

$$y^{(2)} = a_1 a_2 y_{12} + a_1 a_4 y_{14} + a_2 a_3 y_{23} + a_3 a_4 y_{34} + a_2 \kappa y_{2\kappa} + a_4 \kappa y_{4\kappa}. \quad (8.10)$$

The secondary fundamental rays are linear combinations of the eikonal coefficients. We find the rays for the horizontal section as

$$x_{\mu\nu} = x_{\mu\nu}(z) = (2 - \delta_{\mu\nu}) (x_1 A_{\mu\nu 3} - x_3 A_{\mu\nu 1}), \quad \mu, \nu = 1, 3 \\ x_{\mu\nu} = x_{\mu\nu}(z) = (2 - \delta_{\mu\nu}) (x_1 B_{\mu\nu 3} - x_3 B_{\mu\nu 1}), \quad \mu, \nu = 2, 4 \\ x_{\mu\kappa} = x_{\mu\kappa}(z) = x_1 C_{\mu\kappa 3} - x_3 C_{\mu\kappa 1}, \quad \mu = 1, 3. \quad (8.11)$$

and for the vertical section as

$$y_{\mu\nu} = y_{\mu\nu}(z) = 2(y_2 B_{\mu\nu 4} - y_4 B_{\mu\nu 2}), \quad \mu = 1, 3, \quad \nu = 2, 4; \\ y_{\mu\kappa} = y_{\mu\kappa}(z) = y_2 C_{\mu 4\kappa} - y_4 C_{\mu 2\kappa}, \quad \mu = 2, 4. \quad (8.12)$$

The ray

$$x_{\kappa\kappa} = x_{\kappa\kappa}(z) = x_1 D_{3\kappa\kappa} - x_3 D_{1\kappa\kappa} \quad (8.13)$$

represents the second-degree dispersion ray. Fifteen complex second-rank fundamental rays $w_{\mu\nu} = x_{\mu\nu} + i y_{\mu\nu}$ exist in general systems, resulting in 30 real components. Owing to the imposed midsection symmetry, 15 components are zero. The x -component (8.9) of the second-rank path deviation does only possess ray parameters with odd index defining the x -component of the paraxial ray. The opposite behavior holds for the y -component (8.10), which contains exclusively parameters with even and odd indices. Each second-rank fundamental ray satisfies the initial condition:

$$x_{\mu\nu}(z_0) = 0, \quad x'_{\mu\nu}(z_0) = 0, \\ y_{\mu\nu}(z_0) = 0, \quad y'_{\mu\nu}(z_0) = 0. \quad (8.14)$$

Hence, the second-rank deviation $w^{(2)} = x^{(2)} + i y^{(2)}$ originates within the system with slope zero in the same way as the dispersion ray x_κ . In systems with midsection symmetry, all skew multipole components are zero (7.70). Moreover, the axial potential is constant ($\Phi' = 0$) within most systems of practical interest, such as energy filters, beam separators, or the cells of an accelerator. Assuming these conditions, we obtain with $\Phi = \Phi_0$ from (7.49), (7.50), and (8.5) for the eikonal coefficients the integral expressions

$$\begin{aligned}
A_{\lambda\mu\nu} = & \left[\left(\frac{3\gamma_0\Phi_{1c}}{\Phi_o^*} - \frac{e}{8q_o}\Psi_{2s} \right) (x'_\mu x_\nu x_\lambda + x'_\nu x_\lambda x_\mu + x'_\lambda x_\mu x_\nu) \right]_{z_o}^z \\
& + \int_{z_o}^z \left\{ (g_1 + g_2) (x'_\nu x'_\mu x_\lambda + x'_\lambda x'_\mu x_\nu + x'_\lambda x'_\nu x_\mu) + 3(g_3 + g_4)x_\lambda x_\mu x_\nu \right\} dz, \\
& \lambda, \mu, \nu = 1, 3, \kappa,
\end{aligned} \tag{8.15}$$

$$\begin{aligned}
B_{\lambda\sigma\tau} = & \left[\left\{ \frac{\gamma_0\Phi_{1c}}{\Phi_o^*} \left(\frac{x'_\lambda}{x_\lambda} + \frac{y'_\sigma}{y_\sigma} + \frac{y'_\tau}{y_\tau} \right) + \frac{e\Psi_{2s}}{8q_o} \left(\frac{y'_\sigma}{y_\sigma} + \frac{y'_\tau}{y_\tau} - 3\frac{x'_\lambda}{x_\lambda} \right) \right\} x_\lambda y_\sigma y_\tau \right]_{z_o}^z \\
& + \int_{z_o}^z \left\{ (g_1 - g_2)(y'_\sigma y_\tau + y'_\tau y_\sigma)x_\lambda + (g_1 + 3g_2)x'_\lambda y_\sigma y_\tau \right. \\
& \left. + (g_4 - 3g_3)x_\lambda y_\sigma y_\tau \right\} dz, \quad \sigma, \tau = 2, 4,
\end{aligned} \tag{8.16}$$

$$C_{\mu\nu\kappa} = 2A_{\mu\nu\kappa} + \frac{1}{1 + \gamma_0} \int_{z_o}^z \left\{ \gamma_0 x'_\mu x'_\nu + g_5 x_\mu x_\nu \right\} dz, \tag{8.17}$$

$$C_{\sigma\tau\kappa} = 2B_{\sigma\tau\kappa} + \frac{1}{1 + \gamma_0} \int_{z_o}^z \left\{ \gamma_0 y'_\sigma y'_\tau + g_6 y_\sigma y_\tau \right\} dz, \tag{8.18}$$

$$\begin{aligned}
D_{\mu\kappa\kappa} = & A_{\mu\kappa\kappa} + \frac{1}{1 + \gamma_0} \int_{z_o}^z \left\{ \gamma_0 x'_\mu x'_\kappa + g_5 x_\mu x_\kappa \right\} dz \\
& + \frac{1}{(1 + \gamma_0)^2} \int_{z_o}^z \left(\frac{\gamma_0\Phi_{1c}}{\Phi_o^*} - \frac{e}{2q_o}\Psi_{1s} \right) x_\mu dz.
\end{aligned} \tag{8.19}$$

The functions $g_m, m = 1, \dots, 6$, are abbreviations defined by

$$g_2 = -2g_1 = \frac{\gamma_0}{4} \frac{\Phi_{1c}}{\Phi_o^*} - \frac{e}{2q_o} \Psi_{1s}, \tag{8.20}$$

$$\begin{aligned}
g_3 = & \frac{\gamma_0}{2} \frac{\Phi_{3c}}{\Phi_o^*} - \frac{2 + 3\gamma_0^2}{16} \frac{\Phi_{2c}\Phi_{1c}}{\Phi_o^{*2}} + \frac{\gamma_0(5 + 2\gamma_0^2)}{128} \frac{\Phi_{1c}^3}{\Phi_o^{*3}} \\
& - \frac{e}{q_o} \left(\Psi_{3s} - \frac{\gamma_0}{8} \frac{\Phi_{2c}}{\Phi_o^*} \Psi_{1s} - \frac{7\gamma_0}{24} \frac{\Phi_{1c}}{\Phi_o^*} \Psi_{2s} + \frac{1 + \gamma_0^2}{64} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} \Psi_{1s} \right) \\
& - \frac{e^2}{q_o^2} \left(\frac{1}{12} \Psi_{2s} \Psi_{1s} + \frac{\gamma_0}{16} \frac{\Phi_{1c}}{\Phi_o^*} \Psi_{1s}^2 \right) + \frac{e^3}{16q_o^3} \Psi_{1s}^3,
\end{aligned} \tag{8.21}$$

$$\begin{aligned}
g_4 = & -\frac{2 + 3\gamma_0^2}{16} \frac{\Phi_{2c}\Phi_{1c}}{\Phi_o^{*2}} + \frac{\gamma_0(13 + 6\gamma_0^2)}{128} \frac{\Phi_{1c}^3}{\Phi_o^{*3}} \\
& + \frac{e}{q_o} \left(\frac{3\gamma_0}{8} \frac{\Phi_{2c}}{\Phi_o^*} \Psi_{1s} + \frac{3\gamma_0}{8} \frac{\Phi_{1c}}{\Phi_o^*} \Psi_{2s} - \frac{5(1 + 3\gamma_0^2)}{64} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} \Psi_{1s} \right) \\
& - \frac{3e^2}{4q_o^2} \left(\Psi_{2s}\Psi_{1s} - \frac{\gamma_0}{2} \frac{\Phi_{1c}}{\Phi_o^*} \Psi_{1s}^2 \right) - \frac{3e^3}{16q_o^3} \Psi_{1s}^3, \tag{8.22}
\end{aligned}$$

$$\begin{aligned}
g_5 = & -\frac{\Phi_{2c}}{\Phi_o^*} + \frac{3\gamma_0(7 + \gamma_0^2)}{16} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} - \frac{(7 + 4\gamma_0^2)e}{8q_o} \Psi_{1s} \frac{\Phi_{1c}}{\Phi_o^*} + \frac{\gamma_0 e^2}{4q_o^2} \Psi_{1s}^2, \tag{8.23}
\end{aligned}$$

$$\begin{aligned}
g_6 = & \frac{\Phi_{2c}}{\Phi_o^*} - \frac{\gamma_0(1 - \gamma_0^2)}{16} \frac{\Phi_{1c}^2}{\Phi_o^{*2}} - \frac{(1 + 4\gamma_0^2)e}{8q_o} \Psi_{1s} \frac{\Phi_{1c}}{\Phi_o^*} + \frac{3\gamma_0 e^2}{4q_o^2} \Psi_{1s}^2. \tag{8.24}
\end{aligned}$$

We have derived the first term on the right-hand side of formulas (8.15) and (8.16) by partial integration. These terms vanish if the initial plane z_o and the recording plane $z = z_r$ are located outside the multipole fields. We obtain in this case the second-rank aberrations at the image z_e of the diffraction plane z_d and the image z_i the object plane z_o by substituting the expression (8.8) of the third-rank perturbation eikonal $L_o^{(3)}(z_e) = L_o^{(3)}(z_i)$ for the eikonal $E_f^{(3)}$ in (8.3). The coefficients $A_{\lambda\mu\nu}$ (8.15) and $B_{\lambda\sigma\tau}$ (8.16) are of entirely geometric nature for $\lambda, \mu, \nu = 1, 3$ and $\sigma, \tau = 2, 4$, resulting in ten geometric eikonal terms. The six coefficients $C_{\mu\nu\kappa}$ (8.17) and $C_{\sigma\tau\kappa}$ (8.18) describe the primary chromatic aberration composed of the axial chromatic aberration ($\mu = \nu, \sigma = \tau$) and the chromatic distortion ($\mu \neq \nu, \sigma \neq \tau$). These aberrations limit the contrast and the resolution of low-voltage electron microscopes and of electron microscopes corrected for spherical aberration. The remaining two coefficients $D_{\mu\kappa\kappa}$ determine the second-degree dispersion.

It follows from (8.3) that any eikonal monomial results in two different types of aberrations, one observed at the image plane and the other at an image of the diffraction plane, which is generally the energy-selection plane in an analytical transmission electron microscope. We have illustrated this behavior in Table 8.1. At the image plane, the geometric aberrations that depend exclusively on the slope parameters $a_1 = \alpha$ and $a_2 = \beta$ represent the second-order axial aberrations. The mixed second-order aberrations are bilinear in either α and $a_3 = x_o$ or in β and $a_4 = y_o$, respectively. These aberrations cause an inclination of the image field and second-order field astigmatism. The eikonal terms that are quadratic in the field parameters x_o or y_o and linear in the slope parameters α or β produce the second-order distortions.

On going from the image plane to the image of the diffraction plane, we must exchange the meaning of the geometric ray parameters, because the fundamental field rays intersect the optic axis at the diffraction plane and, therefore, represent the fundamental axial rays with respect to this plane. Hence, x_o/f and y_o/f form in this case the initial slope components, where f is the focal length of the

Table 8.1 Properties of the coefficients of the third-rank eikonal monomials at the image of the object plane and the image of the diffraction plane, respectively

Image plane						
Type	Geometric aberrations			Chromatic aberrations		
	Aperture aberration	Distortion	Mixed aberration	Axial chromatic aberration	Chromatic distortion	Second-degree distortion
Coefficient	A_{111}	A_{133}	A_{113}	C_{11k}	C_{13k}	D_{1kk}
	B_{122}	B_{243}	B_{124}	C_{22k}	C_{24k}	
		B_{144}	B_{223}			
Image of the diffraction plane						
Type	Geometric aberrations			Chromatic aberrations		
	Aperture aberration	Distortion	Mixed aberration	Axial chromatic aberration	Chromatic distortion	Second-degree distortion
Coefficient	A_{333}	A_{133}	A_{133}	C_{33k}	C_{13k}	D_{3kk}
	B_{344}	B_{124}	B_{243}	C_{44k}	C_{24k}	
		B_{223}	B_{144}			

objective lens. Correspondingly, the parameters αf and βf determine the field of view of the diffraction plane. The same behavior holds for the chromatic aberrations, as demonstrated in Table 8.1.

The integrals (8.17) and (8.18) for the chromatic coefficients do not vanish in the absence of multipole fields because all round lenses located between the object plane and the observation plane also contribute to the chromatic aberrations. This behavior becomes obvious if we reshape by partial integrations the integrals with integrands $x'_\mu x'_\nu$ and $y'_\sigma y'_\tau$, respectively. The eikonal terms, which depend solely on the field parameters a_3 and a_4 and the chromatic parameter κ , do not contribute to the aberrations in the image, yet they determine the axial aberrations at the image z_e of the diffraction plane z_d . Conversely, the eikonal terms containing exclusively the slope parameters $a_1 = \alpha$ and $a_2 = \beta$ do not contribute to the aberrations at the plane z_e , which is the energy-selection plane in an energy-filtering electron microscope. The eikonal terms producing the mixed field aberrations at the image plane cause axial aberrations at the energy-selection plane, preventing *isochromatic* energy filtering [74]. As a result, the energy selected by the slit aperture depends on the lateral position of the individual object elements. This behavior may falsify the information about the chemical composition of the object because material with a characteristic energy loss will only become visible in the inner region of the object, although it may be present in its outer region and vice versa.

8.1.2 Systems with Straight Optic Axis

The fields of systems with straight optic axis do not possess dipole components apart from the Wien filter. If we require in addition that the systems do not contain field components with odd multiplicity, the resulting second-rank aberrations comprise

only chromatic aberrations of first order and first degree. Chromatic aberrations limit the contrast and the resolution of low-voltage electron microscopes and of electron microscopes corrected for spherical aberration. Employing the rotating $u - z$ coordinate system (4.21) and assuming distortion-free stigmatic imaging, $u_4(z_i) = iu_3(z_i)$, we obtain from (7.69) the chromatic aberrations at the image plane $u_1(z_i) = u_2(z_i) = 0$ in the simple form:

$$u_c^{(2)}(z_i) = -2u_{3i} \frac{\partial E_{ci}^{(3)}}{\partial \bar{\omega}}. \quad (8.25)$$

We have derived the modified third-rank eikonal

$$\begin{aligned} E_{ci}^{(3)} &= E_c^{(3)}(z_i) = \int_{z_0}^{z_i} \mu_{c1}^{(3)}(z) dz \\ &= \frac{\kappa \Phi_0}{4\Phi_0^*} \operatorname{Re} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \left[\gamma_0 u^{(1)'} \bar{u}^{(1)'} + \left(\frac{1}{4} \frac{\Phi''}{\Phi^*} + \gamma_0 \chi'^2 \right) u^{(1)} \bar{u}^{(1)} \right. \\ &\quad \left. - \frac{\Phi_2}{\Phi^*} u^{-(1)2} + i \chi' (u^{(1)} u^{-(1)'} - u^{(1)'} u^{-(1)}) \right] dz \end{aligned} \quad (8.26)$$

by substituting the paraxial ray $w^{(1)}(z) = u^{(1)}(z) \exp(i\chi)$ for the true ray w in the expression (7.92) for the third-rank variational polynomial $\mu_c^{(3)}$; ε is the relativistic parameter (4.4) and $\chi = \chi(z)$ the angle of rotation (4.24) of the u -coordinate system. To survey the structure of the integrand in more detail, we reshape the first term by partial integration giving

$$\int_{z_0}^{z_i} \frac{\gamma_0}{\sqrt{\Phi^*}} u^{(1)'} \bar{u}^{(1)'} dz = \frac{\gamma_0}{\sqrt{\Phi^*}} u^{(1)'} \bar{u}^{(1)} \Big|_{z_0}^{z_i} - \int_{z_0}^{z_i} \left[\frac{\gamma_0}{\sqrt{\Phi^*}} \bar{u}^{(1)} u^{(1)''} - \frac{\Phi'}{2\Phi^{3/2}} \bar{u}^{(1)} u^{(1)'} \right] dz. \quad (8.27)$$

Subsequently, we eliminate the second derivative $u^{(1)''}$ by means of the paraxial path equation (4.25) and insert the result into the expression (8.26). Finally, we remove the second derivative of the axial potential by partial integration to give

$$\begin{aligned} E_{ci}^{(3)} &= \frac{\kappa}{4} \Phi_0 \left(\frac{\gamma_{0i}}{\Phi_1^*} - \frac{\gamma_{0o}}{\Phi_0^*} \right) \operatorname{Re} (\bar{\omega} w_o) - \frac{\kappa}{2(1 + \varepsilon \Phi_0)} \int_{z_0}^{z_i} \gamma_0 \sqrt{\frac{\Phi_0^*}{\Phi^*}} \chi' \operatorname{Im} (u^{(1)} \bar{u}^{(1)'}) dz \\ &\quad + \frac{\kappa \Phi_0}{\Phi_0^*} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \left[\left(\frac{\gamma_0 e B^2}{16 m_e \Phi^*} + \frac{\gamma_0 (2 + \gamma_0)}{16} \frac{\Phi'^2}{\Phi^{*2}} \right) \bar{u}^{(1)} u^{(1)} \right. \\ &\quad \left. - \operatorname{Re} \left\{ \left(\frac{1 + \gamma_0^2}{4} \frac{\Phi_2}{\Phi^*} + i \frac{\gamma_0 e}{2q} \Psi_2 \right) e^{-2i\chi} \bar{u}^{(1)2} \right\} \right] dz. \end{aligned} \quad (8.28)$$

We have derived the first term by employing the expression (8.6) for the paraxial ray and the Helmholtz–Lagrange relations for the fundamental rays.

8.1.3 Axial Chromatic Aberration and Chromatic Distortion

The total second-rank chromatic aberration at the image plane of an electron microscope consists of two kinds: the axial chromatic aberration and the chromatic distortion. The first component affects the resolution, while the second component represents the chromaticity of the magnification. The axial chromatic aberration of arbitrary systems with straight axis is composed of the chromatic defocus and the axial chromatic astigmatism. The chromatic distortion has also two terms, which we characterize as round lens or *regular chromatic distortion* and as *odd chromatic distortion*. We derive the components of the chromatic aberration by inserting the representation (4.225) $u^{(1)} = \omega u_\omega + \bar{\omega} u_{\bar{\omega}} + \rho w_\rho + \bar{\rho} w_{\bar{\rho}}$ for the axial trajectory into the integrands of the integrals (8.28). We write the result in the form

$$E_{\text{ci}}^{(3)} = \kappa \left[\frac{1}{2} C_c \omega \bar{\omega} + \text{Re} \left(\frac{1}{2} A_c \bar{\omega}^2 + D_{\text{cr}} \bar{\omega} \rho + D_{\text{ce}} \bar{\omega} \bar{\rho} \right) \right]. \quad (8.29)$$

Here, we have omitted terms, which solely depend on the complex off-axis coordinates $\rho = a_3 + i a_4 = w_o$ and $\bar{\rho}$ because they do not contribute to the chromatic aberration (8.25) at the image plane. Substituting the right-hand side of (8.29) for $E_{\text{ci}}^{(3)}$ in (8.25) and considering that $\rho = u_o = w_o$, we find the chromatic aberration as

$$u_{\text{ci}}^{(2)} = -u_{\text{ci}} \kappa (C_c \omega + A_c \bar{\omega} + D_{\text{cr}} w_o + D_{\text{ce}} \bar{w}_o). \quad (8.30)$$

The aberration coefficient C_c of the chromatic defocus is real, whereas the coefficient A_c of the axial chromatic astigmatism, the coefficients D_{cr} of the chromatic round-lens distortion, and the coefficient D_{ce} of the elliptical chromatic distortion are generally complex. We obtain the coefficients of the axial chromatic aberrations as

$$\begin{aligned} C_c &= \frac{1}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} [T_c (u_\omega \bar{u}_\omega + u_{\bar{\omega}} \bar{u}_{\bar{\omega}}) - 4 \text{Re} (G_c \bar{u}_\omega \bar{u}_{\bar{\omega}}) \\ &\quad - \gamma_0 \chi' \text{Im} (u_\omega \bar{u}'_{\bar{\omega}} + u_{\bar{\omega}} \bar{u}'_\omega)] dz, \\ A_c &= \frac{2}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} [T_c u_\omega u_{\bar{\omega}} - G_c \bar{u}_\omega^2 - \bar{G}_c u_\omega^2 - i \gamma_0 \chi' (u_\omega \bar{u}'_{\bar{\omega}} - u_{\bar{\omega}} \bar{u}'_\omega)] dz, \end{aligned} \quad (8.31)$$

where

$$T_c = \frac{\gamma_0 e B^2}{8 m_e \Phi^*} + \frac{\gamma_0 (2 + \gamma_0)}{8} \frac{\Phi'^2}{\Phi^{*2}}, \quad (8.32)$$

$$G_c = \frac{\gamma_0}{4} G + \frac{1}{4} \frac{\Phi_2}{\Phi^*} e^{-2i\chi} = \left[\frac{1 + \gamma_0^2}{4} \frac{\Phi_2}{\Phi^*} + i \frac{\gamma_0 e}{2q} \Psi_2 \right] e^{-2i\chi}. \quad (8.33)$$

The coefficient C_c of the chromatic defocus is real, whereas the coefficient A_c of the axial chromatic astigmatism is generally complex. By changing the focal length of the objective lens, we can eliminate the chromatic defocus for any fixed value of the chromatic parameter κ . However, any realistic electron beam has a continuous energy spread. Therefore, we can only minimize the chromatic defocus by adjusting the focal length of the objective lens in such a way that it focuses sharply the electrons with mean energy. The same considerations hold for the axial astigmatism, which we can compensate for any fixed energy deviation by means of a quadrupole stigmator. The last term of the first integrand vanishes for orthogonal systems since in this case the paraxial pseudo rays (4.227) are real. In the presence of an axial chromatic astigmatism, the two terms (8.31) create an elliptical aberration disc in the Gaussian image plane. The shape of this disc degenerates into a circle in the case of rotational symmetry, where we have $u_{\bar{\omega}} = 0, u_{\omega} = \bar{u}_{\omega} = u_1 = u_{\varepsilon}, G_c = 0$ giving $A_c = 0$ and

$$C_c = \frac{1}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \gamma_0 \sqrt{\frac{\Phi_0^*}{\Phi^*}} \left(\frac{e B^2}{8 m_e \Phi^*} + \frac{2 + \gamma_0}{8} \frac{\Phi'^2}{\Phi^{*2}} \right) u_{\alpha}^2 dz > 0. \quad (8.34)$$

Hence, the axial chromatic aberration of round lenses is unavoidable. In particular, these lenses always focus particles with energy deviation $\Delta E > 0$ less strongly and particles with $\Delta E < 0$ more strongly than particles with nominal energy ($\Delta E = 0$), as illustrated in Fig. 8.2. This statement is part of the Scherzer theorem.

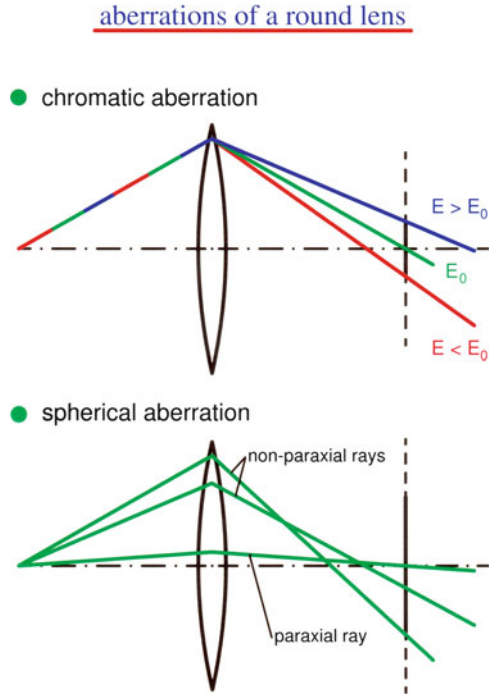
The coefficient of the chromatic round-lens distortion has the form

$$\begin{aligned} D_{cr} = & \frac{1}{4(1 + \varepsilon \Phi_0)} \left(\gamma_{0i} \frac{\Phi_0^*}{\Phi_i^*} - \gamma_{0o} \right) + \frac{1}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \{ T_c (u_{\rho} \bar{u}_{\omega} + u_{\bar{\omega}} \bar{u}_{\rho}) \\ & - 2 G_c \bar{u}_{\omega} \bar{u}_{\rho} - 2 \bar{G}_c u_{\bar{\omega}} u_{\rho} + i \gamma_0 \chi' (u_{\rho} \bar{u}'_{\omega} + u_{\bar{\omega}} \bar{u}'_{\rho} - \bar{u}_{\omega} u'_{\rho} - \bar{u}_{\rho} u'_{\bar{\omega}}) / 2 \} dz. \end{aligned} \quad (8.35)$$

The coefficient of the elliptical chromatic distortion is given by

$$\begin{aligned} D_{ce} = & \frac{1}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \{ T_c (u_{\bar{\rho}} \bar{u}_{\omega} + u_{\bar{\omega}} \bar{u}_{\rho}) - 2 G_c \bar{u}_{\bar{\omega}} \bar{u}_{\rho} - 2 \bar{G}_c u_{\bar{\omega}} u_{\rho} \\ & + i \gamma_0 \chi' (u_{\bar{\rho}} \bar{u}'_{\omega} + u_{\bar{\omega}} \bar{u}'_{\rho} - \bar{u}_{\rho} u'_{\bar{\omega}} - \bar{u}_{\omega} u'_{\bar{\rho}}) / 2 \} dz. \end{aligned} \quad (8.36)$$

Fig. 8.2 Axial rays passing through an imperfect lens causing axial chromatic and spherical aberration at the Gaussian image plane



In the special case of rotationally symmetric fields, the coefficient (8.36) of the elliptical chromatic distortion is zero since $u_{\bar{\omega}} = u_{\bar{\rho}} = 0$ and $G_c = 0$. In addition, we have $u_{\omega} = \bar{u}_{\omega} = u_{\alpha} = u_1$ and $u_{\rho} = \bar{u}_{\rho} = u_{\gamma} = u_3$. Considering the Helmholtz–Lagrange relation $\sqrt{\Phi^*}(u_{\gamma}u'_{\alpha} - u_{\alpha}u'_{\gamma}) = \sqrt{\Phi_0^*}$ for these rays, the coefficient (8.35) of the regular chromatic distortion adopts for round lenses the form

$$D_{cr} = \Phi_0 \left(\frac{\gamma_{0i}}{\Phi_i^*} - \frac{\gamma_{0o}}{\Phi_o^*} \right) + \frac{i}{2} \sqrt{\frac{e}{8m_e}} \int_{z_0}^{z_1} \frac{\gamma_0 \Phi_0 B}{\Phi^{*3/2}} dz + \int_{z_0}^{z_1} \sqrt{\frac{\Phi_o^2}{\Phi^* \Phi_o^*}} T_c u_{\alpha} u_{\gamma} dz. \quad (8.37)$$

This coefficient is complex for magnetic lenses. The real part accounts for the isotropic or radial chromatic distortion and the imaginary part for the anisotropic or azimuthal chromatic distortion. The latter distortion results from the fact that the angle of Larmor rotation χ depends on the axial potential Φ or on the energy of the particles, respectively. Therefore, a change of their energy causes an image rotation for magnetic round lenses.

By setting $A_c = D_{co} = 0$ in (8.30), we obtain the chromatic aberration of rotationally symmetric systems as

$$u_{ci}^{(2)} = -M\kappa (C_c \omega + D_{cr} w_o). \quad (8.38)$$

The magnification $M = u_{\gamma i}$ of the final image is positive if the number of intermediate images between the object and the final image plane is odd and negative if it is even. The real part of the distortion coefficient (8.37) produces a shift of the image point in the direction of w_o , which is the radial direction of the image point in the rotated image. The observed chromatic aberration at any point in the image consists of a superposition of deviations of many rays originating from the conjugate object point w_o with different slope angle ω and energy deviation κ . As a result, the Gaussian image point becomes a spot reducing contrast and resolution in high-performance electron microscopes.

The objective lens contributes the most to the chromatic defocus of an electron microscope because the integrand of C_c (8.34) is proportional to the square of the fundamental axial ray u_ω , which is relatively large within this lens. Owing to this quadratic factor, the contribution of the subsequent lenses decreases in proportion to $1/M_n^2$ with increasing magnification M_n of the image in front of the n th intermediate lens. Since these magnifications are large compared to unity, the contribution of the intermediate lenses and the projector lens to the axial chromatic aberration is negligibly small. However, all lenses contribute with about the same order of magnitude to the coefficient (8.37) of the chromatic distortion because the field ray u_ρ increases if the axial ray u_ω decreases such that their product does not vary appreciably within the constituent lenses of the microscope. The integrand of D_{cr} and is proportional to this product, which can be positive or negative. Therefore, it is possible to eliminate the radial chromatic distortion by arranging and exciting the intermediate lenses appropriately. The azimuthal chromatic distortion depends on the sign of the magnetic field. Hence, by alternating the directions of the currents within the coils of consecutive magnetic lenses, we can compensate for the chromatic image rotation. One utilizes these possibilities for compensating the chromatic distortion in high-performance transmission electron microscopes.

The expressions (8.31), (8.35), and (8.36) for the chromatic aberration coefficients reduce considerably if we employ orthogonal quadrupole systems ($G = \overline{G}$) and require that the paraxial path of rays is stigmatic within the field of the round lenses ($u_{\bar{\omega}} = u_{\bar{\rho}} = 0$ if $B \neq 0, \Phi' \neq 0$). We can satisfy this condition by placing anastigmatic quadrupole systems in the regions between the round lenses. In order that the path of rays is staying stigmatic within the entire region of the round lenses, their fringing fields must not overlap those of the quadrupoles. In this case, the fundamental pseudo rays $u_\omega, u_{\bar{\omega}}, u_\rho$ and $u_{\bar{\rho}}$ are real so that the total chromatic aberration is composed of the chromatic aberration of conventional round lenses and that of orthogonal quadrupole systems having real aberration coefficients. We obtain the aberration coefficients for the latter systems by setting $\overline{G}_c = G_c$ and $T_c = 0, \chi' = 0$ in the formulas (8.31), (8.35), and (8.36). These requirements simplify considerably the addition of the aberrations introduced by the constituent subsystems because we obtain the aberration coefficients of each subsystem by assuming stigmatic initial conditions. In this case, we need only to consider the change in magnification for adding equivalent aberration coefficients of subsequent compounds.

8.2 Third-Order Aberrations of Systems with Straight Axis

In most electron optical systems with straight optic axis, the chromatic parameter κ is small compared with the geometrical beam parameters. In the absence of dipole and hexapole fields, the chromatic aberrations of third rank are negligibly small compared with the geometric third-order aberrations because the chromatic part (7.55) of the fourth-rank variational function has only terms that are quadratic in the chromatic parameter κ . If we assume in addition the absence of hexapole fields, we need to consider only the primary third-order aberrations. We obtain these geometric aberrations in the first step of the iteration procedure from the modified fourth-order eikonal:

$$E_{\text{go}}^{(4)}(z_i) = E_{\text{gi}}^{(4)} = \int_{z_0}^{z_i} \mu_{\text{gl}}^{(4)}(z) dz. \quad (8.39)$$

We derive the geometric fourth-order variational function $\mu_{\text{g}}^{(4)}$ of systems with straight optic axis and even multipoles from (7.54) by setting $\Gamma = 0$, $\Phi_1 = \Phi_3 = 0$, and $\Psi_1 = \Psi_3 = 0$. As a result, we obtain

$$\begin{aligned} \mu_{\text{g}}^{(4)} = & \sqrt{\frac{\Phi^*}{\Phi_0^*}} \operatorname{Re} \left\{ -\frac{1}{8} w'^2 \bar{w}'^2 - \frac{\gamma_0}{16} \frac{\Phi''}{\Phi^*} w' \bar{w}' w \bar{w} + i \frac{e}{16q} B'' \bar{w}' \bar{w} w^2 \right. \\ & + \frac{\gamma_0}{4} \frac{\Phi_2}{\Phi^*} w' \bar{w}' \bar{w}^2 + i \frac{e}{4q} \Psi_2' \bar{w}' w \bar{w}^2 - i \frac{e}{12q} \Psi_2' w' \bar{w}^3 + \left[\frac{\gamma_0}{2} \frac{\Phi_4}{\Phi^*} - \frac{1}{16} \frac{\Phi_2^2}{\Phi^{*2}} + i \frac{e}{q} \Psi_4 \right] \bar{w}^4 \\ & \left. + \left[\frac{1}{16} \frac{\Phi_2 \Phi''}{\Phi^{*2}} - \frac{\gamma_0}{24} \frac{\Phi_2''}{\Phi^*} \right] w \bar{w}^3 + \left[\frac{1}{128} \frac{\Phi''''}{\Phi^*} - \frac{1}{128} \frac{\Phi''^2}{\Phi^{*2}} - \frac{1}{16} \frac{\Phi_2 \bar{\Phi}_2}{\Phi^{*2}} \right] w^2 \bar{w}^2 \right\}. \end{aligned} \quad (8.40)$$

This expression is rather simple compared with that for arbitrary systems (7.54). However, it results in numerous aberration monomials with rather involved coefficients if the paraxial path equation does not decouple. In order to minimize the number of aberrations from the very beginning, we restrict our investigations to orthogonal systems and require that the axial magnetic field does not overlap the quadrupole and octopole fields. We satisfy this requirement by encapsulating the solenoids inside of rotationally symmetric iron pole pieces. This procedure confines the magnetic field within a short region centered in the gap between the pole pieces. All magnetic round lenses of electron microscopes are constructed in this way.

8.2.1 Structure of the Geometrical Eikonal Polynomials

In accordance with the light-optical convention, we classify the geometrical aberrations with respect to their Seidel order n . Since we obtain these aberrations from

polynomials of the perturbation eikonal, not all coefficients of the individual aberration monomials are independent from each other. A multipole with multiplicity m only affects eikonal polynomials whose rank is equal or higher than m . To survey the effect of the multipoles on the individual terms of an eikonal polynomial, it is advantageous to represent the polynomials as a sum of monomials in the four complex ray parameters $\omega, \bar{\omega}, \rho,$ and $\bar{\rho}$. If we define the ray at the object plane, we obtain most conveniently all aberrations from the polynomials

$$L_{\text{gi}}^{(r)} = \hat{E}_{\text{gi}}^{(r)}, \quad r = n \geq 3 \quad (8.41)$$

of the modified perturbation eikonal. To elucidate optimally the nature of the geometrical polynomials of the perturbation eikonal, we separate them according to the “parity” of their order n . Polynomials of order $n = 2s, s = 1, 2, \dots$, have *even* parity; those of order $n = 2s + 1$ have *odd* parity. This separation enables us to write each polynomial as a sum of sub-polynomials with different multiplicities $m = 2\mu$ and $m = 2\mu + 1$, respectively:

$$L_{\text{gi}}^{(n)} = L_{\text{gi}}^{(2s)} = \sum_{\mu=0}^s L_{\text{gi}}^{(2s, 2\mu)},$$

$$L_{\text{gi}}^{(2s, 2\mu)} = \text{Re} \sum_{\nu=0}^{s+\mu} \sum_{\lambda=0}^{s-\mu} L_{\nu\lambda}^{(2s, 2\mu)} \bar{\omega}^{s+\mu-\nu} \omega^{s-\mu-\lambda} \bar{\rho}^{\nu} \rho^{\lambda}, \quad (8.42)$$

$$L_{\text{gi}}^{(n)} = L_{\text{gi}}^{(2s+1)} = \sum_{\mu=0}^s L_{\text{gi}}^{(2s+1, 2\mu+1)},$$

$$L_{\text{gi}}^{(2s+1, 2\mu+1)} = \text{Re} \sum_{\nu=0}^{s+\mu+1} \sum_{\lambda=0}^{s-\mu} L_{\nu\lambda}^{(2s+1, 2\mu+1)} \bar{\omega}^{s+\mu-\nu+1} \omega^{s-\mu-\lambda} \bar{\rho}^{\nu} \rho^{\lambda}. \quad (8.43)$$

These representations reveal that polynomials with even order $n = 2s$ only contain sub-polynomials with even multiplicity $m = 2\mu \leq 2s$, whereas those of odd order $n = 2s + 1$ have solely sub-polynomials with odd multiplicity $m = 2\mu + 1 \leq 2s + 1$. If we place a $2N$ -multipole with N -fold symmetry of its field in the stigmatic paraxial region, this element produces among others a primary eikonal term of order N and multiplicity $m = N$ and a secondary rotationally symmetric term ($m = 0$) of rank $n = 2N - 2$. We shall utilize this unexpected term for eliminating the third-order spherical aberration of round lenses by means of sextupoles ($N = 3$).

The eikonal coefficients $L_{\nu\lambda}^{(n, m)}$ are generally complex if m is nonzero. The coefficients with multiplicity $m = 0$ satisfy the relation

$$L_{\nu\lambda}^{(2s, 0)} = \overline{L}_{\lambda\nu}^{(2s, 0)}. \quad (8.44)$$

Accordingly, all coefficients $L_{\lambda\lambda}^{(2s, 0)}$ are real. Owing to the nonlinearity of the variational function, these rotationally symmetric terms result not only from

round lenses but also from multipoles. This behavior is partly demonstrated by the representations (7.54) and (8.40) of the fourth-order variational polynomial. We can also directly create a fourth-rank rotationally symmetric eikonal term by placing an octopole element in the astigmatic paraxial region. For example, one exploits this possibility for correcting the unavoidable third-order spherical aberration of round lenses by means of a corrector consisting of quadrupoles and octopoles.

The eikonal of rotationally symmetric systems has only expansion terms with multiplicity zero. Moreover, systems composed of multipole elements with even-fold symmetry do not introduce eikonal polynomials with odd multiplicity. If we align azimuthally the constituent multipole elements in such a way that their principal sections coincide, all eikonal coefficients are real ($L_{\nu\lambda}^{(2s,2\mu)} = \overline{L_{\nu\lambda}^{(2s,2\mu)}}$). This is the case for orthogonal systems with plane-section symmetry. In the most general case, the polynomial $L_{\text{gi}}^{(n)}$ has $(n+1)(n+2)(n+3)/6$ linearly independent real coefficients. Each of its sub-polynomials $L_{\text{gi}}^{(n,m)}$ has

$$N_m^{(r)} = \frac{(n+2)^2 - m^2}{4} \quad (8.45)$$

complex coefficients for $m \neq 0$. The rotationally symmetric polynomial $L_{\text{gi}}^{(2s,0)}$ has $s+1$ real and at most $s(s+1)/2$ complex coefficients totaling $(s+1)^2 = (n+2)^2/4$ real coefficients. Accordingly, the fourth-order eikonal polynomial of rotationally symmetric systems has three real coefficients and three complex coefficients. Their imaginary parts originate from the Larmor rotation. Hence, they are zero in the case for electrostatic round lenses.

To keep the number of monomials of any given polynomial as small as possible, it is advantageous to impose symmetry conditions on the system as a whole and on individual parts of it. For example, this procedure enables one to eliminate the sub-polynomials with multiplicity $m=2$ of special orthogonal systems. In most cases, we fix the ray by its complex lateral position $\rho = u_o = w_o$ at the object plane z_o and by the complex aperture angle ω . The eikonal monomials, which are independent of $\rho = w_o$ and $\bar{\rho} = \bar{w}_o$, produce aperture aberrations in the image plane of the object. The monomials, which depend solely on the object coordinates, do not contribute to the aberrations in the image plane, but they cause aperture aberrations in images of the diffraction plane, as we have discussed in the context of the second-order aberrations of imaging energy filters. In the following, we discuss in detail the connection of the fourth-order eikonal monomials with the third-order aberrations of round lenses.

8.3 Geometrical Aberrations of Round Lenses

The primary geometrical aberrations of round lenses are of third order because in the case of rotational symmetry, the first nonvanishing polynomial of the perturbation eikonal is of fourth order. We obtain this polynomial from the formula (8.42) by

setting $s = 2$ and $\rho = w_0$ and considering that only monomials with multiplicity $m = 2\mu = 0$ contribute to the fourth-order perturbation eikonal of round lenses as

$$\begin{aligned} L_{\text{gR}}^{(4)} &= L_{\text{gi}}^{(4,0)} = \text{Re} \sum_{\nu=0}^2 \sum_{\lambda=0}^2 L_{\nu\lambda}^{(4,0)} \bar{\omega}^{2-\nu} \omega^{2-\lambda} \bar{w}_0^\nu w_0^\lambda \\ &= \text{Re} \sum_{\lambda=0}^2 \sum_{\nu=0}^{\lambda} \frac{2}{1 + \delta_{\nu\lambda}} L_{\nu\lambda}^{(4,0)} \bar{\omega}^{2-\nu} \omega^{2-\lambda} \bar{w}_0^\nu w_0^\lambda. \end{aligned} \quad (8.46)$$

By employing this representation, we readily derive the total third-order aberration at the stigmatic Gaussian image plane:

$$u_i^{(3)} = -2u_{\gamma i} \frac{\partial L_{\text{gR}}^{(4)}}{\partial \bar{\omega}} = -2u_{\gamma i} \sum_{\lambda=0}^2 \sum_{\nu=0}^1 (2-\nu) L_{\nu\lambda}^{(4,0)} \bar{\omega}^{1-\nu} \omega^{2-\lambda} \bar{w}_0^\nu w_0^\lambda. \quad (8.47)$$

The sum consists of six terms, which account for five types of aberration. In electron optics, one refers the total third-order aberration of round lenses back to the object plane and defines it by the representation

$$u_i^{(3)}/u_{\gamma i} = C_3 \omega^2 \bar{\omega} + 2K_3 \omega \bar{\omega} w_0 + \bar{K}_3 \omega^2 \bar{w}_0 + F_3 \omega w_0 \bar{w}_0 + A_{\text{f3}} \bar{\omega} w_0^2 + D_3 w_0^2 \bar{w}_0. \quad (8.48)$$

We have chosen the notation of the coefficients due to mnemonic reasons apart from the coefficient C_3 of the third-order spherical aberration. In many books on electron microscopy, this coefficient is denoted as C_s . We do not follow this widely used notation because it may lead to the wrong conclusion that the entire spherical aberration vanishes if one nullifies this coefficient. The comparison of the representation (8.48) with the result (8.47) obtained from the perturbation eikonal gives the following relations between the aberration coefficients and the coefficients of the eikonal:

$$C_3 = -4L_{00}^{(4,0)}, \quad K_3 = -2L_{01}^{(4,0)}, \quad F_3 = -2L_{11}^{(4,0)}, \quad A_{\text{f3}} = -4L_{02}^{(4,0)}, \quad D_3 = -2L_{12}^{(4,0)}. \quad (8.49)$$

These relations reveal that the aberration coefficients have the opposite sign with respect to the associated eikonal coefficients. This confusing stipulation goes back to the early days of electron optics and was chosen primarily to obtain a positive coefficient $C_3 = C_s$ for the spherical aberration. The real eikonal coefficient $L_{22}^{(4,0)} = \bar{L}_{22}^{(4,0)}$ does not affect the aberration at the image plane, but it produces a spherical aberration at the image of the diffraction plane. The coefficient C_3 of the spherical aberration and the coefficient F_3 of the *field curvature* are always real, as follows from (8.49) and (8.44). The coefficients K_3 , A_{f3} and D_3 are complex for magnetic round lenses because the Larmor rotation of the outer rays differs from that of the paraxial rays. This difference causes a rotation of the aberration figures

of *off-axial coma* (K_3), *field astigmatism* (A_{f3}), and *third-order distortion* (D_3). We use the notation A_{f3} for the coefficient of the third-order field astigmatism in order to differ it from that (A_3) of the third-order axial astigmatism. The names of the different terms of the third-order aberration are chosen in accordance with the definition of light optics. Each name characterizes a characteristic feature of the associated aberration figure. These figures depend in different ways on the complex object position w_0 and aperture angle ω . In order to discuss each term in detail, we need to find the structure of the aberration coefficients or the structure of the coefficients of the fourth-order eikonal polynomial, respectively. We readily obtain the eikonal coefficients $L_{\nu\lambda}^{(4,0)}$ from the integral expression (8.39) of the modified fourth-order eikonal by setting $\Phi_2 = \Psi_2 = \Phi_4 = \Psi_4 = 0$ in the integrand (8.40) resulting in

$$L_{gR}^{(4)} = - \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left\{ \frac{1}{8} w'^2 \bar{w}'^2 + \frac{\gamma_0}{16} \frac{\Phi''}{\Phi^*} w' \bar{w}' w \bar{w} + \frac{1}{128} \left(\frac{\Phi''^2}{\Phi^{*2}} - \gamma_0 \frac{\Phi''''}{\Phi^*} \right) w^2 \bar{w}^2 \right\} dz \\ - \frac{e}{16q_0} \int_{z_0}^{z_i} \text{Im}(B'' \bar{w}' w^2 \bar{w}) dz. \quad (8.50)$$

We must substitute the paraxial approximation

$$w^{(1)} = e^{i\chi} u^{(1)}, \quad u^{(1)} = \omega u_\alpha + w_0 u_\gamma, \quad (8.51)$$

for the true ray w in the integrands. Since the fundamental rays $u_1 = u_\alpha$ and $u_3 = u_\gamma$ are real for axially symmetric systems, it is advantageous to express the integrand in terms of the familiar rays u_α and u_γ . Introducing the rotating coordinate system with the aid of (8.51) and neglecting for simplicity the superscripts (1), we find

$$w' \bar{w}' = u' \bar{u}' + \chi'^2 u \bar{u} + i \chi' (u \bar{u}' - \bar{u} u') \\ = \omega \bar{\omega} (u_\alpha'^2 + \chi'^2 u_\alpha^2) + (\omega \bar{\omega}_0 + \bar{\omega} w_0) (u_\alpha' u_\gamma' + \chi'^2 u_\alpha u_\gamma) + w_0 \bar{w}_0 (u_\gamma'^2 + \chi'^2 u_\gamma^2) \\ - 2 \sqrt{\frac{\Phi_0^*}{\Phi^*}} \chi' \text{Im}(\bar{\omega} w_0). \quad (8.52)$$

We have derived the last term by employing the Lagrange–Helmholtz relation $\sqrt{\Phi^*} (u_\gamma u_\alpha' - u_\alpha u_\gamma') = \sqrt{\Phi_0^*}$ of the fundamental rays u_α and u_γ . For numerical accuracy, terms with high derivatives of B and/or Φ are generally undesirable. We may also wish to find out whether certain aberration coefficients can change sign. In this case, we inquire whether or not the integrand of the aberration integral can be written as a sum of squared terms. By means of partial integrations, we can reduce high derivatives of the axial field strengths. The resulting second derivatives u'' are eliminated by means of the paraxial path equation. Using this method for eliminating the second derivative of B and considering

$$\text{Im}(u\bar{u}') = \sqrt{\frac{\Phi_o^*}{\Phi^*}} \text{Im}(\bar{\omega}w_o), \quad \chi'' = \frac{e}{2q} \left(B' - \frac{\gamma_0}{2} \frac{\Phi'}{\Phi^*} B \right), \quad (8.53)$$

we obtain

$$\begin{aligned} \text{Im} \int_{z_0}^{z_i} B'' \bar{w}' w^2 \bar{w} dz &= \text{Im} \int_{z_0}^{z_i} B'' (\bar{u}' u^2 \bar{u} - i \chi' u^2 \bar{u}^2) dz = B' [\text{Im}(u\bar{u}') - \chi' u\bar{u}] u\bar{u} \Big|_{z_0}^{z_i} \\ &\quad - \int_{z_0}^{z_i} B' \text{Im} (2u\bar{u}\bar{u}' u' + u^2 \bar{u}'^2 + u^2 \bar{u}\bar{u}'') dz + \int_{z_0}^{z_i} [B' \chi'' u^2 \bar{u}^2 \\ &\quad + 2B' \chi' u\bar{u}(\bar{u}u' + u\bar{u}')] dz \\ &= B' \left(\sqrt{\Phi_o^*/\Phi^*} \text{Im}(\bar{\omega}w_o) - \chi' u\bar{u} \right) u\bar{u} \Big|_{z_0}^{z_i} \\ &\quad - \text{Im}(\bar{\omega}w_o) \int_{z_0}^{z_i} \sqrt{\frac{\Phi_o^*}{\Phi^*}} B' \left(u\bar{u}' + \bar{u}u' - \frac{\gamma_0 \Phi'}{2\Phi^*} u\bar{u} \right) dz \\ &\quad + \int_{z_0}^{z_i} \left[\frac{e}{2q} B'^2 - \frac{e\gamma_0}{4q} \frac{\Phi'}{\Phi^*} BB' + \frac{e}{q} BB' \left(\frac{u'}{u} + \frac{\bar{u}'}{\bar{u}} \right) \right] u^2 \bar{u}^2 dz. \end{aligned} \quad (8.54)$$

The first term of the final relation vanishes if the magnetic field gradient is zero at the object and the image plane. Otherwise, it contributes to the distortion yet not to the other aberrations because the fundamental axial ray u_α is zero at the object and image planes. The second term of the final relation contributes to the anisotropic field aberrations, while the remaining integral adds a term to all isotropic aberrations or to the real part of each eikonal coefficient, respectively.

By inserting (8.51)–(8.54) into (8.50) and ordering the result in a sum of monomials, we eventually derive the following expressions for the coefficients of the eikonal polynomials:

$$L_{00}^{(4,0)} = -\frac{1}{16} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_o^*}} (2u_\alpha'^4 + h_1 u_\alpha'^4 + h_2 u_\alpha'^2 u_\gamma'^2 + 2h_3 u_\alpha'^3 u_\gamma') dz, \quad (8.55)$$

$$\begin{aligned} L_{11}^{(4,0)} &= -\frac{1}{4} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_o^*}} \left[2 \frac{u_\alpha'^2 u_\gamma'^2}{u_\alpha'^2 u_\gamma'^2} + h_1 + \frac{h_2}{4} \left(\frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right)^2 \right. \\ &\quad \left. + h_3 \left(\frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right) \right] u_\alpha'^2 u_\gamma'^2 dz - \frac{1}{4} \int_{z_0}^{z_i} h_4 dz, \end{aligned} \quad (8.56)$$

$$\begin{aligned}
L_{01}^{(4,0)} = & -\frac{1}{16} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left[4 \frac{u_\alpha'^3 u_\gamma'}{u_\alpha^3 u_\gamma} + 2h_1 + h_2 \left(\frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right) \frac{u_\alpha'}{u_\alpha} \right. \\
& \left. + h_3 \left(3 \frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right) \right] u_\alpha^3 u_\gamma dz \\
& - \frac{i}{8} \int_{z_0}^{z_i} \left(h_5 + h_6 \frac{u_\alpha'^2}{u_\alpha^2} + h_7 \frac{u_\alpha'}{u_\alpha} \right) u_\alpha^2 dz,
\end{aligned} \tag{8.57}$$

$$\begin{aligned}
L_{02}^{(4,0)} = & -\frac{1}{16} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left[2 \frac{u_\alpha'^2 u_\gamma'^2}{u_\alpha^2 u_\gamma^2} + h_1 + h_2 \frac{u_\alpha' u_\gamma'}{u_\alpha u_\gamma} + h_3 \left(\frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right) \right] u_\alpha^2 u_\gamma^2 dz \\
& + \frac{1}{8} \int_{z_0}^{z_i} h_4 dz \\
& - \frac{i}{8} \int_{z_0}^{z_i} \left[h_5 + h_6 \frac{u_\alpha' u_\gamma'}{u_\alpha u_\gamma} + h_7 \left(\frac{u_\alpha'}{u_\alpha} + \frac{u_\gamma'}{u_\gamma} \right) \right] u_\alpha u_\gamma dz.
\end{aligned} \tag{8.58}$$

We derive the eikonal coefficient $L_{12}^{(4,0)}$ from the expression (8.57) by exchanging the indices α and γ and by adding the term $i e [M^2 B'(z_i) - B'(z_0)]/32$ obtained by partial integration (8.54). This term contributes to the anisotropic or *azimuthal* distortion. By substituting the index γ for α in (8.55), we obtain the integral representation of the eikonal coefficient $L_{22}^{(4,0)}$.

The factors h_μ , $\mu = 1, 2, \dots, 7$ define the functions

$$h_1 = \frac{1}{8} \left(\frac{\Phi''^2}{\Phi^{*2}} - \gamma_0 \frac{\Phi''''}{\Phi^*} \right) + \frac{e^2}{2q^2} \left(B'^2 + \frac{e^2}{4q^2} B^4 - 2\gamma_0 B B' \frac{\Phi'}{\Phi^*} + \frac{\gamma_0}{2} B^2 \frac{\Phi''}{\Phi^*} \right), \tag{8.59}$$

$$h_2 = \frac{e^2}{q^2} B^2 + \gamma_0 \frac{\Phi''}{\Phi^*}, \quad h_3 = \frac{e^2}{q^2} B B', \quad h_4 = \sqrt{\frac{\Phi_0^*}{\Phi^*}} \chi'^2 = \frac{e^2}{4q^2} \sqrt{\frac{\Phi_0^*}{\Phi^*}} B^2, \tag{8.60}$$

$$h_5 = \frac{e^3}{4q^3} B^3 - \frac{\gamma_0 e}{4q} B' \frac{\Phi'}{\Phi^*} + \frac{\gamma_0 e}{2q} B \frac{\Phi''}{\Phi^*}, \quad h_6 = \frac{e}{q} B, \quad h_7 = \frac{e}{2q} B'. \tag{8.61}$$

In practice, the superposition of electric and magnetic fields is only used in low-voltage electron microscopes and photoemission electron microscopes [136, 138]. The electric field serves primarily for decelerating or accelerating the electrons close to the object, whereas one uses the magnetic field for focusing. Since purely electrostatic lenses have appreciably larger aberrations than magnetic lenses and because of severe limitations of the tolerable maximum electric field strength, one

employs such lenses primarily for focusing ions. The eikonal coefficients of these lenses are real because their imaginary parts result from the Larmor rotation caused by the axial magnetic field B . Transmission electron microscopes use exclusively magnetic lenses, in which case we have $\Phi^* = \Phi_0^*$, $\Phi' = \Phi'' = 0$. Considering these relations, the expressions (5.59), (5.60), and (5.61) simplify considerably. Inserting the results into the integrands of the eikonal coefficients (8.55) and (8.56) and rearranging the resulting terms, we eventually obtain

$$L_{00}^{(4,0)} = -\frac{1}{8} \int_{z_0}^{z_i} \left[(u_\alpha'^2 - \chi'^2 u_\alpha^2)^2 + (\chi'' u_\alpha^2 + 2\chi' u_\alpha u_\alpha')^2 \right] dz, \quad \chi' = \sqrt{\frac{e}{8m_e \Phi_0^*}} B, \quad (8.62)$$

$$L_{11}^{(4,0)} = -\frac{1}{2} \int_{z_0}^{z_i} \left[(u_\alpha' u_\gamma' - \chi'^2 u_\alpha u_\gamma)^2 + (\chi'' u_\alpha u_\gamma + \chi' (u_\alpha u_\gamma' + u_\gamma u_\alpha'))^2 + \chi'^2 / 2 \right] dz. \quad (8.63)$$

Replacing the index α by γ in the integrand of (8.62), we obtain the equivalent expression for the eikonal coefficient $L_{22}^{(4,0)}$. Since each integrand consists of a sum of squared terms, the coefficients of spherical aberration $C_3 = -4L_{00}^{(4,0)}$ and of image curvature $F_3 = -2L_{11}^{(4,0)}$ are both positive definite and can never change sign.

The comparison of the integral expressions (8.56) and (8.58) suggests that the coefficients of image curvature and field astigmatism $A_{f3} = -4L_{02}^{(4,0)}$ are related with each other in some way. In order to prove this conjecture, we utilize the so-called Petzval curvature of light-optical round lenses defined as

$$\frac{1}{R_P} = \frac{1}{2} F_3 - A_{f3} = 4L_{02}^{(4,0)} - L_{11}^{(4,0)} = \frac{1}{16} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \left(\frac{2e}{m_e} \frac{B^2}{\Phi^*} + \gamma_0 \frac{\Phi''}{\Phi^*} \right) dz. \quad (8.64)$$

We have derived this expression by employing the relations (8.56), (8.58), and (8.60) and by utilizing the Helmholtz–Lagrange relation for the fundamental rays u_α and u_γ . Employing partial integration, the Petzval curvature adopts the form

$$\frac{1}{R_P} = \frac{1}{16} \left(\gamma_{0i} \sqrt{\frac{\Phi_0^*}{\Phi_i^*}} \frac{\Phi_i'}{\Phi_i^*} - \gamma_{0o} \frac{\Phi_o'}{\Phi_o^*} \right) + \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} \left(\frac{e}{8m_e} \frac{B^2}{\Phi^*} + \frac{1 + 2\gamma_0^2}{32} \frac{\Phi'^2}{\Phi^{*2}} \right) dz. \quad (8.65)$$

Hence, the Petzval curvature is always positive definite for rotationally symmetric electromagnetic fields if the electric field strength is zero ($\Phi'(z_i) = \Phi_i' = \Phi'(z_o) = \Phi_o' = 0$) at the object and image plane. The coefficient of the field astigmatism A_{3f} can change its sign and, therefore, can be zero. However, the image curvature cannot change sign if the electric field strength vanishes at the image and object plane.

In the case of short lenses, the Petzval curvature is related with the focal lengths of the lenses located between the object and image plane. It follows from the relation (4.94) that the Petzval curvature of short magnetic electron lenses equals the sum of the reciprocal focal lengths of all lenses. In the case of short electric lenses (493), the Petzval curvature is half of this sum.

8.3.1 Scherzer Theorem

The Scherzer theorem is the only named and well-established theorem in charged-particle optics [8]. This theorem is of central importance in electron microscopy because it limits the attainable resolution of any electron microscope employing rotationally symmetric lenses. In particular, the theorem states: “spherical aberration and axial chromatic aberration are unavoidable for static rotationally symmetric electron lenses free of space charges” ($\partial/\partial t = 0, \partial/\partial\theta = 0, \rho_e = 0, \Phi > 0$). The validity of this theorem implies that the object and image are real. Mirrors do not belong to this class of lenses because the axial electric potential Φ changes sign within the mirror, resulting in a reversion of the direction of flight of the particles. We have already proven the validity of the Scherzer theorem for the axial chromatic aberration and for the spherical aberration of magnetic round lenses by demonstrating that the coefficient C_3 of their spherical aberration is positive definite. Hence, we only need to demonstrate that $C_3 > 0$ holds also for the general case of arbitrary electromagnetic round lenses. For this purpose, we must transform by partial integrations the integrand of the aberration integral (8.55) in a sum of squared terms with positive sign. The representation of the integrand by a sum of squared terms is not unique because we can form squared terms of different structure. Scherzer’s original prove was nonrelativistic. Nevertheless, the theorem is also valid in the relativistic case [64, 139]. The retention of the relativistic effects in the presence of electrostatic fields renders the calculations very elaborate. We eventually find the representation

$$\begin{aligned}
 C_3 = & \frac{1}{32} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left\{ \frac{\Phi'^2}{\Phi^{*2}} \left(\frac{u'_\alpha}{u_\alpha} + \frac{5\gamma_0}{6} \frac{\Phi'}{\Phi^*} \right)^2 + \frac{3}{2} \left(\frac{\Phi''}{\Phi^*} + \frac{\Phi'}{\Phi^*} \frac{u'_\alpha}{u_\alpha} - \gamma_0 \frac{\Phi'^2}{\Phi^{*2}} \right)^2 \right. \\
 & + \frac{2\gamma_0^2 - 1}{36} \frac{\Phi'^4}{\Phi^{*4}} \\
 & + \frac{\Phi'^2}{\Phi^{*2}} \left(\gamma_0 \frac{u'_\alpha}{u_\alpha} + \frac{2 + 3\gamma_0^2}{6} \frac{\Phi'}{\Phi^*} \right)^2 + \left(\gamma_0 \frac{\Phi''}{\Phi^*} + \gamma_0 \frac{\Phi'}{\Phi^*} \frac{u'_\alpha}{u_\alpha} - \frac{3 + 2\gamma_0^2}{4} \frac{\Phi'^2}{\Phi^{*2}} \right. \\
 & \left. \left. + \frac{eB^2}{2m_e\Phi^*} \right)^2 \right\} dz
 \end{aligned}$$

$$\begin{aligned}
& + \frac{2eB^2}{m_e \Phi^*} \left[\left(\frac{u'_\alpha}{u_\alpha} + \frac{B'}{B} - \frac{3\gamma_0}{4} \frac{\Phi'}{\Phi^*} \right)^2 + \left(\frac{u'_\alpha}{u_\alpha} + \frac{\gamma_0}{2} \frac{\Phi'}{\Phi^*} \right)^2 \right] \\
& + \frac{21 + 2\gamma_0^2}{16} \frac{eB^2}{m_e \Phi^*} \frac{\Phi'^2}{\Phi^{*2}} + \frac{e^2 B^4}{4m_e^2 \Phi^{*2}} \left\} u_\alpha^4 dz. \tag{8.66}
\end{aligned}$$

In the nonrelativistic case, $\gamma_0 = 1$, $\Phi^* = \Phi$, the resulting expression for C_3 does not coincide with Scherzer's original formula. The reason is that more quadratic representations exist in the nonrelativistic case than in the rather complicated relativistic case.

A positive coefficient C_3 implies that the outer zones of a round electron lens refract the rays more strongly toward the axis than the paraxial zone close to the optic axis, as demonstrated in Fig. 8.3. The reason for this behavior originates from the Laplace equation, which puts a constraint on the spatial distribution of the electric and the scalar magnetic potentials. The solutions of the Laplace equation adopt extrema at the boundaries. In the case of round lenses, the extrema do not depend on the azimuthal angle. Since the index of refraction for charged particles depends on the electromagnetic potentials, the outer zones of a round lens focus the rays always more strongly than the paraxial zone. This behavior differs from that of a multipole because its potential depends on the azimuthal angle. Therefore, it adopts at the boundary alternately a maximum and a minimum depending on the polarity of the potential at the electrode or pole piece, respectively. For this reason, multipoles are able to compensate for the unavoidable aberrations of round lenses.

The paraxial approximation $S^{(0)} + S^{(2)}$ of the eikonal or surface of constant action forms a rotationally symmetric paraboloid in the case of round lenses. Its curvature at a given location of the apex coincides with that of the true eikonal S . In the ideal case, the eikonal is a sphere in the field-free image space centered at the image point. The real eikonal for the ray emanating from the object point $w_0 = 0$ forms a rotationally symmetric surface about the optic axis located between the ideal sphere and the paraxial paraboloid because the optical path length $L = S/m_e c$ of the true ray is shorter than the optical path length $L^{(0)} + L^{(2)}$ of the paraxial ray. The path length difference is $L^{(4)} = -C_3 \omega^2 \bar{\omega}^2 / 4$ in fourth-order approximation. The ideal eikonal, the true eikonal, and its parabolic approximation touch each other at the same point on the axis in front of the image point. Since we can choose the location of the touching point arbitrarily, we can construct constant eikonals forming a set of surfaces by varying the optical path length in discrete steps. In the absence of the magnetic field, the rays are the orthogonal trajectories of this set of surfaces.

8.3.2 Spherical Aberration and Disc of Least Confusion

The spherical aberration is the only third-order aberration, which does not vanish at the center of the image plane. If we limit the beam by a circular aperture, the

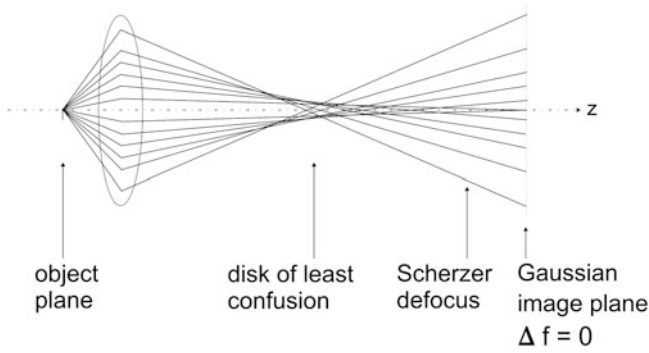


Fig. 8.3 Path of axial rays illustrating the formation of the disk of least confusion

spherical aberration broadens each Gaussian image point to a circular spot with radius $r_s = C_3 \vartheta_0^3$ referred back to the object plane; $\vartheta_0 = |\omega|_{\max}$ is the maximum aperture angle. A ray starting from the center of the object plane with angle $\omega = \bar{\omega} = \alpha$ has the form

$$u = \alpha u'_{\alpha i} (z - z_i) + u_{\gamma i} C_3 \alpha^3 \quad (8.67)$$

in the field-free region near the Gaussian image plane $z = z_i$, where the image is recorded by means of a photographic plate or a CCD camera. Since C_3 is positive, this ray intersects the optic axis before it reaches the Gaussian image plane at a distance:

$$z_i - z = \frac{u_{\gamma i}^2}{u_{\gamma i} u'_{\alpha i}} C_3 \alpha^2 = \sqrt{\frac{\Phi^*}{\Phi_o^*}} M^2 C_3 \omega \bar{\omega} = M_1 C_3 \omega \bar{\omega}. \quad (8.68)$$

Here, M_1 denotes the longitudinal magnification (4.65). One defines the distance for the maximum aperture angle $\alpha = \vartheta_0$ as the longitudinal spherical aberration. This distance corresponds to a change of the focal length of the objective lens by the amount:

$$\Delta f_o = \frac{z_i - z}{M_1} = C_3 \vartheta_0^2. \quad (8.69)$$

The path of the rays (8.67) shown in Fig. 8.3 suggests that the waist of the beam is the smallest in some plane in front of the Gaussian image plane. In order to find this plane, we first determine the plane at which the marginal ray ($\alpha = \vartheta_0$) and a general ray ($\alpha < \vartheta_0$) coincide, yielding

$$z_i - z = M_1 C_3 \frac{\vartheta_0^3 - \alpha^3}{\vartheta_0 - \alpha} = \frac{u_{\gamma i}}{u'_{\alpha i}} C_3 (\vartheta_0^2 + \vartheta_0 \alpha + \alpha^2). \quad (8.70)$$

Substituting this expression for $z_i - z$ in (8.67), we obtain

$$u = -u_{yi}C_3(\vartheta_0^2\alpha + \vartheta_0\alpha^2). \quad (8.71)$$

The lateral distance (8.71) is smallest at the plane for which $du/d\alpha = 0$ giving $\alpha = -\vartheta_0/2$. Substituting this value for α in (8.70) and (8.71), we find that the radius $|u/u_{yi}| = C_3\vartheta_0^3/4$ of the disc of least confusion referred back to the object plane is only one quarter of that of the disc of spherical aberration at the recording plane $z = z_i$.

The location of the plane of least confusion $z = z_{lc}$ is at a distance

$$z_{lc} - z_i = -\frac{3}{4}M_1C_3\vartheta_0^2 \quad (8.72)$$

in front of the recording plane. We can place the plane of least confusion into the recording plane by changing the focal length of the objective lens by the amount

$$\Delta f_{lc} = \frac{z_i - z_{lc}}{M_1} = \frac{3}{4}C_3\vartheta_0^2, \quad (8.73)$$

which is defined as the *defocus of least confusion*. This defocus is most suitable for electron holography in uncorrected electron microscopes.

8.3.3 Coma

After the spherical aberration, the (off-axial) *coma* is the next important third-order aberration in high-performance electron microscopes because it affects the resolution of off-axis points located within the imaged object centered about the optic axis. We have shown in Sect. 5.6.1 that one must eliminate both spherical aberration and coma in order to satisfy the Abbe sine condition, which guarantees that the optical system images perfectly all points of a small central object area. The word coma originates from the Greek, meaning hair. The opticians have chosen this name because the aberration figure resembles that of a comet. The third-order coma

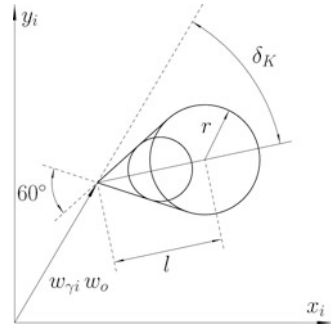
$$u_K^{(3)} = u_{yi}(2K_3\omega\bar{\omega}w_o + \bar{K}_3\omega^2\bar{w}_o) \quad (8.74)$$

is composed of the *coma streak* with length $l = |2u_{yi}K_3\omega\bar{\omega}w_o|$ and the *coma circle* with radius $|u_{yi}\bar{K}_3\omega^2\bar{w}_o|$, as shown in Fig. 8.4.

The coefficient $K_3 = K_{3r} + iK_{3i}$ of the coma is complex for magnetic round lenses. The real part is associated with the *radial* coma component, which points in the direction of the Gaussian image point, as it is the case for glass lenses. The imaginary part K_{3i} results from the Larmor rotation, which depends on the aperture angle. Because the Larmor rotation of the marginal rays is larger than that of the paraxial rays, we obtain a coma figure whose coma streak is perpendicular to the radius vector of the Gaussian image point if the radial (isotropic) coma

Fig. 8.4 Formation of the drop-like coma spot by superposition of coma circles whose centers are shifted from the Gaussian image point by the length of the coma streak

$$l = 2r = 2 |MK_3 w_o| \omega \bar{\omega}$$



vanishes. Therefore, one defines the component connected with K_{3i} as *azimuthal* or *anisotropic* coma. The coma streak shifts the image point by the distance l from its Gaussian image point in a direction, which encloses the angle

$$\delta_K = \arctan(K_{3i}/K_{3r}) \quad (8.75)$$

with the radius vector $w_{\gamma i} w_o$ of the Gaussian image point. If we rotate a ray starting from the object point with slope $\omega = |\omega| e^{i\phi_o}$ on a cone with fixed cone angle $|\omega|$, the image point describes twice the circle of the coma disc, while the coma streak remains unaffected. Because the homocentric pencil of rays originating from the object point fills the entire cone accepted by the aperture, the resulting coma figure is a superposition of coma circles and coma streaks, each of which is attributed to a distinct angle $|\omega| < \vartheta_0$. The tangents to the circles originate from the Gaussian image point and enclose an angle of $2\psi_K = 60^\circ$ with each other because the length of the coma streak is twice the length of the radius of the coma disc giving $\sin \psi_K = 1/2$. The resulting aberration figure has the shape of a comet or of a tail of hairs, respectively.

8.3.3.1 Coma-Free Aperture

The integrand of the real part of the eikonal coefficient $L_{01}^{(4,0)} = -K_3/2$ (8.57) depends linearly on the fundamental field ray $u_3 = u_\gamma$. If we assume that this ray originates from the center of the effective source, the ray is parallel to the optic axis at the object plane if we image the effective source into the back focal plane z_F of the objective lens. Usually, one places the beam-defining aperture at the image of the effective source to avoid vignetting. In this case, the field ray u_γ coincides with the principal ray u_π introduced in Sect. 4.3.1. Let us assume that we image the effective source into some other plane $z = z_K \neq z_F$, where we then place the aperture. As a result, the field ray will differ from the principal ray. Since any paraxial ray is a linear combination of the axial ray and the principal ray, we may write the field ray as

$$u_\gamma = u_\pi + a u_\alpha. \quad (8.76)$$

Substituting this expression for u_γ in the integrand of the eikonal coefficient (8.57), we obtain

$$K_{3r} = -2 \operatorname{Re} L_{01}^{(4,0)} = -2 \operatorname{Re} \tilde{L}_{01}^{(4,0)} - 4a L_{00}^{(4,0)} = \tilde{K}_{3r} + a C_3. \quad (8.77)$$

Here, the tilde indicates the coefficient obtained by substituting the principal ray u_π for the field ray u_γ in the integrand of (8.57). By imposing the condition that the radial coma vanishes, we get

$$a = -\tilde{K}_{3r}/C_3. \quad (8.78)$$

Inserting this expression into (8.76) and considering that the field ray vanishes at the plane z_K , we obtain for the location of this plane the implicit equation

$$u_\pi(z_K)C_3 - u_\alpha(z_K)\tilde{K}_{3r} = 0. \quad (8.79)$$

This coma-free aperture plane is located within the field of the objective lens between its center and the back focal plane. We define the point on the axis of the coma-free plane as the *coma-free* point of the lens because if we place the pivot point of the beam in a scanning electron microscope in this point, the scanning does not produce isotropic coma in the image.

The coefficient of the azimuthal coma K_{3i} is independent of the location of the aperture because the integrand of the imaginary part of the eikonal coefficient (8.57) does not depend on the field ray. Since this coefficient arises solely in the presence of an axial magnetic field, we restate K_{3i} only for purely magnetic lenses ($\Phi' = 0$, $\Phi^* = \Phi_o^*$) in a concise form. Eliminating the derivative of B in the term h_7 (8.61) of (8.57) by partial integration, we obtain

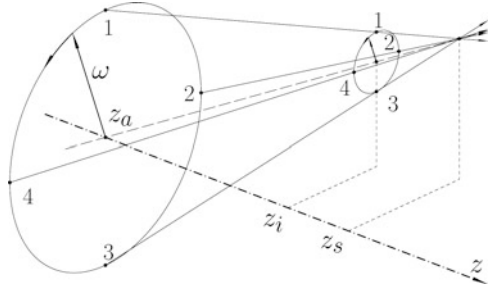
$$K_{3i} = -2 \operatorname{Im} L_{01}^{(4,0)} = \frac{1}{4} \int_{z_o}^{z_i} (\chi' u_\alpha'^2 + 3 \chi'^3 u_\alpha'^2) dz. \quad (8.80)$$

This relation demonstrates that we can eliminate the azimuthal coma of magnetic round lenses only if the magnetic field $B = 2q_o \chi'/e$ changes its sign. Standard magnetic round lenses with a single gap do not meet this requirement because they are immersion lenses with respect to the scalar magnetic potential. Changing the sign of the magnetic field requires two coils with opposite direction of their currents. Hence, such a lens is composed of two standard magnetic round lenses.

8.3.4 Image Curvature

The image curvature $u_i^{(3)} = u_{\gamma i} F_3 \omega w_o \bar{w}_o$ bulges the image field so that the stigmatic image points are located on a rotationally symmetric paraboloid, which touches the central region of the Gaussian image plane. In this plane, the Gaussian image point

Fig. 8.5 Image curvature forming a sharp image spot of the object point w_o at the point $z_s = z_i - M_1 F_3 w_o \bar{w}_o$ located on a paraboloid touching the Gaussian image plane at its center for the case $F_3 < 0$



broadens to a disc, whose radius $|u_i^{(3)}|$ is proportional to the square of the lateral distance of the conjugate object point. Image curvature is of little importance in electron microscopes because (a) the image object area is very small and (b) each lens contributes roughly the same amount to the total image curvature regardless of the position of the lenses. This behavior results from the fact that the integrand of the corresponding eikonal coefficient (8.56) is depending quadratic on the field ray and the axial ray. Since the field ray increases and the axial ray decreases in proportion to the intermediate magnification, their product has about the same order of magnitude within each lens of the microscope.

The situation differs in electron lithography where image curvature and field astigmatism are the most disturbing aberrations because they decisively limit the usable area of the mask. In the presence of image curvature, the position of a ray at a plane shifted by the small distance $z - z_i$ from the Gaussian image plane is

$$u(z) = w_o u_{\gamma i} + \omega [u'_{\alpha i}(z - z_i) + u_{\gamma i} F_3 w_o \bar{w}_o]. \quad (8.81)$$

We obtain a stigmatic image point at the position

$$z_{st} = z_i - \frac{u_{\gamma i}}{u'_{\alpha i}} F_3 w_o \bar{w}_o = z_i - \sqrt{\frac{\Phi_i^*}{\Phi_o^*}} F_3 |M w_o|^2, \quad (8.82)$$

as illustrated in Fig. 8.5. The figure demonstrates that we can conceive the field curvature as a defocus, which depends quadratically on the off-axial distance $|w_o|$ of the object point.

Equation (8.82) describes a paraboloid of revolution about the optic axis. The radius of curvature of the paraboloid at its apex is

$$r_F = \frac{1}{2F_3} \sqrt{\frac{\Phi_o^*}{\Phi_i^*}}. \quad (8.83)$$

Therefore, the sharp image point is located in front of the Gaussian image plane if $F_3 > 0$. We encounter this situation if the electric field strength vanishes at both the

object and the Gaussian image plane. In this case, the image curvature results in a convex image field for an observer who is looking in the direction of the source. This is the reason why the screen of original TV sets had a convex curvature. Here, the point source is located at a fixed position in front of the focusing lens. The off-axial image points referred back to the object plane are formed by a deflection element placed behind the lens. Deflecting the beam by the deflection element is equivalent to a lateral shift of the source.

8.3.5 Field Astigmatism

We may conceive the physical origin of third-order field astigmatism by considering that an observer at an off-axial position is seeing the projection of the round lens. Since the projected lens has an elliptical shape, we can describe it as a superposition of a quadrupole with a round lens whose axis points toward the observer. We have seen that a quadrupole centered on the optic axis splits the stigmatic image point in two lines, one located in front of the other behind the stigmatic image. Therefore, we can assume that the third-order field astigmatism splits each off-axial image point in two astigmatic lines whose separation distance increases quadratic with the distance Mw_0 . To prove this conjecture, we discuss the course of rays originating from a distinct object point in the field-free region around the Gaussian image plane. In the presence of field astigmatism, the position of an arbitrary ray in the image space is given by

$$u^{(3)}(z) = w_0 u_\gamma + \omega u'_{\alpha i}(z - z_i) + u_{\gamma i} A_{f3} \bar{\omega} w_0^2. \quad (8.84)$$

The field astigmatism produces a circle at the Gaussian image plane, as does the image curvature. However, this circle is described in opposite direction if we vary the azimuthal angle ϕ_ω of the complex slope $\omega = |\omega| e^{i\phi_\omega}$ from zero to 2π , as shown in Fig. 8.6. As a result, the initially circular beam becomes astigmatic such that its cross-section forms an ellipse whose shape varies appreciably in the neighborhood of the Gaussian image plane. The ellipse collapses to a line at two distinct planes, one located in front of the other behind the Gaussian image plane. The lines are perpendicular to each other and equally distant from the Gaussian image plane. The imaginary part $A_{f3,i}$ of the complex astigmatism coefficient causes a rotation of the ellipses about the central ray $w_0 u_\gamma$ by the angle:

$$\delta_A = \arctan \left(\frac{A_{3f,i}}{A_{3f,r}} \right). \quad (8.85)$$

If this angle is zero, the tangential or meridional line focus is embedded in the meridional section formed by the central ray and the optic axis. The conjugate sagittal line focus is perpendicular to this section. In the presence of an axial magnetic field, the

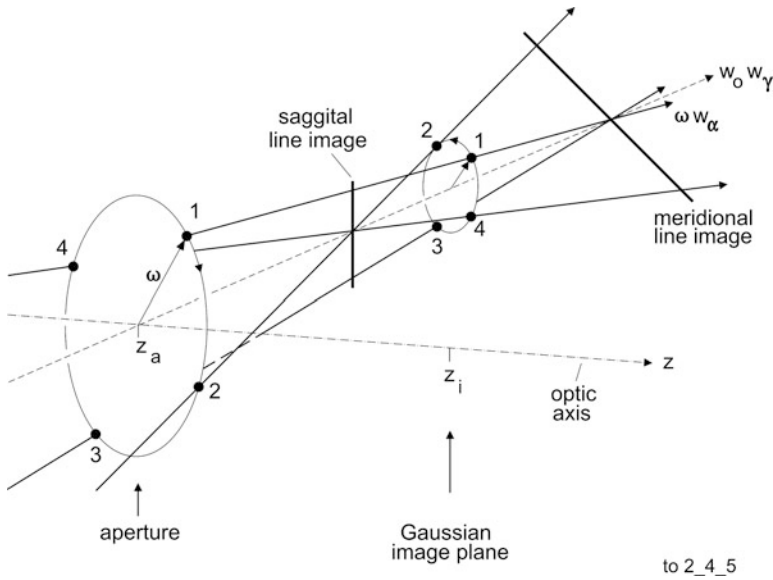


Fig. 8.6 Formation of the meridional and sagittal image lines by field astigmatism

imaginary part of the astigmatism coefficient is nonzero. As a result, the two line foci are rotated with respect to the meridional section by the angle (8.85).

We find the azimuthal orientation and the planes of the line foci from the condition that the second and third terms in (8.84) cancel out, giving

$$z - z_i = -\frac{u_{yi}}{u'_{\alpha i}} A_{i3} \frac{\bar{\omega}}{\omega} w_0^2 = -|A_{i3}| \sqrt{\frac{\Phi_i^*}{\Phi_o^*}} |M w_0|^2 e^{i(\delta_A - 2\phi_\omega + 2\phi_o)}. \quad (8.86)$$

The azimuthal angle ϕ_o indicates the orientation of the position vector $w_0 = |w_0| e^{i\phi_o}$ with respect to the x -axis. Since the distance $\Delta = z - z_i$ is real, we can satisfy the requirement (8.86) only if the exponent equals zero or π :

$$\begin{aligned} \phi_{\omega m} &= \phi_o + \delta_A/2, \\ \phi_{\omega s} &= \phi_o + \delta_A/2 + \pi/2. \end{aligned} \quad (8.87)$$

The indices “m” and “s” indicate the meridional section and the sagittal section, respectively. The loci of the meridional line foci and the sagittal line foci are paraboloids of revolution about the optic axis and tangent to the Gaussian image plane at its center. The curvatures of the two paraboloids are opposite in sign such that the meridional paraboloid is convex and the sagittal paraboloid concave with respect to an observer looking in the direction of the source. One defines the distance

between any pair of conjugate line foci

$$\Delta_a = \Delta_s - \Delta_m = 2\Delta_s = 2|A_{f3}| \sqrt{\frac{\Phi_i^*}{\Phi_o^*}} |M w_o|^2 \quad (8.88)$$

as the *astigmatic difference*.

The combination of field astigmatism and image curvature changes the curvatures of the meridional and the sagittal paraboloid in such a way that conjugate line foci are formed at distances $\pm\Delta_m$ from the corresponding image point (8.82) situated on the paraboloid of image curvature. The field astigmatism broadens this point to a circle. The line foci are located in front of the Gaussian image plane z_i in the case $F_3 > |A_{f3}|$. Then the meridional and the sagittal paraboloids have curvatures with the same sign.

Light-optical systems free of image curvature and field astigmatism are known as anastigmats. Rotationally symmetric electron optical anastigmats do not exist if the electric field strength is zero at the object and image planes because the Petzval curvature (8.65) is unavoidable in this case.

8.3.6 Distortion

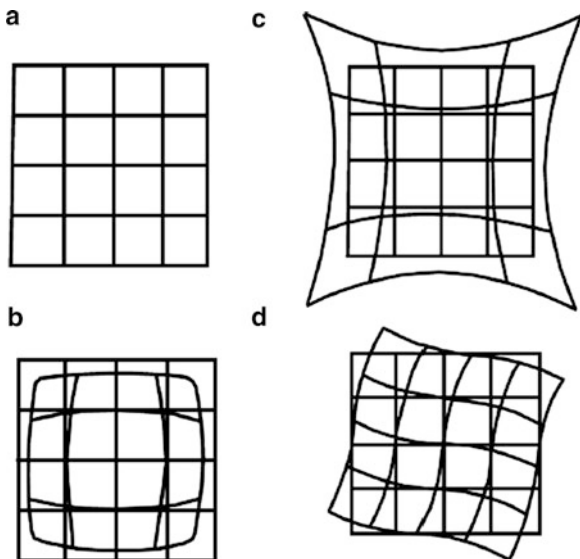
The third-order distortion at the Gaussian image plane

$$u_i^{(3)} = u_{\gamma i} D_3 w_o^2 \bar{w}_o \quad (8.89)$$

does not broaden the image points but destroys the paraxial proportionality $u_i^{(1)} = u_{\gamma i} w_o$ between the position vectors w_o and u_i of conjugate object and image points. The contribution of the objective lens to the total distortion in the image of an electron microscope is negligible due to the very small object area transferred at high magnifications.

Then the distortion is the dominant defect of the projector lenses because of the largely increased lateral distances of the field rays within these lenses. Fortunately, we can eliminate the distortion in principle, and one keeps it sufficiently small in an actual electron microscope by proper design of the projector system. As a rule of thumb, we can state that one must correct the objective lens of a high-performance microscope for spherical aberration and coma and the projector system for distortion. The distortion (8.89) shifts the image point radially relative to its paraxial position if the distortion coefficient $D_3 = D_{3r} + iD_{3i}$ is real ($D_{3i} = 0$). The shift is in azimuthal direction if the coefficient is imaginary ($D_{3r} = 0$). In the general case, the distortion is composed of the *radial* or *isotropic distortion* and the *azimuthal* or *anisotropic distortion*, which is often referred to as *spiral distortion*. Since this distortion results from the Larmor rotation, it vanishes for electrostatic round lenses.

Fig. 8.7 Distortion of the image of a square grid in the cases of (a) ideal imaging, (b) barrel distortion ($D_3 = D_{3r} < 0$), (c) pincushion distortion ($D_3 = D_{3r} > 0$), and (d) spiral or azimuthal distortion ($D_3 = -\overline{D}_3 = iD_{3i}$)



To illustrate the effect of the distortion, we consider a square grid in the object plane. First, we assume that the distortion coefficient is real and negative ($D_3 = \overline{D}_3 = D_{3r} < 0$). Then the outer region of the grid image shrinks, as depicted in Fig. 8.7b. According to its characteristic shape, one denotes this distortion as *pincushion distortion*. We obtain a distended image exhibiting a *barrel distortion* if $D_{3r} > 0$, as illustrated in Fig. 8.7c. The image is spirally warped if the distortion coefficient is imaginary, as shown in Fig. 8.7d. The azimuthal direction of the spiral deformation depends on the sign of D_{3i} . The twist is right-handed referred to the direction of flight if $D_{3i} > 0$.

8.4 Primary Aberrations of Glaser's Magnetic Model Lens

The Glaser bell-shaped field (4.96) approximates sufficiently well the field of a strong symmetric magnetic lens employed in most electron microscopes. Glaser's treatment of the bell-shaped field has provided a qualitative understanding of the properties of magnetic lenses, especially at a time when computers were not available. The Glaser field provides the coefficients of the primary aberrations in analytical form as functions of the object plane z_0 and of the dimensionless lens strength (4.97). The primary aberrations of electron lenses are composed of the third-order geometrical aberrations discussed in Sect. 8.3 and the second-rank chromatic aberrations. In electron microscopy, we are concerned primarily with the aberrations of the objective lens. Therefore, we restrict our considerations to the so-called real aberrations whose coefficients are expressed in terms of the object and aperture coordinates. Moreover, for determining the decisive imaging properties

of the objective lens, it suffices to discuss the coefficients of its axial chromatic aberration, spherical aberration and coma, respectively.

In Sect. 8.3, we have expressed the total third-order aberration of round lenses as a polynomial of degree 3 in the complex ray parameters ω , $\bar{\omega}$ and $w_o = u_o$, $\bar{w}_o = \bar{u}_o$. In some cases, one uses the complex ray coordinate u_a at the aperture plane z_a instead of the initial slope ω . In paraxial approximation, these ray parameters satisfy the relation

$$u_a = \omega u_a(\phi_a) = \omega \frac{a}{\sigma} \frac{\sin \sigma(\phi_o - \phi_a)}{\sin \phi_o \sin \phi_a}. \quad (8.90)$$

The coefficient C_3 and $K_3 = K_{3r} + iK_{3i}$ of the spherical aberration and the coma of the Glaser lens expressed in terms of the position and slope coordinates at the object plane (8.48) have the form

$$\frac{C_3}{a} = \left\{ \frac{\pi k^2}{4\sigma^2} - \frac{4k^2-3}{8(4k^2+3)} \sin \frac{\pi}{\sigma} \cos(2\phi_o - \pi/\sigma) \right\} \frac{1}{\sin^4 \phi_o}, \quad (8.91)$$

$$K_{3r} = \frac{1}{4\sigma^2 \sin^2 \phi_o \sin \sigma(\phi_a - \phi_o)} \left\{ \frac{\pi k^2 \cos \sigma(\phi_a - \phi_o)}{-\frac{\sigma^2 \sin \pi \sigma}{4\sigma^2-1} \left[(4\sigma^2-7) \cos(\phi_o + \phi_i) \cos \sigma(\phi_a - \phi_o) \right] + 3 \sin(\phi_o + \phi_i) \sin \sigma(\phi_a - \phi_o)} \right\}, \quad (8.92)$$

$$K_{3i} = \frac{\pi k}{4 \sin^2 \phi_o} \left\{ \frac{2\sigma^2-1}{2\sigma^3} - \frac{2\sigma^2-1}{k^2} \sin \frac{\pi}{\sigma} \cos(\phi_o + \phi_i) \right\}. \quad (8.93)$$

We obtain the position $z_{aK} = \cot \phi_{aK}$ of the coma-free aperture by setting $K_r = 0$. The angle ϕ_{aK} is the solution of the resulting equation

$$\tan \sigma(\phi_{aK} - \phi_o) = \frac{\pi \sigma k^2}{3 \sin(\phi_o + \phi_i)} \left\{ \frac{4\sigma^2-1}{\sigma^3 \sin(\pi/\sigma)} - \frac{4\sigma^2-7}{\pi k^2} \cos(\phi_o + \phi_i) \right\}. \quad (8.94)$$

The coefficient of axial chromatic aberration (8.34) is given by

$$\frac{C_c}{a} = \frac{\pi k^2}{2\sigma^3} \frac{1}{\sin^2 \phi_o}. \quad (8.95)$$

The coefficient (8.35) of the chromatic round-lens distortion reduces considerably for magnetic round lenses. For the Glaser field, the complex coefficient is given by

$$D_{cr} = \frac{1}{1 + \varepsilon \Phi_0} \left\{ \frac{\pi k^2}{2\sigma^2} \cot \sigma(\phi_a - \phi_o) + i \frac{1 + 2\varepsilon \Phi_0}{2} \chi_i \right\}. \quad (8.96)$$

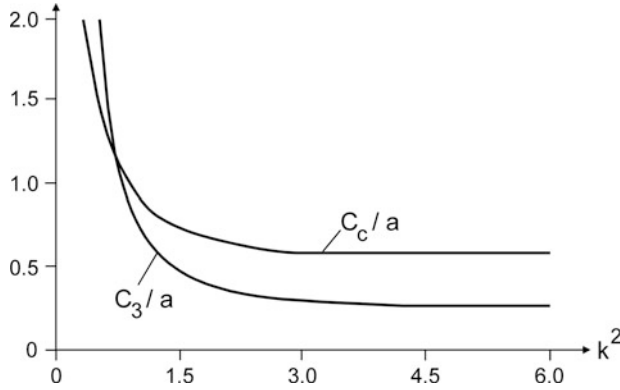


Fig. 8.8 Normalized aberration coefficients C_c/a and C_3/a as functions of the strength k^2 of the bell-shaped field (4.96) in the case of high magnification ($M \rightarrow \infty$)

The real part vanishes if an image of the effective source is placed at the *achromatic aperture*:

$$z_{ac} = a \cot \phi_{ac} = a \cot(\phi_o - \pi/2\sigma). \quad (8.97)$$

Most objective lenses operate with high magnification ($M \gg 1$). In the limit $M \rightarrow \infty$, we have $\phi_i = 0$, $\phi_o = \pi/\sigma$. For this case, we have plotted in Fig. 8.8 the normalized aberration coefficients C_c/a and C_3/a as functions of the lens strength k^2 . The curves show that both coefficients are close to their minimum for the symmetric condenser-objective lens with $k^2 = 3$.

For this special case, we have $\sigma = 2$, $\phi_o = \pi/2$, $\phi_i = 0$. The angular position of the coma-free aperture is found from (8.94) as

$$\phi_{aK} = \frac{1}{2} \arctan \frac{15\pi}{4} = 0.743, \quad (8.98)$$

giving $z_{aK} = 1.09a$. The position $z_{ac} = a$ of the achromatic aperture (8.97) coincides with the back focal plane $z_{FP} = a$ of the post-field of the condenser-objective lens, whereas the coma-free plane is slightly shifted away from this plane toward the image.

8.5 Geometrical Aberrations of Quadrupole–Octopole Systems

Systems composed of magnetic quadrupoles and octopoles are favorable for focusing relativistic electrons. The magnetic quadrupoles yield strong paraxial focusing, whereas the octopoles provide third-order focusing. At moderate energies, such systems are primarily used as correctors compensating for the unavoidable chromatic and spherical aberration of round lenses. Although the principle of this

type of correction is sound from the theoretical point of view, it took almost 50 years of intense effort to surpass the resolution of a high-quality round lens by means of a corrector in practice. The reasons for this long lasting struggle are the extremely high requirements on mechanical and electrical stability and the complexity of the systems. The precise adjustment of their numerous elements has become possible only recently by means of high-speed computers and microprocessors and by procedures enabling a fast determination of the state of alignment.

In order to facilitate the alignment and to keep the number of additional aberrations as small as possible, the quadrupole and octopole fields must not overlap the field of the magnetic round lenses. In this case, the fourth-order eikonal of the entire system consists of a term produced by the round lenses and a term formed exclusively by the quadrupoles and octopoles. Since quadrupole systems can provide stigmatic focusing, the presence of round lenses is not mandatory for obtaining a stigmatic image.

We readily obtain the fourth-order variational polynomial of quadrupole octopole systems by setting $\Phi' = B = 0$ in the expression (8.40) for systems with straight optic axis. In order to account for immersion systems, the axial electric potential Φ within the quadrupole system may differ from that at the object plane. Moreover, we assume that the quadrupole fields vanish at the object and image planes, as it is always the case for correctors. In this case, we can recast the fourth-order perturbation eikonal

$$L_{\text{gi}}^{(4)} = -\text{Re} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left\{ \frac{1}{8} w'^2 \bar{w}'^2 + \frac{1}{16} \frac{\Phi_2 \bar{\Phi}_2}{\Phi^{*2}} w^2 \bar{w}^2 - \frac{\gamma_0}{4} \frac{\Phi_2}{\Phi^*} w' \bar{w}' \bar{w}^2 + \frac{\gamma_0}{24} \frac{\Phi_2''}{\Phi^*} w \bar{w}^3 \right. \\ \left. - i \frac{e}{4q} \Psi_2' \bar{w}' w \bar{w}^2 + i \frac{e}{12q} \Psi_2' w' \bar{w}^3 - \left(\frac{\gamma_0}{2} \frac{\Phi_4}{\Phi^*} + i \frac{e}{q} \Psi_4 - \frac{1}{16} \frac{\Phi_2^2}{\Phi^{*2}} \right) \bar{w}^4 \right\} dz \quad (8.99)$$

by partial integrations without obtaining contributions at the boundaries. We thus eliminate all derivatives of the quadrupole strengths and replace second derivatives of the paraxial ray $w = w^{(1)}$ by the paraxial equation

$$w'' = \left(\frac{\gamma_0 \Phi_2}{\Phi^*} + 2i \frac{e}{q} \Psi_2 \right) \bar{w} \quad (8.100)$$

and its conjugate complex, respectively. The straightforward calculation eventually gives

$$L_{\text{gi}}^{(4)} = -\text{Re} \int_{z_0}^{z_i} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \left\{ \frac{1}{8} w'^2 \bar{w}'^2 + \frac{1}{8} \left[\frac{\Phi_2 \bar{\Phi}_2}{2\Phi^{*2}} + \left| \gamma_0 \frac{\Phi_2}{\Phi^*} + i \frac{2e}{q} \Psi_2 \right|^2 \right] w^2 \bar{w}^2 \right. \\ \left. + \frac{1}{4} \left(\frac{\gamma_0 \Phi_2}{\Phi^*} + i \frac{2e}{q} \Psi_2 \right) \bar{w}'^2 w \bar{w} + \left[\frac{1}{16} \frac{\Phi_2^2}{\Phi^{*2}} + \frac{1}{24} \left(\frac{\gamma_0^2 \Phi_2^2}{\Phi^{*2}} + \frac{4e^2}{q^2} \Psi_2^2 \right) \right. \right. \\ \left. \left. - \frac{\gamma_0 \Phi_4}{2\Phi^*} - i \frac{e}{q} \Psi_4 \right] \bar{w}^4 \right\} dz. \quad (8.101)$$

To obtain the fourth-order eikonal polynomial, we must substitute the paraxial ray $w^{(1)}$ for w in this expression, which holds for arbitrary azimuthal orientations of the quadrupoles and octopoles. The structure of the integrand reveals that quadrupole–octopole systems introduce fourth-order eikonal terms with multiplicity $m = 0, 2, 4$. However, only monomials of the complex ray parameters with multiplicity $m = 0$ can compensate for the corresponding monomials of round lenses. In order to reduce the number of aberrations with multiplicity $m \neq 0$, we assume in the following regular quadrupoles with strengths $\Phi_2 = \Phi_{2c}$, $\Psi_2 = i\Psi_{2s}$ forming a system with a pair of orthogonal principal sections. We do not impose this condition on the octopoles to allow for complex aberration coefficients enabling the correction of both the isotropic and the anisotropic components of the field aberrations of magnetic round lenses, anisotropic coma in particular.

For an orthogonal quadrupole system, the complex paraxial path equation decouples yielding two real equations: one for the x -component and the other for the y -component (4.102). Hence, the fundamental pseudo rays (4.223) are real (4.227). We obtain directly the integral expressions for the eikonal coefficients $L_{\nu\lambda}^{(4,2\mu)} = \bar{L}_{\nu\lambda}^{(4,2\mu)}$ of an orthogonal quadrupole system by inserting into the integrand of the integral (8.92) the representation

$$w^{(1)} = \omega w_\omega + \bar{\omega} w_{\bar{\omega}} + \rho w_\rho + \bar{\rho} w_{\bar{\rho}} \quad (8.102)$$

for the paraxial ray $w = w^{(1)}$ and setting $\Phi_4 = \Psi_4 = 0$.

8.5.1 Aperture Aberration of Stigmatic Orthogonal Quadrupole Systems

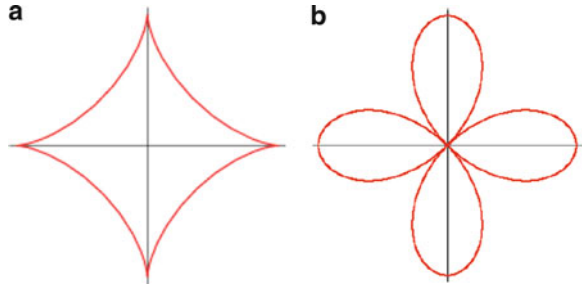
We find the general form of the third-order aperture aberration of a stigmatic orthogonal quadrupole system by employing the expression (8.42) and setting $w_0 = \rho = 0$. Assuming distortion-free paraxial imaging, $w_{\bar{\rho}}(z_i) = 0$, $w_\omega(z_i) = w_\gamma(z_i) = x_{\gamma i}$, we find

$$w_a^{(3)}(z_i) = -2w_{\gamma i} \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}} = -x_{\gamma i} \left(4L_{00}^{(4,0)} \omega^2 \bar{\omega} + L_{00}^{(4,2)} [3\bar{\omega}^2 \omega + \omega^3] + 4L_{00}^{(4,4)} \bar{\omega}^3 \right). \quad (8.103)$$

We readily identify the first term as spherical aberration. The aberration figure of the second term is an astroid shown in Fig. 8.9. Therefore, one defines this aberration as *star aberration*. Although this aberration is fourfold symmetric at the Gaussian image plane, we obtain a twofold aberration figure if we superpose spherical aberration or observe the aberration $w^{(3)}(z) = \omega w'_{\alpha i}(z - z_i) + w_a^{(3)}(z_i)$ at a slightly defocused plane $z \neq z_i$.

The last term on the right-hand side of (8.103) accounts for the fourfold axial astigmatism. Its aberration figure forms a “rosette” at distinct defocused planes and

Fig. 8.9 Aberration figure of (a) the axial star aberration at the Gaussian image plane and (b) the fourfold axial astigmatism forming a rosette at two characteristic planes



degenerates to a circle in the Gaussian image plane. If we rotate the axial ray on the margin of the circular aperture by an angle 2π , the ray describes in opposite azimuthal direction three times the image circle.

In order to determine if the aperture aberration of orthogonal quadrupole systems can change sign, it suffices to consider an axial ray propagating in the $x - z$ principal section. For this ray, we have $\omega = \bar{\omega} = a_1 = \alpha$ resulting in $w^{(1)} = a_1(w_\omega + w_{\bar{\omega}}) = a_1 w_\alpha = \alpha x_\alpha$ and

$$w_a^{(3)}(z_i) = x_a^{(3)}(z_i) = x_{yi} C_{\alpha\alpha\alpha} \alpha^3 = -4x_{yi} \left(L_{00}^{(4,0)} + L_{00}^{(4,2)} + L_{00}^{(4,4)} \right) \alpha^3. \quad (8.104)$$

We obtain the aberration coefficient $C_{\alpha\alpha\alpha} = -4 \left(L_{00}^{(4,0)} + L_{00}^{(4,2)} + L_{00}^{(4,4)} \right)$ by substituting the axial ray αx_α for the ray w in the integrand of the fourth-order eikonal (8.101). By partial integration of a part of the first term of this integral, we find the representation

$$C_{\alpha\alpha\alpha} = \frac{1}{12} \sqrt{\frac{\Phi^*}{\Phi_0^*}} \int_{z_0}^{z_i} \left\{ 2 \frac{x_\alpha^4}{x_\alpha^4} + \left(3\gamma_0 \frac{\Phi_{2c}}{\Phi^*} - 4 \frac{e}{q} \Psi_{2s} \right)^2 + (6 - \gamma_0^2) \frac{\Phi_{2c}^2}{\Phi^{*2}} \right\} x_\alpha^4 dz. \quad (8.105)$$

The integrand consists of a sum of positive squared terms in the cases of purely magnetic or purely electric quadrupoles and for mixed systems if $\gamma_0 \leq \sqrt{6}$, that is, when the acceleration voltage is smaller than about 0.74 MV. If the acceleration voltage exceeds this value, it is possible to make the coefficient (8.105) zero. We can achieve this by choosing the electric and magnetic quadrupole strengths such that the second term in the integrand vanishes. The absolute value of the negative $(6 - \gamma_0^2 < 0)$ third term surpasses the first term in the case of a strong short quadrupole. This behavior becomes obvious when Φ_{2c} approaches a delta function. Then the third term diverges, whereas the first term stays finite.

We can nullify the coefficient of the star aberration by introducing a symmetry plane, such that the quadrupole fields are antisymmetric and one of the axial pseudo rays $w_\omega = (x_\alpha + y_\beta)/2$ and $w_{\bar{\omega}} = (x_\alpha - y_\beta)/2$ is symmetric, the other antisymmetric with respect to this plane. The antisymmetric quadrupole quadruplet shown in Fig. 4.33 satisfies this condition.

8.5.2 Aberrations Introduced by Octopoles

Octopoles affect neither the paraxial ray nor the second-order path deviation. They primarily introduce a third-order path deviation resulting in a fourfold third-order deformation of a rotationally symmetric pencil of rays. The deformation gets an additional twofold component and a rotationally symmetric component if we place octopoles within the astigmatic paraxial region formed by the quadrupoles. To minimize the number of aberrations introduced by these elements, we impose that they possess mutually orthogonal plane principal sections. However, we allow for arbitrary azimuthal orientations of the octopoles in order to compensate for the anisotropic eikonal components introduced by the magnetic round lenses.

The imaginary components of the octopole strengths

$$\Phi_4 = \Phi_4(z) = \Phi_{4c}(z) + i\Phi_{4s}(z), \quad \Psi_4 = \Psi_4(z) = \Psi_{4c}(z) + i\Psi_{4s}(z) \quad (8.106)$$

are zero if the electrodes and the pole pieces are centered along the x - and y -axes and along the diagonals. The real parts of the complex octopole strengths (8.106) vanish if we rotate the octopoles by 22.5° with respect to the orientation for which the imaginary parts are zero. We derive the contribution of the octopoles to the fourth-order eikonal from (8.101). In order to obtain monomials in the complex ray parameters, we must choose the representation (4.225) for the paraxial ray. Introducing the modified total octopole strength

$$O(z) = O_r(z) + iO_i(z) = \frac{\gamma_0}{2} \frac{\Phi_4}{\Phi^*} + i\frac{e}{q}\Psi_4, \quad (8.107)$$

we obtain for the coefficient of the fourth-order perturbation eikonal introduced by the octopoles the expression

$$L_{\nu\lambda}^{(4,2\mu)} = \frac{4!}{(2+\mu-\nu)!(2-\mu-\lambda)!} \int_{z_0}^{z_i} \left\{ \frac{O_r}{1+\delta_{\mu 0}} \left(w_\omega^{2+\mu-\nu} w_\omega^{2-\mu-\lambda} w_\rho^\nu w_\rho^\lambda + w_\omega^{2+\mu-\nu} w_\omega^{2-\mu-\lambda} w_\rho^\nu w_\rho^\lambda \right) \right. \\ \left. + iO_i \left(w_\omega^{2+\mu-\nu} w_\omega^{2-\mu-\lambda} w_\rho^\nu w_\rho^\lambda - w_\omega^{2+\mu-\nu} w_\omega^{2-\mu-\lambda} w_\rho^\nu w_\rho^\lambda \right) \right\} dz. \quad (8.108)$$

The fundamental pseudo rays are real for orthogonal quadrupole systems. Hence, the imaginary part of the eikonal coefficients (8.108) results solely from the imaginary (skew) component

$$O_i = \frac{\gamma_0}{2} \frac{\Phi_{4s}}{\Phi^*} - \frac{e}{q}\Psi_{4s} \quad (8.109)$$

of the total octopole strength (8.107). The integral representation (8.108) demonstrates that it is possible to produce all fourth-order eikonal coefficients by octopoles provided they are located within the astigmatic paraxial domain where all pseudo fundamental rays are nonzero apart from distinct planes. In order to guarantee

that each octopole affects every aberration differently, the value of the products of the pseudo rays in the integrand of the integral (8.123) must be different for each location of the octopoles. To meet this condition in a feasible way, it is advantageous to form astigmatic and strongly first-order distorted stigmatic images of both the object plane and the diffraction plane within the astigmatic paraxial domain, as is the case for the orthogonal quadrupole system shown in Fig. 4.35. The fundamental pseudo rays $w_{\bar{\omega}} = (x_{\alpha} - y_{\beta})/2$ and $w_{\bar{\rho}} = (x_{\gamma} - y_{\delta})/2$ vanish for rotationally symmetric systems or within the stigmatic paraxial domains of systems composed of round lenses and quadrupoles. As a result, octopoles placed within the stigmatic domains can only induce fourth-order eikonal polynomials with multiplicity $m = 2\mu = 4$. Therefore, it is not possible in this case to correct any of the rotationally symmetric aberrations by octopoles. To correct for these aberrations, we must place the octopoles within regions of the astigmatic domain, where $w_{\omega} \approx w_{\bar{\omega}}$ and $w_{\rho} \approx w_{\bar{\rho}}$.

8.5.3 Third-Order Aberrations of Systems with Threefold Symmetry Corrected for Second-Order Aberrations

Systems with threefold symmetry are composed of round lenses and multipole elements consisting of $6N$, $N = 0, 1, \dots$, electrodes or pole pieces, respectively. One generally employs sextupoles to compensate for second-order aberrations of systems with curved axis, such as accelerators or spectrometers and energy filters. However, we can utilize these elements also for correcting third-order aberrations of round lenses if the primary second-order path deviations cancel out, as happens for the system depicted in Figs. 7.1 and 7.2. Since rotationally symmetric fields are invariant with respect to the azimuthal angle, they possess all multiplicities including threefold multiplicity. Owing to this behavior, systems with threefold symmetry possess field components having multiplicities $m = 0, 3, 6, 9, \dots$, in the most general case. In order to achieve a threefold symmetric system, we must adjust the azimuthal orientation of the multipoles such that their principal sections coincide. Assuming that the primary second-order aberrations are eliminated and employing the relation (7.29) for the integrand $m_E^{(4)}$, we obtain the fourth-order perturbation eikonal as

$$L_{\text{gi}}^{(4)} = \frac{m_e c}{q_0} \int_{z_0}^{z_1} \left(\mu_{\text{g1}}^{(4)} + \frac{1}{2} D^{(2)} \mu_{\text{g1}}^{(3)} \right) dz = L_{\text{R}}^{(4)} + L_{\text{H}}^{(4)}, \quad (8.110)$$

$$L_{\text{R}}^{(4)} = \frac{m_e c}{q_0} \int_{z_0}^{z_1} \mu_{\text{g1}}^{(4)} dz, \quad L_{\text{H}}^{(4)} = \frac{3}{2} \int_{z_0}^{z_1} \text{Re} (H \bar{u}^{(1)2} \bar{u}^{(2)}) dz. \quad (8.111)$$

The round lenses produce the term $L_R^{(4)}$, whereas the sextupoles account for the second term $L_H^{(4)}$. We derive this term by substituting the expression (7.91) for $\mu_{g1}^{(3)}$ and introduce the modified total strength H (7.96) of the hexapole fields. These fields produce in first approximation the second-order path deviation $u^{(2)}$. Therefore, they do not affect the course of the paraxial ray $u^{(1)} = \omega u_\alpha + \rho u_\gamma$ defined by the round lenses. Substituting the expression (7.103) for $u^{(2)}$ in (8.126) and considering that the fundamental paraxial rays u_α and u_γ are real, we rewrite the eikonal polynomial $L_H^{(4)}$ in the form

$$L_H^{(4)} = \frac{3}{2} \operatorname{Re} \int_{z_0}^{z_1} \bar{H} (\omega u_\alpha + \rho u_\gamma)^2 (\bar{\omega}^2 u_{11} + \bar{\omega} \bar{\rho} u_{12} + \bar{\rho}^2 u_{22}) dz. \quad (8.112)$$

The secondary fundamental rays $u_{\mu\nu}$ are always complex if the sextupoles have different azimuthal orientations with respect to the rotating coordinate system. In this case, we need more than two sextupoles in order that the second-order path deviation vanishes behind the system. The representation (8.112) reveals that the fourth-order eikonal polynomial produced by the hexapole fields contains exclusively monomials with multiplicity zero, as is the case for round lenses. To demonstrate this equivalence, we recast (8.112) in the form

$$L_H^{(4)} = \operatorname{Re} \sum_{\lambda=0}^2 \sum_{\nu=0}^{\lambda} \frac{2}{1 + \delta_{\nu\lambda}} L_{\nu\lambda}^{(4,0)} \bar{\omega}^{2-\nu} \omega^{2-\lambda} \bar{\rho}^\nu \rho^\lambda. \quad (8.113)$$

The eikonal coefficients $L_{\nu\lambda}^{(4,0)}$ have the form

$$L_{00}^{(4,0)} = \frac{3}{2} \operatorname{Re} \int_{z_0}^{z_1} \bar{H} u_\alpha^2 u_{11} dz, \quad L_{01}^{(4,0)} = \frac{3}{4} \int_{z_0}^{z_1} (2\bar{H} u_\alpha u_\gamma u_{11} + H u_\alpha^2 \bar{u}_{12}) dz, \quad (8.114)$$

$$L_{11}^{(4,0)} = 3 \operatorname{Re} \int_{z_0}^{z_1} \bar{H} u_\alpha u_\gamma u_{12} dz, \quad L_{02}^{(4,0)} = \frac{3}{4} \int_{z_0}^{z_1} (\bar{H} u_\gamma^2 u_{11} + H u_\alpha^2 \bar{u}_{22}) dz, \quad (8.115)$$

$$L_{12}^{(4,0)} = \frac{3}{4} \int_{z_0}^{z_1} (\bar{H} u_\gamma^2 u_{12} + 2H u_\alpha u_\gamma \bar{u}_{22}) dz, \quad L_{22}^{(4,0)} = \frac{3}{2} \operatorname{Re} \int_{z_0}^{z_1} \bar{H} u_\gamma^2 u_{22} dz. \quad (8.116)$$

Because the secondary fundamental rays (7.104) depend linearly (7.105) on the hexapole strength H , the coefficients are proportional to the square of the excitation of the sextupoles. Accordingly, replacing H by $-H$ does not change the value of

the eikonal coefficients. Therefore, the coefficient $L_{00}^{(4,0)}$ is positive definite if the secondary axial ray u_{11} does not change sign, as it is the case for the hexapole corrector shown in Fig. 7.2. By varying the hexapole strength, we can adjust the value of this coefficient to compensate for the corresponding negative eikonal coefficient of the round lenses.

8.5.4 Parasitic Aberrations

So far, we have always assumed ideal elements perfectly aligned along the optic axis. In practice, the electromagnetic fields will deviate from the ideal symmetry due to unavoidable mechanical inaccuracies in the construction and alignment of the elements and due to small inhomogeneities of the permeability within the magnetic pole pieces. These static defects will generate additional aberrations, which we define as *coherent parasitic aberrations*. The most encountered parasitic aberration in electron microscopes is the twofold axial astigmatism, which arises in round lenses due to small deviations from rotational symmetry. This deviation generates primarily a weak quadrupole field causing a first-order astigmatism, which one cancels routinely by means of a stigmator. This element produces a quadrupole field. One adjusts its strength and azimuthal orientation in such a way that the stigmator compensates for the ellipticity of the paraxial path of rays.

Coherent aberrations falsify the transfer of the spatial object frequencies in an electron microscope resulting in a “coded” image, as it is the case in holography. Since the information about the object structure is not lost, we can retrieve the correct object structure from the image by appropriate restoration procedures, at least in principle. Such a restoration is not possible for the *incoherent parasitic aberrations* resulting from random mechanical and electromagnetic instabilities. These stochastic time-dependent perturbations suppress the transfer of the high spatial frequencies, thereby limiting the attainable resolution in an electron microscope. Since this holds also true for the chromatic aberrations, we must also conceive them as incoherent. The incoherent aberrations define the so-called *information limit* of an electron microscope. Therefore, correction of the static lens defects improves the actual resolution of the microscope at most up to the information limit.

The effect of the parasitic aberrations on the performance of the instrument is the stronger the lower the order of the aberration apart from the static zero-order aberration, which merely shifts the image in lateral direction. However, a varying zero-order aberration is most deleterious since it blurs the image, reducing contrast and resolution. In order to achieve a sub-Angstrom information limit in an electron microscope at acceleration voltages between 150 and 300 kV, it is necessary to stabilize the electromagnetic fields with a relative accuracy smaller than 0.1 ppm, to suppress the deleterious mechanical vibrations, and to reduce the energy width of the incident electrons by means of a monochromator below 0.2 eV.

In the absence of incoherent aberrations, the coherent axial aberrations determine entirely the resolution of the imaging system. In the most general case, all axial aberrations will arise, partly from misalignment and partly from the inherent static defects of the focusing elements. In order to compensate for these aberrations, we must know their properties. We obtain conveniently the individual aberrations from the monomials of the power series expansion of the axial part of the perturbation eikonal:

$$L_a(z_i) = \operatorname{Re} \sum_{\nu=0}^{\infty} \sum_{\mu=0}^{\nu} \left(L_{00}^{(2\nu, 2\mu)} + L_{00}^{(2\nu+1, 2\mu+1)} \bar{\omega} \right) \bar{\omega}^{\nu+\mu} \omega^{\nu-\mu}. \quad (8.117)$$

In the case of rotational symmetry, all eikonal coefficients $L_{00}^{(n,m)}$ with multiplicity $m \neq 0$ and those with odd order $n = 2\nu + 1$ arise from misalignments. Although these coefficients are generally small, the attributed aberrations may be dominant for low orders due to the small absolute value $|\omega| \leq 0.02$ of the slope parameter ω in a medium-voltage electron microscope. At present, aberration-corrected electron microscopes utilize aperture angles up to $|\omega| \approx 0.05$.

Assuming perfect stigmatic imaging for the ideally aligned system, we obtain the total axial aberration at the image plane $z = z_i$ of the real system as

$$\begin{aligned} \Delta u_a(z_i) &= -2u_{\gamma i} \frac{\partial L_a(z_i)}{\partial \bar{\omega}} = -u_{\gamma i} \left\{ L_{00}^{(1,1)} + 2L_{00}^{(2,0)} \omega + 2L_{00}^{(2,2)} \bar{\omega} + 2L_{00}^{(3,1)} \omega \bar{\omega} + \bar{L}_{00}^{(3,1)} \omega^2 \right. \\ &\quad \left. + 3L_{00}^{(3,3)} \bar{\omega}^2 + 4L_{00}^{(4,0)} \omega^2 \bar{\omega} + 3L_{00}^{(4,2)} \bar{\omega}^2 \omega + \bar{L}_{00}^{(4,2)} \omega^3 + 4L_{00}^{(4,4)} \bar{\omega}^3 + \dots \right\}. \end{aligned} \quad (8.118)$$

For stigmatically imaging systems consisting of round lenses and multipoles with even multiplicity $m = 2\mu$, all coefficients with odd order and those with $n = 2$ arise from misalignment. The first term on the right-hand side defines the lateral displacement of the image points. The coefficients $L_{00}^{(2,0)}$ and $L_{00}^{(2,2)}$ of the first-order aberrations account for defocusing and twofold axial astigmatism, respectively. One eliminates the defocus by adjusting the current of the objective lens and the twofold astigmatism by means of a quadrupole stigmator. The third and the fourth term describe the axial coma. This second-order aberration arises primarily if the axes of the round lenses and/or multipoles are tilted with respect to the straight optic axis. Since this aberration has multiplicity $m = 1$, it cannot be eliminated by a sextupole stigmator as long as the paraxial path of rays of the ideal system is rotationally symmetric. In this case, the sextupole can only compensate for the threefold axial astigmatism $-3u_{\gamma i} L_{00}^{(3,3)} \bar{\omega}^2$. To compensate for the parasitic axial coma of misaligned round lens, one must employ dipoles [140]. The combination of the zero-order deflection of the dipoles with the third-order axial aberration of the round lenses produces the appropriate second-order combination aberration with multiplicity $m = 1$. We can also eliminate the axial coma by means of a sextupole by placing it at a position where the paraxial path of rays is astigmatic. This possibility

follows readily from the integral representation of the third-order polynomial of the perturbation eikonal (7.97) induced by sextupoles.

Substituting the representation (4.225) for the paraxial ray $u^{(1)}(z)$ and setting $\rho = \bar{\rho} = 0$, we obtain the third-order perturbation eikonal:

$$L_s^{(3)}(z_i, \rho = 0) = \operatorname{Re} \int_{z_0}^{z_i} H (\bar{\omega} \bar{u}_\omega + \omega \bar{u}_{\bar{\omega}})^3 dz = \operatorname{Re} \left(L_{00}^{(3,3)} \bar{\omega}^3 + L_{00}^{(3,1)} \bar{\omega}^2 \omega \right). \quad (8.119)$$

The integral representations of the complex eikonal coefficients

$$L_{00}^{(3,3)} = \int_{z_0}^{z_i} (H \bar{u}_\omega^3 + \bar{H} u_\omega^3) dz, \quad L_{00}^{(3,1)} = 3 \int_{z_0}^{z_i} (H \bar{u}_\omega^2 \bar{u}_{\bar{\omega}} + \bar{H} u_\omega^2 u_\omega) dz \quad (8.120)$$

show that the eikonal coefficient $L_{00}^{(3,1)}$ of the axial coma vanishes in the case of rotational symmetry ($u_{\bar{\omega}} = 0, u_\omega = u_\alpha = \bar{u}_\alpha$), whereas the coefficient $L_{00}^{(3,3)}$ of the threefold axial astigmatism stays finite. Hence, by placing a sextupole at a position $u_\omega \approx u_{\bar{\omega}} \neq 0$, that is, in the region of an astigmatic image, we introduce an axial coma and a threefold axial astigmatism. If we adjust the hexapole strength H to cancel the parasitic axial coma of the system, we introduce an additional threefold axial astigmatism. By placing another sextupole at a position within the stigmatic paraxial region ($u_{\bar{\omega}} = 0$), we compensate subsequently for the threefold astigmatism without affecting the preceding correction of the axial coma.

The term $-4u_{\gamma i} L_{00}^{(4,0)} \omega^2 \bar{\omega}$ describes the third-order spherical aberration, which does not vanish for perfect alignment. The remaining terms in the expression (8.118) account for the axial star aberration and the fourfold axial astigmatism shown in Fig. 8.10 for the special case of real eikonal coefficients $L_{00}^{(4,2)} = \bar{L}_{00}^{(4,2)}$ and $L_{00}^{(4,4)} = \bar{L}_{00}^{(4,4)}$. These aberrations are parasitic aberrations for rotationally symmetric systems. In this case, they are negligibly small compared with the spherical aberration. For systems composed of round lenses, quadrupoles, and octopoles, the axial star aberration and the fourfold axial astigmatism are not of parasitic nature since their magnitude is the same as that of the spherical aberration. In the presence of these aberrations, we need only to consider the parasitic aberrations of first and second order. However, if we compensate for the third-order aberrations of round lenses by means of a corrector, we must also provide means to nullify the residual azimuthal components of the star aberration and the fourfold axial astigmatism in order to improve the resolution appreciably. These parasitic components arise from azimuthal misalignment of the multipole elements. To compensate for the resolution-limiting coherent parasitic aberrations of aberration-corrected electron microscopes, one has developed a computer-assisted iterative alignment procedure, which measures the residual parasitic aberrations and cancels them by means of various stigmator elements placed at proper positions within the corrector. Only due

to this strategy has it become possible to push the resolution limit of aberration-corrected electron microscopes below 1 Å.

So far, we have only considered axial parasitic aberrations. However, in order to transfer a large field of view without distortion and all points with the same resolution, we must also keep the field aberrations sufficiently small. If we have compensated for the dominant third-order off-axial aberrations, the parasitic second-order field aberrations may become dominant, as it is the case in systems with curved axis. We derive the general representation of the second-order aberrations most conveniently from the expression (8.43) for the eikonal term $L_{\text{gi}}^{(2s+1)}$ by putting $s = 1$ and $\rho = w_o$. By neglecting the axial terms, we obtain for the third-order eikonal the representation

$$\begin{aligned} L_{\text{gi}}^{(3)} = \text{Re} \bigg\{ & L_{10}^{(3,3)} \bar{\omega}^2 \bar{w}_o + L_{01}^{(3,1)} \bar{\omega}^2 w_o + L_{10}^{(3,1)} \bar{\omega} \omega \bar{w}_o \\ & + L_{20}^{(3,3)} \bar{\omega} \bar{w}_o^2 + L_{20}^{(3,1)} \omega \bar{w}_o^2 + L_{11}^{(3,1)} \bar{\omega} w_o \bar{w}_o \\ & + L_{30}^{(3,3)} \bar{w}_o^3 + L_{21}^{(3,1)} \bar{w}_o^2 w_o \bigg\}. \end{aligned} \quad (8.121)$$

The terms of the third row do not contribute to the aberrations at the image plane because they do not depend on the complex aperture parameter ω . Assuming stigmatic and distortion-free paraxial imaging, we obtain for the second-order field aberrations the most general expression

$$\begin{aligned} -u_{\text{fi}}^{(2)}/u_{\gamma i} = 2 \frac{\partial L_{\text{gi}}^{(3)}}{\partial \bar{\omega}} = & 2L_{10}^{(3,3)} \bar{\omega} \bar{w}_o + 2L_{01}^{(3,1)} \bar{\omega} w_o + 2\omega \text{Re} \left(L_{10}^{(3,1)} \bar{w}_o \right) \\ & + L_{20}^{(3,3)} \bar{w}_o^2 + \bar{L}_{20}^{(3,1)} w_o^2 + L_{11}^{(3,1)} w_o \bar{w}_o. \end{aligned} \quad (8.122)$$

The terms of the second row describe the second-order distortions. They do not affect the resolution but distort the image, as illustrated in Fig. 8.10 for an object consisting of concentric circles. In the most general case, the distortion is composed of the *conchoidal* distortion $\bar{L}_{20}^{(3,1)} w_o^2$, the *cross-eye* distortion $L_{11}^{(3,1)} w_o \bar{w}_o$ and the *trilobe* distortion $L_{20}^{(3,3)} \bar{w}_o^2$.

The first two terms $2\bar{\omega} \left(L_{10}^{(3,3)} \bar{w}_o + L_{01}^{(3,1)} w_o \right)$ define the field astigmatism. This aberration broadens the lateral image point to a circular spot whose radius depends on the maximum aperture angle $|\omega_{\text{max}}|$, the position $w_i = w_{\gamma i} w_o$ of the image point, and the values of the complex eikonal coefficients $L_{10}^{(3,3)}$ and $L_{01}^{(3,1)}$.

The term $2\omega \text{Re} \left(L_{10}^{(3,1)} \bar{w}_o \right)$ describes the image tilt. This aberration broadens the lateral image points at the Gaussian image plane z_i to circular discs in one direction, as depicted in Fig. 8.11. The diameters of the discs depend on the azimuthal and radial coordinates of the object points. We obtain a sharp image at a plane tilted with respect to the optic axis, as shown in Fig. 8.12. The aberration figure of the field astigmatism becomes an ellipse at defocused planes. The ellipse degenerates

Fig. 8.10 Effect of the individual second-order distortions on the image of object points located on circles in the object plane

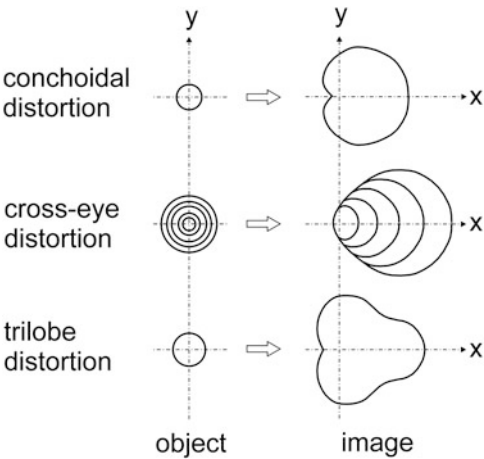


Fig. 8.11 Illustration of the broadening of image points by image tilt and field astigmatism

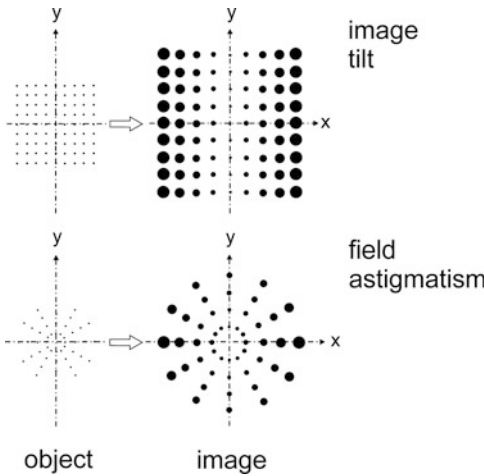
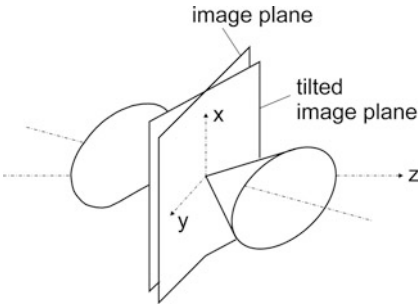


Fig. 8.12 Image tilt



to a straight line at two special planes: one located in front and the other behind the Gaussian image plane. The second-order aberrations become dominant in systems with curved axis, for example, imaging energy filters discussed in Sect. 8.1.1. One compensates most appropriately for the second-order distortions by imposing symmetry conditions on the system. We shall outline this procedure in more detail in Chap. 13.

8.6 Fifth-Order Aberrations of Orthogonal Systems with Straight Axis

The primary geometrical and chromatic aberrations determine the performance of most systems of charged-particle optics. The primary geometrical aberrations of orthogonal systems with a straight optic axis are of third order. In conventional electron microscopy, the spherical aberration and the coma are the dominant third-order lens defects because they limit the resolution and the field of view. In order to overcome these limitations, aberration correctors have been developed which compensate for the spherical aberration and the coma of the round objective lens. Systems corrected for these aberrations represent electron optical aplanats. To keep the number of non-rotationally symmetric aberrations introduced by the corrector as small as possible, we assume that electric potential φ of the corrector is symmetric with respect to two perpendicular plane sections, which we choose as the x, z -section and the y, z -section, respectively. The scalar magnetic potential ψ of the system is supposed to be antisymmetric with respect to these sections. Then the potentials satisfy the relations

$$\begin{aligned}\varphi(x, y, z) &= \varphi(x, -y, z) = \varphi(-x, y, z) = \varphi(-x, -y, z), \\ \psi(x, y, z) &= -\psi(x, -y, z) = -\psi(-x, y, z) = \psi(-x, -y, z).\end{aligned}\quad (8.123)$$

Accordingly, the multipole expansion of the potentials comprises only multipole terms with even multiplicity. The condition (8.1) for the electric potential implies that its power series expansion is composed of even powers in the lateral coordinates x and y . This is the case if the electric multipole strengths (3.26) are real: $\Phi_{2m} = \bar{\Phi}_{2m} = \Phi_{2m,c}, \Phi_{2m,s} = 0$. We satisfy the condition for the scalar magnetic potential by choosing imaginary multipole strengths: $\Psi_{2m} = -\bar{\Psi}_{2m} = i\Psi_{2m,c}, \Psi_{2m,s} = 0$. With these assumptions, we readily obtain from (3.24) and (3.31) the multipole expansions

$$\varphi = \sum_{m=0}^{\infty} \sum_{\lambda=0}^{\infty} (-1)^{\lambda} \frac{(2m)!}{\lambda!(\lambda+2m)!} \left(\frac{x^2 + y^2}{4} \right)^{\lambda} \operatorname{Re}(x - iy)^{2m} \Phi_{2m,c}^{[2\lambda]}(z), \quad (8.124)$$

$$\psi = - \sum_{m=0}^{\infty} \sum_{\lambda=0}^{\infty} (-1)^{\lambda} \frac{(2m)!}{\lambda!(\lambda+2m)!} \left(\frac{x^2 + y^2}{4} \right)^{\lambda} \operatorname{Im}(x - iy)^{2m} \Psi_{2m,s}^{[2\lambda]}(z). \quad (8.125)$$

The corrected objective lens consists of the standard magnetic objective lens and the corrector. To avoid the formation of anisotropic aberrations, we suppose that the axial magnetic field of the round lens $B(z) = -\Psi'_0(z)$ does not overlap the multipole fields of the corrector:

$$\Phi_{2m}(z)B(z) = 0, \quad \Psi_{2m}(z)B(z) = 0. \quad (8.126)$$

In this case, at least one field-free plane $z = z_s$ exists which separates the round-lens field from the multipole fields of the corrector.

We assume that the trajectories are fixed at the object plane $z = z_o$ by their complex slope $\omega = \alpha + i\beta$ and by their lateral position $w_o = x_o + iy_o$. Employing these boundary conditions, we derive the fifth-order aberrations of orthogonal systems by the same procedure that has been used in Sect. 7.2.2 for obtaining the fourth-order aberrations of systems which contain sextupole fields and/or dipole fields, respectively. Assuming stigmatic imaging and the substitutions discussed in Sect. 7.2.3, we eventually obtain for the geometrical fifth-order path deviation at the Gaussian image plane $z = z_i$ the relation

$$\begin{aligned} w^{(5)}(z_i) &= -2w_{yi} \frac{\partial L_{gi}^{(6)}}{\partial \bar{\omega}} - w_{yi} \left\{ \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}}, L_{gi}^{(4)} \right\}, \\ \left\{ \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}}, L_{gi}^{(4)} \right\} &= \left\{ \frac{\partial L_g^{(4)}(z_i)}{\partial \bar{\omega}}, L_g^{(4)}(z_i) \right\} = 2 \left[\frac{\partial^2 L_{gi}^{(4)}}{\partial \omega \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \bar{w}_o} - \frac{\partial^2 L_{gi}^{(4)}}{\partial \bar{w}_o \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \omega} \right. \\ &\quad \left. + \frac{\partial^2 L_{gi}^{(4)}}{\partial \bar{\omega}^2} \frac{\partial L_{gi}^{(4)}}{\partial w_o} - \frac{\partial^2 L_{gi}^{(4)}}{\partial w_o \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}} \right]. \end{aligned} \quad (8.127)$$

The sixth-order eikonal term is a polynomial of degree six in the complex ray parameters and given by the integral expression

$$L_{gi}^{(6)} = L_g^{(6)}(z_i) = \int_{z_o}^{z_i} \left[\mu_1^{(6)} - \frac{q_o}{q} \left| \frac{\partial \mu_1^{(4)}}{\partial w^{(1)}} \right| + \frac{1}{2} \left\{ \mu_1^{(4)}, L_g^{(4)} \right\} \right] dz. \quad (8.128)$$

The Poisson bracket in (8.127) accounts for the so-called fifth-order combination aberrations, whose coefficients are products of two fourth-order eikonal coefficients apart from an integer factor.

Because the functions $\mu_1^{(4)} = \mu_1^{(4)}(z)$ and $L_g^{(4)} = L_g^{(4)}(z)$ of the Poisson bracket in the integrand of (8.128) are real, we can write this bracket as

$$\left\{ \mu_1^{(4)}, L_g^{(4)}(z) \right\} = 2 \operatorname{Re} \left[\frac{\partial \mu_1^{(4)}}{\partial \omega} \int_{z_o}^z \frac{\partial \mu_1^{(4)}}{\partial \bar{w}_o} dz - \frac{\partial \mu_1^{(4)}}{\partial \bar{w}_o} \int_{z_o}^z \frac{\partial \mu_1^{(4)}}{\partial \omega} dz \right]. \quad (8.129)$$

The polynomials $L_g^{(2s)}$, $s = 2, 3, \dots$ given by the relation (8.42) are composed of sub-polynomials with even multiplicity $m = 2\mu \leq 2s$. In particular, we obtain

$$\begin{aligned}
 L_g^{(4)}(z) &= \text{Re} \sum_{\mu=0}^2 \sum_{v=0}^{2+\mu} \sum_{\lambda=0}^{2-\mu} L_{v\lambda}^{(4,2\mu)}(z) \bar{\omega}^{2+\mu-v} \omega^{2-\mu-\lambda} \bar{w}_0^v w_0^\lambda \\
 &= \text{Re} \sum_{\mu=0}^2 \left[L_{00}^{(4,2\mu)} \bar{\omega}^{2+\mu} \omega^{2-\mu} + L_{10}^{(4,2\mu)} \bar{\omega}^{1+\mu} \omega^{2-\mu} \bar{w}_0 \right. \\
 &\quad \left. + L_{01}^{(4,2\mu)} \bar{\omega}^{2+\mu} \omega^{1-\mu} w_0 + \dots \right]. \tag{8.130}
 \end{aligned}$$

We have omitted off-axial terms which depend nonlinearly and/or bilinearly on the ray parameters w_0 and $\bar{w}_0$ because they do not contribute to the fifth-order aperture aberrations. For orthogonal systems with double plane-section symmetry, all eikonal coefficients $L_{v\lambda}^{(2s,2\mu)}$ are real. However, they may be complex for magnetic round lenses due to the Larmor rotation. Because the fields of the round lenses are spatially separated from those of the multipoles, a field-free region exists between any neighboring round lens and multipole element.

8.6.1 Fifth-Order Aperture Aberration of Achromatic Third-Order Aplanats

The fifth-order aperture aberration limits the resolution of electron microscopes corrected for primary chromatic aberration and third-order spherical aberration if we have eliminated or sufficiently suppressed the deleterious parasitic aberrations. In order to minimize or nullify the components of the fifth-order aberration of orthogonal systems, it is advantageous to elucidate the structure of the constituent aberration coefficients or eikonal coefficients, respectively. The fifth-order combination aperture aberration vanishes for orthogonal systems corrected for third-order aperture aberration, as follows from the Poisson bracket representation

$$\begin{aligned}
 \left\{ \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}}, L_{gi}^{(4)} \right\}_{w_0=0} &= 2 \left[\frac{\partial^2 L_{gi}^{(4)}}{\partial \omega \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \bar{w}_0} - \frac{\partial^2 L_{gi}^{(4)}}{\partial \bar{w}_0 \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \omega} + \frac{\partial^2 L_{gi}^{(4)}}{\partial \bar{\omega}^2} \frac{\partial L_{gi}^{(4)}}{\partial w_0} \right. \\
 &\quad \left. - \frac{\partial^2 L_{gi}^{(4)}}{\partial w_0 \partial \bar{\omega}} \frac{\partial L_{gi}^{(4)}}{\partial \bar{\omega}} \right]_{w_0=0} = 0. \tag{8.131}
 \end{aligned}$$

In this case, the fifth-order aperture aberration $w_a^{(5)} = w^{(5)}\omega, \bar{\omega}; z_i)$ is determined entirely by the sixth-order eikonal component:

$$L_{ga}^{(6)} = L_g^{(6)}(\omega, \bar{\omega}; w_o = \bar{w}_o = 0, z_i) = \int_{z_o}^{z_i} \left[\mu_1^{(6)} - 2 \frac{q_o}{q} \left| \frac{\partial \mu_1^{(4)}}{\partial w^{(1)}} \right| + \frac{1}{2} \left\{ \mu_1^{(4)}, L_g^{(4)} \right\} \right]_{w_o=0} dz. \quad (8.132)$$

Orthogonal systems serve as correctors in high-resolution electron microscopes compensating for the primary aberrations of the magnetic objective lens. The axial component (8.132) of the sixth-order eikonal of the combined system depends strongly on the distance between the objective lens and the corrector. We assume that their fields vanish at the front nodal plane z_{N1} of the telescopic corrector. In this case, we can separate the contribution of the objective lens to the fifth-order aperture aberration from that of the corrector as follows:

$$\begin{aligned} L_{ga}^{(6)} &= L_{la}^{(6)} + L_{ca}^{(6)} + L_{lca}^{(6)}, \\ L_{la}^{(6)} &= \int_{z_o}^{z_{N1}} \left[\mu_1^{(6)} - 2 \frac{q_o}{q} \left| \frac{\partial \mu_1^{(4)}}{\partial w^{(1)}} \right| + \frac{1}{2} \left\{ \mu_1^{(4)}, \int_{z_o}^z \mu_1^{(4)} dz \right\} \right]_{w_o=0} dz, \\ L_{ca}^{(6)} &= \int_{z_{N1}}^{z_i} \left[\mu_1^{(6)} - 2 \frac{q_o}{q} \left| \frac{\partial \mu_1^{(4)}}{\partial w^{(1)}} \right| + \frac{1}{2} \left\{ \mu_1^{(4)}, \int_{z_{N1}}^z \mu_1^{(4)} dz \right\} \right]_{w_o=0} dz, \\ L_{lca}^{(6)} &= \frac{1}{2} \left\{ \int_{z_{N1}}^{z_i} \mu_1^{(4)} dz, \int_{z_o}^{z_{N1}} \mu_1^{(4)} dz \right\}_{\rho=0} = \frac{1}{2} \left\{ L_c^{(4)}, L_{ol}^{(4)} \right\}_{w_o=0}. \end{aligned} \quad (8.133)$$

The terms $L_{la}^{(6)}$ and $L_{ca}^{(6)}$ describe the contributions of the objective lens and the corrector to the sixth-order aperture eikonal. The round-lens contribution is given by

$$L_{la}^{(6)} = -\frac{1}{6} C_{5r} \bar{\omega}^3 \omega^3, \quad (8.134)$$

where $C_{5r} = \overline{C}_{5r}$ denotes the real coefficient of the fifth-order spherical aberration of the third-order corrected objective lens. The contribution of the corrector is given by

$$\begin{aligned} L_{ca}^{(6)} &= \sum_{\mu=0}^3 L_{00}^{(6,2\mu)} \bar{\omega}^{3+\mu} \omega^{3-\mu} \\ &= -\text{Re} \left(\frac{1}{6} C_{5c} \bar{\omega}^3 \omega^3 + S_5 \bar{\omega}^4 \omega^2 + R_5 \bar{\omega}^5 \omega + \frac{1}{6} A_5 \bar{\omega}^6 \right). \end{aligned} \quad (8.135)$$

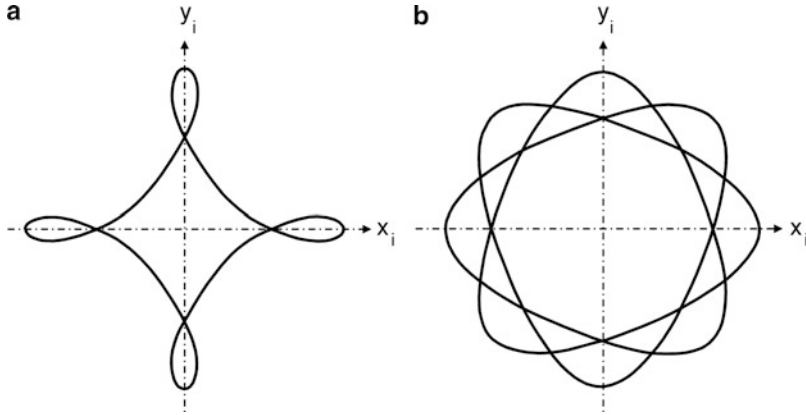


Fig. 8.13 Fifth-order aberration figures at the Gaussian image plane $z = z_i$ of (a) the axial star aberration and (b) the axial rosette aberration

The coefficients $L_{00}^{(6,0)} = -C_5/6$, $L_{00}^{(6,2)} = -S_5$, $L_{00}^{(6,4)} = -R_5$, $L_{00}^{(6,6)} = -A_5/6$ are real for orthogonal systems with double plane-section symmetry. The first term on the right-hand side produces the fifth-order spherical aberration, the second term the star aberration, the third term the rosette aberration, and the last term the sixfold axial astigmatism. The aberration figures of the star aberration and the rosette aberration at the Gaussian image plane are depicted in Fig. 8.13.

The star aberration and the sixfold axial astigmatism vanish for correctors with double symmetry such as the ultracorrector and the TEAM corrector, respectively. These systems are very suitable in practice because they introduce only a small number of additional non-rotationally symmetric aberrations. The term $L_{lca}^{(6)}$ in (8.133) accounts for the fifth-order axial combination aberration produced by the third-order spherical aberration and the off-axial coma of the objective lens and the corrector. We rewrite this eikonal term in the general form

$$\begin{aligned} L_{lca}^{(6)} &= \frac{1}{2} \{L_c^{(4)}, L_o^{(4)}\}_{\rho=0} = 2 \operatorname{Re} \left[\frac{\partial L_c^{(4)}}{\partial \bar{\omega}} \frac{\partial L_o^{(4)}}{\partial w_o} - \frac{\partial L_c^{(4)}}{\partial w_o} \frac{\partial L_o^{(4)}}{\partial \bar{\omega}} \right]_{w_o=0} \\ &= C_{3r} \operatorname{Re} \left[\omega^2 \bar{\omega} \left(\frac{\partial L_c^{(4)}}{\partial w_o} + \frac{\partial L_o^{(4)}}{\partial w_o} \right) \right]_{w_o=0}. \end{aligned} \quad (8.136)$$

Feasible correctors have a coma-free entrance plane which coincides with their front nodal plane z_{N1} . This behavior implies that the corrector does not introduce a third-order off-axial coma if we place an image of the effective source at the front nodal plane. In this case, the coma results entirely from the objective lens because

$$\left. \frac{\partial L_c^{(4)}}{\partial w_o} \right|_{w_o=0} = 0. \quad (8.137)$$

In order to obtain a third-order aplanat, the coma introduced by the objective lens must vanish too. To satisfy this condition, we must either place the front nodal plane z_{N1} of the corrector at the coma-free plane of the objective lens or image this plane into the plane z_{N1} without introducing coma. If the two planes are spatially separated by the distance $z_{N1} - z_c$, we obtain from (8.47) and (8.49)

$$\begin{aligned} \left. \frac{\partial L_o^{(4)}}{\partial w_o} \right|_{w_o=0} &= L_{01}^{(4,0)} \bar{\omega}^2 \omega = -\frac{1}{2} K_3 \bar{\omega}^2 \omega, \\ \text{Re } K_3 &= \frac{C_{3r}}{f_r^2} (z_{N1} - z_c). \end{aligned} \quad (8.138)$$

Inserting this result into (8.136) gives

$$\begin{aligned} L_{lca}^{(6)} &= -\frac{1}{6} C_{5com} \bar{\omega}^3 \omega^3, \\ C_{5com} &= -\frac{C_{3r}^2}{2f_r^2} (z_{N1} - z_c). \end{aligned} \quad (8.139)$$

The imaginary part of $K_{3i} = \text{Im } K_3$ of the coma coefficient of the objective lens does not contribute to $L_{lca}^{(6)}$. The coefficient C_{5com} of the fifth-order axial combination aberration is generally negative because the entrance nodal plane z_{N1} of the corrector can usually not be placed in front of the coma-free plane z_c of the objective lens ($z_{N1} > z_c$). We may nullify a positive fifth-order spherical aberration, at least in principle, by moving the corrector away from the objective lens. However, this procedure is generally accompanied by an unduly large increase of the third-order off-axial coma. The expression (8.139) is valid for any spherical-aberration-corrected system consisting of a round lens and a coma-free corrector, for example, the hexapole corrector outlined in Sects. 9.2.2 and 9.2.3.

Chapter 9

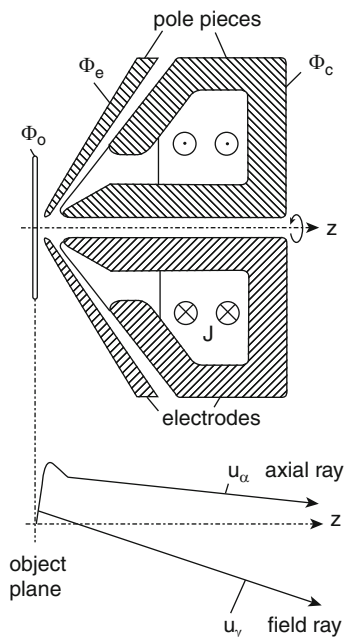
Correction of Aberrations

Abstract Several multipole correctors compensating for the unavoidable axial aberrations of round lenses are outlined. It is shown that the correction of chromatic aberration requires superposed electric and magnetic quadrupole fields which act as first-order Wien filters. The hexapole semi-aplanat which eliminates third-order spherical aberration and the radial off-axial coma is discussed in detail as well as quadrupole–octopole correctors compensating for spherical and chromatic aberration.

The Scherzer theorem imposes limitations on the performance of electron microscopes and other instruments employing round lenses. However, a positive-definite integrand of the integrals of the coefficients of spherical and chromatic aberration does not suffice to draw conclusions about the performance of round lenses. Although we cannot nullify the aberration coefficients, it may be possible to minimize the aberrations by skillful design to such an extent that their effect on the resolution is negligibly small. Unfortunately, this conjecture does not hold true because constraints exist for the design of realistic lenses. These limits are, for example, the maximum strength of the electric field, the magnetic saturation, a field-free working distance, and restrictions in realizing the required configurations of the electrodes and pole pieces. As a result, the relative flux density gradient B'/B of magnetic lenses cannot exceed a maximum value. By taking into account this constraint and employing the calculus of variations, Tretner optimized round lenses and derived minimum attainable values for their chromatic and spherical aberration coefficients [141, 142]. In the important case of magnetic round lenses, he found

$$f \geq f_{\min} = 0.8 \left| \frac{B}{B'} \right|_{\min}, \quad C_c \geq C_{c,\min} = \frac{1}{2} \left| \frac{B}{B'} \right|_{\min}, \quad C_3 \geq C_{3,\min} = \frac{1}{4} \left| \frac{B}{B'} \right|_{\min}. \quad (9.1)$$

Fig. 9.1 Path of the fundamental rays and scheme of an electromagnetic compound immersion lens for low-voltage electron microscopes



Due to magnetic saturation of the pole pieces, the minimum values of $|B/B'|$ are rather large at voltages employed in transmission electron microscopy. Therefore, the minimum achievable values for the focal length and the aberration coefficients C_c and C_3 are larger than about 1 mm for voltages above 100 kV. Moses performed extensive analytical and numerical investigations to find magnetic lenses with smallest spherical aberration if the object is located in field-free space [143]. Present magnetic and electrostatic lenses are designed in such a way that their fields are close to the optimum. Unfortunately, we cannot utilize the results obtained from the optimization of conventional lenses for determining the configuration and performance of optimum compound lenses. In order to optimize these lenses, Preikszas and Rose [64] have developed a computer-aided semi-analytical procedure for determining magnetic and electrostatic compound lenses with minimum chromatic and spherical aberration for various constraints. As an example of the calculations, we have depicted in Fig. 9.1 the geometry the fundamental rays of an optimum immersion compound lens employed in low-energy electron microscopy.

Diffraction and spherical aberration limit the resolution of a conventional uncorrected electron microscope [144]. The resulting resolution limit is

$$d \approx 0.6 \sqrt[4]{C_3 \lambda^3}. \quad (9.2)$$

Because this limit is proportional to the fourth root of the coefficient C_3 , we must largely reduce this coefficient in order to increase appreciably the resolution $1/d$ for a fixed wavelength λ . Owing to magnetic saturation of the pole pieces, a significant reduction of the coefficient C_3 of the objective lens is not possible. Therefore, an

appreciable improvement in resolution is only possible by means of a corrector compensating for the resolution-limiting aberrations of the objective lens. Spherical aberration limits the resolution as long as the acceleration voltage Φ_0 at the object is larger than about 10 kV. For lower voltages, chromatic aberration becomes the dominant limitation of the resolution because the chromatic parameter $\kappa = \Delta E / E_0$ increases with decreasing energy $E_0 = e\Phi_0$ of the electrons. Therefore, in order to improve appreciably the resolution of low-voltage and photoemission electron microscopes, it is mandatory to compensate for both chromatic and spherical aberration [136, 145].

By lifting any of the constraints of the Scherzer theorem, it is possible to correct for spherical and chromatic aberration. Scherzer showed this possibility as early as 1947 and sketched for each relaxation means for correction. So far, the most successful ways of correcting spherical and chromatic aberration are the departure from rotational symmetry and the incorporation of a tetrode mirror. The problem encountered with mirrors is that we must find means to separate the incident beam from the reflected beam without introducing harmful dispersion and second-rank aberrations, which would negate the elimination of the original aberrations. Fortunately, we can achieve a nondispersive and largely aberration-free splitting of the beams by means of highly symmetric beam separators.

The incorporation of space charges necessitates the introduction of a foil [146, 147]. However, scattering of the electrons by the atoms within the foil produces a frosted-glass effect, which has prevented an improvement of resolution so far. By illuminating the object with a pulsed beam and reducing the potential applied to the central electrode of an electric einzel lens, it is possible to focus the slower electrons, which arrive somewhat later, in the same way as the faster electrons. This chromatic correction also reduces the spherical aberration because the non-paraxial electrons, which travel a longer distance than the paraxial electrons, arrive a little later and encounter a weaker focusing of the outer zone of the lens. In the ideal case, this reduction equalizes the focusing of the inner and outer zones of the lens, thus providing spherical correction. This time-dependent correction procedure may become useful for dynamic electron microscopy. Employing this method, Schoenhense and Spieker [148] have achieved chromatic correction for a laser-pulsed photoemission electron microscope. They utilize the fact that the faster electrons of the pulse are ahead of the slower ones. They achieve chromatic correction by decreasing the focal length of the electrostatic objective lens in such a way that all electrons are focused at the same plane regardless of their energy. The required frequencies for reducing rapidly enough the potential of the central electrode are in the range of gigahertz.

In the following, we shall discuss the various correction procedures offered by abandoning rotational symmetry because this avenue has proven most successful. We shall investigate separately the correction of chromatic and spherical aberration by mirrors in Chap. 10. The reason is that we must treat the optics of electron mirrors differently because the assumptions for the validity of the paraxial approximation break down close to the turning point where the gradients of the rays become very large. Early attempts and trends to correct aberrations up to 1966 are summarized and discussed extensively in the review article by Septier [149].

9.1 Correction of Chromatic Aberrations

The primary chromatic aberration is a second-rank aberration of first order and first degree. For systems with straight optic axis, we can only eliminate this aberration by means of electric quadrupoles in combination with either a magnetic quadrupole or an axial electric field. Since these elements also affect the paraxial path of electrons with nominal energy, we must find means, which allow us to adjust the chromatic correction without affecting the paraxial path of the electrons with nominal energy. We can readily achieve such a correction in systems with curved axis by means of sextupoles placed at positions where the dispersion is large. The sextupole does not affect the paraxial regime but couples the chromatic parameter of the dispersion with the geometric parameters of the paraxial rays resulting in a second-rank chromatic aberration. Since its aberration coefficient depends linearly on the hexapole strength, we can adjust it to cancel the chromatic aberration of the entire system.

9.1.1 First-Order Wien Filter

The conventional Wien filter consists of crossed electric and magnetic dipole fields perpendicular to the optic axis whose strengths are adjusted in such a way that the Lorentz force

$$\vec{F} = -e \left(\vec{E} + \vec{v} \times \vec{B} \right) \quad (9.3)$$

is zero for electrons with nominal velocity $\vec{v} = \vec{v}_n = \vec{e}_z v_n$. Since the dipole fields introduce dispersion, we call the filter a zero-order Wien filter. We can generalize the Wien filter by considering arbitrary mixed electric and magnetic multipole fields.

The primary action of the quadrupole fields is paraxial focusing. Employing crossed electric magnetic quadrupoles shown in Fig. 9.2, we can nullify their total focusing strength for a given velocity by counterbalancing the electric and the magnetic forces. Using complex notation, the total lateral force on an electron with velocity parallel to the optic axis is

$$F = F_x + iF_y = -e \left(E_x + iE_y + v_z[-B_y + iB_x] \right) = 2e \left(\frac{\partial \varphi}{\partial \bar{w}} + i v_z \frac{\partial \psi}{\partial \bar{w}} \right). \quad (9.4)$$

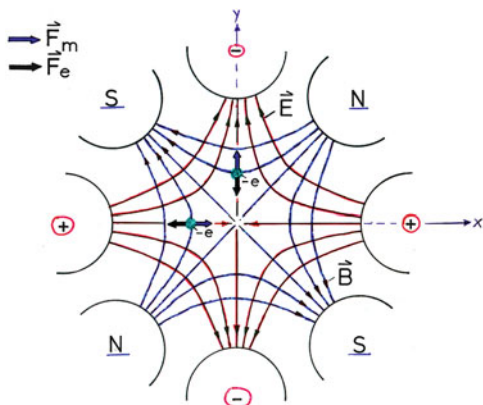
An electron whose energy differs from the nominal energy E_n by ΔE has the velocity

$$v_z = v_n + \Delta v \approx v_n \left(1 + \frac{\Delta E}{2E_n} \right) = v_n (1 + \kappa/2). \quad (9.5)$$

In the paraxial domain, the electric and magnetic potentials of crossed quadrupoles have the form

$$\varphi = \varphi_2 \approx \text{Re} \left(\Phi_{2c} \bar{w}^2 \right), \quad \psi = \psi_2 \approx \text{Re} \left(i \Psi_{2s} \bar{w}^2 \right). \quad (9.6)$$

Fig. 9.2 Electric and magnetic forces $\vec{F}_e$ and $\vec{F}_m$ acting on electrons within a first-order Wien filter composed of crossed electric and magnetic quadrupoles



Substituting these expressions for the potentials φ and ψ in (9.4) and the expression (9.5) for v_z , we obtain

$$F \approx -2e \left(v_n \Psi_{2s} - \Phi_{2c} + \frac{\kappa}{2} v_n \Psi_{2s} \right) \bar{w}. \quad (9.7)$$

Imposing the Wien condition $F = 0$ for $\kappa = 0$, we find that the components of the force

$$F_x = -\kappa e \Phi_{2c} x, \quad F_y \approx \kappa e \Phi_{2c} y \quad (9.8)$$

depend linearly on the position coordinates and the chromatic parameter κ . Therefore, the first-order Wien filter introduces a second-rank aberration of first order and first degree. We can adjust the force components by varying the strengths $\Phi_{2c} = v_n \Psi_{2s}$ of the electric and magnetic quadrupoles.

Relations (9.8) demonstrate that the first-order Wien filter is focusing in one principal section and defocusing in the other. In order to compensate for the axial chromatic aberration of round lenses, we need to deflect the faster electrons with $\kappa > 0$ toward the optic axis and the slower electrons away from the axis. We can achieve this for the x - z section by choosing a proper positive value for the electric quadrupole strength Φ_{2c} . Unfortunately, we double the y -component of the axial chromatic aberration if we place the first-order Wien filter within the stigmatic paraxial domain ($x_\alpha = y_\beta$). However, we can achieve chromatic correction by means of two filters each of which must be placed at one of two orthogonal line images within the astigmatic domain of a quadrupole corrector. The filter centered at the astigmatic line image $z_\alpha (x_\alpha(z_\alpha) = 0, y_\beta(z_\alpha) \neq 0)$ compensates for the y -component of the axial chromatic aberration, whereas the filter located at the line image z_β cancels independently the x -component.

The quadrupole quadruplet shown in Fig. 4.33 furnishes an appropriate corrector for eliminating the axial chromatic aberration of a scanning electron microscope if we substitute crossed electric magnetic quadrupoles for the two inner quadrupoles [102, 145]. We have depicted the course of the axial rays within the system consisting of the quadrupole corrector and the round lens in Fig. 9.3 for electrons

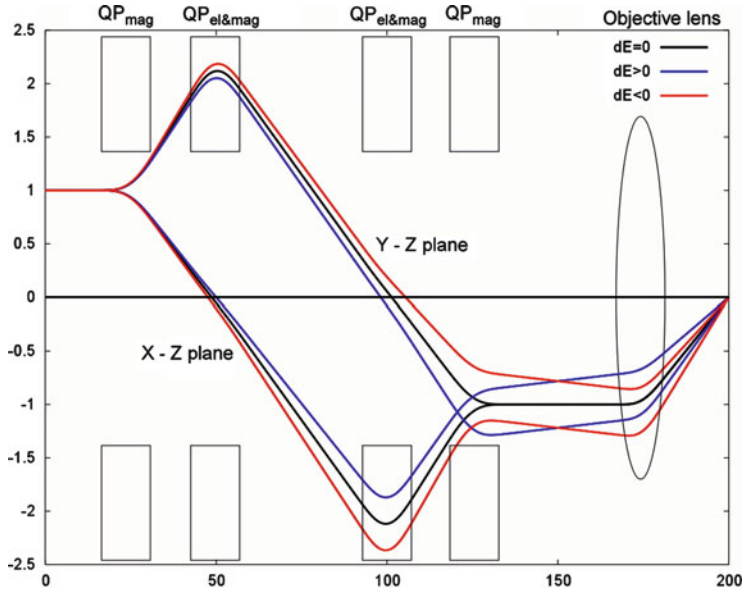


Fig. 9.3 Correction of the axial chromatic aberration by a quadrupole corrector composed of two magnetic outer quadrupoles and two crossed electric and magnetic inner quadrupoles. These elements act simultaneously as quadrupoles and as first-order Wien filters compensating for the axial chromatic aberration of the round lens

with nominal energy $\Delta E = 0$, with energy deviation $\Delta E > 0$ and with deviation $\Delta E < 0$. The figure illustrates that the slope of the rays in front of the object plane depends on the energy deviation. Applying the Helmholtz–Lagrange relation, we find that the corrector only compensates for the axial chromatic aberration but introduces elliptic (twofold) chromatic distortion. Hence, we cannot use this corrector for canceling the chromatic aberration of a conventional electron microscope.

We can also demonstrate formally the correction of chromatic aberration by means of the formulas (8.31) for the coefficients C_c and A_c of chromatic defocus and axial chromatic astigmatism, respectively, of arbitrary systems with straight optic axis. Assuming an orthogonal telescopic quadrupole system whose fields do not overlap with the magnetic field of the round lenses, the formulas (8.31) adopt the simple forms

$$C_c = C_{cR} + C_{cQ}, \quad C_{cQ} = -\frac{4}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_1} \sqrt{\frac{\Phi_0^*}{\Phi^*}} G_c u_\omega u_{\bar{\omega}} dz, \quad (9.9)$$

$$A_c = A_{cQ} = -\frac{2}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_1} \sqrt{\frac{\Phi_0^*}{\Phi^*}} G_c (u_\omega^2 + u_{\bar{\omega}}^2) dz, \quad G_c = \frac{\gamma_0}{4} G + \frac{1}{4} \frac{\Phi_{2c}}{\Phi^*}. \quad (9.10)$$

The part C_{cR} of the chromatic coefficient (9.9) denotes the contribution of round lenses to the chromatic defocus. The quadrupole strength $G = \gamma_0 (\Phi_{2c} - v_n \Psi_{2s}) / \Phi^*$, which acts on the paraxial electrons with nominal velocity v_n , vanishes if the crossed electric and magnetic quadrupoles satisfy the Wien condition. Imposing this condition, we vary $G_c = \Phi_{2c} / 4\Phi^*$ without affecting the paraxial path of rays of the electrons with nominal energy. By placing a Wien filter at the position $u_\omega = u_{\bar{\omega}}$ and another with opposite strength $G_{c2} = -G_{c1}$ at position $u_\omega = -u_{\bar{\omega}}$, we nullify the coefficient of chromatic defocus (9.9) without introducing an axial chromatic astigmatism (9.10). The relations (4.227) for the fundamental pseudo rays reveal that we precisely fulfill the first requirement at the astigmatic image plane $z = z_\beta$ and the other condition at the astigmatic image plane $z = z_\alpha$. Hence, our formal treatment leads to the same correction scheme as that obtained by means of intuitive physical considerations.

9.1.2 Correction of Chromatic Distortions

Chromatic distortion shifts the Gaussian image point in proportion to the energy deviation by the distance $u_{\text{ci}}^{(2)} = u_{\gamma i} k (D_{\text{cr}} w_o + D_{\text{ce}} \bar{w}_o)$. Since the image-forming electron beam has a continuous energy spread, the chromatic distortion transforms the Gaussian image point in a streak whose direction depends on the location of the Gaussian image point and on the coefficients of the chromatic distortion. Therefore, the chromatic distortion reduces the resolution of the object points with increasing lateral distance. This behavior differs from that of the third-order geometrical distortion, which does not affect the resolution. Therefore, we must also compensate for the chromatic distortion in order that all points of the transferred object area will be imaged with the same resolution. To minimize the number of aberrations introduced by the corrector, we require that it does not introduce paraxial astigmatism outside of the corrector. In this case, the round lenses do not contribute to the elliptical chromatic distortion. Hence, we must design the corrector in such a way that the elliptical chromatic distortion of its constituent elements cancels out. Because this is not the case for the corrector shown in Fig. 9.3, this corrector is not useful for a fixed-beam electron microscope. Assuming regular azimuthal orientations of the quadrupoles, the expression (8.36) for the coefficient of the elliptical chromatic distortion adopts the simple form

$$D_{\text{ce}} = \bar{D}_{\text{ce}} = -\frac{2}{1 + \varepsilon \Phi_o} \int_{z_o}^{z_i} \sqrt{\frac{\Phi_o^*}{\Phi^*}} G_c (u_\omega u_\rho + u_{\bar{\omega}} u_{\bar{\rho}}) dz. \quad (9.11)$$

We nullify this coefficient most appropriately by making the integrand an antisymmetric function. We can achieve this in two different ways by making the product of the fundamental pseudo rays of each pair u_ω, u_ρ and $u_{\bar{\omega}}, u_{\bar{\rho}}$ either symmetric or

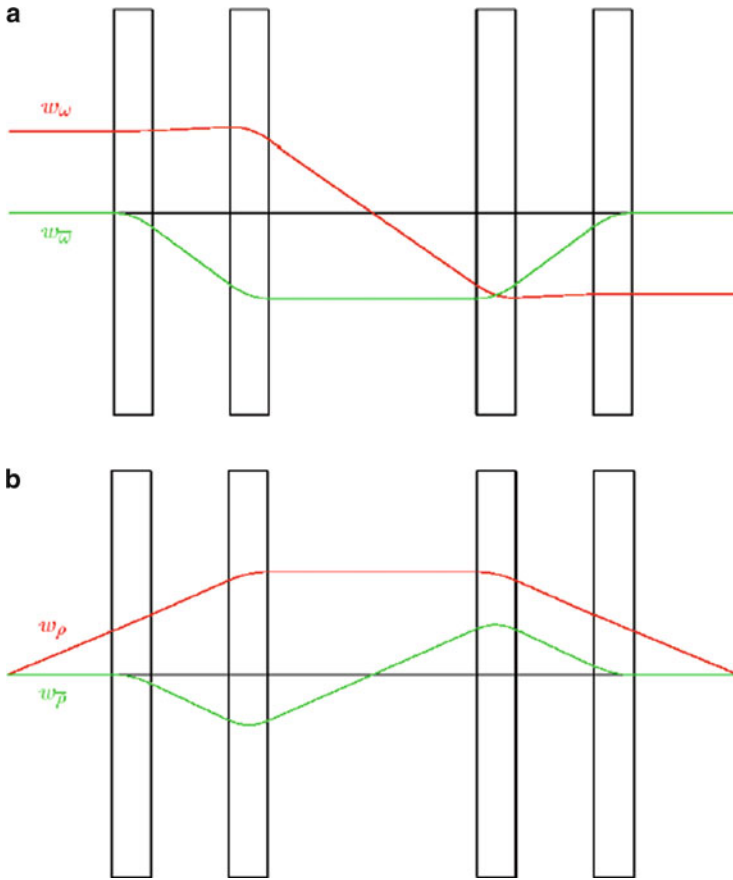


Fig. 9.4 Course of (a) the axial pseudo rays w_ω , $w_{\bar{\omega}}$ and (b) the pseudo field rays w_ρ , $w_{\bar{\rho}}$ within the telescopic antisymmetric quadrupole quadruplet; the corresponding fundamental rays are shown in Fig. 4.33

antisymmetric with respect to the midplane of the system or with respect to the central plane of each subsystem. In the first case, the chromatic quadrupole function G_c must be symmetric and in the second case, antisymmetric with respect to these planes. Since the products of the fundamental pseudo rays are antisymmetric with respect to the midplane of the antisymmetric quadrupole quadruplet shown in Figs. 4.33 and 9.4, the chromatic distortion of this system does not vanish.

The fundamental pseudo rays $u_{\bar{\omega}}$ and $u_{\bar{\rho}}$ are zero in the region of the round lenses, and the nonvanishing rays $u_\omega = u_\alpha$ and $u_\rho = u_\gamma$ coincide with the axial fundamental ray and the field ray, respectively. Hence, in this degenerate case, each fundamental pseudoray represents a possible ray. Considering these relations and employing the Helmholtz–Lagrange relation for the nonvanishing rays, the coefficient (8.35) of the chromatic round-lens distortion adopts the form

$$D_{\text{cr}} = D_{\text{cR}} + D_{\text{cr,Q}}. \quad (9.12)$$

The round lenses contribute the complex term

$$D_{\text{cR}} = \frac{1}{4(1 + \varepsilon \Phi_0)} \left(\gamma_{0i} \frac{\Phi_0^*}{\Phi_i^*} - \gamma_{0o} \right) + \frac{1}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} T_c u_\alpha u_\gamma dz \\ + \frac{i}{2} \int_{z_0}^{z_i} \frac{\Phi_0}{\Phi^*} \sqrt{\frac{e}{8m_e \Phi^*}} B dz \quad (9.13)$$

and the orthogonal quadrupole system the real term

$$D_{\text{cr,Q}} = \overline{D}_{\text{cr,Q}} = -\frac{2}{1 + \varepsilon \Phi_0} \int_{z_0}^{z_i} \sqrt{\frac{\Phi_0^*}{\Phi^*}} G_c (u_\omega u_{\bar{\rho}} + u_{\bar{\omega}} u_\rho) dz \quad (9.14)$$

to the coefficient (9.12) of the chromatic round-lens distortion of the total system.

The coefficient $D_{\text{cr,Q}}$ vanishes for the quadrupole quadruplet shown in Fig. 9.3, because the integrand of the integral (9.14) is antisymmetric for this system. The imaginary part of the distortion coefficient (9.13) equals half the angle of Larmor rotation between object and image in the absence of electrostatic round lenses ($\Phi' = 0$). Hence, this component vanishes in the case of rotation-free imaging.

One achieves this situation in a standard electron microscope by changing the directions of the currents in the coils of the constituent magnetic lenses. In the presence of axial chromatic aberration, we can affect the real part (9.13) of the distortion coefficient (9.12) by changing the direction of illumination or the location of the beam-limiting aperture. However, this kind of correction fails if we have eliminated the axial chromatic aberration. In this case, changing the field ray $u_\gamma \rightarrow u_\gamma + a u_\alpha$ does not alter the distortion coefficient. However, we can eliminate the chromatic distortion of electron microscopes or systems consisting of several lenses by exciting these lenses appropriately. To illustrate this possibility, we assume a large magnification $M_v \approx d_v/f_v$ for each lens; d_v is the distance of the intermediate image from the image principal plane of the lens v . The total magnification at the image plane formed by the N lenses of the system is

$$M_i \approx \prod_{v=1}^N \frac{d_v}{f_v}. \quad (9.15)$$

In the case of high magnification, the shift of the principal plane $\delta z_{Pv} = -\delta d_v$ caused by a change $\delta E = e\delta\Phi$ of the electron energy is small compared with the image distance d_v . Therefore, we can consider the distances d_v as constant with a sufficient degree of accuracy. With this assumption, we obtain

$$\frac{\partial M_i}{\partial \Phi} = M_i \sum_{v=1}^N \left(\frac{1}{d_v} \frac{\partial d_v}{\partial \Phi} - \frac{1}{f_v} \frac{\partial f_v}{\partial \Phi} \right) \approx -M_i \sum_{v=1}^N \frac{1}{f_v} \frac{\partial f_v}{\partial \Phi}. \quad (9.16)$$

The derivative of the focal length of a lens with respect to the axial potential is positive for weak excitations and zero at highest refracting power or shortest focal length, respectively. If we further increase the excitation of the lens, the derivative becomes negative. This change of sign enables us to nullify the energy dependence of the magnification (9.16) by choosing the excitations of the lenses appropriately. We can also vary the chromatic distortion by means of a field lens placed at an intermediate image. This lens does neither affect the magnification of the final image nor its location, yet it changes the course of the field rays and, hence, the chromatic distortion.

9.1.3 Electrostatic Correction of Chromatic Aberration

Contrary to magnetic systems, we can eliminate the chromatic aberration of electrostatic systems with straight axis. To prove this statement, we set $\chi = \chi' = 0$, $u = w$ in the expression (8.26) of the chromatic part of the third-rank perturbation eikonal. We eliminate in the integrand the electric quadrupole strength by means of the paraxial path equation

$$\frac{\Phi_2}{\Phi^*} \bar{w} = \frac{1}{\gamma_0} w'' + \frac{1}{2} \frac{\Phi'}{\Phi^*} w' + \frac{1}{4} \frac{\Phi''}{\Phi^*} w \quad (9.17)$$

and remove the second derivative of w by partial integration. As a result, we eventually find the representation

$$L_{ci}^{(3)} = \frac{\kappa \Phi_o}{4\Phi_o^*} \text{Re} \left\{ -\frac{1}{\gamma_0} \sqrt{\frac{\Phi_o^*}{\Phi^*}} \bar{w} w' \Big|_{z_o}^{z_i} + \int_{z_o}^{z_i} \frac{1}{\gamma_0} \sqrt{\frac{\Phi_o^*}{\Phi^*}} \left[(1 + \gamma_0^2) w' \bar{w}' + \frac{1 - 3\gamma_0^2}{2\gamma_0} \frac{\Phi'}{\Phi^*} \bar{w} w' \right] dz \right\}. \quad (9.18)$$

To facilitate our investigation, we assume regular quadrupoles ($\Phi_2 = \bar{\Phi}_2 = \Phi_{2c}$) so that the complex path equation (9.17) decouples in two real equations, one for the x -coordinate and the other for the y -coordinate. We obtain the x -component of the axial chromatic aberration by partial differentiation of the eikonal (9.18) with respect to the real slope parameter $a_1 = \alpha$ as

$$x_{ci}^{(2)} = -x_{yi} \frac{\partial L_{ci}^{(3)}}{\partial \alpha} = -x_{yi} \kappa (\alpha C_{c\alpha} + x_o C_{cy}). \quad (9.19)$$

We readily derive the coefficient $C_{c\alpha}$ from the expression (9.18) by substituting αx_α for w and considering that the fundamental axial ray $w_1 = x_\alpha$ vanishes at the object and image planes. By recasting the terms of the integrand, we eventually find

$$C_{c\alpha} = \frac{\Phi_o}{2\Phi_o^*} \int_{z_o}^{z_i} \frac{1}{\gamma_o} \sqrt{\frac{\Phi_o^*}{\Phi^*}} \left\{ (1 + \gamma_o^2) \left(\frac{x'_\alpha}{x_\alpha} - \frac{3\gamma_o^2 - 1}{4\gamma_o^2(1 + \gamma_o^2)} \frac{\Phi^{*'}}{\Phi^*} \right)^2 - \frac{(3\gamma_o^2 - 1)^2}{16\gamma_o^2(1 + \gamma_o^2)} \frac{\Phi'^2}{\Phi^{*2}} \right\} x_\alpha^2 dz. \quad (9.20)$$

We obtain the coefficient of the y -component of the axial chromatic aberration by replacing x_α by the fundamental axial ray y_β . The integrand consists of two squared terms with opposite sign. In order to compensate for the positive coefficient C_{cR} of the round lens, we need a correcting element whose coefficient is negative. The form of the integrand of the integral (9.20) reveals that such an element exists if we choose

$$\Phi_{2c}(z) = \frac{1}{4} \Phi'' + \frac{x'_\alpha}{2x_\alpha} \Phi' + \frac{x''_\alpha}{\gamma_o x_\alpha} \Phi^* \quad (9.21)$$

in such a way that the first term of the integrand is zero. This condition leads to the differential equation

$$4 \frac{x'_\alpha}{x_\alpha} = \frac{3\gamma_o^2 - 1}{\gamma_o^2(1 + \gamma_o^2)} \frac{\Phi^{*'}}{\Phi^*} = \frac{1 + 6\varepsilon\Phi^*}{(1 + 2\varepsilon\Phi^*)(1 + 4\varepsilon\Phi^*)} \frac{\Phi^{*'}}{\Phi^*}. \quad (9.22)$$

We solve this differential equation for x_α by decomposing the last fraction into parts. Integration of the resulting equation from the starting plane $z = z_s$ to the plane z gives

$$x_\alpha = x_{\alpha 1} \left(\frac{\gamma_o^2 (1 + \gamma_{os}^2)^2 \Phi^*}{\gamma_{os}^2 (1 + \gamma_o^2)^2 \Phi_s^*} \right)^{\frac{1}{4}}. \quad (9.23)$$

By substituting this expression for x_α in (9.21), we find the relation between the quadrupole strength and the axial potential as

$$\Phi_{2c} = \frac{\gamma_o^4 + 4\gamma_o^2 - 1}{4\gamma_o^2(1 + \gamma_o^2)} \Phi'' - \frac{12\gamma_o^2(\gamma_o^2 - 1)^2 + (1 + \gamma_o^2)^2 \Phi'^2}{16\gamma_o^3(1 + \gamma_o^2)^2} \frac{\Phi'^2}{\Phi^*}. \quad (9.24)$$

In the nonrelativistic limit $\gamma_o = 1 + 2\varepsilon\Phi \rightarrow 1$, $e\Phi \ll m_e c^2$, one calls the resulting relation

$$\Phi_{2c} = \frac{1}{2} \Phi'' - \frac{1}{16} \frac{\Phi'^2}{\Phi} \quad (9.25)$$

the *Scherzer condition* [16]. It relates the quadrupole strength of the correction unit to its axial potential. This element represents a straight-vision prism for electrons propagating in the x - z section with nominal energy if the axial potential Φ_∞ on the far side coincides with that at the starting plane in front of the prism. We can realize such an element with a sufficient degree of accuracy by means of three quadrupoles.

Purely electrostatic correctors are especially suitable for ion-optical instruments because the velocity of ions is very small in comparison to that of electrons for a given accelerating voltage. Owing to the large mass of the ions, we do not need to consider relativistic effects. Relation (9.25) shows that we must form the quadrupole field primarily in regions where the curvature of the axial potential is large, as it is the case at the locations of the aperture electrodes of an electrostatic einzel lens. Although a three-electrode element acts as a straight-vision prism with respect to the x - z section, it represents a strong focusing lens for the y - z section. In order to obtain a correction element, which is telescopic for both principal sections, we must place additional quadrupoles in front of and behind the correction element [150, 151].

The central element consists of three quadrupoles superposed with a decelerating axial field. This unit serves as the actual correction device. Its central quadrupole is diverging in the x - z section producing a negative axial chromatic aberration in this section. We can adjust this aberration by varying the axial potential in combination with that of the quadrupoles. To obtain a large negative chromatic aberration, one must maximize the lateral distance of the axial ray x_α and minimize the distance of the axial ray y_β in the other convergent section because it produces a positive chromatic aberration. We can satisfy this requirement only by means of additional quadrupoles which produce a strongly astigmatic path of rays in the region of the correction unit.

A suitable symmetric correcting element, which satisfies these conditions, consists of seven quadrupoles. The central quadrupole is at a lower average potential $\Phi_m = \Phi(z_m)$ than the other quadrupoles, thus producing an adjustable symmetric axial potential, as shown in Fig. 9.5. This correcting element acts like a thick telescopic lens for electrons with nominal energy. It only deflects electrons whose energies differ from the nominal energy, as it is the case for the first-order Wien filter. The axial rays x_α and y_β are symmetric to the midplane z_m of the correction element, while the field rays x_γ and y_δ are antisymmetric, as illustrated in Figs. 9.7 and 9.8, respectively.

The axial field overlaps only with the field of the three inner quadrupoles. We adjust these fields to compensate for the x -component of the axial chromatic aberration. The two outer quadrupoles $Q_{0,1}$ and $Q_{0,2}$ on each side of central correction unit provide a strongly distorted image of the diffraction plane at the central plane z_m . Only then is the distance of the axial ray y_β within the correction unit sufficiently small so that it does not produce an appreciable positive chromatic aberration in the y - z section. We have optimized the arrangement such that the quadrupole strength $\Phi_2(z) = \Phi_{2c}(z)$ satisfies the Scherzer condition with a sufficient degree of accuracy to give a negative chromatic aberration. To demonstrate this behavior, we have depicted in Fig. 9.6 the course of the real quadrupole strength and the course of the ideal quadrupole strength $\Phi_{2,id} = \Phi_{2,Scherz}$, which satisfies the Scherzer condition (9.25).

Because we must largely reduce the axial potential at the central quadrupole, this element strongly defocuses in the x - z section and strongly focuses in the y - z section forming two astigmatic images: one located in front and the other behind the

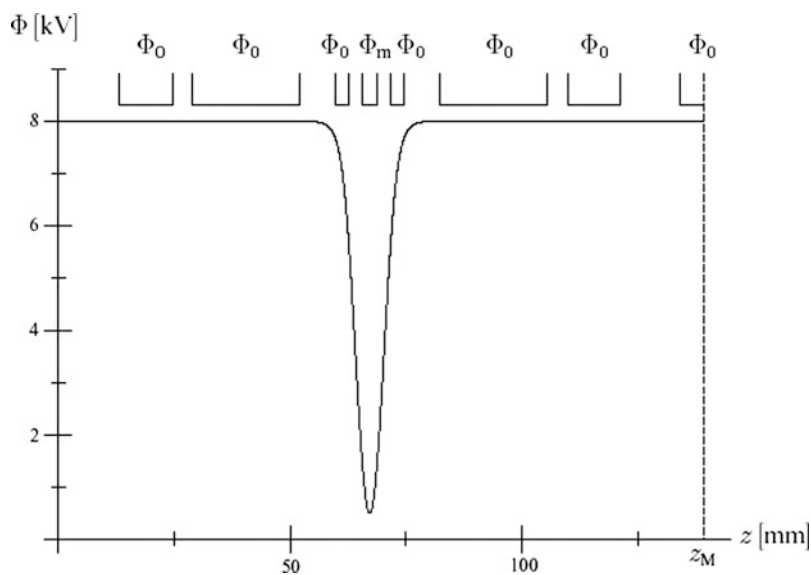


Fig. 9.5 Course of the axial potential $\Phi = \Phi(z)$ within the correcting element

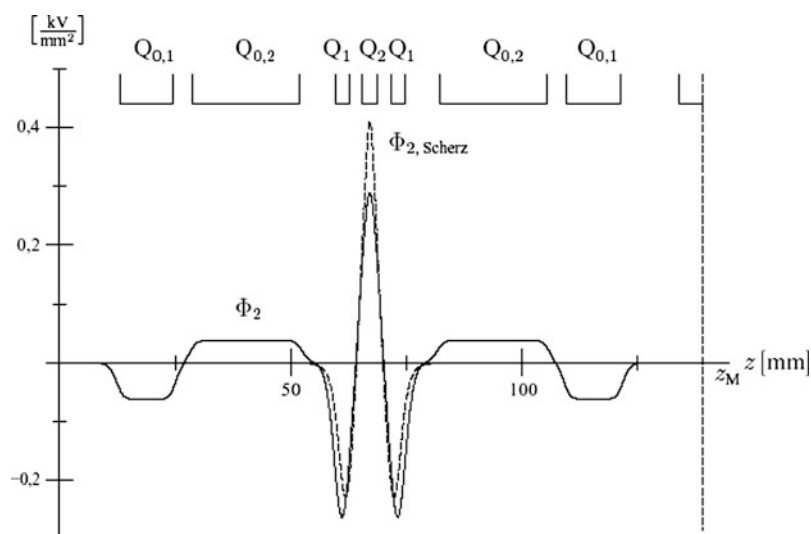


Fig. 9.6 Quadrupole strength $\Phi_2(z)$ along the optic axis within the correction element and optimum strength $\Phi_{2,\text{Scherz}}(z)$ of the ideal field providing largest negative chromatic aberration

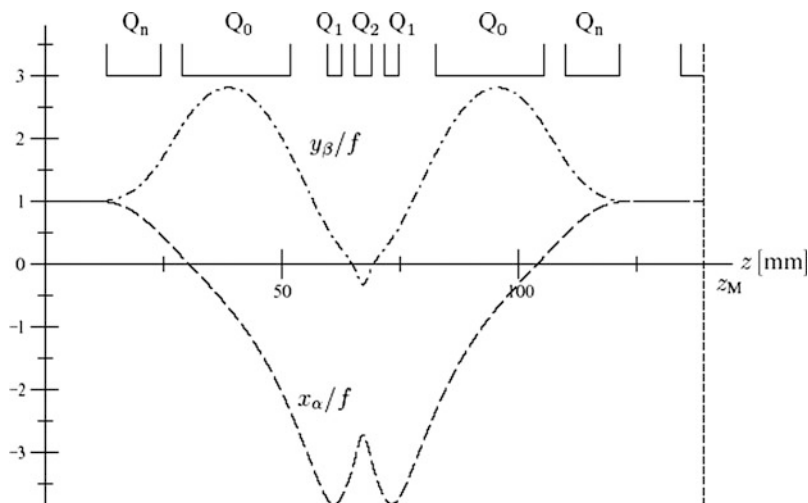


Fig. 9.7 Course of the normalized fundamental axial rays within the first subunit of electrostatic corrector; f is the focal length of the objective lens

quadrupole. In order to achieve chromatic correction in both principal sections, we need two correcting elements whose quadrupoles are excited with opposite polarity. The elements are separated by a distance such that a first-order distortion-free image of the diffraction plane is located at the midline z_M of the corrector. In this case, the corrector does not introduce chromatic distortion, third-order coma, and aberrations with twofold symmetry.

The advantage of the electrostatic corrector is that it allows a fast and reproducible alignment and its suitability for focusing ions regardless of their mass. Its disadvantages are the large number of quadrupoles and the extreme stability requirements for the central correction units. Therefore, we can state that electrostatic correction of chromatic aberration shown in Fig. 9.7 is primarily suitable for ion-optical instruments operating at low voltages, whereas crossed electric and magnetic quadrupoles are most feasible in the case of electrons.

After correction of the first-order chromatic aberration, the third-order aperture aberration limits the resolution. We eliminate this aberration by means of three octopole fields, one placed at the midplane z_M between the two correction elements and one octopole field at the central plane z_m of each of the two correction units. These fields are excited together with the quadrupole fields within the central multipole element of the correction units. In order to avoid large dodecapole field components, these elements must consist of twelve electrodes. Because the octopole fields are located at stigmatic images of the diffraction plane, they do not introduce any field aberrations. The octopole at the midplane z_M eliminates the fourfold axial astigmatism, while those placed at the central planes of the correction units compensate for the spherical aberration without introducing any other third-order aberration.

So far, we have only considered chromatic correction of systems with straight axis. In this case, we need to incorporate correcting elements consisting of electrostatic quadrupoles in combination with magnetic quadrupoles or superposition with an axial electric field. These elements directly affect the paraxial path of rays. However, in systems with curved axis, it is possible to eliminate the chromatic aberration by sextupoles without affecting the paraxial rays.

9.1.4 Chromatic Correction of Systems with Curved Axis

Dipole fields deflect charged particles and, hence, they are the basic multipole components in systems with curved optic axis. This axis is usually the central trajectory of the beam and formed by a particle with nominal energy. The course of a particle, which moves initially along the optic axis with different energy, deviates from this axis. The resulting primary path deviation is the dispersion. If we center a sextupole about the optic axis at a position with nonvanishing dispersion, it produces second-order geometric aberrations and second-rank chromatic aberrations. We obtain the chromatic aberrations of first order and first rank most easily from the third-rank eikonal polynomial linear in the chromatic parameter κ . Employing the notation (7.96) for the total hexapole strength and the representation $u^{(1)} = \omega u_\omega + \bar{\omega} u_{\bar{\omega}} + \rho u_\rho + \bar{\rho} u_{\bar{\rho}} + \kappa u_\kappa$ for the paraxial ray, we find the contribution of the hexapole fields to this polynomial as

$$L_{\text{CH}}^{(3)} = 3\kappa \text{Re} \int_{z_0}^{z_1} \bar{H} u_\kappa (\omega u_\omega + \bar{\omega} u_{\bar{\omega}} + \rho u_\rho + \bar{\rho} u_{\bar{\rho}})^2 dz. \quad (9.26)$$

It follows from the expression (9.26) that the constituent monomials are of second order having even multiplicity with respect to the geometrical ray parameters. This change of multiplicity results from the dispersion because it laterally shifts the axis of a beam whose energy differs from the nominal energy. Since the course of the dispersion ray differs from that of the geometrical fundamental rays, we can eliminate in principle the chromatic aberrations by means of sextupoles without introducing second-order aberrations. For example, such a correction is possible in systems with double symmetry with respect to the field distribution and the course of the geometrical fundamental rays. One has applied these principles for designing so-called second-rank achromats, which are required to transport the beam in long beam lines or storage rings without an appreciable loss of particles. Second-rank achromats define systems corrected for all second-rank aberrations.

To obtain integral expressions for the first-degree chromatic aberration coefficients of sextupoles, we rewrite the expression (9.28) in the form

$$L_{\text{CH}}^{(3)} = \kappa \text{Re} \sum_{\mu=0}^1 \sum_{v=0}^{1+\mu} \sum_{\lambda=0}^{1-\mu} L_{v\lambda,1}^{(2,2\mu)} \bar{\omega}^{1+\mu-v} \omega^{1-\mu-\lambda} \bar{\rho}^v \rho^\lambda. \quad (9.27)$$

We use the notation $L_{\nu\lambda,l}^{(n,m)}$ for the coefficients of the chromatic monomials, where n denotes the order and m the multiplicity of the geometrical part of the monomials. The rank of the monomial $r = n + l$ is the sum of the order and degree l , which defines the exponent of the chromatic parameter κ . The polynomial (9.27) consists of seven monomials:

$$L_{00,1}^{(2,0)} = C_{\text{CH}}/2 = 6\text{Re} \int_{z_0}^{z_i} \bar{H} u_{\kappa} u_{\omega} u_{\bar{\omega}} dz, \quad L_{11,1}^{(2,0)} = 6\text{Re} \int_{z_0}^{z_i} \bar{H} u_{\kappa} u_{\rho} u_{\bar{\rho}} dz, \quad (9.28)$$

$$L_{01,1}^{(2,0)} = \bar{L}_{10,1}^{(2,0)} = D_{\text{cr}}/2 = 3 \int_{z_0}^{z_i} (H \bar{u}_{\kappa} \bar{u}_{\omega} \bar{u}_{\bar{\rho}} + \bar{H} u_{\kappa} u_{\bar{\omega}} u_{\rho}) dz, \quad (9.29)$$

$$L_{10,1}^{(2,2)} = D_{\text{ce}} = 3 \int_{z_0}^{z_i} (\bar{H} u_{\kappa} u_{\bar{\omega}} u_{\bar{\rho}} + H \bar{u}_{\kappa} \bar{u}_{\omega} \bar{u}_{\rho}) dz, \quad (9.30)$$

$$L_{00,1}^{(2,2)} = A_{\text{c}}/2 = 3 \int_{z_0}^{z_i} (\bar{H} u_{\kappa} u_{\bar{\omega}}^2 + H \bar{u}_{\kappa} \bar{u}_{\omega}^2) dz, \quad L_{20,1}^{(2,2)} = 3 \int_{z_0}^{z_i} (\bar{H} u_{\kappa} u_{\bar{\rho}}^2 + H \bar{u}_{\kappa} \bar{u}_{\rho}^2) dz. \quad (9.31)$$

The coefficients (9.28) of the monomials with multiplicity $m = 2\mu = 0$ and lower indices $\lambda = \nu$ are real, whereas the coefficients of the other monomials can be complex. They are real for systems with plane midsection symmetry. The coefficients (9.29) and (9.30) of the elliptic chromatic distortion and the chromatic round-lens distortion vanish if the hexapole strength H , the dispersion ray u_{κ} , and the axial pseudo rays $u_{\omega}, u_{\bar{\omega}}$ are symmetric and the field pseudo rays $u_{\rho}, u_{\bar{\rho}}$ are antisymmetric with respect to the midplane of the system. This behavior prevails if we exchange the symmetry properties of the axial rays and the field rays. Since the axial pseudo rays and the field pseudo rays have opposite symmetry, the integrands of the chromatic distortion coefficients (9.29) and (9.30) are antisymmetric in both cases. By placing a sextupole at a distortion-free stigmatic image $z = z_{\rho}$ of the diffraction plane located in the dispersive region, we only affect the coefficient A_{c} of the chromatic axial astigmatism since at this plane, we have $u_{\kappa}(z_{\rho}) \neq 0, u_{\rho}(z_{\rho}) = u_{\bar{\rho}}(z_{\rho}) = 0, u_{\bar{\omega}}(z_{\rho}) = 0, u_{\omega}(z_{\rho}) \neq 0$. In the same way, we can adjust the coefficient $L_{20,1}^{(2,2)}$ by inserting a sextupole at a distortion-free stigmatic image $z = z_{\omega}$ of the object plane without affecting the other coefficients. In order to compensate for the coefficient C_{c} of the rotationally symmetric component of the axial chromatic aberration, we must place sextupoles within the dispersive region at positions where the Gaussian ray path is strongly astigmatic, as it is the case at stigmatic line images of the object plane. At these planes, we have $u_{\omega} = u_{\bar{\omega}}$ or $u_{\omega} = -u_{\bar{\omega}}$. By placing a sextupole at each of these two line images and by exciting the sextupoles with opposite polarity, we can eliminate C_{c} and A_{c} . This correction procedure

corresponds to that for systems with straight axis employing first-order Wien filters discussed in Sect. 9.1.1 and illustrated in Fig. 9.3.

The correction of the primary chromatic aberrations also influences the second-order geometrical aberrations and vice versa. For example, in symmetric systems, which are free of second-order aberrations, the chromatic distortions vanish as well [79]. For these systems, we find from (8.17) and (8.18) the relations

$$C_{13\kappa} = 2A_{13\kappa}, \quad C_{24\kappa} = 2B_{24\kappa}, \quad (9.32)$$

where the coefficients $A_{13\kappa}$ and $B_{24\kappa}$ are given by the integral expressions (8.15) and (8.16), respectively.

Assuming that the fundamental ray $x_3 = x_\gamma$ is symmetric and the axial ray $x_1 = x_\alpha$ antisymmetric with respect to the symmetry plane z_M , we split up the primary dispersion ray (4.248) $x_\kappa = x_{\kappa s} + x_{\kappa a}$ into the symmetric part

$$x_{\kappa s} = x_1 \int_{z_M}^z \Lambda_c x_3 dz - x_3 \int_{-\infty}^z \Lambda_c x_1 dz, \quad \Lambda_c = -\frac{e}{q_0} \Psi_{1s} \quad (9.33)$$

and an antisymmetric part

$$x_{\kappa a} = x_1 \int_{-\infty}^{z_M} \Lambda_c x_3 dz = x_1 C_{3\kappa}/2. \quad (9.34)$$

Since the magnetic field and the fundamental rays $x_3 = x_\gamma, y_4 = y_\delta$ are symmetric and the axial rays $x_1 = x_\alpha, y_2 = y_\beta$ are antisymmetric with respect to the midplane z_M , we can rewrite each integrand of the coefficients (8.15) and (8.16) as a sum of an antisymmetric term containing the symmetric part (9.33) of the dispersion ray and a symmetric term formed by the antisymmetric part (9.34). The integrals of the antisymmetric integrands vanish. Hence, only the antisymmetric part $x_{\kappa a}$ of the dispersion ray gives a contribution to the eikonal coefficients. Since this part is proportional to the axial ray, the chromatic coefficients (9.32) relate linearly with the coefficients of the geometrical aberrations as

$$C_{13\kappa} = A_{113} C_{3\kappa}, \quad C_{24\kappa} = B_{124} C_{3\kappa}. \quad (9.35)$$

Accordingly, the chromatic distortions vanish if either the dispersion ray is symmetric with respect to the central plane ($C_{3\kappa} = 0$) or the coefficients of the mixed geometrical aberrations are zero ($A_{113} = B_{124} = 0$). In order that we can utilize the system as an imaging energy filter for analytical electron microscopes, we must maximize the dispersion at the energy selection plane behind the filter. This condition is necessary to enable imaging of the energy-loss spectrum. Then the chromatic second-rank path deviations do not vanish outside the filter even if we eliminate the second-order geometrical deviations. Fortunately, the remaining chromatic deviations either vanish or are negligibly small at the energy selection

plane and the final image plane if we place the filter at a position where the intermediate magnification is sufficiently large.

One utilizes symmetric dispersive systems as monochromators, which reduce the energy width of the beam by allowing only electrons within a given energy window to pass the energy selection slit. We must place this slit at an image of the source located at the midplane of the system where the dispersion has a maximum [83, 152]. The optic axis of a feasible monochromator is Ω -shaped to form a straight-vision element because standard electron microscopes possess a straight optic axis. Hence, the slope of the dispersion ray must be zero at the midplane z_M . Therefore, the dispersion ray u_κ is symmetric with respect to the midplane so that the system is nondispersive as a whole in first degree. To allow for accurate energy filtering, the aberrations at the midplane must stay small in the direction of the dispersion. One places the monochromator behind the electron source at a position where the potential does not exceed 3–5 kV in order to obtain a dispersion of about $30 \mu\text{m}/\text{eV}$. Such a high dispersion is necessary for reducing the energy width below 0.2 eV with reasonable slit widths of several micrometers. Due to the high voltage to ground (100–300 kV) in most transmission electron microscopes, a purely electrostatic design is best suited. A dispersion-free monochromator corrected for second-order aberrations preserves the diameter and the emission characteristic of the effective source. Conserving the reduced brightness is important to maintain a sufficiently high current density within the spot when operating the microscope in the scanning mode. Since the relative energy width is very small behind the monochromator, we can tolerate the secondary dispersion. Elimination of the secondary dispersion by imposing midplane symmetry is only possible if the primary dispersion ray is zero at this plane and antisymmetric with respect to the central planes of each half of the monochromator, thus preventing energy selection.

Correction of all second-rank aberrations is achievable by employing repetitive symmetry [153]. One exploits this possibility in beam lines with curved axis consisting of identical cells, which form a one-dimensional lattice. In the simplest case, four cells suffice to yield a second-rank achromat such as the one shown in Fig. 9.8. Each cell of this achromat consists of a dipole D, two quadrupoles Q_1 , Q_2 , and two sextupoles S_1 , S_2 . The dipoles curve the axis and introduce dispersion, whereas the quadrupoles provide paraxial focusing. The dispersion $\kappa u_\kappa(z)$ serves as the optic axis for electrons whose energy $E = E_0 + \Delta E$ differs from the nominal energy E_0 by the relative energy deviation $\kappa = \Delta E/E_0$.

We have decomposed in Sect. 3.3 the electromagnetic potentials in multipole components centered about the curved optic axis. Here, we assume that the axis represents the trajectory of a distinct electron with nominal energy. Accordingly, we call this axis the nominal optic axis. If the energy of the electron differs from the nominal energy, this electron forms a different optic axis. As a result, the symmetry axes of the multipoles are laterally displaced from the new optic axis. The dispersion represents the lateral displacement of the shifted axis from the nominal optic axis.

By laterally displacing a multipole field with multiplicity m , we introduce in first approximation additional multipole fields with multiplicity $m = \pm 1$ with respect to the fixed optic axis. For example, the third-order terms of the dipole

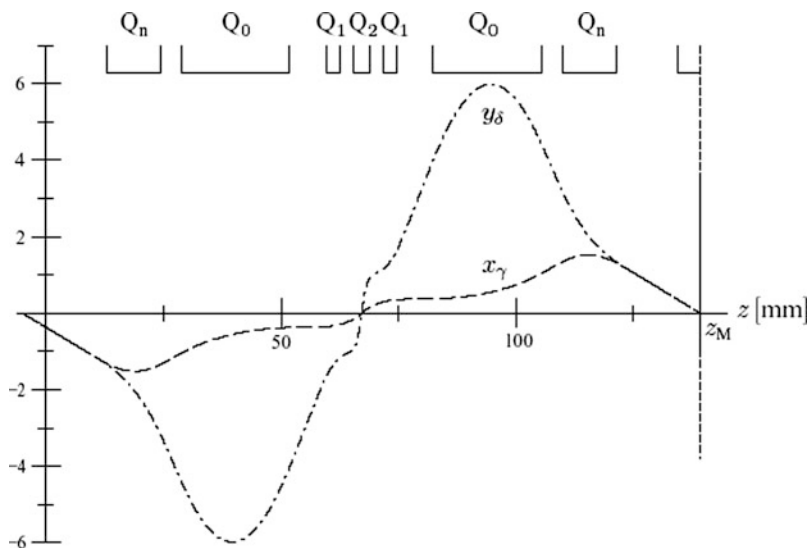


Fig. 9.8 Course of the fundamental field rays within the first half of the electrostatic corrector

potential produce additional quadrupole and round-lens fields with respect to the new optic axis. Their strengths are proportional to the energy deviation. Hence, in order to compensate for the resulting defocus and first-order astigmatism, we need two additional quadrupoles with opposite polarity per cell whose strengths are proportional to the dispersion. We obtain the required quadrupole fields by adjusting the strengths of the sextupoles S_1 and S_2 appropriately. Compensating for the defocus and the astigmatism by means of the quadrupole fields induced within the sextupoles by the chromatic displacement of the optic axis cancels simultaneously the first-order distortion at the exit plane of the system. This behavior is a consequence of the Helmholtz–Lagrange relation for the paraxial rays in the new coordinate system. Referred to the nominal coordinate system, these first-order relations correspond to the correction of both the axial chromatic aberration and the chromatic distortion.

By requiring that the first-order transfer matrix of the entire four-cell system equals the identity matrix, the fundamental rays satisfy the periodicity relation $u_\mu(z + 2l) = -u_\mu(z)$. The multipole fields have repetitive symmetry $\Psi_{ms}(z + l) = \Psi_{ms}(z)$. Therefore, we can conceive this system as a curved-axis analogue of a light-optical telescopic system with magnification $M=1$ consisting of four identical lenses. The cell length l corresponds to the distance $2f$ between the principal planes of two adjacent lenses. Since, the integrand of the third-order eikonal coefficients (8.15) and (8.16), contains odd powers of one of the fundamental rays, the contribution of the third cell compensates for that of the first cell and the contribution of the fourth cell cancels that of the second cell, as illustrated in Fig. 9.9. The same holds true for the paraxial dispersion ray:

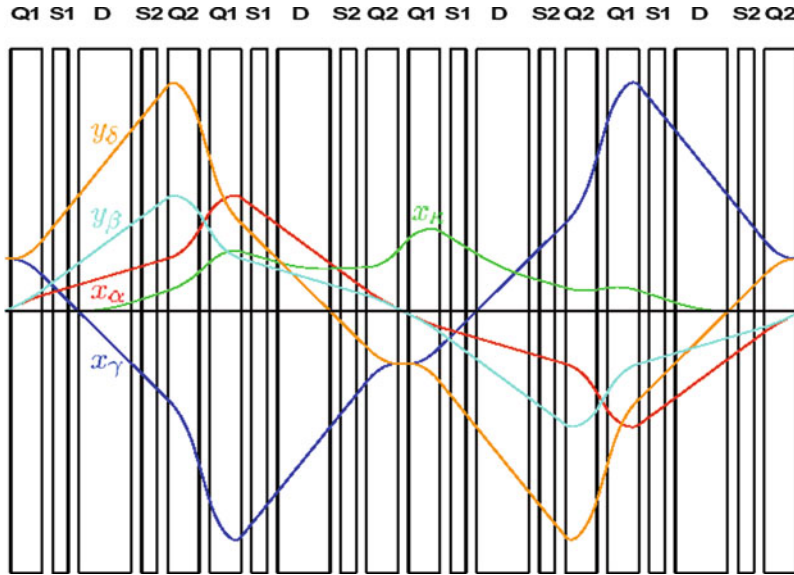


Fig. 9.9 Course of the fundamental rays along the straightened optic axis within the telescopic second-rank achromat consisting of four identical cells, each of which is composed of a dipole D, two quadrupoles Q₁, Q₂, and two sextupoles S₁, S₂

$$u_k = x_k = \frac{\gamma_0}{1 + \gamma_0} \frac{e}{q} \left(x_\alpha \int_{z_0}^z \Psi_{1s} x_\gamma dz - x_\gamma \int_{z_0}^z \Psi_{1s} x_\alpha dz \right). \quad (9.36)$$

We readily obtain this formula from Eq. (4.263) by setting $\Phi_1 = \Phi' = 0$, $\Phi_o = \Phi$. Surprisingly, the repetitive symmetry also nullifies the second-degree dispersion if we eliminate the other chromatic aberrations.

Equation (9.36) demonstrates that correction of the third-order geometrical aberrations by means of symmetries imposed on the fields and the course of the fundamental rays simultaneously eliminates the first-degree dispersion outside of the system. Hence, if we require dispersion outside of the system, we must abandon repetitive or double symmetry, respectively. It seems quite remarkable that all eight chromatic second-rank eikonal coefficients vanish simultaneously with the introduction of only two sextupoles per cell. However, because we have four cells and the dispersion ray differs from zero at the position of six of the eight sextupoles, we have six sextupoles acting on the eight coefficients of the total chromatic aberration. It is due to the repetitive symmetry that each four of the eight sextupoles have the same strength. The first and the last sextupole only affect the geometric second-order aberrations because the dispersion vanishes at the location of these elements. Considering that the geometrical aberration coefficients are related with the chromatic coefficients, for example, by (9.35), we need six additional variables

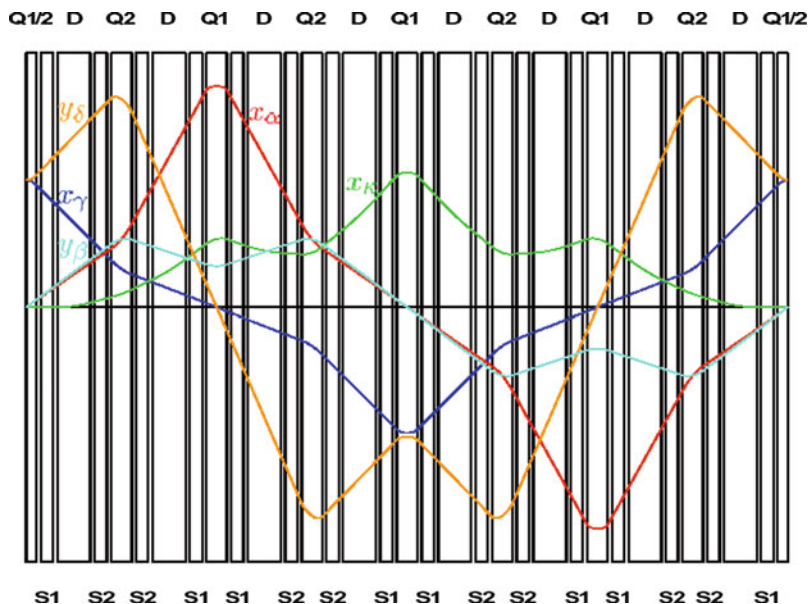


Fig. 9.10 Path of the fundamental rays along the straightened axis within the doubly symmetric second-rank achromat consisting of four symmetric cells, each of which is composed of two dipoles D; four quadrupoles, two $(1/2)Q_1$ and two Q_2 ; and four sextupoles, two S_1 and two S_2

to compensate for the eight chromatic coefficients of the second-rank component of the eikonal. We cannot reduce this number by imposing symmetry conditions, yet we can equalize each four of the eight sextupole strengths, thus reducing the requirements on the power supply considerably.

Systems with double symmetry illustrate the correction of aberrations by symmetry even better than systems with repetitive symmetry. To demonstrate convincingly this behavior, we choose a system composed of four identical symmetric cells shown in Fig. 9.10. This system has double symmetry as well as repetitive symmetry. Each cell consists of a dipole, two quadrupoles, and two sextupoles located symmetrically about the midplane of the central dipole. We arrange the cells in such a way that we introduce three symmetry planes for the fundamental rays and the multipole fields. Since the fundamental rays are linearly independent, two of these rays are symmetric, one for each principal section, while the other rays are antisymmetric about each of the symmetry planes. The rays, which are symmetric to the midplane z_M of the entire systems, are antisymmetric with respect to the midplanes z_{m1} , z_{m2} of each half of the system and vice versa. As a result, the integrands of the geometrical aberration coefficients become antisymmetric functions, either with respect to z_M or z_{m1} and z_{m2} . The same holds true for the integrands in (9.36) for the dispersion ray. Accordingly, the course of this ray must be symmetric with respect to the midplane z_M of the total system. The dispersion ray starts from the optic axis with zero slope at the entrance plane

of the first dipole magnet and vanishes in the same way at the exit plane of the last dipole. According to this symmetry, the coefficients of the second-order aberrations and the coefficients $C_{\kappa\alpha\gamma}$, $C_{\kappa\alpha\delta}$ of the chromatic distortion cancel regardless of the presence of sextupoles. We use these elements for eliminating the axial chromatic aberrations. Owing to the symmetry, the correction of two of the remaining five chromatic second-rank eikonal coefficients suffices to compensate for all the others.

9.1.5 Correction of Second-Degree Dispersion

The integral expressions (8.17) and (8.18) for the coefficients of the second-rank chromatic aberrations show that all imaging elements located between the object plane and the observation plane contribute to these aberrations. By transforming the integrals by partial integrations, we obtain the contribution of each element to the aberration coefficients. We subdivide the second-rank chromatic aberrations into the first-order and first-degree chromatic aberrations and the second-degree dispersion. Procedures for correcting the axial chromatic aberrations and the chromatic distortions have been discussed extensively in the preceding sections of this chapter. The second degree dispersion arises only in systems with a curved optic axis, for example, in an energy-filtering electron microscope equipped with an imaging omega filter. In this case, the second-degree dispersion at the image plane consists of two contributions, one originating within the filter, the other resulting from the combination of the first-degree dispersion with the chromatic aberrations of the lenses located between the filter and the image plane. However, only the first lens behind the filter contributes predominantly to the chromatic combination aberration because the axial ray u_α decreases with increasing magnification of the achromatic image located behind the filter. As a result, the contribution of each subsequent lens to the second-degree dispersion decreases inversely proportional to the square of the magnification of the intermediate image located in front of the lens.

At medium voltages and small energy windows of a few eV, it suffices to compensate for the first-degree dispersion at the image plane. However, at low voltages and large energy windows, the correction of all second-rank chromatic aberrations is necessary for achieving atomic resolution. So far, one has only accomplished the correction of the chromatic aberrations of first order and first degree. In order to compensate most conveniently for the second-degree dispersion, we insert two first-order (quadrupole) Wien filters (Fig. 9.2) within the region $z_1 \leq z \leq z_2$, as depicted in Fig. 9.11. In the ideal case, one quadrupole Wien filter is centered at the diffraction image D_1 in front of the energy filter, the other at the energy selection plane D_4 behind the energy filter such that the course of the fundamental rays u_α and u_γ is the same in both Wien filters apart from the sign. In this case, the symmetric omega filter images the front quadrupole Wien filter W_1 with unit magnification onto the back quadrupole Wien filter $W_2 = -W_1$. The electric quadrupole strength Φ_2 and the magnetic quadrupole strength Ψ_2 of each first-order Wien filter are chosen in such a way that the filters do not affect the

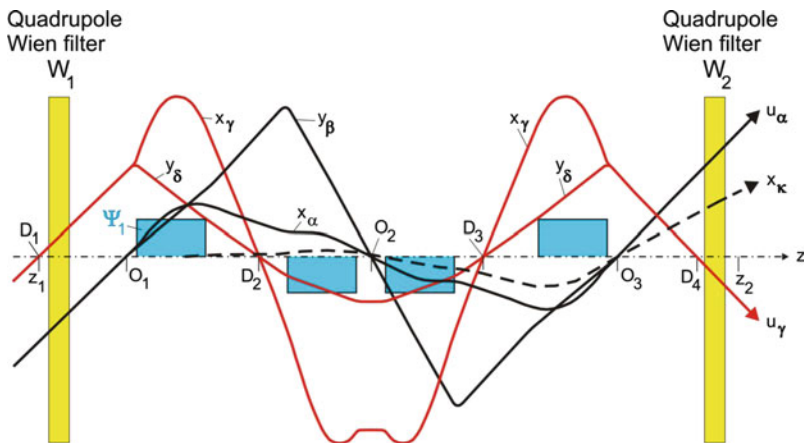


Fig. 9.11 Arrangement of the first-order (quadrupole) Wien filters W_1 and W_2 in the TEM equipped with the imaging omega filter. The Wien filters compensate for the second-degree dispersion at the image plane without introducing other second-rank aberrations. The fundamental axial rays x_α , y_β and the fundamental field rays x_γ , y_δ coincide in the region outside the deflection magnets of the energy filter ($x_\alpha = y_\beta = u_\alpha$, $x_\gamma = y_\delta = u_\gamma$). The blue boxes show the dipole strength $\Psi_1 = \Psi_1(z)$ of the deflection magnets along the straightened optic axis. Stigmatic intermediate images of the object plane z_0 are formed at the planes O_1 , O_2 , and O_3 ; those of the diffraction plane z_d are located at D_1 , D_2 , D_3 , and D_4 where the energy-loss spectrum is formed

paraxial course of the electrons with nominal energy $E_0 = e\Phi_0$. In this case, the quadrupole focal strength (4.100) vanishes ($G = 0$), giving

$$\Psi_2 = \gamma_0 \sqrt{\frac{m_e}{2e\Phi_0^*}} \Phi_2. \quad (9.37)$$

In practice, we must slightly shift the Wien filters from their ideal position, as shown in Fig. 9.11, because the energy selection slit is placed at the plane D_4 , thus preventing the incorporation of the Wien at this plane.

The first-order Wien filters introduce at the image plane the second-rank deviation

$$u^{(2)}(z_i) = u_{ci}^{(2)} = -u_{\gamma i} (C_{\alpha\kappa\kappa} \kappa^2 + A_c \kappa \bar{\omega} + D_{c\kappa} \bar{\omega}_0), \quad (9.38)$$

$$C_{\alpha\kappa\kappa} = \frac{1}{1 + \gamma_0} \int_{z_1}^{z_2} \frac{\Phi_2}{\Phi_0^*} e^{-2i\chi} u_\alpha u_\kappa dz, \quad (9.39)$$

$$A_c = \frac{1}{1 + \gamma_0} \int_{z_1}^{z_2} \frac{\Phi_2}{\Phi_0^*} e^{-2i\chi} u_\alpha^2 dz, \quad D_{c\kappa} = \frac{1}{1 + \gamma_0} \int_{z_1}^{z_2} \frac{\Phi_2}{\Phi_0^*} e^{-2i\chi} u_\alpha u_\gamma dz, \quad (9.40)$$

The front Wien filter W_1 does not contribute to the coefficient $C_{\alpha\kappa\kappa}$ of the second-degree dispersion because the dispersion ray $u_\kappa = x_\kappa$ starts from zero within the first deflection magnet of the energy filter and intersects the optic axis behind the filter at the achromatic image plane O_3 , as depicted in Fig. 9.11. The secondary dispersion

ray $x_{\kappa\kappa}$ usually intersects the optic axis at a plane which differs from the plane z_{ia} . In this case, the ray $u_{\kappa\kappa}$ is a linear combination of the axial ray u_α and the field ray u_γ in the post-filter region. We nullify the coefficients A_c and D_{ce} by choosing the quadrupole strengths Φ_2 and Ψ_2 of the Wien filters such they have opposite sign and equal absolute values ($W_2 = -W_1$). We compensate for the second-degree dispersion of the entire system by adjusting the quadrupole strengths of the Wien filters in such a way that the coefficient $C_{\alpha\kappa\kappa,W}$ introduced by the Wien filter W_2 is the negative of the coefficient $C_{\alpha\kappa\kappa,S}$ of the uncorrected system.

To judge the effect of the residual chromaticity, we must calculate the third-degree dispersion at the recording plane. This aberration consists of the third-degree dispersion of the filter and of two combination aberrations: one resulting from the combination of the first-degree dispersion with the spherical aberration of the projector lenses, the other from the combination of the second-degree dispersion with the first-rank chromatic aberration of the projector system. The latter combination aberration consists in general of a radial component and an azimuthal component produced by the combination of the second-degree dispersion with the azimuthal (anisotropic) chromatic distortion of the projector lenses. The azimuthal component vanishes if we nullify the total Larmor rotation of these lenses.

9.2 Correction of Geometrical Aberrations

Unlike the ideal instrument, real systems are initially always misaligned and suffer from mechanical inaccuracies and inhomogeneous magnetization of the pole pieces. These static defects cause parasitic aberrations of any order. Since the effect of these aberrations on the performance of the instrument is the stronger the lower the order of the aberrations, we must first compensate for parasitic aberrations whose orders are lower than that of the primary geometrical aberration of the ideally aligned instrument. For systems with curved optic axis, the primary geometrical aberrations are of second order, while they are of third order for system with straight axis containing exclusively multipole fields with even multiplicity. Therefore, we must first achieve perfect alignment up to the order of the primary aberrations in order that their correction improves the performance of the instrument. In systems with curved optic axis, the parasitic aberrations are of first order, whereas they are of first and second order in systems with a straight axis. Parasitic aberrations are resolution-limiting aberrations, which arise from mechanical imperfections and misalignments.

9.2.1 Correction of Second-Order Aberrations

In an electron microscope, one compensates in first-order only for the axial aberrations because the residual distortion is tolerable and does not impair the resolution.

The primary aberrations of the constituent round lenses are of third order, whereas those of the imaging energy filter are of second order. Therefore, we must design and place a suitable energy filter in such a way that many second-rank aberrations cancel and the third-order aberrations are small compared with those of the round lenses. Since the dispersion must not vanish behind the filter, we can only eliminate the second-order axial aberration and the distortion by imposing midplane symmetry. We need to eliminate the other second-order aberrations by means of sextupoles. To achieve an effective correction of these aberrations, the paraxial fundamental rays must differ substantially at all planes of the correcting sextupole elements. In this case, we decouple sufficiently the effects of the sextupoles so that each element affects primarily a single eikonal coefficient, thus preventing the formation of large third-order combination aberrations. We eliminate largely independently the nonvanishing eikonal coefficients $A_{\alpha\alpha\gamma}$, $B_{\alpha\beta\delta}$, and $B_{\gamma\beta\beta}$ of the field astigmatism and image tilt by placing sextupoles at astigmatic images of both the object plane and the diffraction plane. Since we have canceled half of the second-order aberrations by symmetry, it is necessary to incorporate the sextupoles in pairs placed symmetrically about the central symmetry plane of the filter. A sextupole centered at this plane does not need to be split because it automatically satisfies the symmetry condition. The correction of these aberrations introduces axial aberrations at the energy selection plane. Hence, we must eliminate subsequently their coefficients $A_{\gamma\gamma\gamma}$ and $B_{\gamma\delta\delta}$ without affecting the preceding correction of the other aberrations. We can satisfy this requirement by forming a strongly distorted image ($x_\gamma \ll y_\delta$) of the object plane at the midplane z_M of the filter. In this case, we need only three sextupoles, one placed at the midplane and the two others at undistorted conjugate images of the object plane, one located in front of and the other behind the filter. The smaller the ratio $x_\gamma(z_M)/y_\delta(z_M)$ is, the less the correction of $B_{\gamma\delta\delta}$ will affect the other coefficient $A_{\gamma\gamma\gamma}$ of the axial aberration at the energy-selection plane. Although the elimination of these coefficients is slightly coupled, it has the great advantage not to introduce any other second-order aberration.

Our example shows that we need nine sextupoles for correcting rather independently five second-order aberrations, one more than we needed for the second-order achromats. If we tolerate that the mixed aberrations are not eliminated independently, we need only two sextupole pairs to compensate for the mixed coefficients $A_{\alpha\alpha\gamma}$, $B_{\alpha\beta\delta}$, and $B_{\gamma\beta\beta}$. However, we must place the sextupoles at distinct locations in order to compensate for three coefficients with two adjustable sextupole strengths. All present corrected energy filters utilize this possibility. This example convincingly demonstrates the increase in complexity of aberration correction by abandoning repetitive or double symmetry.

The advantage of double symmetry for eliminating second-order aberrations and most of the chromatic second-rank aberrations holds even if we cannot incorporate sextupoles. We convincingly demonstrate this behavior by means of the beam separator shown in Fig. 9.12. This system is part of the SMART mirror corrector and separates the incident beam from the deflected beam [154]. The beam separator consists of two plane-parallel iron plates containing loop-shaped coils inserted into grooves on the inner surfaces of the plates. These grooves form the boundaries of

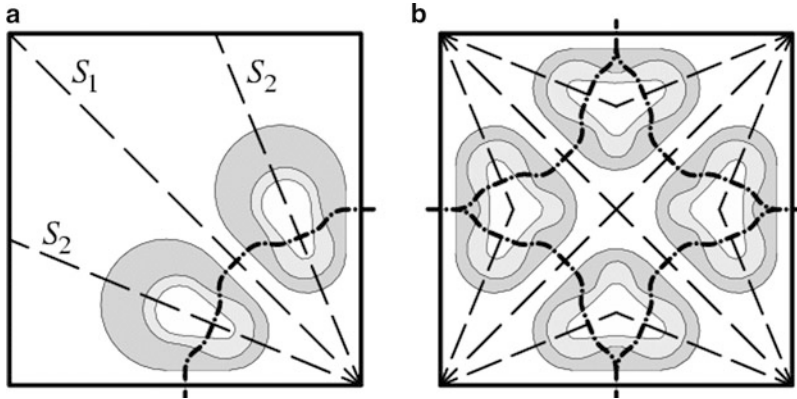


Fig. 9.12 Cross section of (a) the fourth quarter of the beam separator showing the double symmetry of the fields and the curved optic axis and of (b) the entire separator. The *shaded areas* represent the regions of the dipole field perpendicular to the pole plates. The sign and the strength of the dipole field differ for regions with different shading; the *dash-dotted curve* represents the optic axis

the shaded areas in Fig. 9.12. The total magnetic field consists of four identical quadrants each forming a system with double symmetry. Contrary to the second-rank achromat shown in Fig. 9.9, each of the four cells of a quadrant has two dipole components with opposite polarity forming a meandering curved optic axis.

The diagonal plane S_1 represents the midplane of the four-cell beam guiding system formed by the fourth quadrant. Each of the two planes S_2 represents a symmetry plane of each half of the quadrant.

We achieve focusing in the vertical y - z section, by means of the edge quadrupoles formed in the region of the fringing fields by tilting the boundaries of the magnetic dipole field with respect to the direction of the optic axis, as illustrated in Fig. 9.12. The strength of the fringe quadrupole and the derivative of the magnetic dipole strength are related by

$$\Psi_{2s} = -\frac{1}{2}\Psi'_{1s} \tan \theta(z). \quad (9.41)$$

Here, $\theta(z)$ is the angle enclosed by the direction of the optic axis and the normal to the isoinduction lines $B_y(x, z, y = 0) = -\partial\psi_1/\partial y|_{y=0} = \text{const.}$ along the optic axis ($x = 0, y = 0$). The quadrupole component vanishes if the optic axis is perpendicular ($\theta = 0$) to the isoinduction lines within the entire region of the fringing fields. These fields also introduce a hexapole component about the optic axis [96]. Its strength

$$\Psi_{3s}(z) = \frac{1 + 3 \sin^2 \theta}{24 \cos^3 \theta} (\Psi''_{1s} \cos \theta + \Gamma \Psi'_{1s} \sin \theta) - \frac{\Psi'_{1s} \cos 2\theta}{6\rho_m \cos^3 \theta} \quad (9.42)$$

depends on the tilt angle $\theta = \theta(z)$ and the local radius of curvature $\rho_m = \rho_m(z)$ of the magnetic isoinduction lines along the optic axis. We introduce this curvature by curving the boundary faces of the magnetic dipole fields in the region of the optic axis. Its curvature $\Gamma = \Gamma(z)$ must satisfy the relation (4.14) in order that the optic axis forms a possible trajectory. The hexapole strength (9.42) does not vanish if the isoinduction lines are straight ($\rho_m = \infty$), as it is the case for the beam separator shown in Fig. 9.10. We conceive this surprising result, if we consider that in the case $\theta = 0$ the remaining hexapole strength $\Psi_{3s} = \Psi'_{1s}/24$ guarantees that the third-order term of the scalar magnetic potential ψ will be independent of the x -coordinate. This condition must be satisfied because in the case of infinitely extended straight boundaries, the scalar magnetic potential $\psi = \psi(y, z)$ is two dimensional. The radius of curvature ρ_m of the magnetic field lines is positive if the isoinduction lines are convex with respect to the direction of flight of the axial electron and negative if this curvature is concave.

By adjusting the tilt angle of the grooves for the coils and the currents of the beam separator appropriately, we obtain a doubly symmetric course of the fundamental rays and of the dispersion ray for each quadrant of the beam separator, as demonstrated in Fig. 9.13. In order to obtain this double symmetry for a total deflection angle of 90° per quadrant, we need to introduce regions with opposite direction of the magnetic dipole field producing the meandering optic axis shown in Fig. 9.12.

To eliminate the chromatic aberrations of the beam separator, we must introduce adjustable hexapole fields within each of the four cells of the doubly symmetric quadrant. According to the relation (9.42), we can produce in principle the required hexapole strengths by properly curving the grooves, which define the field boundaries. However, a slight deviation of the optic axis from its nominal path produces parasitic quadrupole fields, which misalign the paraxial trajectories. Since we cannot adjust the quadrupole fields without varying the dipole fields, we must align the precise course of the optic axis by additional stigmators placed best at the symmetry planes. Unfortunately, experiments have shown that curving the pole faces of deflection magnets in multielement systems results in chaotic behavior during the alignment of the paraxial path of rays because the misalignment strongly increases along the system. So far, correction of second-order aberration in multielement systems with curved axis has been performed successfully only by means of actual sextupole elements.

The coefficients of the chromatic distortion of the doubly symmetric beam separator are zero because they vanish together with the coefficients of the third-order aberrations, as follows readily from relations (9.35). We best depict the correction of these aberrations by means of the secondary fundamental rays shown in Fig. 9.14. As illustrated in this figure, the geometrical secondary fundamental rays and the dispersion ray of second degree are either symmetric or antisymmetric with respect to the midplane S_1 . Since these rays start with vanishing slope from the axis at the entrance of the first dipole field, they leave the system in the same way. This behavior does not hold for the rays $x_{\gamma\kappa}, y_{\delta\kappa}$ of chromatic distortion and the axial

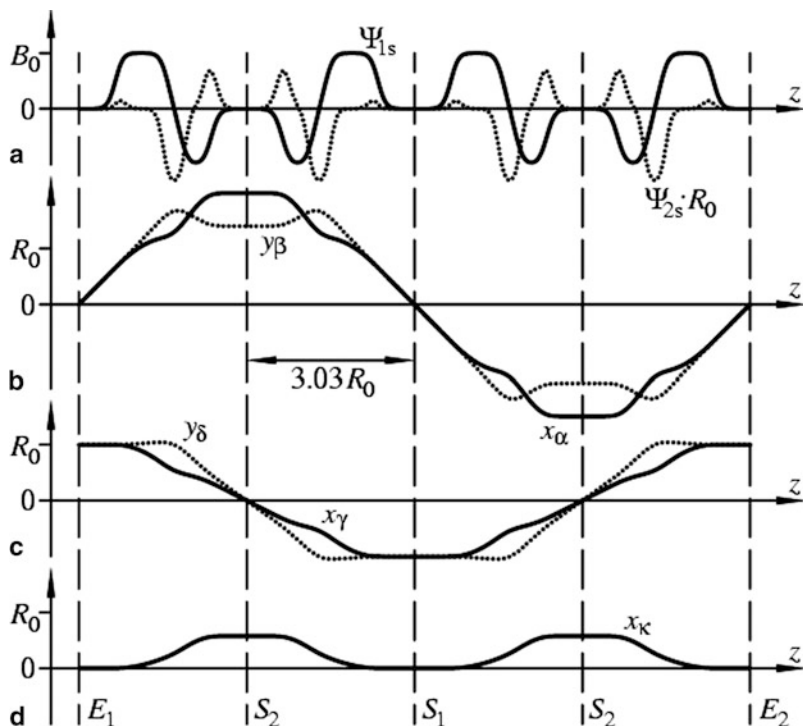


Fig. 9.13 (a) Magnetic dipole and quadrupole strengths, course of (b) the axial rays, (c) the field rays, and (d) the dispersion ray x_κ along the straightened optic axis within one quadrant of the beam separator shown in Fig. 9.12; $1/R_0 = e\Psi_{1s,\max}/q$ is the maximum curvature of the optic axis

chromatic rays $x_{\alpha\kappa}$ and $y_{\beta\kappa}$, which run parallel to the optic axis behind the beam separator (Fig. 9.13b). Since the lateral distances of these rays differ, they introduce chromatic defocus and chromatic axial astigmatism at the image plane. The rays of chromatic distortion intersect the optic axis at an intermediate image of the object plane. Therefore, they are zero at all subsequent conjugate planes including the final image plane.

We compensate for the chromatic axial astigmatism by incorporating a sextupole element at each symmetry plane S_2 . We adjust their strength Ψ_{3s}^{ext} in such a way that the axial chromatic rays coincide outside the beam separator ($x_{\alpha\kappa} = y_{\beta\kappa}$), thus eliminating the chromatic astigmatism (Fig. 9.14c). Because the sextupoles are placed at the symmetry planes S_2 and their fields coincide, they do not affect the course of the other secondary fundamental rays outside the beam separator. We eliminate the remaining chromatic defocus of the beam separator together with that of the round lenses by means of the electrostatic mirror. Hence, the system composed of beam separator and electrostatic mirror does not introduce any second-rank aberration at the image plane apart from an adjustable axial chromatic defocus.

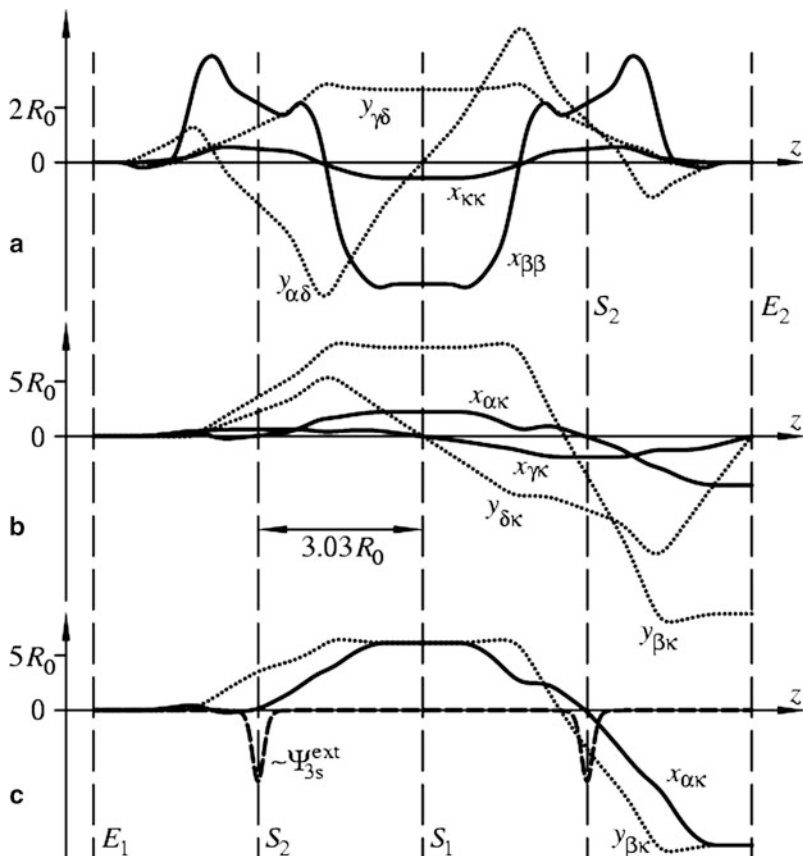


Fig. 9.14 Course of (a) the secondary geometrical fundamental rays $y_{\gamma\delta}$, $y_{\alpha\delta}$, $x_{\beta\beta}$ and the dispersion ray $x_{\kappa\kappa}$ of second-degree, (b) the secondary chromatic fundamental rays, and (c) the secondary axial fundamental rays after adjustment by two external sextupoles each placed at one of the symmetry planes S_2 within each quadrant of the beam separator

Owing to the double symmetry, we need two sextupoles for eliminating the axial chromatic astigmatism within each quadrant without introducing any third-order geometrical aberrations. Because the electrons pass through two quadrants, we have four sextupoles for correcting the axial chromatic astigmatism, exactly as many as required for the system shown in Fig. 9.9 exhibiting repetitive symmetry. Second-rank achromats with repetitive symmetry require at least eight sextupoles, four to eliminate the axial chromatic astigmatism and four to compensate for the chromatic defocus. Our investigations demonstrate that symmetries are an efficient means for canceling third-rank aberrations apart from the axial chromatic aberrations. We cannot compensate for these aberrations merely by imposing symmetries regardless if the optic axis is straight or curved.

9.2.2 *Correction of Third-Order Spherical Aberration*

The primary purpose of correctors is the compensation of the unavoidable aberrations of round lenses. Hence, the compensating aberrations of a corrector must be rotationally symmetric and of opposite sign with respect to those of the round lenses. Hexapole correctors introduce solely rotationally symmetric third-order aberrations, whereas quadrupole–octopole correctors also produce twofold and fourfold aberrations. Such correctors are feasible in practice only if these additional aberrations largely cancel out.

The correction of the third-order spherical aberration improves the resolution only if it is not limited by parasitic lower-order geometrical aberrations as well as mechanical and electromagnetic instabilities. The time-dependent perturbations determine the information limit, which sets a limit to the achievable resolution that cannot be surpassed by compensating the static defects of the lenses. Therefore, it is a condition *sine qua non* to push the information limit beneath the so-called Scherzer limit (9.2) of the non-corrected instrument. The correction of the aberrations in an electron microscope starts by eliminating the first-order axial astigmatism by means of a stigmator. In the next step, we must compensate for the second-order axial aberrations consisting of coma and threefold astigmatism. We achieve this correction by means of dipole and sextupole stigmators placed at appropriate positions within the corrector. After we have eliminated or sufficiently reduced the second-order axial aberrations, the third-order spherical aberration of the object lens becomes the dominant resolution-limiting aberration.

Two different approaches exist for correcting the unavoidable third-order spherical aberration of round lenses. We can nullify this aberration either by the sextupole corrector shown in Fig. 7.1 or by correctors consisting of quadrupoles and octopoles. The sextupole corrector has the advantage that the hexapole fields do not affect the rotationally symmetric paraxial path of rays, whereas the quadrupoles of QO correctors must produce a strongly astigmatic course of the paraxial rays in order that the fourfold symmetric field of the octopoles introduces a rotationally symmetric negative spherical aberration compensating for that of the round lenses. We must eliminate the axial astigmatism of the paraxial rays at the exit of the QO corrector to preserve stigmatic paraxial imaging. This compensation of the twofold first-order aberrations corresponds to that of the threefold second-order aberrations in the sextupole corrector, as illustrated in Fig. 7.2.

We have shown in Sect. 8.4.3 that the secondary third-order aberrations of sextupoles placed in a round-lens system are of the same nature as the primary geometrical aberrations of round lenses. By employing the box approximation (7.107) and inserting the resulting expressions (7.109) for the eikonal coefficients into (7.104), we obtain the secondary fundamental rays in analytical form. Substituting these analytical results for the secondary fundamental rays and expressions (7.106) for the fundamental paraxial rays in the representation (8.103) of the fourth-order eikonal polynomial produced by the hexapole fields, we can evaluate the integral analytically. By comparing the result,

$$\begin{aligned}
L_H^{(4)} &= \operatorname{Re} \sum_{\lambda=0}^2 \sum_{\nu=0}^{\lambda} \frac{2}{1 + \delta_{\lambda\nu}} L_{\nu\lambda}^{(4,0)} \bar{\omega}^{2-\nu} \omega^{2-\lambda} \bar{\rho}^{\nu} \rho^{\lambda} \\
&= \frac{3}{2} H_0^2 l^3 f_0^4 \operatorname{Re} \left(\bar{\omega}^2 \omega^2 - \frac{1}{5} \frac{l^2}{f_0^4} \bar{\omega} \omega \bar{\rho} \rho + \frac{1}{5} \frac{l^2}{f_0^4} \bar{\omega}^2 \rho^2 + \frac{1}{112} \frac{l^4}{f_0^8} \bar{\rho}^2 \rho^2 \right), \quad (9.43)
\end{aligned}$$

with the round-lens representation

$$\begin{aligned}
L_H^{(4)} &= -\operatorname{Re} \left(\frac{1}{4} C_{3H} \omega^2 \bar{\omega}^2 + K_{3H} \bar{\omega}^2 \omega \rho + \frac{1}{2} F_{3H} \bar{\omega} \omega \bar{\rho} \rho + \frac{1}{2} A_{3H} \bar{\omega}^2 \rho^2 + D_{3H} \bar{\omega} \bar{\rho} \rho^2 \right) \\
&\quad + L_{22}^{(4,0)} \bar{\rho}^2 \rho^2, \quad (9.44)
\end{aligned}$$

we obtain the aberration coefficients introduced by the hexapole fields of the sextupole corrector as

$$C_{3H} = -6H_0^2 l^3 f_0^4, \quad F_{3H} = -A_{3H} = \frac{3}{5} H_0^2 l^5, \quad (9.45)$$

$$K_{3H} = 0, \quad D_{3H} = 0, \quad L_{22}^{(4,0)} = \frac{3}{224} H_0^2 \frac{l^7}{f_0^4}. \quad (9.46)$$

According to their secondary nature, the coefficients (9.45) and (9.46) depend quadratic on the hexapole strength H_0 . Owing to the symmetry of the hexapole field, the fundamental rays and the secondary rays with respect to the midplane (Figs. 7.2 and 7.3), the corrector does not introduce off-axis coma and distortion. The coefficient F_{3H} of the field curvature has the same sign as that of the round lenses. Therefore, the sextupole corrector shown in Fig. 7.2 cannot compensate for this aberration. The sign of the other coefficients is opposite to that of the corresponding round-lens coefficients. Therefore, we can nullify the coefficient of spherical aberration $C_3 = C_{3R} + C_{3H}$ of the system consisting of a conventional objective lens and the sextupole corrector by adjusting the hexapole strength H_0 appropriately.

We survey most illustratively the action of the sextupole corrector by considering a bundle of incident rays located on the mantle of a cylinder centered about the optic axis. The corrector transforms these axis-parallel trajectories into a homocentric bundle of rays located on the mantle of a cone whose cone angle is proportional to the third power of the initial cylinder. Since this radius depends linearly on the aperture angle of the axial rays at the object plane, the corrector is able to compensate for the spherical aberration of the objective lens. Figure 8.2 shows that this lens deflects the outer rays more strongly toward the optic axis than the inner paraxial rays, whereas the sextupole corrector deflects the outer rays away from the axis without affecting the paraxial regime, as illustrated in Fig. 9.15. The figure shows that the second order deformation vanishes behind the corrector. Hence, up to the second order inclusively, the path of rays behind the corrector coincides with

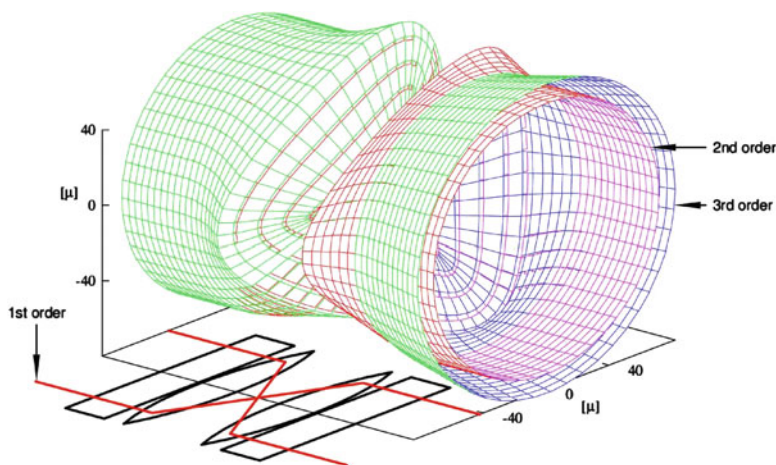


Fig. 9.15 Second- and third-order action of the sextupole corrector on an initially cylindrical bundle of non-paraxial rays, the second-order deformation (*red*) vanishes behind the corrector

that in front of the corrector. We achieve correction of the spherical aberration of the entire system by adjusting the third-order cone angle in such a way that it compensates for the opposite angle introduced by the round objective lens.

9.2.3 Aplanats

Transmission electron microscopes image the central region of the object plane into the image plane with a given magnification. In order that the microscope images all points of the field of view with the same resolution, we must eliminate the off-axial coma. According to the nomenclature of light optics, an aplanat denotes a lens system, which is free of spherical aberration and off-axial coma. The third-order coma of a standard magnetic objective lens has two components because its coefficient K_3 is complex. The real part describes the radial or isotropic coma, the imaginary part the azimuthal, or anisotropic coma. Contrary to the spherical aberration, we can nullify the off-axial coma of rotationally symmetric systems. Each lens has a so-called coma-free point located on the optic axis within the field of the lens. If we place the image of the crossover at the coma-free plane, the lens does not introduce a radial coma [22]. The Larmor rotation of the outer zones of the magnetic lens differs from that of the paraxial regime producing azimuthal coma. Therefore, we can only eliminate the azimuthal coma of a single magnetic round lens if the Larmor rotation and, hence, the axial magnetic field changes its sign. Fortunately, the coefficient of the azimuthal coma of a conventional magnetic objective lens is so small that it allows for a sufficiently large number of equally well-resolved object points at medium voltages as long as the resolution limit is larger than about $1 \text{ \AA}$.

Multipole systems introduce a radial round-lens coma and an additional threefold coma. Both components cancel out if the multipole fields and the course of the paraxial trajectories of the system satisfy specific symmetry properties. For example, this is the case if the multipole and round-lens fields and two of the four fundamental rays are symmetric with respect to the midplane of the system, whereas the two other fundamental rays are antisymmetric. The hexapole corrector shown in Fig. 7.2 is the simplest system, which satisfies these requirements. Because the hexapoles do not affect the paraxial regime, the axial rays and the field rays coincide. The axial ray u_α is antisymmetric and the field ray u_γ symmetric with respect to the midplane of the corrector.

Each symmetric coma-free multipole system has two coma-free points located symmetrically about its midplane. For the hexapole corrector, these points coincide with its nodal points N_1 and N_2 . Contrary to round lenses, the off-axial coma of a spherical-aberration-corrected system does not depend on the location of the crossover or the illumination, respectively. In order to eliminate the radial coma component of this system, we must match the coma-free point N_o of the objective lens with the object nodal point N_1 of the corrector. In this case, the system represents a so-called semi-aplanat [22].

Because the coma-free point of the objective lens and the corresponding point N_1 of the corrector are located within the field of the lens and the center of the first sextupole, respectively, we cannot match these points directly. However, we also achieve a coma-free system by imaging the coma-free plane of the objective lens into the front nodal plane N_1 of the corrector by means of a coma-free optical transfer system. We satisfy this condition by means of another telescopic round lens doublet, as shown in Fig. 9.16. The entire system represents a semi-aplanat because the matching of the coma-free points only eliminates the radial (isotropic) component of the coma. Because the simple two-element hexapole corrector cannot produce an azimuthal (anisotropic) coma, we can only partly compensate for the azimuthal coma of the strong magnetic objective lens by choosing the direction of the currents of the weak lenses of the transfer doublet opposite to that of the objective lens. The required number N_i of equally resolved image points along the diameter of the recorded image defines the tolerable value of the coefficient K_{3i} of the remaining coma for a given maximum aperture angle $|\omega_{\max}| = \theta_o$. To obtain this value conveniently, we refer our consideration back to the object plane and impose the condition that the maximum diameter $3r_o\theta_o^2 K_{3i}$ of the coma of a point located at the edge $r_o = N_i d/2$ of the imaged object area equals the diameter of the Airy disc:

$$2d \approx \frac{5}{4} \frac{\lambda}{\theta_o}. \quad (9.47)$$

According to this definition, we find the number of image points as

$$N_i = \frac{4}{3\theta_o^2 K_{3i}} \approx \frac{7}{2K_{3i}} \frac{d^2}{\lambda^2}. \quad (9.48)$$

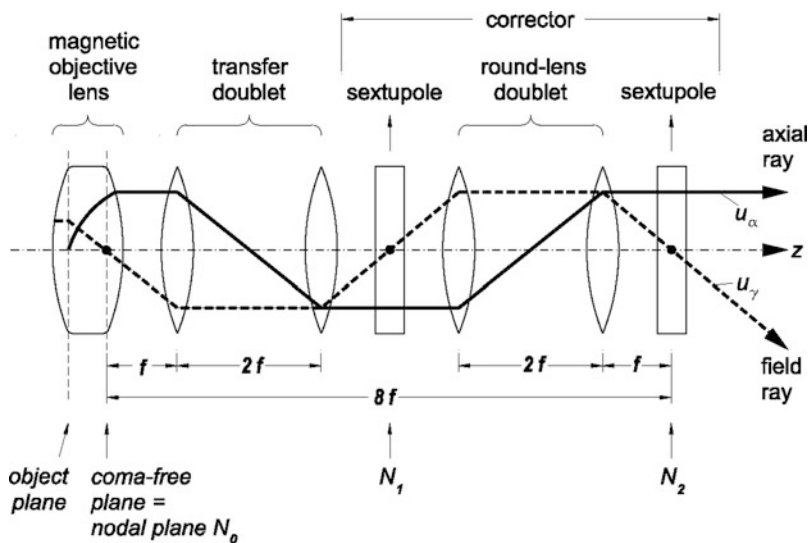


Fig. 9.16 Coma-free arrangement of the objective lens and the sextupole corrector by means of a telescopic transfer doublet

The radius of the Airy disc d defines the instrumental resolution limit of the electron microscope. For an acceleration voltage of 200 kV, the wavelength λ of the electron is 0.025 Å. Assuming an instrumental resolution limit $d = 1$ Å and 1,000 equally resolved image points per diameter, we find from (9.48) that the coefficient K_{3i} of the azimuthal coma of the objective lens must be equal or smaller than about 0.4. The coefficient of the azimuthal coma of high-performance magnetic objective lenses has roughly this value. Hence, if we decrease the resolution limit appreciably and/or want to resolve significantly more object elements per diameter, we need to eliminate the entire coma. This correction is also necessary if we fix the resolution and lower appreciably the accelerating voltage. To obtain a coma-free magnetic objective lens, we replace the conventional objective lens by a compound lens consisting of two spatially separated coils with opposite directions of their currents [102].

The sextupole corrector has the advantage that the hexapole fields do not affect the rotationally symmetric paraxial region. However, we obtain pure hexapole fields if we center the sextupoles very precisely about the optic axis such the transfer doublet exactly images the first sextupole onto the second sextupole with negative unit magnification ($M = -1$). Unfortunately, we cannot satisfy this condition with the necessary precision in practice. Therefore, we must introduce additional alignment dipoles, which compensate for a lateral shift and/or tilt of the sextupoles. After we have aligned these elements with the required accuracy, their currents need not to be stabilized as high as that of the round lenses, because then uncorrelated fluctuations of the sextupole currents produce only second-order aberrations, which

affect the resolution significantly less than the first-order aberrations resulting from fluctuations of the current of the objective lens. The successful incorporation of the hexapole corrector into a high-performance electron microscope has improved significantly resolution and contrast. Nowadays, aberration-corrected microscopes yield sub-Å resolution at voltages of 300 kV. Moreover, the correction of the spherical aberration largely reduces artifacts arising in phase contrast images of nonperiodic objects [155, 156].

9.2.4 Achromatic Aplanats

Apart from the sextupole corrector, we can compensate for the spherical aberration by means of a quadrupole–octopole corrector. Such correctors are mandatory if we want to correct spherical and chromatic aberrations, because the sextupole corrector does not affect the primary chromatic aberration in the case of a straight optic axis. In order to visualize the action of octopoles on the third-order aperture aberrations, we need only to consider the part of the fourth-order perturbation eikonal of the octopoles that depends on the complex slope parameters ω and $\bar{\omega}$ given by

$$L_{\text{Oa}}^{(4)} = \text{Re} \sum_{\mu=0}^2 L_{00}^{(4,2\mu)} \bar{\omega}^{2+\mu} \omega^{2-\mu} = \text{Re} \int_{z_0}^{z_i} \bar{O} (\omega w_{\omega} + \bar{\omega} w_{\bar{\omega}})^4 dz. \quad (9.49)$$

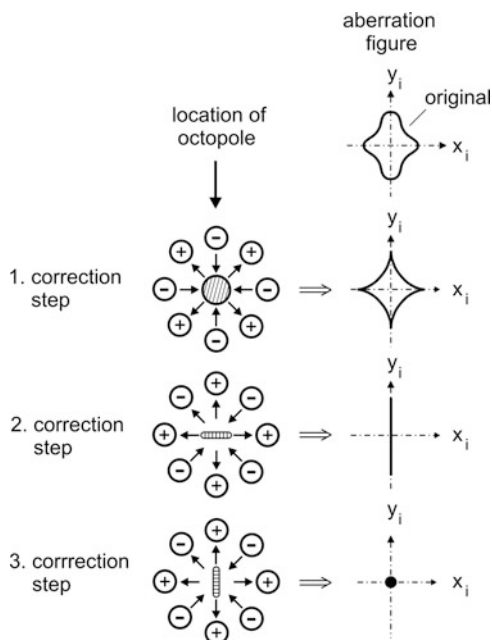
We assume that the correctors are symmetric with respect to two orthogonal plane sections, which we choose as the x – z and the y – z section, respectively. Accordingly, the total octopole strength (8.98) is real ($O = \bar{O} = O_r$). Therefore, the eikonal coefficients in (9.49) are real. We find their integral representations either from the second relation or directly from the general form (8.99) as

$$L_{00}^{(4,0)} = -4C_{30} = 6 \int_{z_0}^{z_i} O_r w_{\omega}^2 w_{\bar{\omega}}^2 dz, \quad (9.50)$$

$$L_{00}^{(4,2)} = 4 \int_{z_0}^{z_i} O_r (w_{\omega}^3 w_{\bar{\omega}} + w_{\bar{\omega}}^3 w_{\omega}) dz, \quad L_{00}^{(4,4)} = \int_{z_0}^{z_i} O_r (w_{\omega}^4 + w_{\bar{\omega}}^4) dz. \quad (9.51)$$

Contrary to the sextupole corrector, the QO corrector introduces aperture aberrations with twofold and fourfold symmetry in addition to the spherical aberration whose coefficient (9.50) is linearly related with the total octopole strength. This direct action of the octopoles minimizes the windings and currents and enables one to adjust the coefficient C_{30} arbitrarily. The disadvantage of the QO corrector is that it introduces additional aberrations, which we must eliminate together with those of the round lenses. We need at least three octopoles to compensate for the spherical

Fig. 9.17 Schematic illustration of the correction of the third-order axial aberration by three octopoles, with the second and third octopole placed at orthogonal astigmatic line images of the object plane; the *arrows* indicate the direction of the third-order force



aberration of the round lenses and for canceling the twofold axial star aberration and the fourfold axial astigmatism introduced by the quadrupole fields in order to improve the resolution of the electron microscope.

We can eliminate the components of the third-order axial aberration almost independently by placing two octopoles at orthogonal astigmatic images of the object plane and the third octopole at a position where the axial beam is rotationally symmetric, as shown in Fig. 9.17. The correction starts by exciting the fields of this octopole such that the aberration disc adopts the figure of a star. In the next step, we contract the star to a line by means of the second octopole centered at the horizontal astigmatic image. The third octopole located at the vertical astigmatic image compensates for the remaining aberration line without affecting the preceding corrections. The sequence of the second and third correction step is arbitrary.

To minimize the correction procedure, it is advantageous to find arrangements of the quadrupoles and octopoles such that the third-order aberrations with twofold symmetry cancel out. The most simple quadrupole system, which satisfies this condition, is the antisymmetric quadrupole quadruplet depicted in Fig. 4.33. Figure 9.4 shows that the course of the corresponding fundamental pseudo rays is either symmetric or antisymmetric with respect to the midplane of the quadruplet. We have used the electric and magnetic quadruplet for eliminating the chromatic aberration of the round lens of a scanning electron microscope (SEM), as demonstrated in Fig. 9.3. In order to utilize the corrector for compensating additionally the third-order aperture aberration, we need to superpose octopole and quadrupole

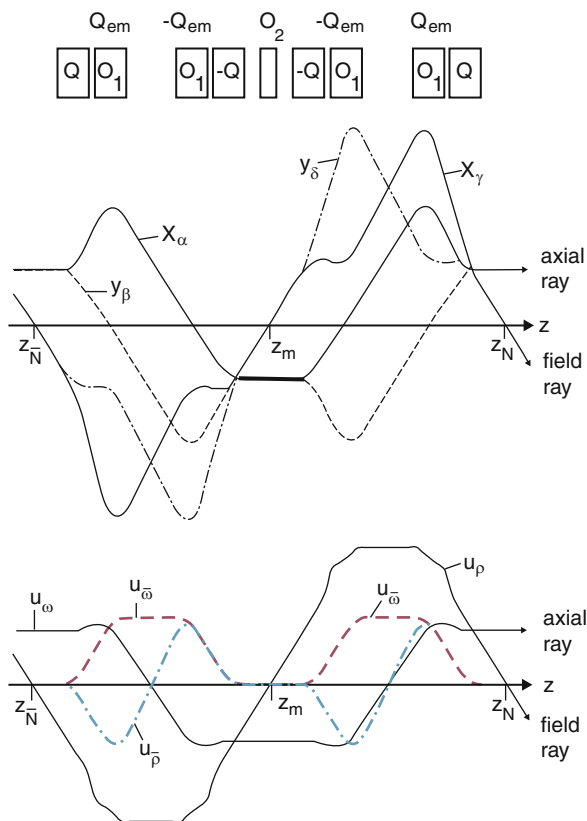
fields. This is possible by employing twelve-pole elements, which allow one to excite multipole components with twofold and fourfold symmetries.

To prevent that the octopole fields introduce a twofold axial aberration, we must excite them symmetrically with respect to the midplane. An octopole centered at this plane satisfies this condition automatically. Because the fundamental axial pseudo ray vanishes at this plane ($w_\omega = 0$), the central octopole only introduces a fourfold axial astigmatism, as follows directly from the representations of the corresponding eikonal coefficients (9.50) and (9.51). In order to compensate most efficiently for the spherical aberration, we must place the two other octopoles at astigmatic line images, where $w_\omega = \pm w_{\bar{\omega}}$. Since each of these images is located at the center plane of one of the two inner quadrupoles, we must replace these elements by dodecapoles. To guarantee that each correction step does not affect the preceding corrections, we must perform the correction procedure in a distinct sequence. Accordingly, we first eliminate the spherical aberrations by the two octopoles placed at the astigmatic images. Because their strengths coincide, we excite them jointly. This correction adds a term to the fourfold axial astigmatism but not to the star aberration, because the integrand of its eikonal coefficient $L_{00}^{(4,2)}$ induced by the correction is antisymmetric (9.51). Subsequently, we eliminate the remaining fourfold axial astigmatism by the central octopole without affecting the correction of the two other aberration components because the integrands of their eikonal coefficients vanish for $w_\omega = 0$.

We cannot use the SEM corrector depicted schematically in Fig. 9.3 for a fixed-beam transmission electron microscope because this corrector has a large off-axial coma. In order to obtain a sufficiently large number of image points, we must employ a corrector, which does not introduce aberrations that are linear in the lateral position coordinates. In addition, the third-order image curvature and field astigmatism of the corrector must not appreciably exceed those of the objective lens to avoid a large fifth-order coma. This aberration results predominantly from the combination of image curvature and field astigmatism of the objective lens with the third-order spherical aberration of the corrector and vice versa. Moreover, we aim for quadrupole and octopole arrangements such that many of their non-rotationally symmetric aberrations cancel out. We have seen in the preceding chapters that doubly symmetric systems possess these properties. Two different types of doubly symmetric systems exist, a symmetric type and an antisymmetric type depending on the symmetry of the quadrupole fields with respect to the midplane of the system. Each system consists of two identical subunits. The quadrupole elements of each subunit are symmetric with respect to its center plane. The quadrupole elements of the symmetric type are symmetrically excited with respect to the midplane of the entire system and antisymmetric with respect to the center plane of each subunit. The excitation is vice versa in the case of the antisymmetric type.

We obtain the simplest doubly symmetric telescopic multipole corrector by combining two antisymmetric quadrupole quadruplets shown in Fig. 9.4 such that the image nodal point of the first quadruplet coincides with the object nodal point of the second quadruplet. The resulting system depicted in Fig. 9.18 consists of

Fig. 9.18 Achroplanator consisting of two symmetrically excited multipole quadruplets; the course of the fundamental pseudo rays $u_\omega, u_{\bar{\omega}}; u_\rho, u_{\bar{\rho}}$ demonstrates the double symmetry. The octopole fields O have the same strength and are excited together with the electric and magnetic quadrupole fields within the inner dodecapole elements of each quadruplet



eight identical quadrupoles, which we arrange and excite symmetrically about the midplane of the system. The intersections of the symmetric pseudo ray $u_\rho(z)$ with the optic axis define the nodal points z_N and z_N of the telescopic octuplet. It is advantageous to combine the corrector with the objective lens in such a way that an image of the diffraction plane is located at the midplane z_m . In this case, the rays u_ρ and $u_{\bar{\rho}}$ represent the pseudo field rays. It follows from Fig. 9.18 that the rays u_ρ and $u_{\bar{\omega}}$ are antisymmetric with respect to the midplane of the entire system and symmetric with respect to the central plane of each constituent antisymmetric quadruplet. The rays $u_{\bar{\rho}}$ and u_ω have the opposite behavior, thus demonstrating the double symmetry of the pseudo fundamental rays within the multipole octuplet.

Owing to the double symmetry, the system does not introduce third-order coma, twofold aberrations, and distortions. Accordingly, its axial chromatic aberration is rotationally symmetric. To compensate for this aberration and that of the round lenses, we substitute mixed electric and magnetic quadrupoles for the second and third quadrupole of each quadruplet. We excite the electric and the magnetic fields of each mixed quadrupoles in such a way that the total quadrupole strength does not change. Therefore, each mixed quadrupole acts as a focusing element

and as a first-order Wien filter. By varying the electric and magnetic quadrupole strengths, we adjust the strength of the first-order Wien filter. Because we center each element about an astigmatic image of the object plane, the quadrupole Wien filter compensates for the chromatic aberration in one principal section without affecting the component in the other section. We eliminate the axial third-order aberrations by means of five octopole fields, four of which are superposed onto the fields of the mixed quadrupoles. It is advantageous to excite the individual multipole fields within twelve-pole elements. The octopole fields O compensate for the spherical aberration of the objective lens and of the corrector without introducing off-axial coma. The fifth octopole O_1 located at the midplane z_m of the achroplanator compensates for the fourfold axial astigmatism. Because the field rays intersect the optic axis at this plane, the octopole does not introduce any field aberrations. As a result, the correction of the third-order axial aberration does not induce appreciable fifth-order combination aberrations.

Instead of combining two antisymmetric quadrupole units in a symmetric way, we can also combine antisymmetric two symmetric quadrupole systems. In this case, the quadrupoles of the subunits are excited symmetrically with respect to their central planes and antisymmetric with respect to the midplane of the aplanator. In order that the octopoles correcting for the axial aberrations do not introduce appreciable off-axial aberrations, it is advantageous to place two octopoles at strongly first-order distorted images of the diffraction plane and one at an undistorted image of this plane. To preserve the symmetry conditions, we must form the anamorphic images at the central planes of the subunits.

The corrector depicted in Fig. 9.19 meets these requirements. It is composed of two identical telescopic multipole subunits. Each subunit consists of a symmetric telescopic quadrupole sextuplet formed by two identical symmetric quadrupole triplets, which are arranged symmetrically about the central plane midway between the triplets. We place an octopole O_1 at this plane within each sextuplet. Owing to the double symmetry of the sextuplet, we need only two different power supplies for exciting the eight quadrupoles Q_1 and the four quadrupoles Q_2 . The sextuplets are separated by a distance such that the back principal plane of the first telescopic sextuplet matches the front principal plane of the second unit. The front principal plane N_1 of the first subunit and the back nodal plane N_2 of the second unit form the nodal planes of the corrector. We excite the quadrupoles of the second sextuplet with opposite polarity with respect to those of the first sextuplet so that the quadrupole field is antisymmetric with respect to the midplane z_m of the entire corrector. In this case, the course of the fundamental rays x_α, x_γ in the x - z section of the second sextuplet coincides with that of the corresponding rays y_β, y_δ in the y - z section of the first sextuplet and vice versa. We characterize this behavior as *exchange symmetry*. Although the corresponding fundamental rays are neither symmetric nor antisymmetric with respect to the midplane z_m of the entire system, the fundamental pseudo rays $w_\omega, w_{\bar{\omega}}, w_\rho$, and $w_{\bar{\rho}}$ shown in Fig. 9.20 exhibit such symmetries. As a result, the corrector does not introduce twofold fourth-order eikonal monomials because the integrands of their coefficients $L_{\nu\lambda}^{(4,2)}$ are antisymmetric either with respect to the midplane

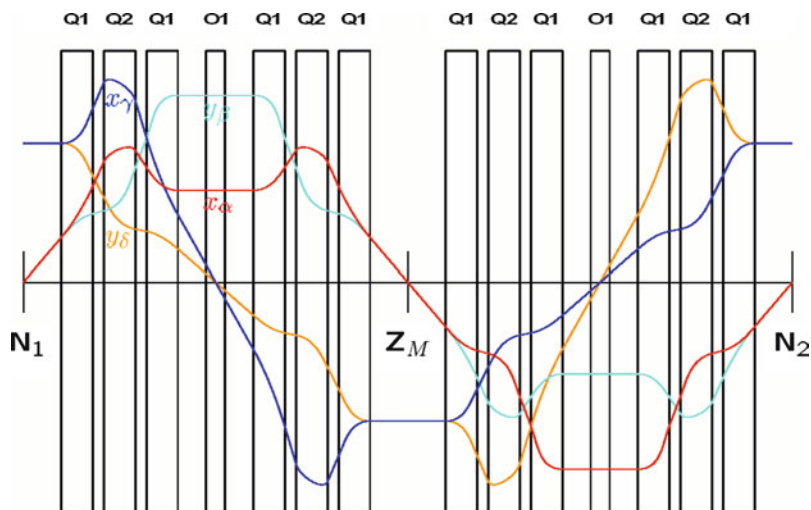


Fig. 9.19 Path of the fundamental rays within the telescopic doubly symmetric quadrupole-octopole corrector composed of four symmetric quadrupole triplets $Q_1 - Q_2 - Q_3$ and two identical octopoles O_1

z_M or with respect to the central plane of each constituent sextuplet. In order that this is also the case for the octopoles, their field must be symmetric with respect to these planes. The doubly symmetric excitation of the octopoles and the symmetric/antisymmetric excitation of the quadrupoles do neither introduce eikonal coefficients with twofold symmetry ($L_{\nu\lambda}^{(4,2)} = 0$) nor coefficients $L_{\nu\lambda}^{(4,0)}$ and $L_{\nu\lambda}^{(4,4)}$ with $\nu + \lambda$ odd. Hence, quadrupole-octopole correctors with exchange symmetry introduce neither third-order comas ($L_{10}^{(4,0)} = L_{10}^{(4,2)} = L_{10}^{(4,4)} = 0$) nor distortions ($L_{12}^{(4,0)} = L_{30}^{(4,2)} = L_{21}^{(4,2)} = L_{30}^{(4,4)} = 0$) if the octopole fields exhibit double symmetry. In order to introduce such aberrations, we must abandon the double symmetry of the octopole fields.

We must combine the corrector with the objective lens in such a way that the isotropic coma of the entire system vanishes in order to obtain a semi-aplanat. Two possibilities exist for obtaining such a system. Either we image the coma-free plane N_0 of the objective lens by means of a telescopic transfer system into the front nodal plane N_1 of the corrector and the object plane at infinity or vice versa. This change does not introduce coma because it only exchanges axial rays and field rays within the corrector without affecting the symmetry properties. We utilized the first method for matching the coma-free points of the objective lens and the sextupole corrector by means of a telescopic transfer doublet. If we employ the QO corrector shown in Fig. 9.19, we need to use the second possibility in order to place images of the diffraction plane at the symmetry planes of the corrector. In this case, we must split up the coma-free transfer doublet and place the lenses symmetrically about the midplane z_M of the corrector such that the back focal plane of the first lens coincides with the front nodal plane N_1 and the front focal plane of the second lens

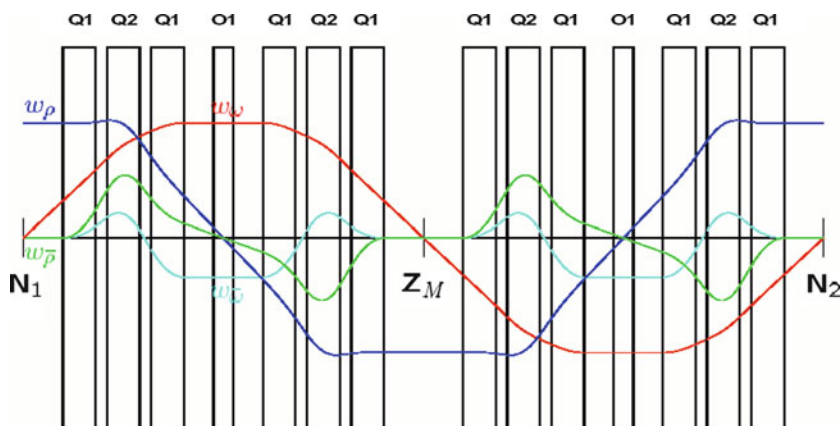


Fig. 9.20 Course of the fundamental pseudo rays within the QO corrector revealing the double symmetry of these rays

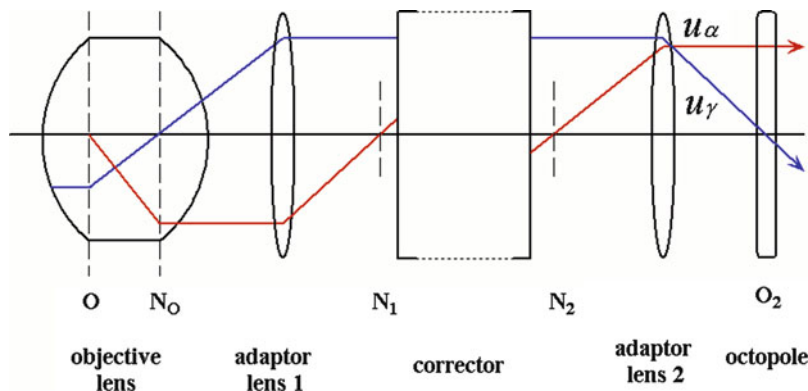


Fig. 9.21 Arrangement of the elements and course of the fundamental rays for a semi-aplanat employing the QO corrector shown in Figs. 9.18 and 9.19, the octopole O_2 compensates for the fourfold axial stigmatism introduced by the corrector

with the back nodal plane N_2 of the corrector. Each lens acts then as an adaptor lens, as illustrated in Fig. 9.21. Because the corrector images the front nodal plane N_1 with unit magnification into the back nodal plane N_2 , the imaging properties of the telescopic doublet are preserved. This doublet images its front nodal plane N_0 with unit magnification into the back focal plane of its second lens where we have placed the octopole. Hence, the outlined insertion of the QO corrector midway between the adaptor lenses does not affect the aberrations of the transfer doublet.

The two octopoles O_1 within the corrector eliminate the spherical aberration of the entire system without introducing appreciable field aberrations because the field rays are zero at the center of the octopoles, as depicted in Figs. 9.19 and 9.20. The same holds true for the octopole O_2 centered at the back focal plane of the adaptor

lens 2. This octopole compensates the fourfold axial astigmatism without affecting the preceding correction of the spherical aberration. We must rotate the principal sections of the octopole O_2 with respect to those of the corrector by the angle of Larmor rotation of the adaptor lens 2 in order to match the azimuthal angle of the fourfold axial astigmatism of the octopole O_2 with that of the corrector. Owing to the symmetric arrangement of the adaptor lenses and the double symmetry of the corrector, the semi-aplanat is free of primary chromatic and geometrical distortions, spherical aberration, isotropic coma, and aberrations with twofold symmetry.

Quadrupole–octopole correctors offer the possibility to correct both chromatic and spherical aberrations of round lenses. In order that the QO corrector compensates for both aberrations, we must substitute crossed electric and magnetic quadrupoles for the inner quadrupoles adjacent to the octopoles. Hence, four of the twelve quadrupoles act as quadrupoles Q_1 and as first-order Wien filters compensating for the axial chromatic aberration in the same way as employed in the SEM corrector shown in Fig. 9.3. Due to the double symmetry, the aplanatic corrector eliminates the axial chromatic aberration of the semi-aplanat without introducing chromatic distortions. Moreover, the number of independent power supplies reduces significantly. We need only six major power supplies, three for the eight magnetic quadrupoles, one for the four electrostatic quadrupoles, and two for the octopoles Q_1 and Q_2 . The power supplies for the quadrupoles must be extremely stable because these elements affect the paraxial path of rays. This stability is not necessary for the octopoles, since they only affect the third-order path deviations. The stability requirements for the stigmators fields are also less stringent provided the misalignments are sufficiently small. Accordingly, it is advantageous to construct the system with the highest achievable mechanical accuracy and to keep the magnetic inhomogeneities as small as possible.

If we allow superposition of the octopole fields with the quadrupole fields, we can combine the two inner quadrupoles of each sextuplet to a single element resulting in the symmetric quadrupole quintuplet discussed in detail in Sect. 4.8. The two quintuplets operate in the telescopic mode. In this case, the nodal points N_1 and N_2 of the quintuplet are located within the first and the fifth quadrupole, respectively. Owing to the imposed symmetry, the central quadrupole of each quintuplet is twice as thick as the outer quadrupoles. In order to correct for the chromatic aberration, the central element of each quintuplet must be a crossed electric and magnetic quadrupole. Since this element consists of eight poles, we can readily superpose an octopole field onto the quadrupole fields. Detailed calculations show that we obtain a feasible corrector compensating for chromatic aberration, spherical aberration, and coma by substituting the telescopic quadrupole quintuplet for each of the sextupoles of the hexapole corrector shown in Fig. 7.1. We eliminate the anisotropic coma of the objective lens by means of four skew octopoles placed symmetrically about the midplane of the corrector depicted schematically in Fig. 9.22. This design serves as the basic concept of the TEAM corrector developed by the company CEOS within the frame of the US “Transmission Electron Aberration-corrected Microscope” project. This project was concerned with the development of a 300 kV

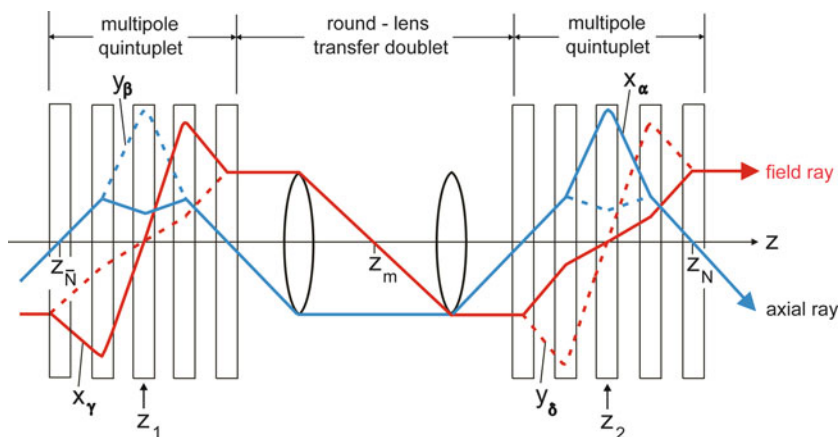


Fig. 9.22 Schematic arrangement of the TEAM corrector and course of the fundamental rays; a strongly anamorphic image of the diffraction plane is located at the center planes z_1 and z_2 of the multipole quintuplets

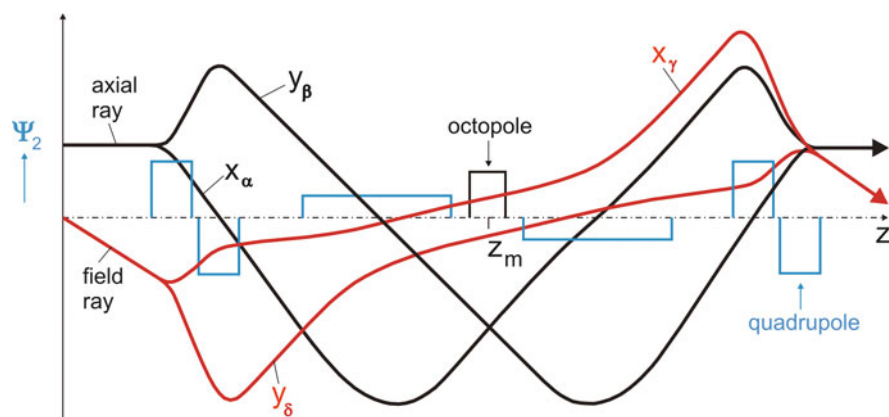


Fig. 9.23 Arrangement of the SALVE corrector and course of the fundamental rays illustrating their exchange symmetry with respect to the midplane z_m , the blue boxes represent the approximated strength $\Psi_2(z)$ of the quadrupoles

atomic-resolution transmission electron microscope, which enables a resolution limit $d = 0.5 \text{ \AA}$.

A promising alternative to the TEAM corrector exists in the case of acceleration voltages below about 100 kV, as employed in the German Sub-Angstrom Low-Voltage Electron microscope (SALVE) project. The SALVE corrector shown in Fig. 9.23 consists of two magnetic quadrupole doublets, two electric magnetic quadrupoles, and at least three magnetic octopoles. One of the octopoles is centered at the midplane z_m of the corrector the others are excited within the inner elements of the quadrupole doublets where astigmatic images of the object plane are formed.

The multipoles, placed symmetrically about the midplane, are excited in such a way that the fundamental paraxial rays exhibit exchange symmetry with respect to this plane. The electric magnetic quadrupoles are located close to the midplane and serve as focusing elements and as first-order Wien filters compensating for the axial chromatic aberration of the round lenses. Owing to the symmetry of the fields and the fundamental rays, the corrector does not introduce any primary field aberrations, which depend linearly on the off-axial distance of the object point. Moreover, the higher-order axial aberrations are reduced to such an extent that aperture angles of up to 50mrad can be used enabling a resolution limit of about 1 Å at 60 kV.

9.2.5 Correction of Third-Order Field Curvature and Field Astigmatism

Image curvature and field astigmatism are the most deleterious aberrations in projection electron lithography because they decisively limit the usable area of the mask. One uses magnetic round lenses to image the mask onto the wafer. We have shown in Sect. 8.3 that the image curvature of round lenses is unavoidable if the axial electric field vanishes at the object and image planes. Hence, rotationally symmetric planar magnetic systems do not exist. A *planar system* is free of off-axis coma, field curvature, and field astigmatism. Hence, we need a corrector to compensate for these aberrations. A corrector composed of two sextupoles cannot compensate for the field curvature because it has the same as that of round lenses, as demonstrated in the preceding chapter. Therefore, the question remains if we can change the sign by increasing the number of sextupoles with the constraint that the second-order path deviations cancel out behind the last sextupole. However, in order to obtain a planar image field, we must also compensate for the field astigmatism. Because its coefficient is complex for magnetic round lenses, we need three free parameters for eliminating simultaneously both aberrations.

The sextupole corrector shown in Fig. 9.24 satisfies this condition. The arrangement consists of four identical round lenses with focal length f forming a telescopic $8f$ -system and five sextupoles placed symmetrically about the midplane z_m . The two outer sextupoles S_1 and $S_5 = S_1$ have the same complex strength (7.96) $H_1 = H_5 = H_3$ as the central sextupole S_3 , which is twice as thick ($l_3 = 2l_1$). Each half of this sextupole is conjugate to one of the outer sextupoles because the first transfer doublet images S_1 onto the first half of S_3 , while the second transfer doublet images the other half of S_3 onto the last sextupole with magnification $M = -1$. The two inner round lenses form another transfer doublet, which images the sextupole S_2 with $M = -1$ onto the fourth sextupole $S_4 = S_2$. Its strength $H_4 = H_2$ and thickness $l_4 = l_2$ coincide with those of the second sextupole. Accordingly, the total second-order path deviation vanishes behind the last sextupole. Since we can choose arbitrarily the phases of the complex hexapole strengths H_1 and H_2 , we have three free parameters $|H_1|$, $|H_2|$ and $\text{Re}(H_1 \bar{H}_2) \leq |H_1||H_2|$ to compensate

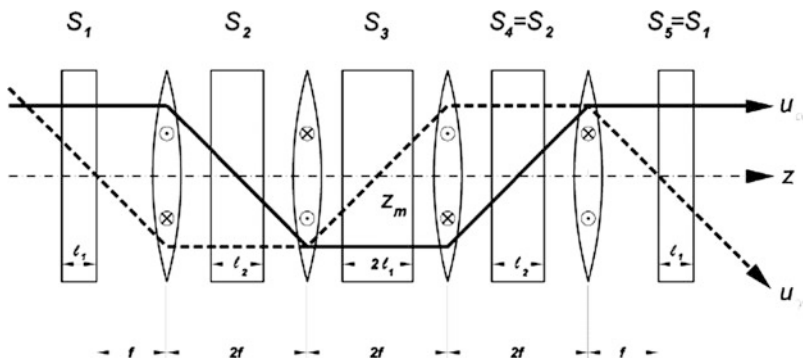


Fig. 9.24 Arrangement of the elements within the sextupole planator compensating for third-order image curvature and field astigmatism

for three coefficients. The phase difference determines the azimuthal orientation of the hexapole fields with respect to each other. Since we cannot choose the parameters arbitrarily, we cannot conclude that the corrector can always provide a planar system.

Image curvature and field astigmatism introduced by the corrector do not depend on its location within the system, whereas the other aberrations depend on the telescopic magnification

$$M_c = u_\alpha(z_c)/f_0 \quad (9.52)$$

in front of the corrector. The telescopic magnification implies that the axial ray u_α runs parallel to the optic axis. Employing the box-shape approximation for the hexapole strengths, we eventually obtain after a lengthy analytical evaluation of the aberration integrals the coefficients of the third-order aberrations generated by the sextupole planator as

$$C_{3P}M_c^{-4} = -12f_0^4l_1^3|H_1|^2 - \frac{3f_0^4}{56f^4}l_2^7|H_2|^2 - 3\frac{f_0^4}{f}l_1l_2^3\text{Re}(H_1\overline{H}_2), \quad (9.53)$$

$$F_{3P} = -\frac{24}{5}l_1^5|H_1|^2 + \frac{3}{5}l_2^5|H_2|^2 + 4\frac{l_1^3l_2^3}{f}\text{Re}(H_1\overline{H}_2), \quad (9.54)$$

$$A_{3P} = -\frac{21}{5}l_1^5|H_1|^2 - \frac{3}{5}l_2^5|H_2|^2 - \frac{l_1^3l_2^3}{f}H_1\overline{H}_2 - 36l_1l_2f^3\overline{H}_1H_2, \quad (9.55)$$

$$K_{3P} = D_{3P} = 0. \quad (9.56)$$

The coefficients of field curvature (9.54) and field astigmatism (9.55) do not depend on the magnification (9.52). Therefore, in order to compensate for these aberrations, we can place the planator at any position within the column. However, the spherical aberration introduced by the planator depends strongly on its location, as it is the case for the round lenses in an electron microscope. The coefficient

$A_{3P} = A_{3Pr} + iA_{3Pi}$ of the field astigmatism is complex if the azimuthal orientation of the sextupoles $S_1 = S_5$ and S_5 differs from that of the two other sextupoles $S_2 = S_4$. The coefficients of field curvature and spherical aberration are always real, whereas the coefficients of coma and distortion (9.56) are zero. The expression for the Petzval curvature of the planator

$$\frac{1}{\rho_P} = F_{3P} - 2\text{Re}(A_{3P}) = \frac{18}{5}l_1^5 |H_1|^2 + \frac{9}{5}l_2^5 |H_2|^2 + 6l_1l_2 \left(\frac{l_1^2 l_2^2}{f} + 12f^3 \right) \text{Re}(H_1 \overline{H_2}) \quad (9.57)$$

reveals that this curvature can only become negative if $\text{Re}(H_1 \overline{H_2})$ is negative. Accordingly, the polarity of the sextupoles S_2 and S_4 must be opposite to that of the three other sextupoles $S_1 = S_5$ and S_3 . The first and second term on the right hand side of (9.57) are positive definite. Since they do not depend on the focal length f of the transfer lenses, we can produce a negative Petzval curvature and a negative coefficient (9.53) of the spherical aberration. By choosing the parameters l_1, l_2, f , and M_c appropriately and by adjusting the hexapole strengths H_1, H_2 and the focal length f_o of the objective lens by electrical means, it should be possible to compensate simultaneously for field astigmatism, image curvature, and spherical aberration.

Finally, we aim to find a system with straight optic axis, which is free of all primary chromatic and geometrical aberrations [157]. Moreover, the corrector must enable a feasible correction of the numerous aberrations such that each subsequent correction step does not affect aberrations eliminated in the preceding correction steps. To meet this requirement, the pseudo fundamental rays must have a distinct course within the corrector. Doubly symmetric systems consisting of 16 quadrupoles are able to satisfy this condition. Figure 9.25 shows the arrangement of the multipoles and the course of the fundamental rays $x_\alpha, y_\beta; x_\gamma, y_\delta$ in for such a system. Owing the double symmetry, the corrector only introduces fourfold and round lens third-order aberrations and rotationally symmetric primary chromatic aberrations. We demonstrate the decoupled correction procedure by means of the fundamental pseudo rays depicted in Fig. 9.26. They degenerate to the fundamental rays outside of the corrector, as shown in Fig. 9.27. In the first step, we compensate for the chromatic aberration by means of the crossed electric and magnetic quadrupoles adjacent to the central octopoles O_4 of each half of the ultracorrector [157].

In order that each correction step does not reintroduce aberrations eliminated in the preceding steps, we must start with the octopoles O_1 compensating for the field curvature. Since none of the pseudo rays vanishes at the locations of the octopoles O_1 , these octopoles contribute to all eikonal coefficients apart from those, which cancel out by symmetry. In the next step, we eliminate the round-lens field astigmatism by means of the octopoles O_2 . Because we have centered these octopoles at planes where the pseudo ray $w_{\tilde{\omega}}$ intersects the optic axis, they do not affect the field curvature eliminated by the octopoles O_1 . Since the system does not introduce aberrations with twofold symmetry, the fourfold field astigmatism with eikonal coefficient $L_{20}^{(4,4)}$ is the only remaining field aberration. We eliminate this

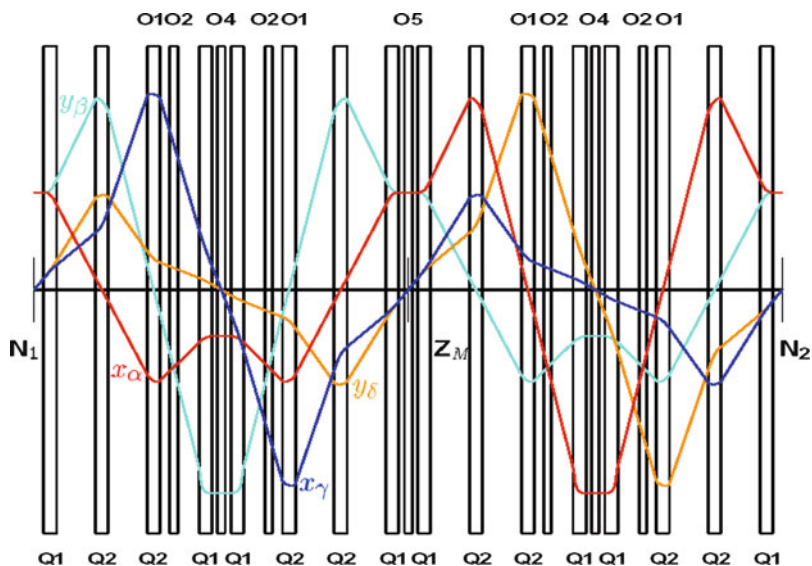


Fig. 9.25 Arrangement of the quadrupoles Q_1 , Q_2 ; the octopoles O_1 , O_2 , O_4 , O_5 ; and course of the fundamental rays within the ultracorrector

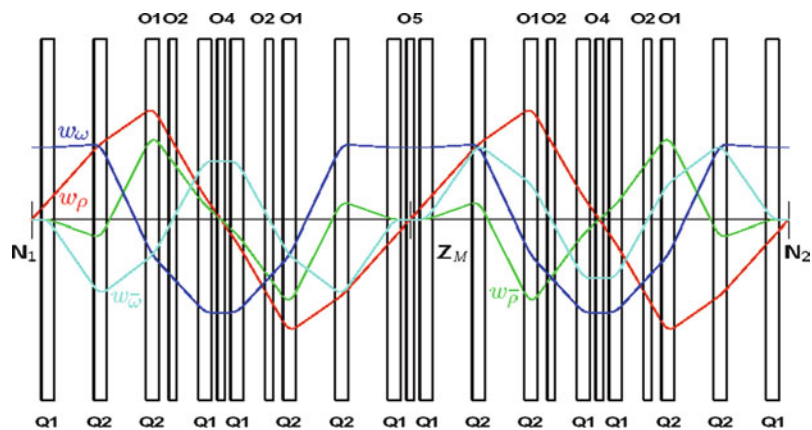


Fig. 9.26 Course of the fundamental pseudo rays w_ω , $w_{\bar{\omega}}$, w_ρ , $w_{\bar{\rho}}$ within the ultracorrector

aberration by means of the octopoles O_3 located outside the corrector. Since we have placed these elements in the stigmatic paraxial domain symmetrically about the midplane of the corrector, they only affect the fourfold field astigmatism and the fourfold axial astigmatism. Hence, we compensate for the fourfold field astigmatism with the octopoles O_3 without affecting the preceding corrections of the round-lens field aberrations. We eliminate the third-order spherical aberration by means of the two octopoles O_4 placed at first-order distorted stigmatic images of the object plane. Each image is located at the center plane of a subunit, as illustrated by

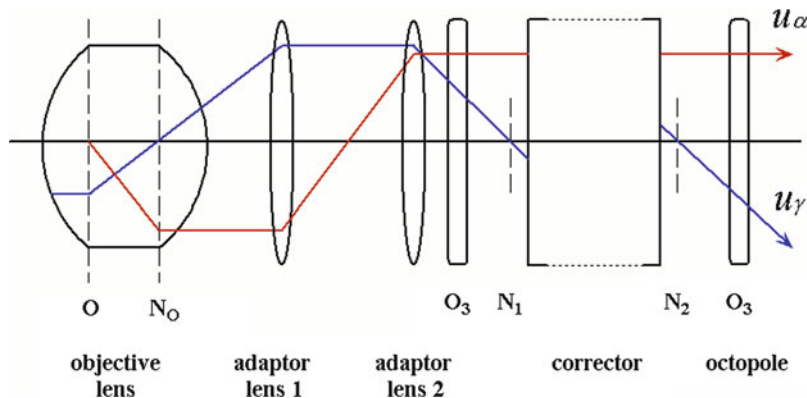


Fig. 9.27 Coma-free matching of the ultracorrector with the objective lens by means of a telescopic round lens doublet, the octopoles O_3 compensate for the fourfold field astigmatism introduced by the corrector multipoles

the fundamental rays depicted in Figs. 9.25 and 9.26, respectively. The correction of the spherical aberration only affects the fourfold axial astigmatism but not the field aberrations. In the last step of the correction procedure, we compensate for the remaining fourfold axial astigmatism by the octopole O_5 placed at the midplane z_M of the ultracorrector. Because an undistorted Gaussian image of the object plane is located at this plane, the octopole O_5 compensates for the fourfold axial astigmatism without affecting the preceding correction.

9.2.6 Correction of Coma

The number of image points (9.48) is inversely proportional to the coma coefficient and proportional to the square of the resolution limit d . Therefore, the diameter of the image field shrinks quadratic with increasing resolution $1/d$. In order to avoid such a drastic reduction of the field of view, we must eliminate the third-order (off-axial) coma. We can eliminate the radial component of the coma by imposing symmetry conditions and by matching the coma-free planes of the objective lens and the corrector. Unfortunately, we cannot nullify the anisotropic coma by such means because it results from the Larmor rotation within the magnetic field of the objective lens. In order that this lens does not introduce an anisotropic coma, it must be a compound lens consisting of two spatially separated coils with opposite direction of their currents. Since the anisotropic coma coefficient of standard magnetic lenses cannot become smaller than about 0.6, it is necessary to eliminate the total coma in order to achieve at least 2,000 image points per diameter at sub-Å resolution and voltages below 300 kV. Sub-Å resolution at such voltages is only possible if we also correct for spherical aberration and sufficiently reduce the chromatic aberration. We

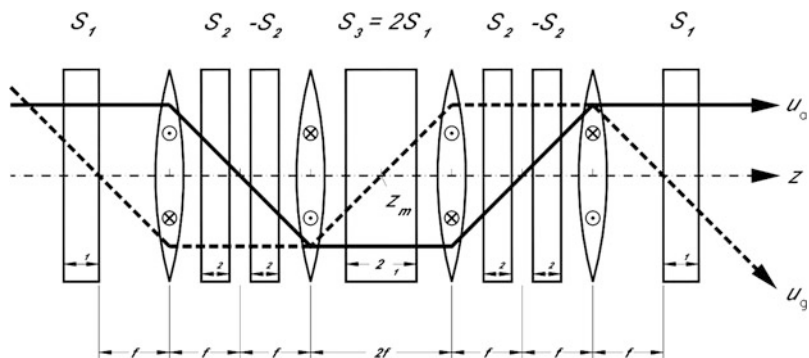


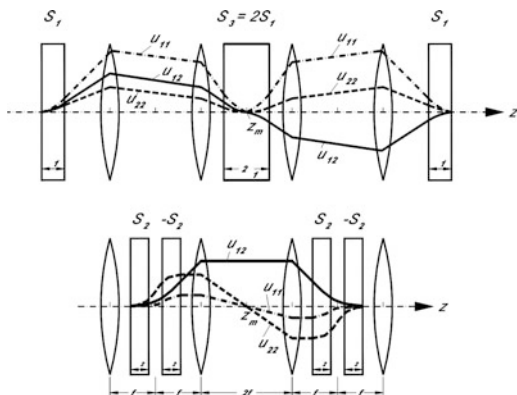
Fig. 9.28 Scheme of the hexapole applanator compensating for third-order spherical aberration, isotropic, and anisotropic coma

can realize this reduction most conveniently by employing a field emission gun and a monochromator. However, we cannot employ a coma-free compound objective lens in this case because its chromatic aberration is almost twice as large as that of a standard objective lens. As a result, the increase of the chromatic aberration largely negates the effect of the monochromator. Therefore, we must either compensate for the chromatic aberration or eliminate the anisotropic coma by means of the corrector.

Unfortunately, it is not possible to produce an anisotropic coma by means of the symmetric hexapole correctors shown in Figs. 9.16 and 9.24, respectively, because symmetric arrangements of the hexapoles produce neither third-order coma nor distortion. Hence, to compensate for spherical aberration and coma, it is necessary to lift the symmetric symmetry requirement. We can eliminate both aberrations by introducing two hexapole subsystems, as shown in Fig. 9.28. One system is symmetric, the other antisymmetric with respect to the midplane of the corrector. In order to eliminate all second-order aberrations and to avoid fourth-order aperture aberration, the hexapole fields must also satisfy the same symmetry relations with respect to the central plane of each half of the corrector. The corrector shown in Fig. 9.28 meets this requirement if we attribute the second half of the central hexapole to the front subsystem and the first half of this hexapole to the back subsystem. Hence, the symmetric and the antisymmetric components of the hexapole field exhibit a double symmetry. This behavior differs from that of the fundamental rays. The axial fundamental ray u_α is symmetric to the midplane of the corrector and antisymmetric with respect to the central plane of each subsystem, whereas the field ray u_γ has the opposite symmetry.

We can calculate the third-order aberrations of the hexapoles with a sufficient degree of accuracy by employing the box-shape approximation because the integrands of the integral representations of the corresponding eikonal coefficients (8.105), (8.106), and (8.107) do not contain derivatives of the hexapole strength H (7.96). This approximation presupposes that the hexapole strength is constant within the hexapole element and zero outside. The approximated field of each

Fig. 9.29 Path of the secondary fundamental rays of the hexapole aplanator introduced by the symmetric hexapole subset S_1 and the antisymmetric subset S_2



hexapole of the symmetric arrangement has strength H_1 and axial extension l_1 . We define the field of the corresponding hexapoles of the antisymmetric arrangement by the quantities H_2 and l_2 . Using this approximation, we can evaluate analytically the second-order path deviations and the integrals of the eikonal coefficients. To illustrate the different behavior of the symmetric and the antisymmetric hexapole subunits, we have plotted in Fig. 9.29 the components of the secondary fundamental rays u_{11} , u_{12} , and u_{22} for each hexapole subunit separately. The total fundamental ray is the sum of these components. The figure demonstrates that the components are either symmetric or antisymmetric with respect to the midplane z_m of the aplanator. The symmetric rays of the first subunit are antisymmetric for the second subunit and vice versa. This opposite behavior results from the opposite symmetries of the hexapole fields with respect to the midplane.

Owing to these different symmetries, the coefficients of third-order coma and distortion coma do not vanish, as it is the case for the coefficients (9.46) of the standard hexapole corrector.

By employing the box approximation for the hexapole strengths and utilizing the relations (8.49) existing between the aberration coefficients and the coefficients of the perturbation eikonal, we obtain the analytical expressions for the third-order aberration coefficients. Since the calculations are rather lengthy yet straightforward, we only state the result for produced by the hexapoles of the aplanator. Neglecting the contributions of the round transfer lenses, we find the aberration coefficients of the hexapole aplanator as

$$\begin{aligned}
 C_{3H} &= -12 |H_1|^2 f_0^4 l_1^3 - \frac{36 f_0^4}{7 f^4} |H_2|^2 l_2^7, \\
 K_{3H} &= 3 l_1 l_2^2 f_0^2 f \left[12 H_1 \overline{H}_2 - \overline{H}_1 H_2 \frac{l_1^2 l_2^2}{f^4} \right], \\
 A_{3H} &= -\frac{21}{5} |H_1|^2 l_1^5 + \frac{54}{5} |H_2|^2 l_2^5,
 \end{aligned} \tag{9.58}$$

$$F_{3H} = -\frac{24}{5} |H_1|^2 l_1^5 - \frac{144}{5} |H_2|^2 l_2^5,$$

$$D_{3H} = 12i \text{Im}(H_2 \overline{H}_1) \frac{l_1^3 l_2^2 f}{f_o^2}.$$

Expressions (9.58) demonstrate that the real coefficients of the spherical aberration and the field curvature are always opposite in sign to those of the round lenses. The coma coefficient K_3 is complex if the azimuthal orientations of the symmetric and antisymmetric hexapole fields differ from each other. We can neglect the second term in the bracket because the focal length f of the transfer lenses is generally more than twice as large as the lengths of the sextupole elements. The same holds for the second term in the expression for C_{3H} . Moreover, we can postulate without loss of generality that the strength $H_1 = \overline{H}_1$ is real. With these assumptions, we find from the correction conditions

$$C_{3R} + C_{3H} = 0, \quad K_{3R} + K_{3H} = 0 \quad (9.59)$$

the hexapole strengths of the aplanator as

$$H_1 = \overline{H}_1 \approx \frac{1}{2f_o^2 l_1} \sqrt{\frac{C_{3R}}{3l_1}}, \quad H_2 \approx -\frac{\overline{K}_{3R}}{6l_2^2 f} \sqrt{\frac{l_1}{3C_{3R}}}. \quad (9.60)$$

It follows from the ratio

$$\left| \frac{H_2}{H_1} \right| \approx \frac{|K_{3R}|}{3C_{3R}} \frac{f_o^2 l_1^2}{f l_2^2} \quad (9.61)$$

and the relation $f > l_1 \approx l_2 \gg f_o$ that the absolute value of the hexapole strength H_2 is small compared with the strength H_1 of the symmetrically arranged hexapoles, which compensate for the spherical aberration. Hence, the correction of the coma does not appreciably affect the preceding correction of the spherical aberration. Because we can eliminate both the isotropic and anisotropic components of the coma, it is not necessary to image the coma-free plane of the objective lens in the front nodal plane of the aplanator. This additional freedom allows us to vary the fifth-order spherical aberration by moving the location z_{Ki} of the image of the coma-free plane of the objective lens away from the front nodal plane z_{N1} of the aplanator. This shift produces a third-order coma and a fifth-order spherical aberration with coefficients [102]

$$K_3 = C_{3R}(z_{N1} - z_{Ki})/f_o^2, \quad C_5 = 3K_3 C_{3R}. \quad (9.62)$$

The coefficient C_5 is positive if we move the corrector toward the image for a fixed location z_{Ki} . We eliminate the induced coma by readjusting the coma correction of the aplanator without producing an appreciable internal fifth-order spherical combination aberration.

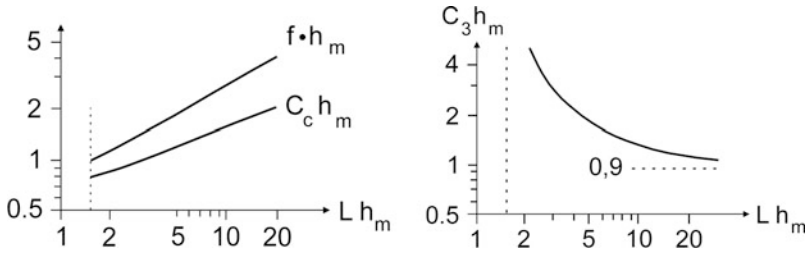
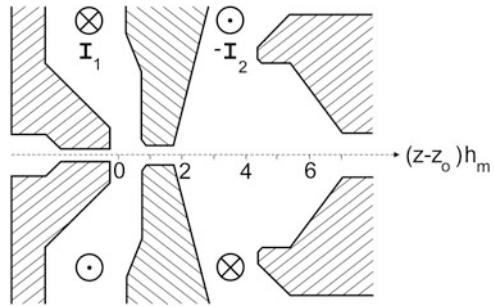


Fig. 9.30 Normalized focal length f and aberration coefficients C_c and C_3 as functions of the normalized axial extension $L h_m$ of a coma-free magnetic lens with minimum spherical aberration

Fig. 9.31 Schematic shape of the pole pieces of a coma-free magnetic lens with an extension

$L = z_e - z_o = 7/h_m$ of the axial induction between its end-plane z_e and the object plane z_o



In order to avoid saturation of the central pole piece of a coma-free magnetic lens and to obtain a sufficiently small focal length, the accelerating voltage must not exceed 100 kV. To achieve sub-Angstrom resolution at these voltages, we must compensate for spherical and chromatic aberration. The smaller the coefficients of these aberrations are, the less stringent are the requirements imposed on the corrector. Therefore, we must design the coma-free lens in such a way that the coefficients of its chromatic and spherical aberrations stay as small as possible. The calculations show that the spherical aberration decreases with increasing extension of the axial magnetic field, whereas the focal length and the coefficient C_c of the axial chromatic aberration increase.

We have plotted in Fig. 9.31 the focal length f and the coefficients C_c and C_3 of a coma-free lens with minimum spherical aberration as a function of the normalized extension $L h_m$ of the axial magnetic field. Here, $L = z_e - z_o$ is the length between the end-plane z_e of the axial magnetic field and the object plane z_o , which is located at maximum attainable induction B_m . The normalization length $L_n = 1/h_m$ is inversely proportional to the maximum induction B_m :

$$h_m = \sqrt{\frac{e}{8m_e \Phi^*}} B_m. \quad (9.63)$$

Figure 9.30 shows that the coefficients of chromatic and spherical aberration have about the same value for a normalized field extension $L h_m = 7$. For this length,

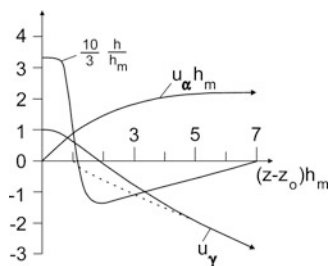


Fig. 9.32 Courses of the normalized axial magnetic induction $B/B_m = h/h_m$ and the fundamental rays u_α and u_γ within a coma-free magnetic lens with minimum spherical aberration in the case $Lh_m = 7$, the intersection of the dotted asymptote of the field ray u_γ with the optic axis defines the asymptotic coma-free plane of the lens

we have plotted in Fig. 9.31 schematically the geometry of the pole pieces and in Fig. 9.32 the corresponding fundamental rays u_α and u_γ . We have fixed the slope of the field ray u_γ at the object plane z_o in such a way that the coefficient of the radial (isotropic) component of the coma vanishes. In this case, the intersection point of the asymptote of this field ray with the optic axis defines the asymptotic coma-free plane z_K of the lens [102]. By matching this point with the coma-free point of the corrector by means of appropriately arranged transfer lenses shown in Fig. 9.21, we obtain an achromatic electron optical aplanat.

The coma-free lens is also useful for high-resolution Lorentz microscopy. For this mode of operation, we turn off the current I_1 and increase the current I_2 such that the object stays fixed at the field-free focal plane. In this case, we can tolerate the induced anisotropic coma because we do not aim for atomic resolution in the images of magnetic structures.

Radiation damage is a major obstacle for high-resolution imaging of objects consisting of light atoms because the threshold electron energy for atom displacement is proportional to the atomic number Z of the specimen atoms. For most biological objects and ceramics, the threshold energy is smaller than about 50 keV. The attainable resolution in the images of these objects is limited by the tolerable dose of the incident electrons rather than by the instrumental resolution of the microscope. In order to utilize as many scattered electrons as possible and to obtain a large field of view at low voltages, we must employ an achromatic aplanat. We realize such a system by combining the coma-free magnetic lens shown in Fig. 9.31 with the apochromator depicted in Fig. 9.18 or by combining a conventional objective lens with the SALVE corrector, which also compensates for the azimuthal coma apart from the axial aberrations. The latter system has been chosen for the SALVE project because this corrector provides a better resolution at low voltages than the first system using the coma-free magnetic objective.

Chapter 10

Electron Mirrors

Abstract A novel treatment of the optics of electron mirrors is presented utilizing the position ζ of the axial reference electron as the independent variable. This procedure enables one to determine the path of electrons with large ray gradients encountered in electron mirrors in the vicinity of turning points. The aberrations of electron mirrors are derived by a perturbation procedure which considers the lateral coordinates $x = x(\zeta)$, $y = y(\zeta)$ and the axial distance $z(\zeta) - \zeta$ of the electron as small quantities. In particular it is shown that a specially designed electrostatic tetrode mirror provides negative chromatic and spherical aberrations, which can be adjusted in such a way to compensate for the corresponding positive aberrations of round lenses, as realized in the SMART microscope operating either as a low-voltage electron microscope or as a photoemission electron microscope.

The conventional theory of electron motion in static fields considers the z -coordinate of a charged particle as the independent variable measured along the optic axis. Usually the optic axis is chosen to coincide with the central trajectory of a bundle of rays regardless of whether this trajectory is straight or curved. We have replaced derivatives with respect to the time in the path equation and in the Lagrange function by derivatives with respect to the z -coordinate utilizing the conservation of energy. As a result, the lateral position coordinates $x = x(z)$, $y = y(z)$ of the electron are functions of the z -coordinate instead of the time t . This approach is valid as long as the axial velocity component does not reverse its direction of flight. If, in addition, the motion is confined to the region near the optic axis, the slope components $x'(z)$ and $y'(z)$ remain sufficiently small. However, large ray gradients do occur in the vicinity of turning points at which the axial direction of flight of the particle changes its sign or near the emitting surfaces of cathodes. Examples for systems with turning points are electron mirrors, ion traps, and the magnetic bottle. Because the components $x'(z)$ and $y'(z)$ of the ray gradient diverge at the turning point, we must describe in this case the position coordinates x , y , and z as functions of an appropriate independent variable τ , which must not necessarily be the time t . Large ray gradients also occur in cathodes.

Although electron mirrors have been studied at the very beginning of electron optics [158], they were not considered as promising elements for correcting aberrations. The reason for this belief may stem from early work by Ramberg [159] who found that the dimension of a mirror must be unrealistically small for correcting aberrations. Later studies showed that this pessimistic view does not hold true [160–162]. An electrostatic mirror had been utilized in electron microscopy only as a reflection element for an imaging energy filter [163, 164]. The long-time negligence of studying thoroughly the correction properties of mirrors may be attributable to the difficulties associated with the violation of the standard conditions for the paraxial trajectories in the region of the turning point where the ray gradient diverges. Revived interest in electron mirrors originated from the work of Rempfer and Mauck [165] in the context to correct the spherical and chromatic aberration of a low-voltage photoemission electron microscope (PEEM) by means of a hyperbolic two-electrode mirror. However, for adjusting the chromatic and spherical aberration of the mirror for a fixed focal length, we must increase the number of electrodes from two to four. Unlike a light-optical mirror, where the reflection occurs at the physical surface, the electron mirror represents a “soft” mirror, which allows the electrons to penetrate into the inhomogeneous refracting medium formed by the electrostatic potential. The depth of penetration depends on the energy and the direction of the electron in front of the mirror. We can conceive the total reflection as the sum of consecutive refractions on a continuous set of electrostatic equipotentials.

Among other constraints, the validity of the Scherzer theorem requires that the velocity of the axial electron does not change sign. This condition is not fulfilled in an electrostatic mirror. As a consequence, the coefficients C_c and C_3 of its chromatic and spherical aberration can be made negative to compensate for the positive coefficients of rotationally symmetric lenses. Owing to this possibility, electron mirrors are employed for correcting the aberrations of low-energy electron microscopes [136].

10.1 Reference Electron

In order to obtain an appropriate independent variable, we introduce the reference electron, which propagates along the optic axis with nominal kinetic energy E_n . The turning point $\zeta_T = z(T)$ of this electron is obtained from the relation $\dot{z}(t = T) = 0$. We measure the position of an arbitrary electron

$$x = x(t), \quad y = y(t), \quad z = \zeta(t) + h(t)y = y(t), z = \zeta(t) + h(t) \quad (10.1)$$

with respect to the corresponding position of the reference electron, as shown in Fig. 10.1. The coordinates of the reference electron are

$$x_r(t) = 0, \quad y_r(t) = 0, \quad z_r(t) = \zeta(t). \quad (10.2)$$

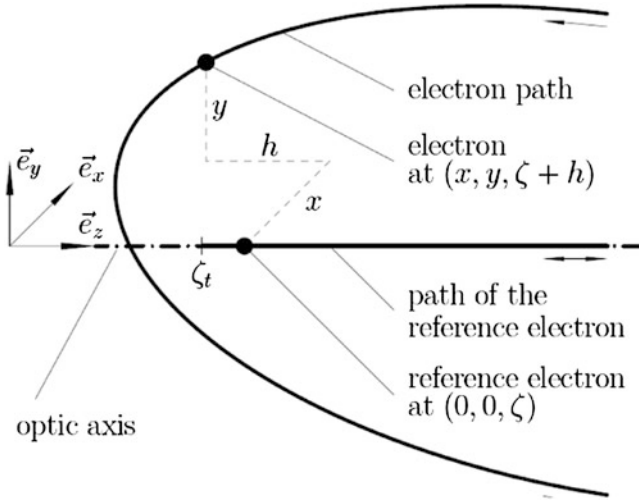


Fig. 10.1 Position (x, y, h) of an electron referred to the position $(x = 0, y = 0, z = \zeta)$ of the reference electron, ζ_T defines its turning point

Since the electrons are confined to the vicinity of the optic axis and because the initial velocities of all electrons including the reference electron differ only slightly, the components x , y and the axial deviation h are small quantities. Therefore, we can expand the equation of motion with respect to these variables.

10.2 Equation of Motion

In the case of a mirror or a cathode, it is advantageous to start from the Lorentz equation of motion (2.3). In paragraph 2.1.2, we have imposed on the electric potential the gauge $\varphi = 0$ at the surface of the cathode. Therefore, the electric potential $\varphi = \varphi(x, y, z)$ must also vanish at the turning point $z = \zeta_T$ of the reference electron which starts with total energy $E_{\text{tot}} = 0$ from the surface of the cathode:

$$\varphi(x = 0, y = 0, z = \zeta_T) = 0. \quad (10.3)$$

Within the field-free column of the microscope, the electric potential has the value $\varphi = \Phi = \Phi_c$. To facilitate the calculations as much as possible, it is advantageous to define the differential time equivalent as

$$d\tau = \frac{\sqrt{2e\Phi_c^*/m_e}}{1 + (e\phi + \delta E)/E_0} dt. \quad (10.4)$$

According to this definition, the generalized time τ has the dimension of a length. The gauge (10.3) sets the total energy of an electron with nominal energy equal to zero. Hence, the total energy of an arbitrary electron equals its energy deviation δE from the total energy (2.8) of the reference electron

$$E_{\text{tot}} = (m - m_e)c^2 - e\varphi = \delta E. \quad (10.5)$$

In the following calculations, dots indicate differentiations with respect to τ , and primes denote differentiations with respect to z , ζ , or h . This notation is reasonable because the three spatial variables are linearly related with each other (10.1). We can rewrite the conservation of energy (10.5) in the form

$$m = m_e\gamma = m_e \{1 + (e\varphi + \delta E)/E_0\} \quad (10.6)$$

for the relativistic mass of an electron with relative energy deviation $\delta E/E_0$. Using the relations (10.4) and (10.5), we transform the path equation (2.3) into the equation

$$\ddot{\vec{r}} = -\frac{\vec{E}}{2\Phi_c^*} \{1 + (e\varphi + \delta E)/E\} - \sqrt{\frac{e}{2m_e\Phi_c^*}} (\dot{\vec{r}} \times \vec{B}). \quad (10.7)$$

We decompose this three-dimensional differential equation into a complex equation for the complex lateral position component $w = x + iy$ and a real equation for the axial deviation h and introduce the complex off-axial field strengths

$$E_w = E_x + iE_y, \quad B_w = B_x + B_y. \quad (10.8)$$

By employing this notation, we eventually obtain the equations

$$\ddot{w} = -\frac{E_w}{2\Phi_c^*} \{1 + (e\varphi + \delta E)/E_0\} - i\sqrt{\frac{e}{2m_e\Phi_c^*}} (B_w\dot{z} - B_z\dot{w}), \quad (10.9)$$

$$\ddot{z} = \ddot{\zeta} + \ddot{h} = -\frac{E_z}{2\Phi_c^*} \{1 + (e\varphi + \delta E)/E_0\} + \sqrt{\frac{e}{2m_e\Phi_c^*}} \text{Im}(B_w\dot{w}). \quad (10.10)$$

Owing to the conservation of energy (10.5), the energy deviation δE is not a free parameter if we fix the trajectory of an electron by its position $w_i = w(\tau_i)$, $z_i = z(\tau_i)$ and its velocity components $\dot{w}_i = \dot{w}(\tau_i)$, $\dot{z}_i = \dot{z}(\tau_i)$ at some initial generalized time $\tau = \tau_i$. Therefore, we can only choose two velocity components arbitrarily for a given gauge of the electric potential and a given δE . Usually, one chooses the lateral velocity components $\dot{x}_i$ and $\dot{y}_i$ as free parameters.

For the reference electron ($\delta E = 0$, $w = 0$, $z = \zeta$, $h = 0$), Eq. (10.10) adopts the simple form

$$\ddot{\zeta} = \frac{\Phi'}{2\Phi_c^*} (1 + 2\varepsilon\Phi). \quad (10.11)$$

Multiplying the resulting equation with $\dot{\zeta}$ and integrating with respect to τ , we find the conservation of energy for the reference electron in the simple form

$$\dot{\zeta}^2 = \frac{\Phi^*}{\Phi_c^*}. \quad (10.12)$$

Within the field-free region of the column, the electric potential is constant. Then the normalized velocity of the reference electron is $\dot{\zeta} = \dot{\zeta}_c = \pm 1$. In this case, differentiations with respect to τ and z are identical apart from the sign. This behavior is a consequence of our special choice (10.4) of the time parameter τ . The negative sign accounts for the propagation toward the mirror and the positive sign for the motion away from the mirror after reflection, as illustrated in Fig. 10.1.

Employing complex notation, the components of the electric and magnetic field strengths are connected with the scalar potentials φ, ψ via the relations

$$\begin{aligned} E_w &= -2 \frac{\partial \varphi}{\partial \bar{w}}, & E_z &= -\frac{\partial \varphi}{\partial h}, \\ B_w &= -2 \frac{\partial \psi}{\partial \bar{w}}, & B_z &= -\frac{\partial \psi}{\partial h}. \end{aligned} \quad (10.13)$$

In the case of a straight optic axis, we readily obtain a power series expansion of the potentials with respect to $w, \bar{w}, h$ by employing the multipole representation (3.37) for the electric potential φ and expanding the multipole strengths $\Phi_\nu(z) = \Phi_\nu(\zeta + h)$ with respect to h , giving

$$\varphi = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{\nu=0}^{\infty} (-)^n \frac{\nu!}{n!(n+\nu)!m!} \left(\frac{w\bar{w}}{4} \right)^n h^m \text{Re} \{ \Phi_\nu^{[2n+m]}(\zeta) \bar{w}^\nu \}. \quad (10.14)$$

We obtain the equivalent expansion for the scalar magnetic potential by substituting ψ for φ and the magnetic multipole strength $\Psi_\nu(\zeta)$ for $\Phi_\nu(\zeta)$. In the case of a curved optic axis, we must employ the representation (3.54) of the electric potential and expand the multipole strengths in the same way as in the case of a straight optic axis. The components $\Phi_\nu(\zeta)$ and $\Psi_\nu(\xi)$ have the symmetry property

$$\Phi_\nu(\zeta(\tau)) = \Phi_\nu(\zeta(2\tau_T - \tau)), \quad \Psi_\nu(\zeta(\tau)) = \Psi_\nu(\xi(2\tau_T - \tau)). \quad (10.15)$$

As a result, symmetric and antisymmetric solutions of the differential equations (10.9) and (10.10) exist with respect to the axial reversal point $\zeta_T = \zeta(\tau_T)$.

10.3 Eikonal Approach

An elegant alternative to the Lorentz equation of motion is the action or eikonal approach. This procedure yields directly the relation (10.4) for the generalized time τ . Starting from the relation (2.10) and considering the expression (10.6) for the conservation of energy, we rewrite the action integral as

$$W = \text{Ex} \int_{\tau_0}^{\tau_r} L(\vec{r}, \dot{\vec{r}}) d\tau. \quad (10.16)$$

The integration starts at a given time τ_0 and ends at the recording time τ_r . Remember that dots indicate derivatives with respect to τ . In order to obtain the path equations (10.9) and (10.10) from the condition $\delta W = 0$, we must construct an appropriate expression for the Lagrangian L . As a guide, we start from the expression (2.14) for the action and utilize the relation

$$m \frac{d\tau}{dt} = m_e \{1 + (e\varphi + \delta E)/E_0\} \frac{d\tau}{dt} = \sqrt{2em_e \Phi_c^*}. \quad (10.17)$$

By employing this relation, we write the action (10.16) in the form

$$W = \text{Ex} \int_{\tau_0}^{\tau_r} \left\{ \sqrt{em_e \Phi_c^*/2} [\dot{\vec{r}}^2 + (\varphi^* + \varphi \delta E/E_0)/\Phi_c^*] - e \dot{\vec{r}} \vec{A} \right\} d\tau. \quad (10.18)$$

Using the eikonal approach has the advantage that we can readily generalize the Lagrangian to systems with curved axis. For this purpose, we must substitute in (10.19) the differential element $g_3 dz$ for dz , as outlined in Sect. 3.3. The metric element g_3 is related with the complex curvature $\Gamma = \Gamma(\zeta)$ of the optic axis via

$$g_3 = 1 - \text{Re} \{ \Gamma \bar{w} \}. \quad (10.19)$$

Substituting $\zeta + h$ for the z -coordinate and introducing complex notation for the lateral components of the vectors $\vec{r}$ and $\vec{A}$, the Lagrangian adopts the form

$$L = \frac{1}{2} \sqrt{2em_e \Phi_c^*} \left\{ \dot{w} \dot{\bar{w}} + g_3^2 (\dot{\zeta} + \dot{h})^2 + \frac{\varphi^* + \varphi \delta E/E_0}{\Phi_c^*} \right\} - e \left\{ (\dot{\zeta} + \dot{h}) g_3 A_z + \text{Re}(\dot{w} A_w) \right\}. \quad (10.20)$$

The function $\zeta = \zeta(\tau)$ is a free parameter, which we choose in such a way that the coordinate $h = h(\tau)$ becomes a small quantity. For example, this is the case if ζ represents the z -coordinate of the axial electron, as realized in Sect. 10.2. Here the deviation h is zero for the axial reference electron. For systems with a straight optic axis ($\Gamma = 0$), we must put $g_3 = 1$.

We derive the path equations from the Euler–Lagrange equations

$$\frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{w}} \right) = \frac{\partial L}{\partial w}, \quad \frac{d}{d\tau} \left(\frac{\partial L}{\partial \dot{h}} \right) = \frac{\partial L}{\partial h}. \quad (10.21)$$

Employing the expression (10.20) for L and putting $g_3 = 1$, we obtain readily Eqs. (10.9) and (10.10). We have chosen the gauge of the magnetic vector potential in Section 3.4 such that the axial component A_z is zero along the optic axis is zero. Taking into account this gauge, we obtain for the Lagrangian of the axial electron the expression

$$L^{(0)} = L(w = 0, \dot{w} = 0, h = 0, \dot{h} = 0; \tau) = \sqrt{\frac{em_e}{2\Phi_c^*}} \Phi^*(\zeta) \left\{ 2 + \frac{\delta E}{(1 + \varepsilon \Phi) E_0} \right\}. \quad (10.22)$$

For the reference electron, the expression reduces to

$$L_r = \Phi^* \sqrt{2em_e / \Phi_c^*}. \quad (10.23)$$

To illustrate the effect of a mirror, we consider the action of the reference electron. The initial starting point is located at position $\zeta_s = z_s > \zeta_T$. After the reflection, the electron is detected at the recording plane $\zeta_r = z_r$. In this case, the electron travels from the initial point ζ_0 to the turning point $\zeta_T = z_T$ and from there back to the point of observation ζ_r . For this electron, the differential elements $d\tau$ and $d\zeta$ are related via

$$d\tau = \pm \sqrt{\frac{\Phi_c^*}{\Phi^*(\zeta)}} d\zeta. \quad (10.24)$$

We must take the minus sign for the motion toward the mirror and the plus sign for the motion away from the mirror. Considering this behavior, we obtain for the action of the reflected reference electron the expression

$$W_r(\zeta_0, \zeta_r) = \int_{\tau_0}^{\tau_r} L_r d\tau = \int_{\zeta_0}^{\zeta_i} \sqrt{2em_e \Phi^*} d\zeta + \int_{\zeta_T}^{\zeta_r} \sqrt{2em_e \Phi^*} d\zeta. \quad (10.25)$$

We have removed the minus sign of the first integral by exchanging its lower and upper limit of integration.

10.4 Rotationally Symmetric Mirrors

In the case of rotationally symmetric fields, it is advantageous to choose the rotating u, z - coordinate system, where the complex off-axial coordinate is given by

$$u = w \exp(-i\chi). \quad (10.26)$$

The angle χ accounts for the Larmor rotation (4.24). Here we choose the slightly modified angle

$$\chi = \sqrt{\frac{e}{8m_e\Phi_c^*}} \int_{\tau_0}^{\tau} B d\tau. \quad (10.27)$$

In accordance with the nomenclature introduced in Sect. 4.1, B denotes the magnetic flux density along the optic axis:

$$B = B_z(x = 0, y = 0, z) = -\Psi'(z). \quad (10.28)$$

Employing the relations (10.26) and (10.27) and considering that $2\ddot{\chi} = \Phi'(1 + \varepsilon\Phi)/\Phi_c^*$, we transform Eqs. (10.9) and (10.10) into the following equations for u and h :

$$\begin{aligned} \ddot{u} + \frac{\gamma_0\Phi''}{4\Phi_c^*} + \frac{eB^2}{8m_e\Phi_c^*}u &= p_u, \\ p_u &= i\sqrt{\frac{e}{8m_e\Phi_c^*}} \left\{ 2(B_z - B)(\dot{u} - \dot{\chi}u) - B'\dot{\chi}u - 2B_w e^{-i\chi}(\dot{\chi} + \dot{h}) \right\} \\ &\quad + \frac{E_w}{2\Phi_c^*} \left(\frac{\delta E}{E_0} - \gamma \right) + \frac{\gamma_0\Phi''}{4\Phi_c^*}, \end{aligned} \quad (10.29)$$

$$\begin{aligned} \ddot{h} - \frac{\Phi''^*}{2\Phi_c^*}h - \frac{\Phi'}{2\Phi_c^*} \frac{\delta E}{E_0} &= p_h, \\ p_h &= \sqrt{\frac{e}{2m_e\Phi_c^*}} \text{Im} \{ \overline{B}_w e^{i\chi} \dot{u} \} + \frac{eB}{4m_e\Phi_c^*} \text{Re} \{ \overline{B}_w u e^{i\chi} \} \\ &\quad - \frac{\gamma_0\Phi' + \Phi''^* + \gamma E_z}{2\Phi_c^*} - \frac{\delta E}{E_0} \frac{(E_z + \Phi')}{2\Phi_c^*}. \end{aligned} \quad (10.30)$$

Here $\gamma_0 = 1 + 2\varepsilon\Phi$ is the relativistic factor for the reference electron. We have written the equations in such a form that the expressions p_u and p_h do not contain terms, which are linear in the variables $u, \bar{u}, \dot{u}, \dot{\bar{u}}$ and δE . This representation allows one to conceive the complex term p_u and the real term p_h as perturbations preventing ideal imaging.

10.4.1 Linear Approximation

We obtain the linear approximation of Eqs. (10.29) and (10.30) by neglecting the nonlinear perturbation terms p_u and p_h . Putting $p_u = 0, p_h = 0$, we readily obtain the Gaussian path equations

$$\ddot{u}^{(1)} + \frac{\gamma_0}{4} \frac{\Phi''}{\Phi_c^*} u^{(1)} + \frac{eB^2}{8m_e\Phi_c^*} u^{(1)} = 0, \quad (10.31)$$

$$\ddot{h}^{(1)} - \frac{\Phi''^*}{2\Phi_c^*} h^{(1)} = \frac{\Phi'}{2\Phi_c^*} \frac{\delta E}{E_0}. \quad (10.32)$$

Employing the relation (10.24), the lateral equation (10.31) adopts the familiar form (4.48) of the paraxial path equation of electromagnetic round lenses. The inhomogeneous term on the right-hand side of (10.32) vanishes in the nonrelativistic approximation $E_0 \rightarrow \infty$.

10.4.2 Lateral Fundamental Rays

The lateral equation (10.31) has two linearly independent solutions. To obtain a symmetric solution u_σ and an antisymmetric trajectory u_v , we impose on these solutions the initial conditions

$$u_\sigma(\tau_T) = -1, \quad \dot{u}_\sigma(\tau_T) = 0, \quad u_v(\tau_T) = 0, \quad \dot{u}_v(\tau_T) = 1. \quad (10.33)$$

With this choice, the Wronskian adopts the simple form

$$u_v \dot{u}_\sigma - u_\sigma \dot{u}_v = 1. \quad (10.34)$$

Owing to the linearity of the paraxial path equations, we can describe an arbitrary paraxial ray as a linear combination of the independent solutions u_σ and u_v . The coefficients are generally complex. Analytical solutions of (10.31) exist only for a few simple systems. Therefore, we must calculate in most cases the fundamental rays numerically. Utilizing the relation (10.24), we can express the fundamental rays $u_\sigma(\tau)$ and $u_v(\tau)$ as functions of the axial coordinate ζ of the reference electron. Using this representation, we obtain the location $\zeta = \zeta_C$ of the center curvature of the mirror from the condition

$$u_\sigma(\zeta_f) = 0. \quad (10.35)$$

If we use the mirror as a corrector compensating for the aberrations of the objective lens of an electron microscope, we must image the object plane into the focal plane of the mirror. In this case, the symmetric fundamental ray u_σ coincides with the axial fundamental ray u_α apart from a constant factor $1/M_m$, where M_m is the magnification of the object plane at the focal plane of the mirror.

10.4.3 Longitudinal Fundamental Deviations

The homogeneous part of the linear differential equation (10.32) for the Gaussian axial deviation $h^{(1)}(\zeta)$ has two linearly independent solutions, one of which is symmetric, the other antisymmetric with respect to the turning plane $\zeta = \zeta_T$. The antisymmetric solution $h_v(\zeta(\tau))$ is given by the function

$$h_v = \dot{\zeta} = \pm \sqrt{\frac{\Phi^*}{\Phi_c^*}}. \quad (10.36)$$

The symbol $\pm$ accounts for the change of the sign at $\tau = \tau_T$ or at the turning plane, respectively. The plus sign has to be taken after the reflection ($\tau > \tau_T$). We have normalized the solution such that $h_v = \pm 1$ within the field-free region of the column. Differentiation of h_v with respect to τ gives

$$\dot{h}_v = \dot{\zeta} h'_v = \frac{\Phi'^*}{2\Phi_c^*} = \frac{\gamma_0 \Phi'}{2\Phi_c^*} = \frac{\Phi'(1 + 2\varepsilon\Phi)}{2\Phi_c^*}. \quad (10.37)$$

The symmetric solution $h_\sigma = h_\sigma(\tau)$ reverses its direction of flight at the turning point $\zeta_T = \zeta(\tau_T)$. We normalize the symmetric solution such that the Wronskian of the fundamental rays has the form

$$h_v \dot{h}_\sigma - h_\sigma \dot{h}_v = 1. \quad (10.38)$$

Utilizing this expression and the relation (10.36), we can express the symmetric axial fundamental ray as

$$h_\sigma = h_v \int_{\tau_\sigma}^{\tau} \frac{1}{h_v^2} d\tau = \Phi_c^* \sqrt{\Phi^*} \int_{\zeta_\sigma}^{\zeta} \frac{d\zeta}{\Phi^{*3/2}}. \quad (10.39)$$

In order that $h_\sigma = h_\sigma(\zeta) = h_\sigma(\tau)$ represents the symmetric solution, we must choose the lower integration limit $\zeta = \zeta_\sigma$ in such a way that $\dot{h}_\sigma(\tau_T) = 0$. Because the slope $h'_\sigma(\zeta)$ changes its sign at the reflection plane ζ_T , the condition $h'_\sigma(\zeta_T) = 0$ must also be fulfilled. We use this condition to determine the integration limit ζ_σ . Differentiating (10.39) with respect to ζ , we obtain

$$h'_\sigma / \Phi_c^* = \frac{1}{\Phi^*} + \frac{\Phi'^*}{2\sqrt{\Phi^*}} \int_{\zeta_\sigma}^{\zeta} \frac{d\zeta}{\Phi^{*3/2}}. \quad (10.40)$$

Here $\Phi'^* = d\Phi^*/d\zeta = \Phi'(1 + 2\varepsilon\Phi)$ denotes the derivative of the relativistic modified axial electric potential $\Phi^* = \Phi^*(\zeta)$. Both terms on the right-hand side of (10.40) diverge at the turning plane ζ_T . In order to avoid this divergence, we transform the integral by partial integration as follows:

$$\int_{\zeta_\sigma}^{\zeta} \frac{d\zeta}{\Phi^{*3/2}} = \frac{2}{\Phi'^*_\sigma \Phi^{*1/2}_\sigma} - \frac{2}{\Phi'^* \Phi^{*1/2}} + 2 \int_{\zeta_\sigma}^{\zeta} \frac{\Phi''^* d\zeta}{\Phi'^{*2} \Phi^{*1/2}}. \quad (10.41)$$

Substituting this relation for the integral in the expression (10.40) and considering the condition $\dot{h}_\sigma(\tau_T) = h'_\sigma(\zeta_T)\dot{\zeta}(\tau_T) = 0$, we obtain the intersection point $\zeta = \zeta_\sigma$ of the symmetric axial fundamental ray with the optic axis from the integral relation

$$\Phi'^*_\sigma \Phi^{*1/2}_\sigma \int_{\zeta_T}^{\zeta_\sigma} \frac{\Phi''^* d\zeta}{(\Phi'^*)^2 \Phi^{*1/2}} = 1. \quad (10.42)$$

The index σ indicates the value taken at the plane $\zeta = \zeta_\sigma$. It readily follows from the representation (10.39) that $h_\sigma(\zeta_\sigma) = 0$. We can determine analytically the value of the symmetric solution at the turning plane from the Wronskian (10.38) by taking into account the expression (10.37) and considering the relations

$$\dot{h}_\sigma(\tau_T) = 0, \quad h_v(\tau_T) = 0. \quad (10.43)$$

As result, we readily obtain

$$h_\sigma(\tau_T) = -1/\dot{h}_v(\tau_T) = -\frac{2\Phi^*_c}{\Phi'^*(\zeta_T)}. \quad (10.44)$$

In the field-free region within the column, the fundamental axial deviations fulfill the relations

$$h_v = \dot{h}_\sigma = \pm 1, \quad \dot{h}_v = 0. \quad (10.45)$$

These properties result from the conservation of energy.

Employing the method of variation of parameters, we obtain for the inhomogeneous solution of the differential equation (10.30) the expression

$$h_{\text{in}} = \kappa h_\kappa = \kappa \varepsilon \left\{ h_\sigma \int_{\tau_T}^{\tau} \Phi' h_v d\tau - h_v \int_{\tau_T}^{\tau} \Phi' h_\sigma d\tau \right\} = \kappa \varepsilon h_v \int_{\tau_T}^{\tau} \frac{d\tau}{1 + \varepsilon \Phi}. \quad (10.46)$$

Here we have introduced the chromatic parameter

$$\kappa = \frac{\delta E}{e \Phi^*_c}. \quad (10.47)$$

We have chosen the lower integration limit such that h_κ vanishes at the turning point ζ_T . In this case, the course of h_κ is symmetric with respect to τ_T because $h_v(\tau)$ and the integral of the second relation on the right-hand side of (10.46) are both antisymmetric with respect to the turning time. We derived the second expression by integrating the integrals of the first relation employing the relation $h_v d\tau = d\zeta$ and substituting the relation (10.39) for h_σ . The inhomogeneous solution (10.45) vanishes in the nonrelativistic limit $\varepsilon \rightarrow 0$.

The general solution of Eq. (10.32) for the axial deviation has the form

$$h_1 = c_\sigma h_\sigma + c_v h_v + \kappa h_\kappa. \quad (10.48)$$

The coefficients c_σ and c_v are real. We postulate that at the starting plane $\zeta = \zeta_0$, the approximation (10.48) and its derivative $\dot{h}_1$ coincide with the exact values $h_0 = h(\tau(\zeta_0)) = h(\zeta_0)$ and $\dot{h}_0 = \dot{h}(\zeta_0)$. In this case, it is advantageous to introduce the fundamental solutions h_α and h_γ . These solutions satisfy the standard initial conditions

$$h_\alpha(\zeta_0) = h_{\alpha 0} = h_{\gamma 0} = 0, \quad \dot{h}_{\alpha 0} = h_{\gamma 0} = 1. \quad (10.49)$$

Considering the Wronskian (10.38), we readily derive the relations

$$\begin{aligned} h_\alpha &= h_{v0} h_\sigma - h_{\sigma 0} h_v = \sqrt{\Phi_0 \Phi} \int_{\tau_0}^{\tau} \frac{d\tau}{\Phi}, \\ h_\gamma &= \dot{h}_{\sigma 0} h_v - h_{v0} h_\sigma = \pm \sqrt{\frac{\Phi}{\Phi_0}} = h_v \sqrt{\frac{\Phi_c}{\Phi_0}}. \end{aligned} \quad (10.50)$$

Employing these solutions, the axial deviation adopts the form

$$h_1 = \dot{h}_0 h_\alpha + h_0 h_\gamma + \kappa \hat{h}_\kappa. \quad (10.51)$$

We obtain the chromatic deviation $\hat{h}_\kappa$ from the expression (10.46) by substituting the starting time τ_0 for the lower integration limit τ_T . As a result, $\hat{h}_\kappa$ vanishes at the starting plane. We assume that the trajectories originate at position w_0 from the plane $z_0 = \zeta_0$. In this case, the initial axial deviation is zero:

$$h_0(z_0) = h_0 = 0. \quad (10.52)$$

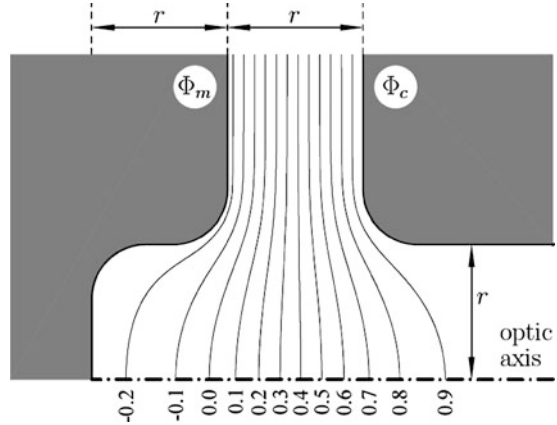
However, the initial deviation of the axial velocity component differs from zero. We obtain this deviation from the conservation of energy. Assuming that the starting plane is located in field-free space at potential $\varphi = \Phi_c$, we derive the relation

$$\begin{aligned} \dot{h}_0 &= [1 + \dot{w}_0 \dot{w}_0 + \kappa(1 + 2\varepsilon \Phi_c) + \kappa^2 \varepsilon \Phi_c^*]^{1/2} - 1 \\ &= \sum_{r=1}^{\infty} \dot{h}_0^{(r)} = \frac{\kappa}{2}(1 + 2\varepsilon \Phi_c) + \frac{\dot{w}_0 \dot{w}_0}{2} - \frac{\kappa^2}{8} + \dots. \end{aligned} \quad (10.53)$$

The index r denotes the sum of the exponents of the coefficients $\dot{w}_0, \ddot{w}_0, \kappa$ and defines the rank of the expansion term $\dot{h}_0^{(r)}$. Equalizing this term with the corresponding term of the last expression in (10.53), we find

$$\dot{h}_0^{(1)} = \frac{\kappa}{2}(1 + 2\varepsilon \Phi_c), \quad \dot{h}_0^{(2)} = \frac{w'_0 \dot{w}_0}{2} - \frac{\kappa^2}{8}. \quad (10.54)$$

Fig. 10.2 Sectional views of the electrodes and of the equipotential surfaces of the diode mirror. The equipotentials are normalized with respect to the column potential Φ_c



The relations (10.52) and (10.53) demonstrate that $\dot{h}_0$ is a function of the expansion parameters $\dot{w}_0$, $\dot{w}_0$, and κ . Since $\dot{h}_0$ does not possess linear terms in $\dot{w}_0$ and $\dot{w}_0$, the first-rank approximation $h^{(1)}$ is of entirely chromatic nature:

$$h^{(1)} = \dot{h}_0^{(1)} h_\alpha = \frac{\kappa}{2} (1 + 2\varepsilon \Phi_c) h_\alpha. \quad (10.55)$$

Hence, for monochromatic electrons ($\kappa = 0$), the axial position ζ of the reference electron represents within the frame of validity of Gaussian optics the z -coordinate of all other electrons which start at the same time $\tau = \tau_0$ from the initial plane $z = z_0 = \zeta_0$.

Example

To illustrate the imaging properties of an electron mirror, we consider an electrostatic mirror consisting of two electrodes. Figure 10.2 shows the arrangement, the shape of the electrodes, and several equipotentials. The cylindrical mirror electrode with bore radius r is put at the voltage $\phi_m = \Phi_m = -0.25\Phi_c$. The radius of the curved edges of the electrode surfaces is $0.4r$. Because the electron velocity is small compared within the velocity of light within the mirror, we assume nonrelativistic conditions ($\varepsilon = 0$). In order to obtain analytical expression for the potential distribution, we employ the charge simulation method, which approximates the true potential by a sum of potential of ring charges located within the electrodes near their surfaces. Employing these potentials, we have numerically solved the linear path equations (10.31) and (10.32). The resulting paths $\zeta = \zeta(\tau)$ of the reference electron, $u_v(\tau) = w_v(\tau)$, $u_\sigma(\tau) = w_\sigma(\tau)$, of the fundamental rays and $h_v(\tau)$, $h_\sigma(\tau)$ of the axial deviations are depicted in Fig. 10.3. The course of these ray components along the optic ζ -axis is shown in Fig. 10.4.

To survey the properties of this mirror, it suffices to assume nonrelativistic conditions. Moreover, we place the origin of the ζ coordinate at the center of the surface of the mirror electrode. In this case, the turning point is located at $\zeta_T = 0.865431r$. The characteristic elements of the mirror are determined by employing

Fig. 10.3 Position $\zeta(\tau)$ of the reference electron and paths of the fundamental rays and of the axial deviations $h_v(\tau)$, $h_\sigma(\tau)$ as functions of τ/r

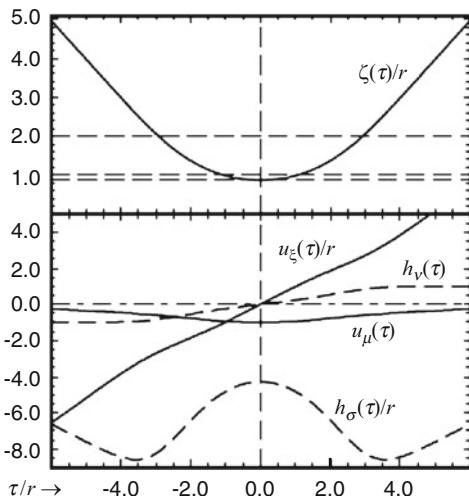
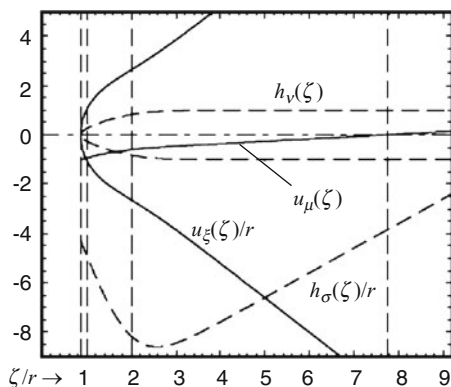


Fig. 10.4 Paths of the fundamental rays $u_v(\zeta)$, $u_\sigma(\zeta)$ and of the axial deviations $u_v(\tau)$, $u_\sigma(\tau)$, $h_v(\zeta)$, $h_\sigma(\zeta)$ as functions of the position ζ of the reference electron



a step-controlled Runge–Kutta method of fourth order and a trapezoidal integration. The results show that the symmetric fundamental ray $u_\sigma = w_\sigma$ intersects the optic axis asymptotically at position $\zeta_C = 7.7376r$, which defines the center of curvature of the mirror. For a convex mirror, the symmetric ray does not intersect the optic axis. In this case, the point of intersection of the asymptote of u_σ with the optic axis can be considered as the center of curvature of the convex mirror.

The mirror acts as a concave mirror if $\Phi'' > 0$ in the region in front of the turning point. Hence, to achieve a positive curvature of the equipotentials in this region, the central surface of the last mirror electrode must be curved toward the electron beam, as it is the case for the dipole mirror shown in Fig. 10.2. We have checked the accuracy of our calculations by observing the convergence with increasing number of ring charges and Runge–Kutta steps. The maximum number of charges was 2,000, and the maximum number of steps was 4,000.

10.5 Path Deviations

We define the differences between the actual position coordinates of the electron and their paraxial approximation as the path deviations

$$\Delta u = u(\tau) - u^{(1)}(\tau) = \sum_{r=2}^{\infty} u^{(r)}(\tau), \quad \Delta h = h(\tau) - h^{(1)}(\tau) = \sum_{r=2}^{\infty} h^{(r)}(\tau). \quad (10.56)$$

The lateral deviations $u^{(r)}(\tau)$ and the longitudinal deviations $h^{(r)}(\tau)$ are polynomials of rank r in the four geometrical ray parameters and the chromatic parameter κ . We derive these deviations most conveniently by transforming the path equations (10.29) and (10.30) into integral equations. For this purpose, we consider the nonlinear perturbation terms p_u and p_h as known functions of the modified time τ . In this case, the path equations represent a set of linear inhomogeneous differential equations which can be solved by applying the method of variation of coefficients. As a result of the somewhat lengthy yet straightforward calculation, we eventually find

$$u = u_1 + u_\alpha \int_{\tau_0}^{\tau} p_u u_\gamma d\tau - u_\gamma \int_{\tau_0}^{\tau} p_u u_\alpha d\tau, \quad (10.57)$$

$$h = h_1 + h_\alpha \int_{\tau_0}^{\tau} p_h h_\gamma d\tau - h_\gamma \int_{\tau_0}^{\tau} p_h d\tau. \quad (10.58)$$

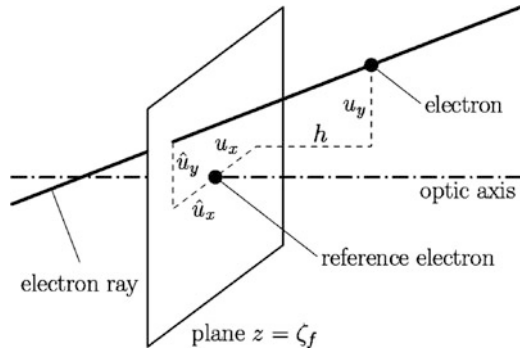
Since p_u and p_h are functions of $u, \bar{u}$ and h , the relations (10.57) and (10.58) represent a set of two coupled inhomogeneous nonlinear integral equations. The inhomogeneous term u_1 is identical with the paraxial solution

$$u_1 = u^{(1)} = \dot{u}_0 u_\alpha + u_0 u_\gamma. \quad (10.59)$$

This behavior does not hold for the inhomogeneous term h_1 , because this term is a function of the ray parameters as follows from the relations (10.50) and (10.52). We solve the integral equations (10.57) and (10.58) most conveniently by an appropriate iteration procedure. To be consistent, we aim for a procedure which yields the deviations according to their rank. For this purpose we substitute the expansions (10.56) for the ray coordinates $u, \bar{u}$ and h in the perturbation functions $p_u(u, \bar{u}, h)$ and $p_h(u, \bar{u}, h)$. Subsequently, we order the resulting expression as a sum of polynomials of equal rank in the ray parameters, giving

$$p_u = \sum_{r=2}^{\infty} p_u^{(r)}, \quad p_h = \sum_{r=2}^{\infty} p_h^{(r)}. \quad (10.60)$$

Fig. 10.5 Transformation of the coordinates $u_x(\zeta_f), u_y(\zeta_f)$ of the electron at position $u(\zeta_f), h(\zeta_f)$ to the coordinates $\hat{u}(z = \zeta_f), \hat{u}_y(z = \zeta_f)$ of the point of the intersection of the electron trajectory with the final plane $z = \zeta_f$



In addition, we write h_1 as a sum of polynomials of rank r in the expansion parameters:

$$h_1 = h^{(1)} + \sum_{r=2}^{\infty} h_1^{(r)}. \quad (10.61)$$

The comparison of terms with equal rank yields the recurrence relations

$$u^{(r)} = u_\alpha \int_{\tau_0}^{\tau} p_u^{(r)} u_\gamma d\tau - u_\gamma \int_{\tau_0}^{\tau} p_u^{(r)} u_\alpha d\tau, \quad (10.62)$$

$$h^{(r)} = h_1^{(r)} + h_\alpha \int_{\tau_0}^{\tau} p_h^{(r)} h_\gamma d\tau - h_\gamma \int_{\tau_0}^{\tau} p_h^{(r)} h_\alpha d\tau. \quad (10.63)$$

Because the perturbation terms of rank r do not contain path deviations of rank higher than $r-1$, the recurrence relations allow one to successively calculate the path deviations in a systematic way up to arbitrary rank.

In the case of static fields, we can consider the modified time τ as an auxiliary variable. To represent the path deviations in the frame of conventional aberration theory, we must transform the time-dependent representation into the standard form $u = \hat{u}(z)$. Owing to the relation $\zeta = \zeta(\tau)$, the axial position of the reference electron can also serve as the independent variable. However, due to the reversal of flight within the mirror, two different values of τ exist for each given ζ , as illustrated in Figs. 10.3 and 10.4. The solution $u(\tau) = u(\tau(\zeta)) := u(\zeta)$ represents the actual lateral position of the electron when the reference electron is at the axial position ζ . Therefore, $u(\zeta)$ does not represent in general the off-axial position $\hat{u}(z)$ of the particle at the plane $z = \zeta$, as shown in Fig. 10.5. Hence, the lateral deviation $u^{(r)}(\zeta)$ of rank r also differs from the corresponding deviation $\hat{u}^{(r)}(z)$ at the plane $z = \zeta$.

In order to derive the relations existing between these different deviations, we consider that the z -coordinate of the particle is given by the relation

$$z = \zeta + h(u_0, \bar{u}_0, \dot{u}_0, \dot{\bar{u}}_0, \kappa; \zeta) = \zeta + \sum_{r=1}^{\infty} h^{(r)}(\zeta). \quad (10.64)$$

The inverse function $\zeta = \zeta(u_0, \bar{u}_0, \dot{u}_0, \dot{\bar{u}}_0, z)$ is a function of the actual z -coordinate of the particle and of its initial ray parameters. Unfortunately, we cannot solve directly the implicit equation (10.64) for ζ . However, we can express the solution as a series by utilizing the Lagrange inversion formula

$$\begin{aligned} \zeta &= z - \sum_{m=1}^{\infty} \frac{(-1)^m}{(m+1)!} \left. \frac{d^m(h^{m+1})}{d\zeta^m} \right|_{\zeta=z} \\ &= z - h(z) + h(z)h'(z) - h^2(z)h''(z)/2 + h(z)h'^2(z) + \dots \end{aligned} \quad (10.65)$$

For obtaining the deviations $\hat{u}^{(r)}(z)$, we substitute the series (10.65) for ζ in the terms $u^{(r)}(\zeta)$ of the representation

$$\hat{u}(z) = u(\zeta(z)) = \sum_{r=1}^{\infty} u^{(r)}(\zeta(z)) = \sum_{r=1}^{\infty} \hat{u}^{(r)}(z). \quad (10.66)$$

We obtain the polynomials $\hat{u}^{(r)}(z)$ of the second sum by expanding each term of the first sum in a Taylor series at the point $\zeta = z$ and rearranging the resulting terms with respect to their rank. This procedure yields the relations

$$\begin{aligned} \hat{u}^{(1)}(z) &= u^{(1)}(\zeta = z) = u^{(1)}(z), \\ \hat{u}^{(2)}(z) &= u^{(2)}(z) - u'^{(1)}(z)h^{(1)}(z), \\ \hat{u}^{(3)}(z) &= u^{(3)}(z) - u''^{(2)}h^{(1)} - u'^{(1)}h^{(2)} + u'^{(1)}h^{(1)}h'^{(1)} - u''^{(1)}h^{(1)^2}/2. \end{aligned} \quad (10.67)$$

To check the validity of our results, we assume that the observation plane is located in the field-free region in front of the mirror. In this case, $h = h(\zeta)$ is a linear function of ζ . Hence, all higher-order derivatives of $h(\zeta)$ vanish. Considering this behavior, we can perform the summation in the inversion formula (10.65), giving the relation

$$\zeta = z - \sum_{m=0}^{\infty} (-)^m h^{(m)} \left. \frac{d\zeta}{d\zeta} \right|_{\zeta=z} = z - \frac{h(z)}{1 + h'(z)} = z - h(z) \frac{d\zeta}{dz}. \quad (10.68)$$

We obtain the final result by considering the relation $z = \zeta + h(\zeta)$.

In the field-free region the trajectory of a particle is a straight line. It readily follows from Fig. 10.5 that the lateral distances $\hat{u}(z) = \hat{u}_x(z) + i\hat{u}_y(z)$ and $u(\zeta = z) = u_x + iu_y$ are linearly connected with each other via

$$u(\zeta = z) = \hat{u}(z) + \hat{u}'(z)h(\zeta = z). \quad (10.69)$$

This expression must coincide with the relation

$$\widehat{u}_z(z) = u(\zeta(z)) = u(z - h\zeta'). \quad (10.70)$$

In order to demonstrate the identity, we expand the expression on the right-hand side in a Taylor series at the plane $\zeta = z$. Considering that all higher-order derivatives of $u(\zeta)$ vanish in the field-free region, we obtain

$$\widehat{u}(z) = u(\zeta = z) - h \left. \frac{du}{d\zeta} \right|_{\zeta=z} = u(\zeta = z) - \widehat{u}'(z)h(\zeta = z). \quad (10.71)$$

This relation coincides with (10.69).

10.6 Electrostatic Mirror

In practice, one always employs purely electrostatic mirrors. Since most mirrors are incorporated in low-voltage systems, we can neglect relativistic effects ($\varepsilon = 0$). Thus we largely reduce the mathematical expenditure. Moreover, we do not have to distinguish between the rotating u -coordinate system and the fixed w -coordinate system because the Larmor rotation vanishes.

10.6.1 Positional Deviations

By employing the nonrelativistic approximation, the perturbation polynomials $p_u^{(r)} = p_w^{(r)}$ and $p_h^{(r)}$ of second- ($r = 2$) and third-rank ($r = 3$) adopt the simple form

$$\begin{aligned} p_w^{(2)} &= -\frac{\Phi'''}{4\Phi_c} w^{(1)} h^{(1)}, \\ p_h^{(2)} &= -\frac{\Phi'''}{8\Phi_c} w^{(1)} \bar{w}^{(1)} + \frac{\Phi'''}{4\Phi_c} h^{(1)2}, \\ p_w^{(3)} &= \frac{\Phi'''}{32\Phi_c} w^{(1)} (w^{(1)} \bar{w}^{(1)} - 4h^{(1)2}) - \frac{\Phi'''}{4\Phi_c} (w^{(1)} h^{(2)} - w^{(2)} h^{(1)}). \end{aligned} \quad (10.72)$$

We have not listed the polynomial $p_h^{(3)}$ because it does not contribute to the third-rank aberrations. The relation (10.54) for $h^{(1)}$ demonstrates that the second-rank deviation is of entirely chromatic nature:

$$w^{(2)}(\tau) = \frac{\kappa}{8\Phi_c} \left\{ w_\gamma \int_{\tau_0}^{\tau} \Phi''' w^{(1)} h_\alpha w_\alpha d\tau - w_\alpha \int_{\tau_0}^{\tau} \Phi''' w^{(1)} h_\alpha w_\gamma d\tau \right\}. \quad (10.73)$$

The second-rank deviation (10.73) and the third-order geometrical deviation $w^{(3)}$ are of prime concern because they produce at the Gaussian image plane the aberrations, which limit the performance of the instrument. The representations (10.62) and (10.63) for the path deviations of arbitrary rank are valid regardless of the presence of a turning point in the region between the initial plane and the plane of observation. Hence, the time-dependent perturbation method is equivalent to the conventional eikonal and trajectory methods, which substitute the z -coordinate for the time t as independent variable. However, an important difference exists between these procedures and the time-dependent approach with respect to the required number of iteration steps. For example, the latter approach necessitates two iteration steps for obtaining the third-rank deviations, whereas the conventional methods need only one step. Moreover, the conventional aberration theory uses the position w_0 and the complex slope w'_0 of the trajectory at the starting plane as expansion parameters. The complex slope is connected with the starting angle θ_0 and the azimuthal angle ϑ_0 by the relation

$$w'_0 = e^{i\vartheta_0} \tan \theta_0. \quad (10.74)$$

On the other hand, the time-dependent procedure defines the direction of the ray at the starting plane by its normalized lateral velocity component

$$\dot{w}(\tau_0) = e^{i\vartheta_0} \sin \theta_0. \quad (10.75)$$

Therefore, the two methods give different coefficients for the higher-order aberrations. Only the coefficients of the primary aberrations coincide. The expansion with respect to $\sin \theta_0$ instead of $\tan \theta_0$ is a direct consequence of choosing the time τ or the coordinate ζ of the reference electron, respectively, as the independent variable.

10.6.2 Axial Aberrations

Each aberration of rank r is a monomial of the lateral deviation $\hat{w}^{(r)}(z)$ taken at the Gaussian image plane $\zeta_i = z_i$. At this plane, the axial fundamental ray intersects the optic axis ($w_\alpha(\zeta_i) = 0$). The value of the field ray determines the magnification $M = w_\gamma(\zeta_i)$ of the image. The magnification is negative if the object is imaged upside down.

The axial aberrations are formed at the image plane by a pencil of rays emanating from the center $w_0 = 0, \zeta_0 = z_0$ of the initial plane. The most important axial aberrations are the second-rank chromatic aberration and the third-order spherical aberration defined as

$$\begin{aligned}\hat{w}_{\text{ca}}^{(2)}(z_i) &= -w_{\gamma i} \kappa C_c \dot{w}_0, \\ \hat{w}_s^{(3)}(z_i) &= w_{\gamma i} C_s \dot{w}_0^2.\end{aligned}\quad (10.76)$$

We derive the integral expression for the coefficient C_c of the primary axial chromatic aberration by starting from the expression (10.67) for the second-rank deviation $\hat{u}^{(2)}(z) = \hat{w}^{(2)}(z)$ taken at the Gaussian image plane $z = z_i$. Considering further the relations (10.55), (10.73), and $w_\alpha(z_i) = 0$, we obtain

$$\hat{w}_{\text{ca}}^{(2)} = \kappa \dot{w}_0 \left\{ w_{\gamma i} \int_{\tau_0}^{\tau_i} \frac{\Phi'''}{8\Phi_c} w_\alpha^2 h_\alpha d\tau - \frac{1}{2} w'_{\alpha i} h_{\alpha i} \right\}. \quad (10.77)$$

We multiply the second term in the parenthesis by the Wronskian $w_{\gamma i} \dot{w}_{\alpha i} = 1$. By equalizing the result with the corresponding expression (10.76), we find

$$C_c = \frac{1}{2} w'_{\alpha i} \dot{w}_{\alpha i} h_{\alpha i} - \frac{1}{8} \int_{\tau_0}^{\tau_i} \frac{\Phi'''}{\Phi_c} w_\alpha^2 h_\alpha d\tau. \quad (10.78)$$

For deriving the aberration coefficient $C_s = C_3$ of the third-order spherical aberration, we must perform two iteration steps because the third-rank perturbation function $p_w^{(3)}$ is a function of the paraxial deviations $w^{(1)}$, $h^{(1)}$ and of the second-rank deviations $w^{(2)}$ and $h^{(2)}$. For an axial trajectory with nominal energy ($w_0 = 0, \kappa = 0$), the fundamental longitudinal deviation $h^{(1)}$ and the lateral deviation $w^{(2)}$ are zero. In this case, the axial deviation $h^{(2)}$ adopts the form

$$\begin{aligned}h^{(2)}(\tau) &= \dot{w}_0 \dot{w}_0 h_{\alpha\bar{\alpha}} = \sin^2 \theta_0 h_{\alpha\bar{\alpha}}, \\ h_{\alpha\bar{\alpha}}(\tau) &= -\frac{1}{2} h_\alpha + \frac{1}{8\Phi_c} \left\{ h_\gamma \int_{\tau_0}^{\tau} \Phi''' w_\alpha^2 h_\alpha d\tau - h_\alpha \int_{\tau_0}^{\tau} \Phi''' w_\alpha^2 h_\gamma d\tau \right\}.\end{aligned}\quad (10.79)$$

The function $h_{\alpha\bar{\alpha}}(\tau)$ represents the secondary fundamental longitudinal deviation. The expression in the parenthesis does not change if we substitute h_γ for h_γ and h_σ for h_α because the Wronskian is the same for each pair of deviations. Moreover, it follows from (10.55) and (10.67) that in the case $\kappa = 0$, the third-rank lateral deviation $\hat{u}^{(3)}(z) = \hat{w}^{(3)}(z)$ simplifies considerably as

$$\hat{w}^{(3)}(z) = w^{(3)}(\zeta = z) - w'^{(1)}(\zeta = z) h^{(2)}(\zeta = z). \quad (10.80)$$

We find the deviation $w^{(3)}(\zeta(\tau)) = w^{(3)}(\tau)$ of the axial ray at time $\tau = \tau_i$ from the relations (10.62), (10.72), and $w_\alpha(\tau_i) = w_{\alpha i} = 0$ as

$$w^{(3)}(\tau_i) = \frac{w_{\gamma i} \dot{w}_0^2 \ddot{w}_0}{32\Phi_c} \int_{\tau_0}^{\tau_i} (8\Phi''' h_{\alpha\bar{\alpha}} - \Phi'''' w_\alpha^2) w_\alpha^2 d\tau. \quad (10.81)$$

We substitute this expression for $w^{(3)}$ and the relation (10.79) for $h^{(2)}$ in (10.80). By comparing the result with the definition (10.76) for the coefficient of the spherical aberration, we eventually obtain the formula

$$C_s = \frac{1}{32\Phi_c} \int_{\tau_0}^{\tau_i} (8\Phi''' h_{\alpha\bar{\alpha}} - \Phi'''' w_\alpha^2) w_\alpha^2 d\tau - w'_{\alpha i} \dot{w}_{\alpha i} h_{\alpha\bar{\alpha}}(\tau_i). \quad (10.82)$$

For transforming the integral expressions of the aberration coefficients, we shall repeatedly utilize the relation

$$\begin{aligned} \int_{\tau_0}^{\tau} \Phi''' w_\alpha^2 h_v d\tau &= \int_{\zeta_0}^{\zeta} \Phi''' w_\alpha^2 d\zeta = \Phi'' w_\alpha^2 - 2 \int_{\zeta_0}^{\zeta} \Phi'' w'_\alpha w_\alpha d\zeta \\ &= \Phi'' w_\alpha^2 + 8\Phi_c \int_{\tau_0}^{\tau} \ddot{w}_\alpha \dot{w}_\alpha d\tau = \Phi'' w_\alpha^2 + 4\Phi w_\alpha'^2 - 4\Phi_c. \end{aligned} \quad (10.83)$$

We obtain the second integral by substituting $\dot{\zeta}$ for h_v in the integrand of the first integral. To verify the final result, we evaluate the second integral by partial integration with respect to Φ''' and by substituting $4\Phi_c \ddot{w}_\alpha$ for $\Phi'' w_\alpha$ in the remaining integral using the paraxial equation (10.31). By employing the relation $w'_\alpha = \dot{w}_\alpha d\tau/d\zeta$, the resulting integrand becomes a total differential with respect to the integration variable τ .

In order to minimize the number of off-axial aberrations, it is advantageous to place the initial plane ζ_0 at the plane ζ_c of the center of curvature of the mirror. In this case, the reflected rays form an image with unit magnification at the location $\zeta_i = \zeta_c = \zeta_0$ of the initial plane. Moreover, we split up each integrand of the aberration coefficients (10.78) and (10.82) into a symmetric part and an antisymmetric part with respect to the turning time. The contribution of the antisymmetric part cancels out. The integral over the symmetric term is twice the integral taken between the object plane $\zeta_0 = \zeta_c$ and the turning plane ζ_T . We perform the separation by substituting the linear combination (10.50) for h_α the integrand of the aberration integral for the chromatic coefficient (10.78). By utilizing the expression (10.50), we find

$$h_\alpha(\zeta_i = \zeta_c) = h_{\alpha i} = 2h_\sigma(\zeta_0 = \zeta_c) = 2h_{\sigma 0}. \quad (10.84)$$

In the remaining symmetric integral, we can directly replace the integration variable τ by ζ without the need to distinguish between the incident and reflected path of the electrons.

To investigate the structure of the aberration coefficients C_c and C_s and their dependence on the properties of the electrostatic field, it is advantageous to perform the integrals in (10.78) and (10.82) with respect to ζ . Considering the symmetry properties of the variables with respect to the turning plane ζ_T and the relation (10.84), we can rewrite the coefficient of the axial chromatic aberration (10.78) in the form

$$\begin{aligned} C_c &= -h_{\sigma C} - \frac{1}{4} \int_{\tau_0}^{\tau_T} \frac{\Phi'''}{\Phi_c} w_\alpha^2 h_\sigma d\tau \\ &= -\Phi_c^{3/2} \int_{\zeta_\sigma}^{\zeta_C} \frac{d\zeta}{\Phi^{3/2}} - \frac{1}{4} \int_{\zeta_T}^{\zeta_C} \frac{\Phi'''}{\Phi_c} w_\alpha^2 h_\sigma d\tau. \end{aligned} \quad (10.85)$$

The symmetric fundamental deviation $h_\sigma(\zeta)$ is negative for $\zeta < \zeta_\sigma$ and positive for $\zeta > \zeta_\sigma$. The center of curvature of the diode mirror shown in Fig. 10.2 is located in front of the plane ζ_σ , as illustrated in Fig. 10.4. Accordingly, $h_{\sigma C}$ is negative for this mirror. Therefore, Φ''' must be negative in the region $\zeta_T \leq \zeta \leq \zeta_C$ in order that the coefficient of the chromatic aberration can be negative. A negative Φ''' in this region implies that the axial curvature of the equipotentials decreases with increasing distance from the mirror electrode, as depicted in Fig. 10.2 for the diode mirror. The negative value $C_c = -0.187461r$ of the chromatic aberration coefficient of this mirror proves our considerations.

To obtain an insight into the structure of C_s for a mirror operating in the symmetric imaging mode, we aim for a formula which corresponds to the representation (10.84) for the coefficient of the axial chromatic aberration. For this purpose, we substitute the relation (10.79) for $h_{\alpha\bar{\alpha}}$ in the expression (10.82). Subsequently, we transform parts of the integral by partial integration utilizing the relation (10.83). As the result of the rather lengthy calculation, we eventually obtain the following integral expression for the coefficient of the third-order spherical aberration of the mirror:

$$\begin{aligned} C_s &= -h_{\sigma 0} - \frac{1}{32} \int_{\tau_0}^{\tau_1} \left\{ \frac{\Phi''''}{\Phi_0} + 2h_\sigma \frac{\Phi'''}{\Phi_0} \left(\frac{\Phi''}{\Phi_0} + 4 \frac{w_\alpha'^2}{w_\alpha^2} \right) \right\} w_\alpha^4 d\tau \\ &= \int_{\zeta_0}^{\zeta_\sigma} \left(\frac{\Phi_0}{\Phi} \right)^{3/2} d\zeta - \frac{\Phi_0^{1/2}}{16} \int_{\zeta_T}^{\zeta_0} \left\{ \frac{\Phi''''}{\Phi^{1/2}} - 2\Phi''' \left(\Phi'' + 4 \frac{w_\epsilon'^2}{w_\alpha^2} \right) \int_{\zeta}^{\zeta_\sigma} \frac{d\tilde{\zeta}}{\Phi^{3/2}} \right\} w_\alpha^4 dz. \end{aligned} \quad (10.86)$$

The structure of the integrands demonstrates that Φ''' should be made positive especially in the region near the turning point where the electric potential is small. The chromatic correction also introduces a negative contribution to the spherical aberration because a focusing mirror with negative axial chromatic aberration must have a positive Φ'' and a negative Φ''' in front of the mirror. These conditions are fulfilled for the diode mirror shown in Fig. 10.2. Accordingly, the coefficient of the spherical aberration of this mirror, $C_s = -0.61629r$, is negative. This behavior does not imply that the curvature of the equipotential surface $\varphi = 0$ determines the properties of the electron mirror, as it is the case for the physical surface of a light mirror.

Unlike a light-optical mirror, where the reflection occurs at the physical surface, the electron optical mirror consists is a “soft” mirror. For such a mirror, the total reflection results from consecutive refractions on a continuous set of equipotentials. The electrons stay a relatively long time in the vicinity of the turning point due to their small axial velocities. Accordingly, the electric potential strongly affects the course of the electrons in this region. Our considerations reveal that the equipotentials of a focusing mirror with negative chromatic and spherical aberration must be concave in the paraxial region and convex in the marginal region viewed in the direction toward the mirror electrode. The correction properties of such a mirror are illustrated schematically in Fig. 10.6. The electrons with energy $E > E_n$ are faster than the electrons with nominal energy. Hence, these electrons penetrate deeper into the mirror and are reflected more strongly because the curvature of the equipotentials increases in the direction toward the mirror indicating that Φ''' is negative, as depicted in Fig. 10.6a. Figure 10.6b illustrates the diverging effect of the convex region of curvature of the equipotentials on the marginal rays. Because the marginal region of the mirror focuses the electrons less than the inner region, the spherical aberration is negative.

The potential Φ_m of the dipole mirror determines the focal length and the aberrations. In order to adjust the focal length, the chromatic aberration, and the spherical aberration independently, we need three free parameters.

By increasing the number of electrodes, which we can put at arbitrary potentials, we provide the proper variability. Figure 10.7 shows an arrangement that consists of four electrodes. Since the potentials Φ_m , Φ_1 , and Φ_2 applied to the electrodes determine the spatial distribution of the potential $\varphi = \varphi(\rho, z)$, it is possible to properly adjust the focal length, the chromatic aberration, and the spherical aberration of the tetrode mirror.

We calculated numerically the properties of the tetrode mirror shown in Fig. 10.7 as functions of the adjustable potentials Φ_m , Φ_1 , and Φ_2 for unit magnification. In this case, the Gaussian object and image planes coincide. They are placed at the center of curvature ξ_C of the mirror located at the distance $\xi_C - \xi_m \approx 21$ mm in front of the mirror electrode. The results demonstrate that negative coefficients for the axial chromatic aberration and the spherical aberration can be adjusted within a wide range for a fixed position of the center of curvature. The adjustable

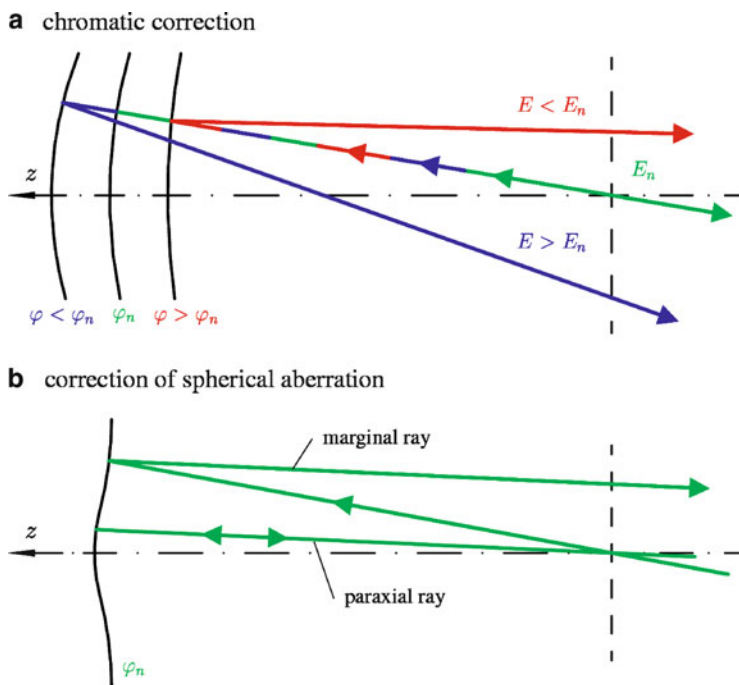
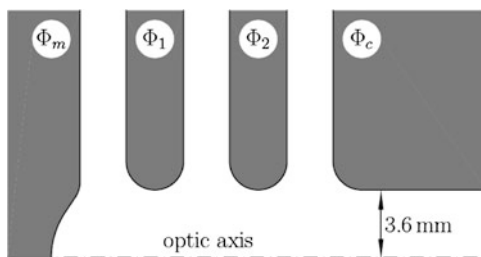


Fig. 10.6 Path of rays illustrating schematically the formation of (a) negative chromatic aberration and (b) negative spherical aberration. The velocity of the nominal electron with energy $E = E_n$ in front of the mirror is zero at the equipotential $\phi_n = 0$

Fig. 10.7 Sectional view of the tetrode mirror; the variable voltages Φ_m , Φ_1 , and Φ_2 enable the adjustment of focal length, chromatic aberration, and spherical aberration



range is sufficiently large to enable correction of the corresponding aberrations of rotationally symmetric lenses for various modes of operation.

The feasible incorporation of an electron mirror into an electron microscope necessitates a beam separator, as depicted schematically in Fig. 10.8. The separator must be placed near the first intermediate image of the objective lens. The magnification should be larger than about 10 to guarantee that the third-order aperture aberrations of the beam separator are negligibly small. Moreover, to significantly increase the resolution and/or the angle of acceptance, the beam separator must be free of dispersion and of all second-order aberrations. The magnetic separator outlined in Sect. 9.2.1 satisfies these conditions.

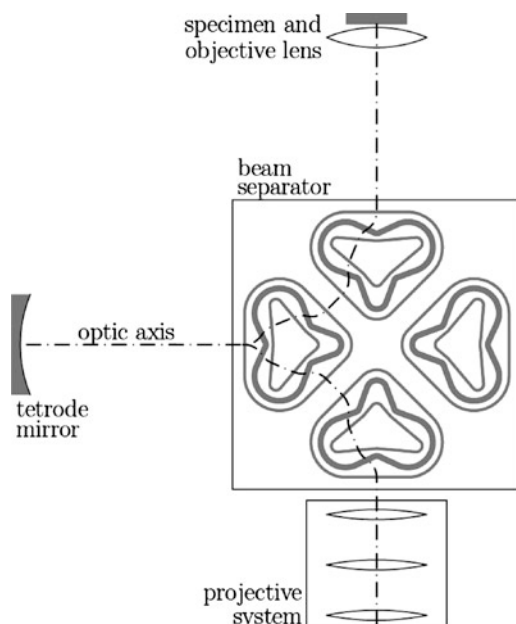


Fig. 10.8 Mirror-corrected objective lens. The tetrode mirror is implemented via a dispersion-free magnetic beam separator. The *thin shaded regions* indicate the induction coils placed at the surface of the pole plates

To precisely eliminate the primary aberrations of this device, the exact evaluation of the magnetic field is necessary. We have solved this intricate problem by means of a special charge simulation method. The calculations showed that on the premises of a precise adjustment, the corrector should improve the resolution up to a factor of 10. The corresponding increase of the angle of acceptance enables one to utilize 100 times more scattered or emitted electrons than without correction. This improvement has been demonstrated experimentally by the mirror-corrected system realized within the frame of the SMART project [136]. The SMART microscope operates as a low-energy electron microscope (LEEM) using reflected or secondary electrons for the image formation or as a photoemission electron microscope (PEEM) using electrons emitted by synchrotron x -rays. The construction of the SMART microscope is shown in Fig. 10.9. Its main components are the electron source, the electric-magnetic immersion objective lens, the beam separator, the mirror corrector, the round-lens transfer system, the imaging energy filter, the projector system, and the camera. The sample can be illuminated either with electrons through the beam separator or with photons from the side.

An important application of mirror correctors is the low-energy region with electron energies below 15 keV because in this case the technological expense for insulation and for voltage supplies for the mirror electrodes is small. The mirror can correct the aberrations of direct imaging low-energy electron microscopes (LEEM)

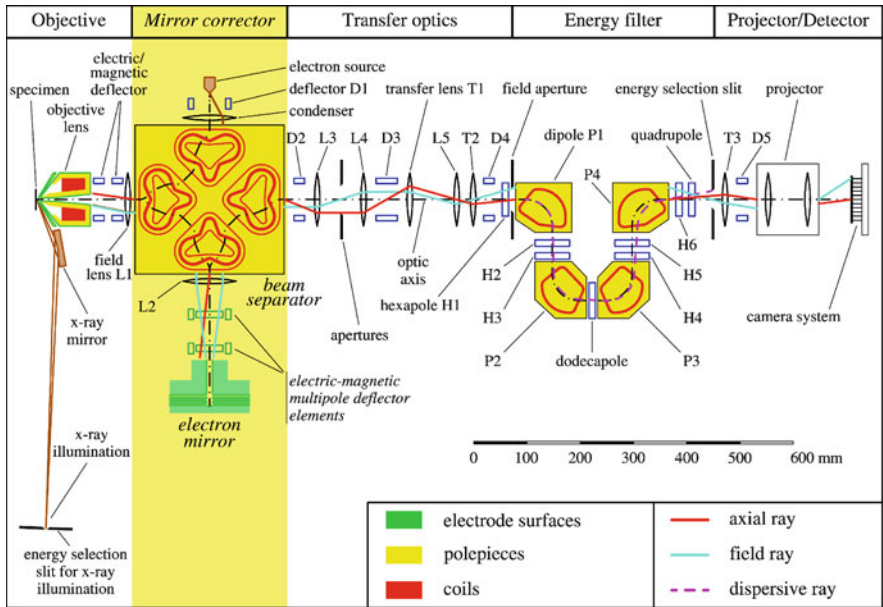


Fig. 10.9 Construction scheme of the SMART electron microscope

as well as those of scanning electron microscopes (SEM). For compensating the chromatic and spherical aberration of a SEM, we must place the mirror in the ray path before the probe-forming objective lens. This means that the electron bundle emitted from the source after a single passage through the beam separator is thereby precompensated at the mirror, so after a further passage through the beam separator, the objective lens will form a smaller probe.

In the LEEM mode, the electrons reflected or emitted at the sample will be imaged by the objective lens with aberrations. They are eliminated by the mirror after a single passage of the electrons through the beam separator. The chromatic and spherical aberration-free image is transferred by the reflected electrons through the beam separator, subsequently magnified by the projector, and imaged onto the detector.

Chapter 11

Optics of Electron Guns

Abstract The calculation procedure developed for electron mirrors is employed for determining the optical properties of field emission guns and cathode lenses. This approach avoids the separation of the system in an accelerating regime and a focusing region used in standard methods. The advantage of the new method is shown by deriving an analytic expression for the spherical aberration at the crossover.

Electron guns are important special cases of systems with large ray gradients. Most electron guns consist of a cathode, a Wehnelt electrode, and an anode. The latter electrode is at positive potential with respect to the cathode. The Wehnelt electrode is held at negative potential, which defines the spatial distribution of the zero-volt equipotential and hence the size of the emitting area of the cathode. Raising the potential of the Wehnelt enlarges the emission area, while a larger negative potential reduces it. The negative Wehnelt potential has a strong focusing effect on the emitted electrons and guarantees that they pass through the hole of the anode electrode. By varying the Wehnelt potential, we can alter the intensity of the emitted beam without changing the anode potential or the cathode temperature.

The shape of the cathode surface largely affects the properties of the electron gun because the curvature of the emitting tip determines the strength of the electric field. The weak electric field of a flat cathode surface cannot immediately remove the thermally emitted electrons resulting in the buildup of an electron cloud. The negative space charge reduces the emission current and broadens the energy width of the emitted electrons. This so-called Boersch effect [111] results from stochastic Coulomb interactions between electrons at regions of high current density within the beam [112]. Systems for imaging surface layers with photoemission electrons (PEEM) or with low-energy emitted electrons also involve large ray gradients. Low-energy electron microscopes (LEEM) use either reflected or secondary electrons for image formation. Accordingly, we can treat the optics of these system like that of cathodes.

11.1 Field Emission Guns

In cold field emitters, the electrons escape through the potential barrier in front of the cathode surface by quantum mechanical tunneling. This emission requires a very high electric field of about 10^8 V/cm. In order to achieve such high field strength, the cathode forms a tip with a very small radius of curvature at the apex. Unfortunately, the emitting area of cold field emitters is not very stable causing a variation in brightness and current. To avoid this drawback, thermal field emitters are largely employed. These so-called Schottky field emitters assist and stabilize the electron emission by heating the cathode. In the absence of space charge effects, we can regard the gun as an accelerating lens system focusing the emitted electrons. Each electron emanates from a given point of the emitting area with specific velocity and direction of flight.

The Wehnelt electrode of a field emitter is always positive with respect to the cathode. Therefore, the electric field strength never vanishes at the surface of the cathode tip. Because the electric field strength is strongest at the emitting apex, the buildup of space charge is prevented. We define the image of the effective source as the smallest waist of the beam formed by all electrons originating from the curved surface of the emitting tip. In order to determine the trajectories of the electrons, we employ the time-dependent procedure developed for electron mirrors and assume that the tip of the cathode has rotational symmetry. Moreover, we consider the surface of the tip as the curved object surface whose apex is located at position $z = \zeta_0$ on the optic axis. In this case, the initial lateral deviation has the form

$$h_0(\tau_0) = h(w_0, \bar{w}_0) = z_0 - \zeta_0 = \sum_{m=1}^{\infty} h_0^{(2m)},$$

$$h_0^{(2m)} = \frac{\Gamma_{2m}}{2} (w_0 \bar{w}_0)^{2m}. \quad (11.1)$$

For needle emitters, the curvature of the apex of the tip $\Gamma_2 = -1/\rho_t$ is negative, and its absolute value represents the inverse of the radius of curvature. The remaining higher-order ($m \geq 2$) curvatures Γ_{2m} describe the deviation of the tip surface from the parabolic shape. The curvature Γ_2 is positive for concave cathode surfaces whose center of curvature is located in the region toward the anode, as it is the case for Pierce guns [166]. For a spherical surface, we find

$$\Gamma_{2m} = \pm \frac{(2m-3)!}{m!(m-2)!} \frac{2}{(2\rho_t)^{2m-1}}. \quad (11.2)$$

The negative sign must be attributed to a pointed cathode. Extensive theoretical studies on cathodes without space charge have been performed using different models [167, 168]. We shall use a different approach based on the theory of electron mirrors outlined in Chap. 10.

The trajectories start perpendicular to the tip surface along the electric field lines for monochromatic electrons with starting velocity $v_0 = 0$. The asymptotes, taken at the cathode surface, form a virtual disc of least confusion, which represents the *effective source*. For a spherical cathode tip, the effective source is a point located at the center of the sphere. However, because monochromatic emission does never exist, the size of the effective source of any field emission gun is finite and primarily determined by the energy spread of the emitted electron beam. The reason for this behavior is due to the fact that the trajectories of electrons with starting velocities $v_0 \neq 0$ may start in any direction with respect to that of the electric field.

The time-dependent formalism is also well suited for calculating the optical properties of electron guns if we can disregard space charge effects. Moreover, we can neglect relativistic effects because the electrons velocity is small compared to the velocity of light in the region of the cathode. In order to minimize the aberrations of field emission guns, compound systems have been proposed consisting of an electric extraction field and a focusing magnetic field. In the case of electron guns, it is advantageous to define the differential modified time as

$$d\tau = \sqrt{\frac{e\Phi_0'^2}{2m_e\Phi_a}} dt. \quad (11.3)$$

The normalization (11.3) makes the modified time τ dimensionless. We formally obtain this relation from the expression (10.4) by substituting $\Phi_0'^2/4\Phi_a$ for the column potential Φ_c and putting $E_0 = \infty$; Φ_a is the potential of the anode. As a consequence of the redefinition (11.3) of the modified time, we must also perform the substitution in the path equations listed in Chap. 10.

In order to obtain the initial ray parameters, we assume that the electron emanates from the surface of the cathode with the initial energy

$$\delta E_c = \frac{1}{2} m_e v_0^2. \quad (11.4)$$

The starting velocity v_0 varies statistically according to the distribution function. In the case of thermal electron emission, we obtain the Maxwell distribution function. By employing the relations (11.3) and (11.4), the nonrelativistic conservation of energy adopts the form

$$\dot{w}\dot{w} + (\dot{\zeta} + \dot{h})^2 = 4 \frac{\Phi_a \varphi}{\Phi_0'^2} + 2 \frac{m_e v_0^2 \Phi_a}{e \Phi_0'^2}. \quad (11.5)$$

This relation simplifies for the reference electron ($w = \dot{w} = 0, v_0 = 0, h = 0, \phi = \Phi$) to

$$\dot{\zeta} = 2 \frac{\sqrt{\Phi \Phi_a}}{\Phi_0'}. \quad (11.6)$$

Because the potential is zero at the starting plane, we cannot consider the starting energy of an electron to be small compared with the average energy of the beam at this plane. As a consequence, the ray parameters of the paraxial electrons will depend on the initial velocity of the electron.

11.2 Gaussian Optics

By choosing the redefined modified time (11.3) as the independent variable, the paraxial path equations (10.30) adopt the modified form

$$\ddot{u}^{(1)} + \frac{\Phi_a \Phi''}{\Phi_0'^2} u^{(1)} + \frac{e \Phi_a B^2}{2m_e \Phi_0'^2} u^{(1)} = 0, \quad (11.7)$$

$$\ddot{h}^{(1)} - 2 \frac{\Phi_a \Phi''}{\Phi_0'^2} h^{(1)} = 0. \quad (11.8)$$

The coefficients of these linear differential equations are dimensionless contrary to those of the corresponding equations (10.30) and (10.31) for the mirror. This difference results from the redefinition (11.3) of the modified time τ . We normalize the fundamental deviations h_α and h_γ in such a way that they satisfy the Wronskian

$$h_\gamma \dot{h}_\alpha - \dot{h}_\gamma h_\alpha = 1. \quad (11.9)$$

Moreover, we postulate that the fundamental deviations are fixed at the plane $z = \zeta_0$ by the initial values

$$\begin{aligned} h_\alpha(\zeta_0) &= \dot{h}_\gamma(\zeta_0) = 0, \\ \dot{h}_\alpha(\zeta_0) &= h_\gamma(\zeta_0) = 1. \end{aligned} \quad (11.10)$$

Considering these constraints, we eventually obtain for h_α and h_γ the expressions

$$h_\alpha(\tau) = \sqrt{\frac{\Phi}{\Phi_a}} = \frac{\Phi_0'}{2\Phi_a} \dot{\zeta}, \quad h_\gamma(\tau) = \sqrt{\frac{\Phi}{\Phi_a}} \int_{\tau_\gamma}^{\tau} \frac{\Phi_a}{2\Phi} d\tau. \quad (11.11)$$

The fundamental deviation h_γ corresponds to the symmetric fundamental deviation h_σ of the mirror. Therefore, we must determine the lower integration limit τ_γ such that $\dot{h}_\gamma(\tau_0) = 0$ as outlined in Chap. 10.

The axial component v_{0z} and the complex lateral component $v_{0w} = v_{0x} + i v_{0y}$ of the initial velocity are given by

$$\begin{aligned} v_{0z} &= v_0 \cos \theta_0, \\ v_{0w} &= v_0 e^{i\vartheta_0} \sin \theta_0. \end{aligned} \quad (11.12)$$

Considering these relations, we obtain the initial components $\dot{w}(\tau_0) = \dot{w}_0$ and $\dot{h}_0 = \dot{h}_0^{(1)}$ from the conservation of energy as

$$\dot{w}_0 T = \eta e^{i\vartheta_0} \sin \theta_0, \quad \dot{h}_0 = \eta \cos \theta_0, \quad (11.13)$$

$$\eta^2 = 2 \frac{m_e v_0^2 \Phi_a}{e \Phi_0^2}. \quad (11.14)$$

The parameter η is the smaller the larger the electric field strength is at the cathode surface. We can consider this parameter as a characteristic length which relates the waist of the beam to the geometrical parameters of the system. The relations (11.1), (11.2), (11.13), and (11.14) demonstrate that h_0 is only a function of the initial lateral position $\bar{w}_0$, whereas $\dot{h}_0$ depends only on the initial velocity and the emission angle θ_0 . Because this angle is in the range between 0 and $\pi/2$, it cannot be considered a small expansion parameter, as in the case of electron lenses. Since h_0 is at least of order two in w_0 and $\bar{w}_0$, this initial longitudinal deviation does not contribute to the paraxial approximation

$$h^{(1)} = \dot{h}_0 h_\alpha = \eta \cos \theta_0 h_\alpha. \quad (11.15)$$

Hence, for monochromatic electrons with starting velocity $v_0 = 0$, the axial position ζ of the reference electron represents within the frame of validity of Gaussian optics the z -coordinate of all other electrons. This behavior becomes obvious if we consider that in paraxial approximation the curved cathode surface is replaced by the tangential plane at the apex.

The lateral paraxial path equation (11.7) has two linearly independent real solutions. We choose as fundamental solutions the rays $u_\alpha(\tau)$ and $u_\gamma(\tau)$ which satisfy the initial conditions

$$\begin{aligned} u_\alpha(\tau_0) &= \dot{u}_\gamma(\tau_0) = 0, \\ \dot{u}_\alpha(\tau_0) &= u_\gamma(\tau_0) = 0. \end{aligned} \quad (11.16)$$

Each trajectory is defined by its initial position $u_0 = w_0$ and its slope

$$\dot{u}_0 = \dot{w}_0 + i \dot{\chi}_0 u_0, \quad \dot{\chi}_0 = \sqrt{\frac{e \Phi_a}{2 m_e}} \frac{B_0}{\Phi_0'}. \quad (11.17)$$

By imposing these initial constraints, we obtain the complex lateral component of the trajectory in paraxial approximation in the standard form

$$u^{(1)} = \dot{u}_0 u_\alpha + u_0 u_\gamma. \quad (11.18)$$

The point of intersection $\zeta = \zeta_c$ of the field ray u_γ with the optic axis ($u_\gamma(\zeta_c) = 0$) defines the location of the crossover, which is the image of the effective source. The zero $\zeta = \zeta_i$ of the axial ray u_α determines the Gaussian image of the tangential plane

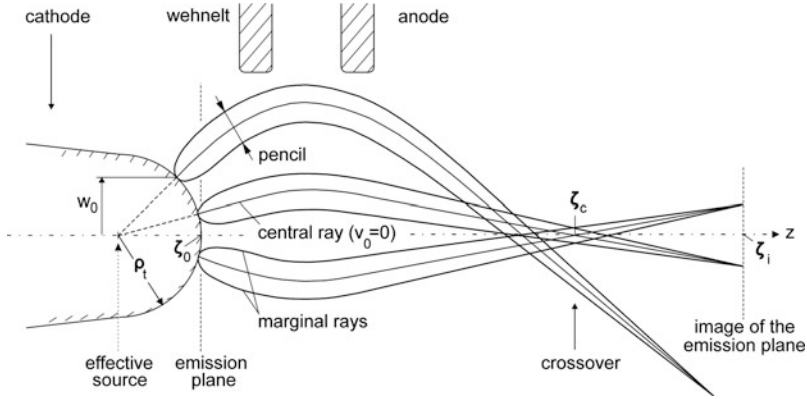


Fig. 11.1 Schematic arrangement of an electron gun and trajectories in the absence of space charge; the marginal rays start tangentially to the cathode surface with maximum initial velocity v_0

$\zeta = \zeta_0$ placed at the apex of the cathode. In paraxial approximation, the crossover forms a round spot with radius

$$\rho_c^{(1)} = u_{ac} |\dot{u}_{0\max}| = u_{ac} \frac{\Phi_a}{\Phi'_0} \sqrt{2 \frac{m_e v_{0\max}^2}{e \Phi_a} + \frac{e B_0^2 \rho_0^2}{2 m_e \Phi_a}}. \quad (11.19)$$

Here, $\rho_0 = |u_{0\max}|$ defines the radius of the emitting area of the cathode. In the absence of a magnetic field at the cathode ($B_0 = 0$), the radius of the disc is determined by the trajectories of the electrons which start with maximum emission velocity $v_{0\max}$ tangential to the cathode surface ($\theta_0 = \pi/2$), as depicted schematically in Fig. 11.1. In order to obtain a small crossover, the magnetic field must be zero at the tip and the electric field strength as large as possible. The effect of the magnetic field on the size of the crossover results from the conservation of the canonical momentum forming a bundle of skew rays.

11.3 Aberrations

Our calculation procedure for determining the optics of electron guns without space charge does not require the separation of the system in an accelerating regime and a focusing region because the time-dependent formalism allows us to treat the system as a whole. Therefore, our procedure yields the aberrations at the crossover plane with a much higher accuracy than the separation method. In order to facilitate the analytical calculations, we only consider purely electrostatic electron guns. This restriction allows us to directly utilize the calculations for electric mirrors outlined in Chap. 10.

11.3.1 Second-Rank Deviations

By considering the redefinition of the modified time, we derive from the relations (10.61) and (10.71) for the second-rank ($r = 2$) lateral path deviation the expression

$$\begin{aligned}
 u^{(2)} = w^{(2)} &= \dot{h}_0 \frac{2\Phi_a}{\Phi_0'^2} \left\{ w_\gamma \int_{\tau_0}^{\tau} \Phi''' w^{(1)} h_\alpha w_\alpha d\tau - w_\alpha \int_{\tau_0}^{\tau} \Phi''' w^{(1)} h_\alpha w_\gamma d\tau \right\} \\
 &= 2 \frac{\dot{h}_0 \Phi}{\Phi_0'} w^{(1)} \left(w_\gamma w'_\alpha - w_\alpha w'_\gamma \right) - \frac{\dot{h}_0 \Phi_0'}{2\Phi_a} \dot{w}_0^{(1)} w_\gamma + \dot{h}_0 w_0 \frac{\Phi_0''}{\Phi_0'} w_\alpha \\
 &= \dot{h}_0 (\dot{w}^{(1)} - w_\gamma \dot{w}_0) \frac{\Phi_0'}{2\Phi_a} + \dot{h}_0 w_0 \frac{\Phi_0''}{\Phi_0'} w_\alpha. \tag{11.20}
 \end{aligned}$$

Substituting this relation for $u^{(2)}$ in (10.66), we find

$$\begin{aligned}
 \hat{u}^{(2)}(z) &= -w_\gamma \frac{\Phi_0'}{2\Phi_a} \dot{h}_0 \dot{w}_0 + w_\alpha \frac{\Phi_0''}{\Phi_0'} \dot{h}_0 w_0 \\
 &= -w_\gamma \frac{m_e v_0^2}{2e \Phi_0'} e^{i\vartheta_0} \sin 2\theta_0 + w_\alpha \sqrt{\frac{2m_e \Phi_a}{e}} \frac{\Phi_0''}{\Phi_0'^2} w_0 v_0 \cos \theta_0. \tag{11.21}
 \end{aligned}$$

The expressions on the right-hand side reveal that the lateral second-rank deviation consists of a chromatic term and a mixed term. The chromatic term is proportional to the initial kinetic energy $m_e v_0^2/2$, whereas the mixed term depends bilinear on the initial velocity v_0 and the lateral position w_0 of the emitted electron. Note that such a deviation does not show up in the standard calculus of systems for which the energy at the object plane is large compared with the energy of the emitted electrons. The first term yields the axial chromatic aberration at the image of the emission plane. Since the axial fundamental ray w_α intersects the optic axis ($w_{\alpha i} = 0$) at the image plane ζ_i , the second term does not contribute to the aberrations at this plane. However, this term produces the second-rank aberration at the crossover plane ζ_c where the fundamental field ray w_γ is zero.

Using the standard representation of the aberrations, we can write the chromatic aberration at the image of the emission plane as

$$\hat{w}_c^{(2)}(\zeta_i) = -w_{\gamma i} \kappa \omega C_c. \tag{11.22}$$

In this representation, the chromatic parameter κ , the angular aperture parameter ω , and the coefficient of the axial chromatic aberration are defined as

$$\kappa = \frac{m_e v_0^2}{2eU_a} = \frac{\delta E}{E_a}, \quad \omega = \frac{e^{i\vartheta_0}}{2} \sin 2\theta_0, \quad C_c = \frac{2U_a}{\Phi'_0} > 0. \quad (11.23)$$

The relation for the angular parameter ω differs from the standard definition which we obtain in the small angle limit $\theta_0 \ll 1$. In particular, ω is zero for the marginal rays which start perpendicular ($\theta_0 = \pi/2$) to the optic axis, as depicted in Fig. 11.1. Although the coefficient of the chromatic aberration depends on the anode voltage U_a , the chromatic aberration does not because U_a cancels out in the product κC_c . Accordingly, we can minimize the chromatic aberration only by choosing the electric field strength at the emission plane as large as possible. In order to nullify this aberration, we must either depart from rotational symmetry or introduce a mirror, which has been realized for an LEEM/PEEM within the frame of the SMART project [136].

The second-rank chromatic deviation vanishes at the crossover plane ζ_c if $\Phi''_0 = 0$. Because the surface of the cathode tip represents an equipotential, the curvature of the equipotential $\varphi_0 = \varphi(w_0 \bar{w}_0, z_0) = 0$ at the position $w_0 = 0$ is related with the curvature Γ_2 of the apex via

$$\Gamma_2 = \frac{\Phi''_0}{2\Phi_0}. \quad (11.24)$$

To determine the third-rank deviations, we must know the second-rank lateral deviation (11.20) and the second-rank longitudinal deviation:

$$\begin{aligned} h^{(2)} = & \frac{\Gamma_2}{2} w_0 \bar{w}_0 h_\gamma + h_\alpha \int_{\tau_0}^{\tau} \left\{ \frac{\Phi \Phi'''}{\Phi_0'^2} \dot{h}_0^2 - \frac{\Phi_a \Phi'''}{2\Phi_0'^2} w^{(1)} \bar{w}^{(1)} \right\} h_\gamma d\tau \\ & - h_\gamma \int_{\tau_0}^{\tau} \left\{ \frac{\Phi \Phi'''}{\Phi_0'^2} \dot{h}_0^2 - \frac{\Phi_a \Phi'''}{2\Phi_0'^2} w^{(1)} \bar{w}^{(1)} \right\} h_\alpha d\tau. \end{aligned} \quad (11.25)$$

We eventually derive this expression by starting from the relation (10.62), considering the redefinition of the modified time (11.3), and by employing the relations (10.71), (11.2), and (11.11). The integration of the first integral on the right-hand side of (11.25) can be performed analytically. Considering in addition the relation (11.22), we obtain

$$\begin{aligned} h^{(2)} = & h_\alpha \int_{\tau_0}^{\tau} \left\{ \frac{\Phi \Phi'''}{\Phi_0'^2} \dot{h}_0^2 - \frac{\Phi_a \Phi'''}{2\Phi_0'^2} w^{(1)} \bar{w}^{(1)} \right\} h_\gamma d\tau \\ & - h_\gamma \left\{ \frac{\dot{h}_0^2}{4} \left[2 \frac{\Phi \Phi''}{\Phi_0'^2} - \frac{\Phi'^2 - \Phi_0'^2}{\Phi_a \Phi_0'} \right] - \frac{\Phi''}{4\Phi_0'} w^{(1)} \bar{w}^{(1)} - \frac{\Phi'_0}{4\Phi_a} (\dot{w}^{(1)} \dot{\bar{w}}^{(1)} - \dot{w}_0 \dot{\bar{w}}_0) \right\}. \end{aligned} \quad (11.26)$$

This expression consists of four terms, which we write as

$$h^{(2)} = w_0 \bar{w}_0 h_{\gamma\gamma} + \text{Re}(w_0 \dot{\bar{w}}_0) h_{\gamma\alpha} + \dot{w}_0 \dot{\bar{w}}_0 h_{\alpha\alpha} + \dot{h}_0^2 h_{\kappa\kappa}. \quad (11.27)$$

The first term describes the second-order longitudinal deviation and the second term the longitudinal chromatic deviation of first order and first degree. The third and the fourth term define the longitudinal chromatic deviation of second degree. We obtain the fundamental second-order longitudinal deviation $h_{\gamma\gamma}$ from the expressions (11.26) and (11.27) by considering only electrons which have initial velocity components $\dot{h}_0 = 0, \dot{w}_0 = 0$. Employing these starting values and substituting the expression (11.26) for the second integral in (11.25), we eventually find

$$h_{\gamma\gamma} = \frac{\Phi''}{4\Phi_0'} w_\gamma^2 h_\gamma + \frac{\Phi_0'}{4\Phi_a} \dot{w}_\gamma^2 h_\gamma - \frac{\Phi_a}{2\Phi_0'^2} h_\alpha \int_{\tau_0}^{\tau} \Phi''' w_\gamma^2 h_\gamma d\tau. \quad (11.28)$$

We shall use this expression for obtaining an analytical expression for the coefficient of the spherical aberration which resembles that of the mirror.

11.3.2 Third-Order Spherical Aberration at the Crossover

The aperture aberrations at the crossover plane are formed by a monochromatic pencil of rays whose asymptotes start from the point-like virtual source. The angular width of this pencil is defined by the radius ρ_0 of the emission spot. Accordingly, the coordinate w_0 adopts the role of the aperture angle of the objective lens. We obtain the geometrical aberrations by setting $\eta = 0$ in all formulas for the higher-rank deviations. The primary geometrical aberrations are of third order because the lateral second-rank deviation (11.20) is of chromatic nature and vanishes for $v_0 = 0$. In this case, the second-rank lateral deviation (11.27) reduces to

$$h^{(2)} = w_0 \bar{w}_0 h_{\gamma\gamma}. \quad (11.29)$$

The geometrical part of the third-rank perturbation function (10.71) is

$$\begin{aligned} p_g^{(3)} &= p^{(3)}(\tau, w_0, \eta = 0) \\ &= w_0^2 \bar{w}_0 \frac{\Phi_a}{8\Phi_0'^2} \left\{ \Phi''' w_\gamma^3 - 8\Phi''' w_\gamma h_{\gamma\gamma} \right\}. \end{aligned} \quad (11.30)$$

Substituting this relation for $p^{(3)}$ in the formula (10.61) and putting $\tau = \tau_c$, we obtain for the third-order lateral deviation at the crossover plane the expression

$$\begin{aligned}
w_g^{(3)}(\zeta_c) &= w_{ac} w_0^2 \bar{w}_0 \frac{\Phi_a}{8\Phi_0'^2} \int_{\tau_0}^{\tau_c} \left\{ \Phi'''' w_\gamma^4 - 8\Phi'''' w_\gamma^2 h_{\gamma\gamma} \right\} d\tau \\
&= w_{ac} \frac{w_0^2 \bar{w}_0}{16\Phi_0'} \int_{\zeta_0}^{\zeta_c} \left\{ \Phi'''' w_\gamma^2 - 2\Phi'''' \left(\frac{\Phi''}{\Phi_0'} w_\gamma^2 + \frac{\Phi_0'}{\Phi_a} \dot{w}_\gamma^2 \right) h_\gamma + \frac{2\Phi'''}{\Phi_0'} h_\alpha \int_{\zeta_0}^{\zeta} \Phi'''' w_\gamma^2 \frac{h_\gamma}{h_\alpha} d\zeta \right\} \frac{w_\gamma^2}{h_\alpha} d\zeta.
\end{aligned} \tag{11.31}$$

We simplify this expression by removing the double integral via partial integration using the relation

$$\int \Phi'''' w_\gamma^2 d\zeta = \Phi'' w_\gamma^2 + \frac{\Phi_0'^2}{\Phi_a} \dot{w}_\gamma^2. \tag{11.32}$$

As a result, we find

$$\begin{aligned}
w_g^{(3)}(\zeta_c) &= -\dot{w}_{\gamma c} \frac{w_0^2 \bar{w}_0}{4\Phi_0'} \int_{\tau_0}^{\tau_c} \Phi'''' w_\gamma^2 h_\gamma d\tau \\
&\quad + w_{ac} \frac{w_0^2 \bar{w}_0}{16\Phi_0'} \int_{\zeta_0}^{\zeta_c} \sqrt{\frac{\Phi_a}{\Phi}} \left\{ \Phi'''' - 4\Phi'''' \left(\frac{\Phi''}{\Phi_0'} + \frac{\Phi_0'}{\Phi_a} \frac{\dot{w}_\gamma^2}{w_\gamma^2} \right) h_\gamma \right\} w_\gamma^4 d\zeta.
\end{aligned} \tag{11.33}$$

We derive the coefficient $\tilde{C}_3$ of the third-order spherical aberration at the crossover plane from the third-order lateral deviation (10.71) in the case $\eta = 0$. By considering the relations (11.11) and (11.32), we eventually obtain

$$\hat{w}_g^{(3)}(\zeta_c) = w_{ac} w_0^2 \bar{w}_0 C_3 = w_g^{(3)}(\zeta_c) - \frac{\Phi_0' w_0^2 \bar{w}_0}{2\sqrt{\Phi_a \Phi_c}} h_{\gamma\gamma}(\zeta_c). \tag{11.34}$$

Substituting in this formula the expression (11.33) for $w_g^{(3)}(\zeta_c)$ and (11.28) for $h_{\gamma\gamma}$ and using the Wronskian $w_{ac} \dot{w}_{\gamma c} = -1$, we find

$$\begin{aligned}
\tilde{C}_3 &= \frac{w_{\gamma c}^3}{4} \int_{\zeta_\gamma}^{\zeta_c} \left(\frac{\Phi_c}{\Phi} \right)^{\frac{3}{2}} d\zeta \\
&\quad + \frac{1}{16\Phi_0'} \int_{\zeta_0}^{\zeta_c} \sqrt{\frac{\Phi_a}{\Phi}} \left\{ \Phi'''' - 4\Phi'''' \left(\frac{\Phi''}{\Phi_0'} + \frac{\Phi_0'}{\Phi_a} \frac{\dot{w}_\gamma^2}{w_\gamma^2} \right) h_\gamma \right\} w_\gamma^4 d\zeta.
\end{aligned} \tag{11.35}$$

The dimension of the aberration coefficient $\tilde{C}_3$ differs from that of the coefficient $C_3 = C_s$ of the third-order spherical aberration of conventional round lenses. In order to account for this difference, we indicate the coefficient (11.35) by a tilde. This coefficient has the dimension cm^{-2} because we define the limiting aperture of the rays intersecting the center of the crossover by the radius $w_{0\text{max}}$ of the emission area rather than by the limiting aperture angle $w_{0\text{max}} w'_{\gamma c}$.

Chapter 12

Confinement of Charged Particles

Abstract The confinement of charged particles is discussed for the Paul trap which employs high-frequency electromagnetic fields and for the combined trap obtained by superposing a homogeneous magnetic field onto the Paul trap.

To investigate the properties of free particles, it is advantageous to confine them in three dimensions. Charged particles can be stabilized by means of high-frequency electromagnetic fields [169] or by a proper superposition of static electric and magnetic fields. Such devices are named *ion traps*, the former type is called *Paul trap*, the latter type is known as *Penning trap* [170]. Ion traps are able to confine charged particles in a small spatial region, mass selectively. This behavior enables high-sensitivity mass spectroscopy and allows one to hold and study single ions or few-particle assemblies at very low kinetic energy.

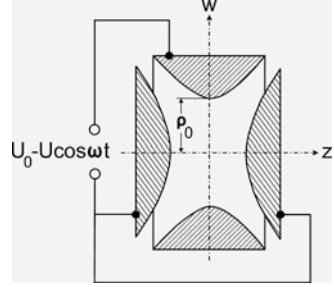
12.1 Paul Trap

We denote the mass and the charge of the particle by m and q , respectively. In the case of electrons, we have $m = m_e$, $q = -e$. In order to avoid a loss of particles, we must prevent an increase of the amplitude of the particle oscillations, as realized in particle accelerators and storage rings. A hyperbolic rotationally symmetric potential meets this requirement because the components of the force are linear with respect to the origin in all directions. The proper electrostatic potential is

$$\varphi = \frac{U_0}{\rho_0^2} (w\bar{w} - 2z^2). \quad (12.1)$$

The equipotential surfaces $\varphi = \text{const}$ form rotational hyperboloids centered about the optic axis, as shown in Fig. 12.1. We can realize the potential by three electrodes. The surface of one electrode is a toroidal hyperboloid; the surfaces of the two others form a rotational hyperboloid of two sheets. The apex of each of these electrodes is

Fig. 12.1 Radial cross section of the electrodes of the rotationally symmetric charged-particle trap; the voltage $V = U_0 - U \cos \omega t$ is applied between the toroidal electrode and the electrodes of the two-sheet hyperboloid centered about the z -axis



located on the z -axis at a distance $\rho_0/\sqrt{2}$ from the origin, whereas the radius of the annular apex of the toroidal electrode is ρ_0 . Employing the potential (12.1) and the complex lateral coordinate $w = x + iy$, we obtain the nonrelativistic equations of motion in the form

$$m \frac{d^2 w}{dt^2} = -2 \frac{q U_0}{\rho_0^2} w, \quad m \frac{d^2 z}{dt^2} = 4 \frac{q U_0}{\rho_0^2} z. \quad (12.2)$$

The equations demonstrate that the motion is unstable in axial direction if it is stable in radial direction and vice versa. Hence, in order to achieve overall stability, the voltage V applied between the electrodes must alternate, as does the polarity of the quadrupoles in strong focusing accelerators. We satisfy this requirement by applying between the electrodes the voltage

$$V = U_0 - U \cos \omega t. \quad (12.3)$$

In this case, the polarity of the voltage $U_0 - V = U \cos \omega t$ oscillates with frequency ω .

Substituting V for U_0 in the equations (12.2) and employing relation (12.3) gives the Mathieu equations

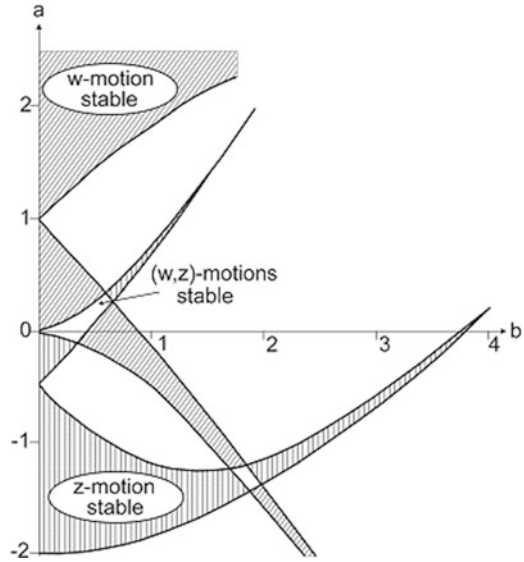
$$\begin{aligned} \frac{d^2 w}{d\zeta^2} + (a - 2b \cos 2\zeta) w &= 0, \\ \frac{d^2 z}{d\zeta^2} - 2(a - 2b \cos 2\zeta) z &= 0. \end{aligned} \quad (12.4)$$

Here, we have introduced normalized quantities defined by

$$\zeta = \frac{\omega t}{2}, \quad a = \frac{4q U_0}{m \rho_0^2 \omega^2}, \quad b = \frac{2q U}{m \rho_0^2 \omega^2}. \quad (12.5)$$

The solutions of the Mathieu equations represent oscillations, which are stable or instable depending on the values of the parameters a, b . The values for stable motion are obtained from the stability chart shown in Fig. 12.2. We have plotted the stability

Fig. 12.2 Stability chart for the radial $w = x + iy$ motion and the axial z -motion; in the overlap regions, the motion of both components is stable



regimes for the radial and axial motions in a single chart. This representation has been achieved by mirroring the stability regimes of the z -motion about the a -axis and by reducing the scale by a factor 2. Stable motion in all directions occurs for values in the overlap region of the stability regimes for the axial and the radial motions. In practice, only the extended overlap region at the origin will be used. Relation (12.5) shows that the parameters a , b are proportional to the ratio q/m . Therefore, we obtain only stable motion for charged particles whose q/m -ratios have values within a distinct range. All other particles perform oscillations with exponentially increasing amplitude and leave the trap. It is not possible to confine a charged particle which is injected from outside the trap. In order to confine an ion, it must be created within the trapping field by ionization of neutral atoms or molecules.

12.2 Combined Trap

The combined trap represents the superposition of the Paul trap with a homogeneous axial magnetic field $\vec{B} = B\vec{e}_z$. This trap has a larger stability region and allows the simultaneous trapping of negatively and positively charged particles. The force of the magnetic field is given by

$$\vec{F}_m = q\dot{\vec{r}} \times \vec{B} = qB(\dot{y}\vec{e}_x - \dot{x}\vec{e}_y). \quad (12.6)$$

By considering this additional force and substituting (12.3) for U_0 in (12.2), we obtain the following equations of motion

$$m \frac{d^2 w}{dt^2} = -2 \frac{q}{\rho_0^2} V(t) w - i q B \dot{w}, \quad m \frac{d^2 z}{dt^2} = 4 \frac{q}{\rho_0^2} V(t) z. \quad (12.7)$$

The real part and the imaginary part of the equation for the complex lateral distance $w = x + iy$ represent two coupled equations in the coordinates x and y . We decouple these equations, as in the case of magnetic round lenses (4.21), by introducing the rotating coordinate system

$$u = w e^{-i\chi(t)}, \quad \chi(t) = \frac{\omega_L}{2} t, \quad \omega_L = \frac{qB}{m}. \quad (12.8)$$

The angle of rotation χ is half the Larmor angle which is proportional to the time t and to the Larmor frequency ω_L . In the frame of the rotating coordinate system, the equations (12.7) adopt the form

$$\begin{aligned} \frac{d^2 u}{d\xi^2} + \left(\frac{\omega_L^2}{\omega^2} + a - 2b \cos 2\xi \right) u &= 0, \\ \frac{d^2 z}{d\xi^2} - 2(a - 2b \cos 2\xi) z &= 0. \end{aligned} \quad (12.9)$$

The comparison of these equations with the equations (12.4) of the Paul trap shows that they differ by the term ω_L^2/ω^2 . Due to this term, it is possible to obtain stable solution in the static case $U = 0$, giving $b = 0$. Moreover, ω becomes a free parameter. The most appropriate choice is $\omega = \omega_L$. The resulting equations represent those of the Penning trap. This trap has stable motions in the case

$$-\frac{1}{a} = -\frac{qB^2}{4mU_0} > 1. \quad (12.10)$$

In order to confine positive ions ($q > 0$), the voltage U_0 must be negative. This situation is realized if the end-cap electrodes are at positive potential with respect to that of the ring electrode. The axial component of the electric field pushes the ions toward the center of the trap causing harmonic oscillations. However, the radial component of the electric field is repulsive and pushes the ions out of the trap. It is the opposite radial force of the magnetic field which forces the ions toward the axis, resulting in their confinement if the condition (12.10) is fulfilled.

The term ω_L^2/ω^2 in the first equation of the combined trap (12.9) enables simultaneous trapping of positively and negatively charged particles and provides a larger stability region than the Paul trap. The simultaneous trapping of positive and negative ions is of great experimental importance because it allows one to study recombination processes.

Chapter 13

Monochromators and Imaging Energy Filters

Abstract The optical properties and the performance of the electrostatic monochromator and different types of imaging energy filters are discussed in detail. These devices enable in combination with aberration correctors analytical electron microscopy with atomic spatial resolution and high-energy resolution of about 0.05 eV. The design criteria for feasible imaging energy filters are illustrated by means of the MANDOLINE filter which has the highest dispersion and transmissivity of all energy filters proposed so far. In addition, the properties of the new W-filter are outlined. This beam-reversing filter is placed at the bottom of the twin-column electron microscope consisting of an object column and a projector column containing the projector lenses and the detection system.

The ultimate goal of high-resolution analytical electron microscopy is the acquisition of detailed information about the atomic structure, the chemical composition, and the local electronic states of real objects whose structure deviates from ideal crystalline periodicity. To obtain detailed information on the interatomic bonding, an energy resolution of about 0.1 eV is necessary. The presently available electron microscopes do not fulfill this requirement because electron sources with a maximum energy spread of 0.1 eV at a sufficiently high current do not yet exist for conventional transmission electron microscopes. The energy width of field emitters lies in the range between 0.3 and 0.8 eV depending on the current. Hence, to enable electron spectroscopy with an energy resolution of 0.1 eV, we must employ a monochromator which filters out the electrons which deviate more than ± 0.05 eV from the most probable energy. A feasible monochromator reduces the energy spread of the beam without affecting the spectral brightness and the effective size of the source. In order to preserve the emission characteristic of the source and to prevent a loss of lateral coherence, the dispersion must vanish on the far side of the monochromator. Moreover, the monochromator should be as compact as possible to avoid an unduly lengthening of the column. These conditions cannot be satisfied satisfactorily by Wien filters. In order that the monochromator does not affect the size and the radiation characteristic of the effective source, the second-order

aberrations and the dispersion must vanish behind the monochromator. Therefore, the energy selection must be performed within the monochromator at a position where the dispersion is at its maximum. Different versions of such dispersion-free energy filters have been proposed [80, 96]. Because the monochromators are placed at high tension, electrostatic designs are most appropriate [82, 83].

13.1 Electrostatic Monochromator

To realize a quasi-monochromatic electron source, we consider a compact electrostatic monochromator, which reduces effectively the energy spread of the illuminating beam without deteriorating the spectral brightness [83]. The Ω -type monochromator is placed behind the gun and removes all electrons whose energies deviate more than ± 0.05 eV from the most probable energy. In the case of a Schottky field emitter, the monochromator takes away about 70 % of the emitted electrons. The monochromator consists of four toroidal sector deflectors which are arranged symmetrically about the midplane. Since this plane is perpendicular to the optic axis of the microscope, the lengthening of the column. The deflection elements introduce a dispersion which adopts a maximum at the center of the filter, as illustrated in Fig. 13.1.

For determining the geometry of the electrodes and the course of the paraxial rays, we start with the SCOFF approximation which neglects the finite extension of the fringing fields. This approximation yields analytical solutions for the paraxial rays, the dispersion, and the coefficients of the second-rank aberrations. The x and y components of the paraxial trajectory are linear combinations of the fundamental rays $x_\alpha, y_\beta, x_\gamma, y_\delta$, and the dispersion ray x_κ . For a ray which emanates at the position x_s, y_s with slope components α, β from the effective source, the ray components are given by

$$\begin{aligned} x &= \alpha x_\alpha + x_s x_\gamma + \kappa x_\kappa, \\ y &= \beta y_\beta + y_s y_\delta. \end{aligned} \quad (13.1)$$

Energy filtering is performed at the symmetry plane z_s where a line image of the source is located ($x_\alpha(z_s) = 0$) and the dispersion is at its maximum. The dispersion ray $x_\kappa(z)$ is the inhomogeneous solution of the paraxial path equation (4.41) for the special case $\Phi = \Phi_0$, and $\kappa = 1$:

$$x'' + \left(\frac{2 + 3\gamma_0^2}{8} \frac{\Phi_{1c}^2}{\Phi_0^{*2}} - \gamma_0 \frac{\Phi_{2c}}{\Phi_0^*} \right) x = -\kappa \frac{1 + \gamma_0^2}{2(1 + \gamma_0)} \frac{\Phi_{1c}}{\Phi_0^*}. \quad (13.2)$$

Using the analytical solutions for the fundamental rays, we have carried out an extensive computer-aided search for finding the optimum system. The arrangement of the electrodes and the course of the fundamental rays of the final solution are shown in Fig. 13.2. The courses of the axial fundamental rays x_α and y_β reveal that only astigmatic vertical and horizontal line images are formed within

Fig. 13.1 View of the toroidal deflection electrodes and illustration of the dispersive properties of the Ω -type monochromator

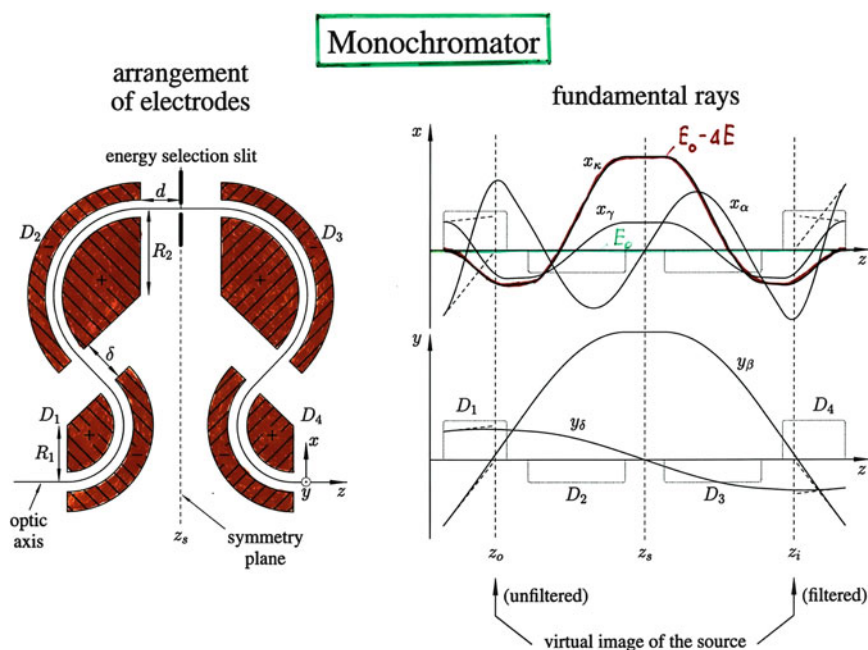
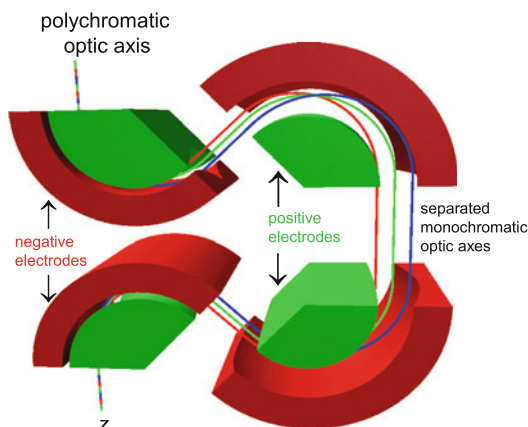


Fig. 13.2 Horizontal x,z -cross section through the omega-shaped electrostatic monochromator and course of the fundamental rays along the straightened optic axis within the horizontal and the vertical sections; z_0 and z_i are the locations of the virtual dispersion-free stigmatic entrance and exit images of the effective source

the monochromator. Therefore, stigmatic images with high current density are completely avoided. The dispersion at the energy selection plane is

$$D = 2.26R_1/E. \quad (13.3)$$

Here, E and R_1 are the energy in front of the monochromator and the radius of the first deflection element, respectively. Choosing $E = 3$ keV and $R_1 = 3$ cm, we obtain a dispersion $D = 22.6 \mu\text{m}/\text{eV}$ which suffices for achieving an energy width of about 0.1 eV.

To find the realistic and accurate geometry of the electrodes, we must consider the finite extension of the fringing fields. We determine the realistic fields by approximating the inner surfaces of the electrodes by triangular meshes. A linearly distributed charge density is assumed for each triangle giving an analytical expression for the potential. Starting from the solution of the SCOFF approximation, we calculate the course of the realistic optic axis and of the fundamental rays by successive iteration. The number of iteration steps depends on the required accuracy [57]. Owing to the symmetry of the fields and the fundamental rays, the monochromator as a whole does not introduce second-order aberrations. The aberrations introduced by the first and second deflectors are compensated by those of the third and fourth deflectors. In order that all electrons with nominal energy pass through the energy selection slit, we must compensate for the second-order aperture aberration at the energy selection plane z_s by means of hexapole fields. We produce these fields by curving appropriately the inner surfaces of the electrodes, as visualized in Fig. 13.1.

In order to fully exploit the capability of the monochromator, it must be combined with a high-performance imaging energy filter. Such a filter must possess (a) a large dispersion to allow for sufficiently small energy windows, (b) no second-order aberrations at the image and the energy selection plane, and (c) a compact geometry to avoid an unduly large lengthening of the microscope column. The latter requirement is especially important in the case of aberration-corrected analytical electron microscopes, because the incorporation of the monochromator and of the energy filter further lengthen the column in addition to the corrector. As a result, the mechanical instabilities increase and may impede an appreciable reduction of the information limit. Recently, this monochromator has been incorporated together with the MANDOLINE filter [130] into the SESAM microscope at the Max Planck Institute in Stuttgart. This high-performance analytical electron microscope enables local electron spectroscopy with an energy resolution of about 0.1 eV, which is necessary for determining local variations of the atomic bonding near interfaces or defects.

13.2 Imaging Energy Filters

The ultimate goal of high-resolution analytical electron microscopy is the acquisition of detailed information about the atomic bonding, the chemical composition, and the local electronic states of nonperiodic objects such as nanoparticles, interfaces, dislocations, and macromolecules. The deviations from the ideal structure affect the electronic properties of nanostructured devices with proceeding miniaturization. Energy filtering offers the possibility (a) to remove the inelastic scattered

electrons from the image-forming beam, (b) to record the energy-loss spectrum from an arbitrary area of the object, and (c) to record images and diffraction patterns with electrons which have suffered a characteristic energy loss. An ideal filter acts like a round lens with respect to the transmitted electrons and like a combination of a round lens and a prism for electrons whose energies differ from the nominal energy of the transmitted electrons. Energy filtering is performed at the energy selection plane z_E located behind the filter. This plane is the dispersion image of the diffraction plane located in front of the filter. Owing to the dispersion, the filter images the polychromatic diffraction pattern into a series of laterally displaced monochromatic spots. In order that the diffraction spots are sufficiently separated from each other and smaller than the energy selection slit, the filter must have a high dispersion, and the diffraction image in front of the filter must be appreciably demagnified. Moreover, the optimum filter should be isochromatic which means that the selected energy does not depend on the lateral position of the object detail. To satisfy these conditions, all second-rank aberrations must be either eliminated or adequately suppressed.

13.2.1 *Types of Imaging Energy Filters*

Imaging energy filters are usually characterized by the shape and the nature of the arrangement. For producing the dispersion, dipole fields are mandatory. Therefore, all filters and spectrometers systems have a curved axis with the exception of the Wien filter outlined in Chap. 7. Electric–magnetic imaging energy filters have the decisive disadvantage that they are limited to accelerating voltages below about 100 kV due to the difficulties in handling large electric field strengths. Imaging energy filters are incorporated in an electron microscope either within the column or as an attachment beneath the viewing screen. The post-column filters bend the optic axis usually by 90° , whereas all present in-column filters are straight-vision systems. In the following, we shall describe briefly the different types of in-column energy filters shown schematically in Fig. 13.3 and subsequently discuss in detail the high-performance MANDOLINE filter and the beam-reversing W-filter [107].

The first imaging energy filter consisting of a triangular magnetic double prism and an electrostatic diode mirror was developed in 1964 by Castaing and Henry [163]. Unfortunately, Henry placed his mirror–prism filter directly behind the first intermediate lens resulting in large second-order aberrations (inclination of the image field and field astigmatism) which decisively limited the field of view. Due to the work of Henkelman and Ottensmeyer [164], this shortcoming was eliminated in the first commercial energy-filtering electron microscope, the Zeiss EM902.

Symmetry principles for correcting second-order aberrations of imaging energy filters were first introduced in 1974 by Rose and Plies, who proposed the first symmetric magnetic equivalent of the prism–mirror–prism system [78]. An improved version of this filter is the partly corrected Omega filter [127]. This compact symmetric system consists of three homogeneous deflection magnets and a sextupole

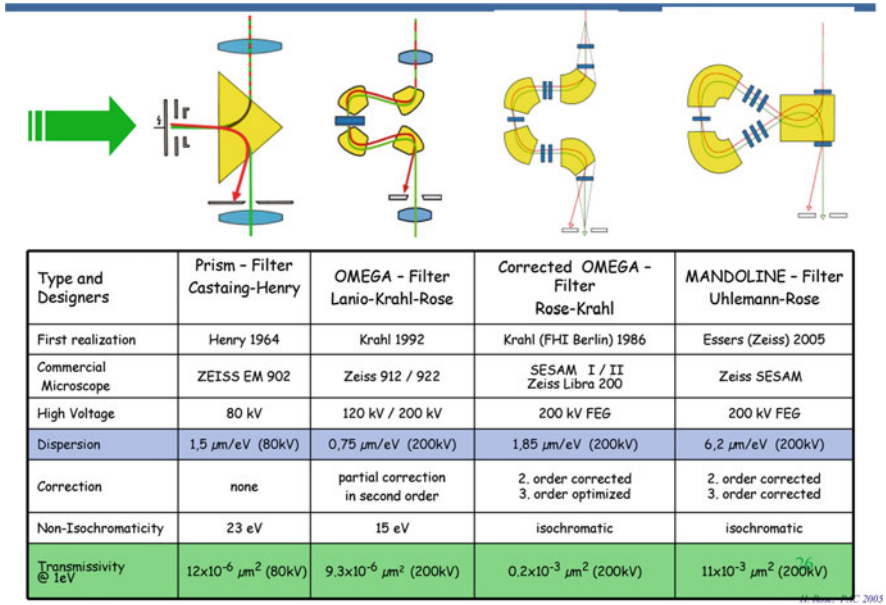


Fig. 13.3 Schematic arrangement and properties of in-column imaging energy filters

placed at the center of the filter. The geometry of the filter has been optimized in such a way that the remaining second-order aberrations are at an overall minimum.

High-performance imaging energy filters must be corrected for second-order aberrations. In order to enable high-resolution imaging of extended objects and very narrow energy windows, we have designed a fully corrected magnetic Omega filter [128]. This filter is part of the Zeiss Libra 200 analytical electron microscope. By imposing midplane symmetry with respect to the magnetic fields and the paraxial rays, half of the second-order aberrations cancel out. We cannot compensate for all aberrations by a single symmetry because two of the four linearly independent rays, one in the xz -section and the other in the yz -section, are symmetric and two are antisymmetric with respect to the midplane. By introducing an additional symmetry plane for each half of the system, the integrands of the aberration integrals that are symmetric with respect to the midplane are then antisymmetric with respect to the central plane of each half of the system. In this case, all second-order aberrations cancel out as well as the dispersion. Unfortunately, such achromatic systems are not suitable as imaging energy filters because they do not allow spectrum imaging. Hence, we must eliminate the second-order field astigmatism and the axial aberration at the energy selection plane by means of hexapole fields because they do not affect the dispersion. We can generate these fields either by curving the entrance and exit faces of the magnets or by sextupole elements. However, curving the pole faces is not suitable for systems composed of several magnets because it results in a chaotic behavior for the alignment of the paraxial rays. Therefore, the incorporation of adjustable sextupole elements is mandatory for such systems.

Simultaneous correction the remaining second-rank aberrations by sextupoles requires a strongly astigmatic path of the paraxial trajectories in the drift spaces between the deflecting magnets. However, not all aberration components can be eliminated independently. Since the coupling hampers the correction procedure, one aims for arrangements in which the correction of the nonvanishing aberration components is as decoupled as possible. We can eliminate the aberration coefficients $A_{\alpha\alpha\gamma}$, $B_{\alpha\beta\delta}$, and $B_{\beta\beta\delta}$ of the image tilt and field astigmatism of energy filters with midsection symmetry largely independently by placing astigmatic images of both the object plane and the diffraction plane in the drift spaces between the magnets. Because half of the geometrical second-order aberrations have been canceled out by symmetry, it is necessary to incorporate the sextupole elements in pairs such that the sextupoles of each pair are placed symmetrically about the midplane z_m of the filter. In this case, each pair introduces neither second-order distortion nor axial aberration regardless of its position [79]. A sextupole centered at the midplane need not to be split up, because it automatically satisfies the symmetry condition.

13.2.2 MANDOLINE Filter

The required properties of a high-performance imaging energy filter are best met by the MANDOLINE filter which has by far the highest dispersion and transmissivity of all energy filters proposed so far. We design this filter by substituting conical magnets for the inner homogeneous deflection magnets of the Ω -filter. The MANDOLINE filter shown in Fig. 13.4 consists of a single homogeneous bending magnet and two inhomogeneous deflection magnets with tapered pole pieces. These elements focus the electrons within their two principle sections toward the optic axis and act as “anamorphic” lenses with a curved axis. Inhomogeneous deflection magnets provide high angular dispersion because the focusing can be made small in the dispersive section and large in the vertical section of the magnets. This behavior differs from that of homogeneous magnets where the focusing is strong in the horizontal section, and the vertical refraction is confined to the short fringing-field regions at the entrance and exit faces of the magnet.

The geometry of the tapered pole pieces of an inhomogeneous sector magnet is shown schematically in Fig. 13.5. Although the MANDOLINE filter realized in the SESAM electron microscope enlarges the column only by about 23 cm, its dispersion is about twice as high as that of the best post-column filter. The asymptotes of the conical inner pole faces intersect each other in the (moving) point

$$x_c = -\frac{D}{2} \cot \delta = -\frac{R}{\nu^2}. \quad (13.4)$$

The distance D denotes the separation of the pole faces taken at the optic axis with radius R ; δ is the inclination angle of each pole with respect to the x -coordinate.

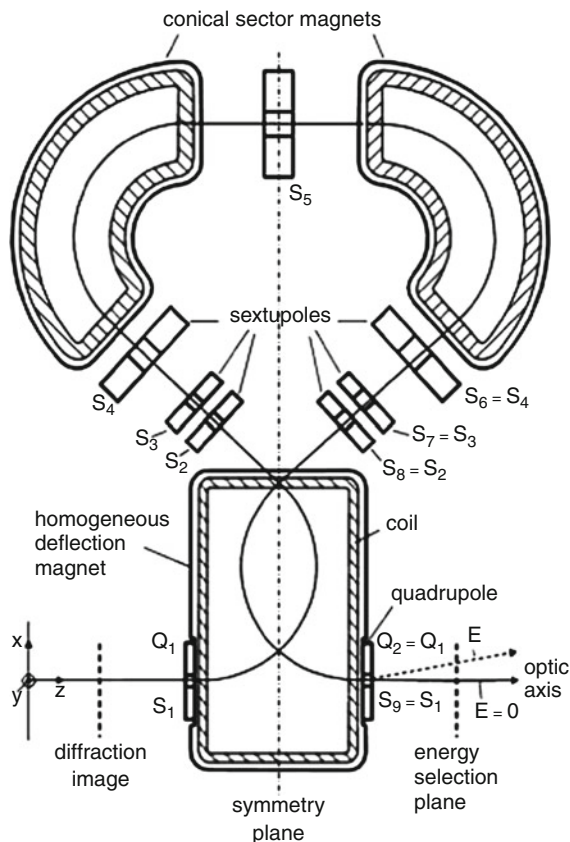


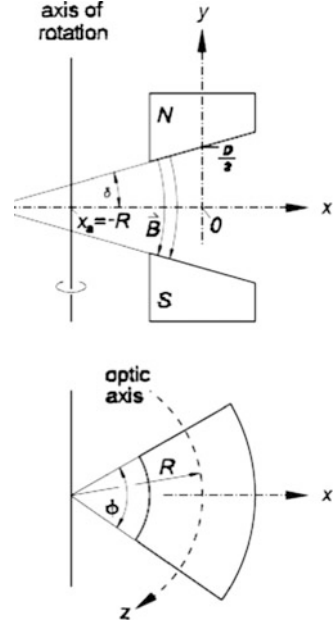
Fig. 13.4 Schematic arrangement of the deflection elements and the sextupoles within the MANDOLINE filter; the distance between the energy selection plane and the diffraction image in front of the filter defines the lengthening of the column

The parameter ν^2 represents the so-called field index. This index and the magnetic dipole strength Ψ_{1s} define the strengths of all other multipole components with higher multiplicity. This index is zero for homogeneous bending magnets with plane-parallel inner pole faces.

Within the frame of validity of the SCOFF approximation, the quadrupole and hexapole strengths are given by the relations

$$\begin{aligned}\Psi_{2s} &= -\frac{\nu^2}{2}\Psi_{1s}^2, \\ \Psi_{3s} &= \left(\frac{\nu^4}{3} - \frac{\nu^2}{24}\right)\Psi_{1s}^3.\end{aligned}\tag{13.5}$$

Fig. 13.5 Top and side view of an inhomogeneous deflection magnet with tapered poles



In the special case of a homogeneous magnet ($\nu = 0$), all multipole components with multiplicity $m \geq 1$ vanish.

The SCOFF approximation assumes box-shaped distributions for the multipole strengths. This approximation is valid with a sufficient degree of accuracy if $D \ll \phi R$ and gives a constant curvature for the optic axis with radius

$$R = \frac{1}{|\eta \Psi_{1s}|}, \quad \eta = \sqrt{\frac{e}{2m_e \Phi_0^*}}. \quad (13.6)$$

In this case, the paraxial path equation (4.37) adopts the form

$$\begin{aligned} R^2 x'' + (1 - \nu^2)x &= \frac{\kappa^*}{2} R \text{sgn}(\psi_{1s}), \\ R^2 y'' + \nu^2 y &= 0. \end{aligned} \quad (13.7)$$

We achieve continuous focusing in both sections in the case $\nu < 1$; the relativistic modified chromatic parameter κ^* is defined by the relation

$$\kappa^* = \frac{1 + e\Phi_0/m_e c^2}{1 + e\Phi_0/2m_e c^2} \kappa, \quad \kappa = \frac{\Delta E}{e\Phi_0}. \quad (13.8)$$

The field ray x_γ runs parallel to the optic axis at the midplane z_m of the filter and intersects the optic axis at all images of the diffraction plane. By considering this

property, we obtain for the lateral displacement of the dispersion ray $x_\kappa(z)$ at the energy selection plane z_E the expression

$$x_\kappa(z_E) = C_{\gamma\kappa} = \frac{1}{2} \int_{z_D}^{z_E} \eta \Psi_{1s} dz. \quad (13.9)$$

The dispersion Δ (displacement per eV) is proportional to the dispersion coefficient $C_{\gamma\kappa}$ and defined as

$$\Delta = C_{\gamma\kappa} \frac{\kappa^*}{\Delta E}. \quad (13.10)$$

In order to obtain a large dispersion, we must adjust the magnetic dipole strength $\psi_{1s}(z)$ and the field ray $x_\gamma(z)$ in such a way that the integrand of the integral (14.9) does not change its sign. Hence, if two adjacent bending magnets deflect the electrons in opposite directions, the field ray must intersect the optic axis in the region between these magnets.

The path equations (13.7) indicate that decreasing the focusing in the horizontal x, z -section increases the focusing in the vertical principal section. To obtain a high dispersion, we must reduce the focusing strength $1 - \nu^2$ in the dispersive horizontal section as much as possible. The small focusing strength allows large deflection angles ϕ giving a high dispersion without enlarging appreciably the lengthening $l = z_E - z_D$ of the column, as illustrated in Fig. 13.4 for the MANDOLINE filter. The lengthening of the column is given by the distance between the energy selection plane z_E and the diffraction plane z_D in front of the filter because it images this plane with unit magnification into the dispersion plane z_E behind the MANDOLINE filter. This filter represents an imaging energy filter with highest performance.

Half of the second-order aberrations are eliminated by the symmetric arrangement of the magnets and the symmetry of the fundamental rays with respect to the midplane of the filter. We employ four sextupole pairs $S_\nu = S_{10-\nu}$, $\nu = 1, 2, 3, 4$ and a single sextupole S_5 at the symmetry to compensate for the nonvanishing second-order aberrations. To compensate simultaneously for the axial aberration at the energy selection plane and for the aberrations (image tilt and field astigmatism) at the final image plane, a strongly astigmatic path of the paraxial rays within the regions between the bending magnets is mandatory [79, 130]. In order to eliminate these aberrations largely independently from each other, we place the sextupoles of one pair at astigmatic images of the object plane and the sextupoles of another pair at images of the diffraction plane [79, 80]. The symmetric correction of all second-order aberrations eliminates simultaneously several third-order aberrations. As a result, the MANDOLINE filter has a high transmissivity and allows isochromatic imaging of large object areas with energy windows as narrow as 0.08 eV. The theoretical predictions are confirmed by the experimental results of the SESAM electron microscope.

13.2.3 *W-Filter*

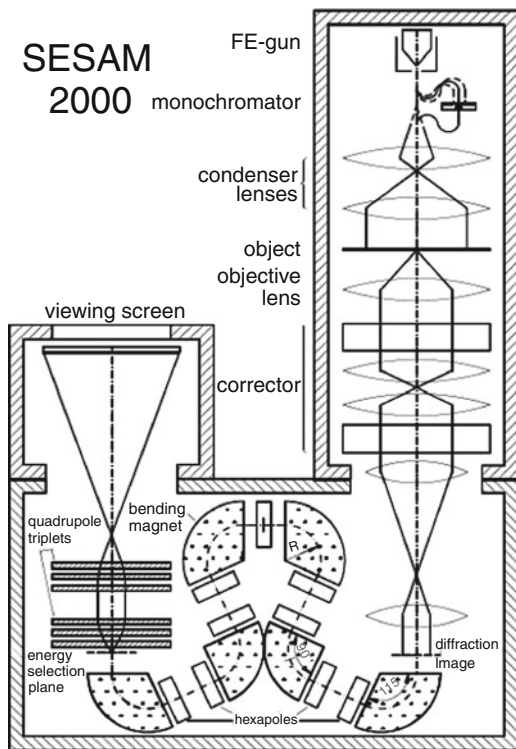
Each additional element incorporated into the electron microscope enlarges the length of the column, increasing its mechanical sensitivity. We can significantly suppress the mechanical instabilities by placing the heavy energy filter at the bottom of the instrument. To achieve a compact and stable microscope, it is advantageous to design it as a twin-column instrument such that the second column contains the projector lenses and the detection system, as depicted schematically in Fig. 13.6. Since the optic axis in the *image* column is parallel to that of the *object* column, the filter must also reverse the direction of flight of the electrons. Hence, the total deflection of the filter must amount to 180° [107], in contrast to the straight-vision in-column filters proposed so far.

Owing to the pronounced W-shaped course of its optic axis, we name our beam-reversing energy filter as *W-filter*. This filter is composed of an Omega filter placed between two bending magnets with equal deflection of the optic axis. Accordingly, the W-filter consists of six deflection magnets, as shown in Fig. 13.7. We minimize the nonvanishing aberrations by equalizing the first and third bending magnets. They are placed in opposite x -direction to reverse the deflection. Owing to the required midplane symmetry, the first magnet coincides with the sixth magnet and the third magnet with the fourth magnet. Since the deflection angles of these bending magnets cancel out, the second and the fifth magnet must each deflect the axis by 90° in order to achieve a total deflection angle of 180° . In this case, mechanical momenta are avoided because the object and image columns rest perpendicular on the filter, which acts as their common base. The two columns can be further stiffened by proper mechanical connections. The resulting twin column will be significantly shorter and less sensitive with respect to mechanical instabilities than the conventional single-column electron microscopes. The separation of the two column axes depends on the radius of curvature R and on the angular deflection angles ϕ_v of the optic axis within the constituent bending magnets of the filter.

To obtain high dispersion, we use conical bending magnets and alternate the curvature of the optic within the W-filter, as illustrated in Fig. 13.7. The deflection magnets are arranged symmetrically about the midplane z_m located midway between the two columns. We avoid large third-order aberrations if the fundamental paraxial rays propagate close to the optic axis within the entire filter. This condition is best achieved for a telescopic filter, where the fundamental axial rays x_α and y_β run parallel to the optic axis in front of and behind the filter. We achieve the alternating curvature and a high dispersion by means of three pairs of conical sector magnets with large deflection angles. For simplicity, we presuppose that all magnets have the same radius of curvature $R_v = R$.

Within the frame of validity of the SCOFF approximation, this radius is constant inside the box-shaped field of each magnet and zero outside. We further impose that astigmatic images of both the object plane and the diffraction plane are located at the midplane of the filter. This is only possible if the two line images are perpendicular to each other. In this case, the fundamental field rays must satisfy the conditions

Fig. 13.6 Scheme of an aberration-corrected twin-column analytical TEM equipped with a W-filter and a quadrupole projection system [107]



$$\begin{aligned} x'_\gamma(z_m) &= 0, & y_\delta(z_m) &= 0, \\ x_\alpha(z_m) &= 0, & y'_\beta(z_m) &= 0. \end{aligned} \quad (13.11)$$

Because the fundamental rays are entirely defined by the initial constraints, we can only meet the additional conditions (13.11) by adjusting four free parameters of the system appropriately. The adjustable parameters are the quadrupole strengths of the conical bending magnets and the spacing between these elements. For the system shown in Fig. 13.7, we have two adjustable field parameters ν_1 and ν_2 . The two other free parameters are the spacing a between the magnets of each half of the filter and the separation distance $2g$ between these halves. Therefore, the proposed doubly symmetric system provides exactly the number of free parameters which are necessary to adjust the required path of the paraxial fundamental rays. In order to obtain a large dispersion for a fixed distance s between the entrance axis and the exit axis, the deflection angles ϕ_1 and $\phi_3 = \phi_1$ of the first and third magnet must be larger than 90° . However, the increase in these angles cannot be made substantially larger than about 115° because the magnets must neither overlap with each other nor with the round lenses of the microscope. The filter depicted in Fig. 13.7 satisfies these design criteria. The SCOFF parameters of this system are

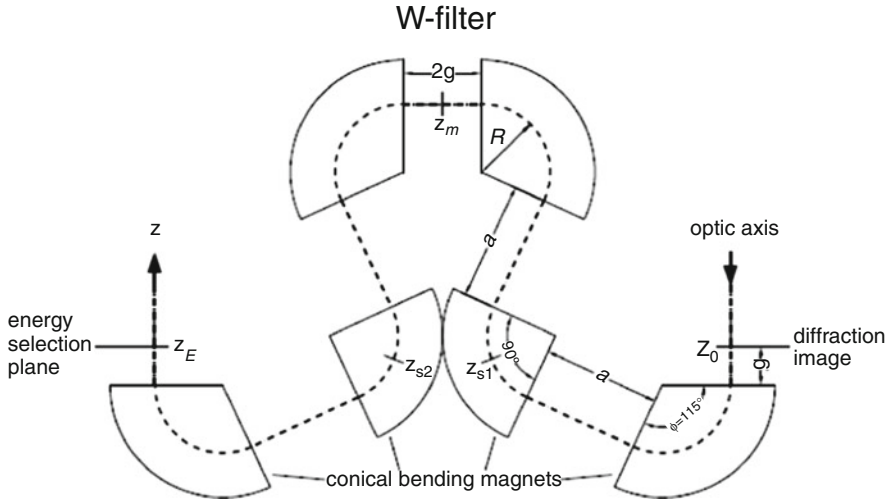


Fig. 13.7 Arrangement of the conical bending magnets of a high-dispersion W-filter

$$\begin{aligned}
 \nu_1 = \nu_3 = \nu_4 = \nu_6 &= 0.7906, & \nu_2 = \nu_5 &= 0.7929, \\
 \phi_1 = \phi_3 = \phi_4 = \phi_6 &= 115^\circ, & \phi_2 = \phi_5 &= 90^\circ, \\
 a &= 1.736R, & g &= 0.581R, & s &= 8.46R.
 \end{aligned}
 \tag{13.12}$$

The dispersion coefficient of this system is

$$C_{\gamma\kappa} = 7.296R = 0.864s. \tag{13.13}$$

Assuming a separation distance $s = 50$ cm and an accelerating voltage of 200 kV, we obtain the dispersion $\Delta \approx 2.2 \mu\text{m}/\text{eV}$. The course of the paraxial rays x_α and x_γ along the straightened optic axis is shown in Fig. 13.8 for one half of the filter. In the vertical yz -section, the axial ray y_β is symmetric and the field ray y_δ antisymmetric with respect to the central plane z_{s1} located in the middle of the second magnet. Since such symmetries do not exist for the corresponding rays x_α and x_γ in the horizontal xz -section, the paraxial path of the rays is largely astigmatic within the entire region of the filter.

This behavior enables one to compensate for the nonvanishing second-rank aberrations by means of sextupoles. These correction elements should be placed at appropriate positions between the bending magnets, in order to eliminate the different aberrations largely independently from each other. Unlike the geometrical fundamental rays, the dispersion ray does not possess any symmetry properties, as can be seen from Fig. 13.9.

By employing the W-filter, the column of future aberration-corrected analytical electron microscopes can be split up into two parts. The object column contains the

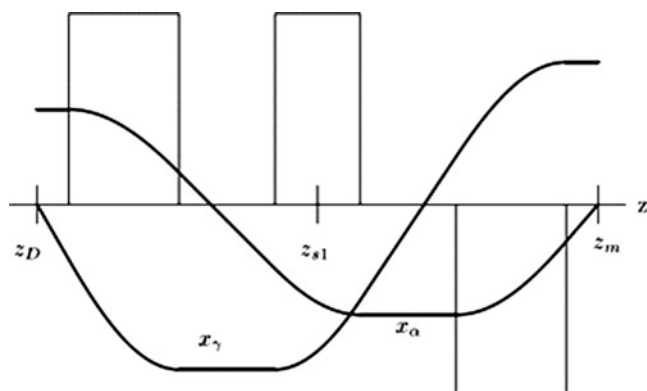


Fig. 13.8 Course of the fundamental rays x_α and x_γ along the straightened optic axis within the first half of the W-filter shown in Fig. 13.7

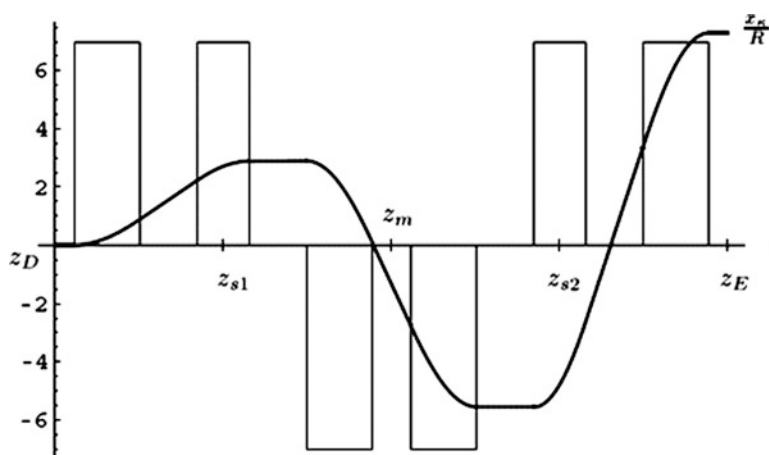


Fig. 13.9 Path of the dispersion ray x_κ within the straightened W-filter

field emission gun, the monochromator, the condenser system, and the objective lens, both of which are aberration corrected. The image column comprises the projector system, the CCD camera, and the viewing screen. The W-filter serves as a common solid base for both columns providing a mechanically stable high-performance instrument.

Chapter 14

Relativistic Electron Motion and Spin Precession

Abstract The effects of spin precession and radiation on the motion of relativistic electrons are derived by employing a novel covariant treatment which is not based on the Dirac equation of relativistic quantum mechanics. By introducing the Lorentz-invariant universal time as the independent variable and by introducing a four-dimensional Lagrangian, which considers Coulomb, spin, and gravitation interactions, we extend the Hamilton–Jacobi formalism of classical mechanics from three to four dimensions. The resulting equations for the spin precession of the electron yield the BMT equations in the special case of homogeneous electromagnetic fields. The Stern–Gerlach experiment is discussed for nonpolarized electrons, and the Lorentz transformations are discussed within the frame of electron motion in Minkowski space.

It is widely believed that the effect of the spin on electron motion cannot be accurately described within the frame of validity of geometrical charged-particle optics. However, there is no convincing reason that prevents one from incorporating the spin into the formalism of relativistic mechanics if an appropriate interaction Hamiltonian is found. To achieve a proper calculation procedure, it is advantageous to describe the relativistic motion and the spin precession of the electron in Minkowski space. By using $x_4 = ict$ as the fourth *spatial* coordinate of the four-dimensional pseudo-Euclidian space, we avoid difficulties in constructing relativistic covariant Lagrangians and Hamiltonians. We describe the motion of the electron by considering its four coordinates $x_\mu(\tau)$ as functions of the independent Lorentz-invariant variable τ , which we conceive as the world time or universal time, respectively. This time increases monotonically, whereas the time-like position coordinate $x_4 = ict$ needs not, contrary to classical mechanics. The four-dimensional Minkowski space is composed of the three-dimensional space with coordinates $x_1 = x, x_2 = y, x_3 = z$ and the imaginary time-like coordinate x_4 . The imaginary character of this coordinate is ultimately connected with certain properties of the time as experienced by men.

Table 14.1 Changes of 3D scalars and vectors to scalars, vectors, and tensors induced by extending the space from three to four dimensions

3D space		4D space
Scalar	$\Rightarrow$	Scalar ($T_0 = G(x_\mu, \tau)$)
Polar vector ($\vec{a}$)	$\Rightarrow$	Polar vector (T_μ), four components
Axial vector ($\vec{a} \times \vec{b}$)	$\Rightarrow$	Antisymmetric tensor ($T_{\mu\nu} = -T_{\nu\mu}$), six components
Pseudo-vector ($\vec{a} \times (\vec{b} \times \vec{c})$)	$\Rightarrow$	Pseudo-vector (C_μ), four components
Pseudo-scalar ($\vec{a} \cdot (\vec{b} \times \vec{c})$)	$\Rightarrow$	Pseudo-scalar (T_{1234})

The extension of the space from three to four dimensions is accompanied by a change of the properties of physical quantities. The transition of different physical quantities is shown in Table 14.1.

The difference between space and time encountered in our perception of the universe gives rise to several questions. Since the time-like coordinate x_4 is imaginary in Minkowski space, we may ask (a) is time real and (b) what is time? Our senses, specialized for interacting with the environment, are not suited for conceiving the time because we cannot see it or feel it. We subdivide the time according to our subjective experience in past, present, and future. However, past and future are nonexistent. Although we can memorize events which have occurred in the past, we do not have access to the past and/or the future. Moreover, we neither know how long the present is nor can we go backward in time.

The Greek philosopher Aristotle conceived time as a sequence of motions, whereas the famous philosopher Kant postulated that time is something how our mind works. He concluded that space and time are a way in which we experience the world, not how it really functions. Newton considered the time as a series of locations in the absolute three-dimensional space. In particular, he considered time as a one-dimensional absolute quantity which increases steadily. Hence, he conceived space geometry and time as independent from each other. Einstein abandoned this concept by considering time and space related with each other via the Lorentz transformations in the absence of acceleration resulting from gravity or other forces. Accordingly, neither an absolute “now” nor an absolute geometrical space does exist. Therefore, there is no universal present moment since the time is measured by a clock in the rest frame of the observer.

Physical measurements of time are made by clocks, which do not move in the inertial system of the observer. Since we can only measure the time in this system, we cannot determine a universal present moment. Therefore, the question remains if a universal time exists, which governs the dynamics of the universe in the four-dimensional Minkowski space. Because this time must be a Lorentz scalar, we can speculate that its conjugate momentum variable is proportional to the rest mass. Hence, without mass, there will be no universal time. As a consequence, the universal time must start with the origin of mass formed at the Big Bang. Moreover,

we can conceive the universal time as a hidden Bell parameter with a realistic physical property [171]. In the past, this time has been considered merely as a convenient calculus parameter. Contrary to this view, we assume the existence of a true universal time that governs the dynamics of particles in four-dimensional space. Our approach allows us to apply Newton's approach to the Minkowski space by considering the universal time as a one-dimensional absolute quantity which represents a series of locations in the absolute four-dimensional space. The assumption of a universal time allows us to formulate a relativistic dynamics of particles in the Minkowski space and to describe their spin precession in a covariant form. We obtain the equation for the particle motion in Minkowski space most conveniently from Hamilton's principle of classical mechanics according to which the action W is an extremum. Variational principles are powerful tools in physics. They date back to the Greek philosophers who applied them for explaining the motion of planets and the reflection of light.

The incorporation of the universal time as the independent Lorentz-invariant variable also avoids the need for statistical or probability descriptions, because it becomes possible to describe the motion of the constituent particles of an ensemble separately as long as interference effects can be neglected. To elucidate this behavior, we consider the classical motion of particles emanating from a point source in a static three-dimensional field. By solving the equations of motion, we obtain the position of each particle as a function of the laboratory time t . Hence, if the position of the point source and the velocity vectors of the particles are given at some initial time $t = t_i$, we can precisely determine their positions at any later time $t > t_i$. Because the forces of static fields are conservative, we can use the relation for the conservation of energy to substitute any spatial coordinate for the time. In this case, the particle ensemble is described by a homocentric bundle of trajectories, each of which represents the path of a particle. However, by using this procedure, we have lost information because we can no longer distinguish particles traveling along the same trajectory or determine the position of the particles at a given time. Using the number of trajectories per unit area as a measure, we can determine the probability to find a particle at a given position if the current density of the source is known. The same situation arises in Minkowski space if we substitute the laboratory time $t = -ix_4c$ for the universal time τ because in this space x_4 has the role of a spatial coordinate.

We consider the electron as a spinning particle in Minkowski space and assume that it has an intrinsic time-like rotation and an intrinsic space-like rotation which is the spin. Moreover, we suppose that the intrinsic time-like angular rotation defines the charge. Hence, if the particle reverses its direction of rotation perpendicular to the time-like two-dimensional hypersurface in Minkowski space, it converts to its antiparticle in the three-dimensional laboratory system. An electron flipping its rotation in Minkowski space represents a positron in the laboratory frame. Reversal of the time-like angular momentum component requires an energy transfer of $2\hbar\omega_C = 2m_e c^2$ which is emitted as a photon in the case of electron-positron annihilation. This process compensates for the time-like rotations (charges), and it adds up the space-like rotations (spins). Accordingly, the quantum number for

the angular momentum of the photons must be 1 and their charge zero. These considerations differ from those of Feynman [172], who considered the positron as an electron flying backward in time. Within the frame of our model, particles flying backward in time represent dark-matter particles.

14.1 Covariant Hamilton Formalism

The Lagrangian treatment of classical mechanics is based on Hamilton's principle $\delta W = 0$. It states that the action along the true path of a particle is an extremum, in general a minimum. In Minkowski space, the action

$$W_4 = Ex \int_{\tau_0}^{\tau} L_4 d\tau, \quad L_4 = T_4 - V_4 \quad (14.1)$$

is the time integral of the four-dimensional Lagrangian L_4 taken along the true path of the particle traveling from its position $x_{0\mu}$, $\mu = 1, 2, 3, 4$ at the initial universal time τ_0 to its position x_μ , $\mu = 1, 2, 3, 4$ at time τ . In the most general case, the four-dimensional action depends on τ_0 , τ and on the initial and final position coordinates:

$$W_4 = W_4(x_\mu, \tau; x_{0\mu}, \tau_0). \quad (14.2)$$

Because the relativistic Lagrangian L_4 is a Lorentz-invariant scalar function, it must contain terms of entire scalar nature, such as a scalar function T_0 , scalar products formed by polar 4-vectors T_μ , second-rank 4-tensors $T_{\mu\nu}$, axial 4-vectors $T_{\mu\nu\lambda}$, and pseudo 4-scalars T_{1234} . We restrict our treatment to scalar interactions and to electromagnetic interactions described by scalar products between 4-vectors and antisymmetric second-rank 4-tensors. In order to be in accordance with nonrelativistic classical mechanics, L_4 must vanish for a particle at rest in Minkowski space. The kinetic energy T_4 is composed of the translational energy T_{4t} , the kinetic energy of rotation T_{4r} , and the radiation energy T_{rad} emitted by the electron:

$$T_4 = T_{4t} + T_{4r} + T_{\text{rad}}. \quad (14.3)$$

The expressions for the three energies must be Lorentz invariant. In order to obtain proper terms, we extend formally the definitions of classical mechanics from three to four dimensions:

$$T_{4t} = \frac{m_e}{2} \sum_{v=1}^4 \dot{x}_v^2, \quad T_{4r} = \frac{1}{2} \sum_{\mu,\nu} S_{\mu\nu} \omega_{\mu\nu}. \quad (14.4)$$

The four-dimensional definition of the translational energy is a straightforward extension of the classical kinetic energy of a point-like particle with rest mass m_e . Dots denote derivatives with respect to the universal time τ . The translational energy

T_{4i} in Minkowski space is always negative regardless of the sign of $\dot{x}_4$. A negative time-like component of the velocity 4-vector describes an electron which travels in Minkowski space backward in time. According to Feynman, this can be conceived as a positron traveling forward in time in the three-dimensional laboratory system. The expression $-imc\dot{x}_4$ defines the total energy of the particle in the conventional three-dimensional space. This energy is positive for the electron and negative for the positron. Within the frame of our considerations, this energy represents the fourth component of the kinetic moment 4-vector apart from a factor c . Although this vector is Lorentz invariant, its components are not. On the other hand, our formulation guarantees that the total energy in Minkowski space is a Lorentz-invariant scalar quantity.

The angular velocity of the electron in Minkowski space is proportional to its intrinsic rotation (spin). This behavior does not hold true for the laboratory system which may rotate with respect to that at which the electron is at rest. Such a rotation occurs when the velocity of the particle changes its direction for what reason ever.

The motion of the electron is governed by the external forces. The acceleration of the electron and the temporal change of its spin necessarily involve the emission of radiation. This radiation carries off energy, momentum, and angular momentum. Consequently the motion of the electron and the precession of its magnetic moment are affected by the emission of the radiation. Because the magnetic dipole radiation of the electron is small compared with the radiation resulting from its accelerated charge, we can neglect the former radiation. We include the reactive effects of radiation by taking into account the power $P(\tau)$ of the radiation emitted by the accelerated charge. We suppose that the covariant expression for the radiation power has the form

$$T_{\text{rad}}(\tau) = \int_{\tau_i}^{\tau} P(\tau) d\tau = \frac{e^2}{6\pi\epsilon_0 c^3} \sum_{\mu=1}^4 \int_{\tau_i}^{\tau} \ddot{x}_\mu^2 d\tau = m_e \tau_r \sum_{\mu=1}^4 \int_{\dot{x}_{\mu i}}^{\dot{x}_\mu} \ddot{x}_\mu d\dot{x}_\mu. \quad (14.5)$$

The characteristic radiation time τ_r is defined by the relation

$$\tau_r = \frac{e^2}{6\pi\epsilon_0 m_e c^3} = \frac{2}{3} \frac{r_e}{c} \approx 6.2 \cdot 10^{-24} \text{ s}. \quad (14.6)$$

Here, $r_e = e^2/4\pi\epsilon_0 m_e c^2$ denotes the classical electron radius.

By going from three to four dimensions, axial vectors are described by anti-symmetric second-rank tensors because a four-dimensional cubic volume has 12 surfaces, twice as many as the three-dimensional cube. Hence, the four-dimensional angular velocity of the particle is defined by its 12 components $\omega_{\mu\nu}(\tau)$, each of which is the projection of the four-dimensional angular velocity onto the normal of the surface element $d\sigma_{\mu\nu}$. Because the normal vectors of conjugate top and bottom surfaces have opposite directions, only six independent components $\omega_{\mu\nu} = -\omega_{\nu\mu}$ exist. The same relations hold true for the components $S_{\mu\nu} = S_{\mu\nu}(\tau)$ of the angular

momentum tensor or spin tensor, respectively. We describe the velocity of the particle in the four-dimensional space by the components of the velocity 4-vector

$$\dot{x}_\mu = \dot{x}_\mu(\tau) = \frac{dx_\mu(\tau)}{d\tau}, \quad \tau = 1, 2, 3, 4. \quad (14.7)$$

The antisymmetric spin tensor with components $S_{\mu\nu} = S_{\mu\nu}(\tau) = -S_{\nu\mu}$ has three real space-like components and three imaginary time-like components:

$$\begin{aligned} S_{12} &= S_{mz}, & S_{13} &= -S_{my}, & S_{23} &= S_{mx}, \\ S_{14} &= iS_{ex}, & S_{24} &= iS_{ey}, & S_{34} &= iS_{ez}. \end{aligned} \quad (14.8)$$

The space-like components are of magnetic nature, whereas the time-like components are of electric nature, as it is the case for the components of the electromagnetic field tensor

$$F_{\mu\nu} = \frac{\partial A_\mu}{\partial x_\nu} - \frac{\partial A_\nu}{\partial x_\mu}, \quad \mu, \nu = 1, 2, 3, 4. \quad (14.9)$$

This equation relates the components A_μ of the four-dimensional magnetic vector potential ($\vec{A}, A_4 = i\varphi/c$) with the components of the electromagnetic field defined as

$$\begin{aligned} F_{12} &= B_z, & F_{13} &= -B_y, & F_{23} &= B_x, \\ F_{14} &= iE_x/c, & F_{24} &= iE_y/c, & F_{34} &= iE_z/c. \end{aligned} \quad (14.10)$$

A particle, which is at rest in the three-dimensional coordinate system, moves in the corresponding four-dimensional Minkowski space with imaginary velocity $\dot{x}_4 \neq 0$. Within the frame of this coordinate system, we have

$$\begin{aligned} \dot{x}_1 &= \dot{x}_2 = \dot{x}_3 = 0, & \dot{x}_4 &= \pm ic, \\ S_{14} &= S_{23} = S_{34} = 0, & S_{23} &= s_z, S_{31} = s_y, S_{12} = s_x. \end{aligned} \quad (14.11)$$

In accordance with Feynman, we attribute the plus sign of the imaginary velocity to the electron and the minus sign to the positron. This definition allows us to consider the positron as an electron, which moves backward in time. In the three-dimensional rest frame of the particle, the imaginary time-like spin components $S_{\mu 4} = -S_{4\mu}$ and the real velocity components are zero. In this case, the spin tensor vector reduces to the conventional three-dimensional spin vector $\vec{s} = \vec{e}_x s_x + \vec{e}_y s_y + \vec{e}_z s_z$. We can conceive the imaginary components of the spin tensor (14.8) as an electric moment induced by the motion of the magnetic moment associated with the spin. Because we can assume the electron as a spherically symmetric particle, its spin and its angular frequency with components $\omega_{\mu\nu}(\tau) = -\omega_{\nu\mu}(\tau)$ have the same direction in Minkowski space.

The absolute value of the spin is a constant of motion satisfying the condition

$$\vec{s}^2 = \frac{1}{2} \sum_{\mu, \nu} S_{\mu\nu}^2 = \frac{\hbar^2}{4}. \quad (14.12)$$

In accordance with the properties of the spin, we assume that the absolute value of the angular velocity of the particle is a constant of motion too. As a consequence, the rotational energy of the electron in Minkowski space is also a constant of motion:

$$T_{4r} = \frac{1}{2} \sum_{\mu, \nu} S_{\mu\nu} \omega_{\mu\nu} = \vec{s} \vec{\omega}_s = \frac{\omega_s \hbar}{2} = a \vec{s}^2 = a \frac{\hbar^2}{4}. \quad (14.13)$$

Here, $\vec{\omega}_s = a \vec{s}$ denotes the three-dimensional angular velocity vector in the rest frame. The absolute value ω_s of the intrinsic angular velocity is obtained from the condition that the total kinetic energy in Minkowski space must be zero in the absence of external fields, giving

$$T_4 = T_{4t} + T_{4r} + T_{\text{rad}} = -m_e c^2 / 2 + \hbar \omega_s / 2 + 0 = 0. \quad (14.14)$$

The result reveals that the absolute value of the intrinsic angular velocity $\omega_s = m_e c^2 / \hbar = k_{CC} = \omega_C$ is identical with the Compton frequency, and the rotational energy is half the rest energy $m_e c^2$. Using this result, we readily find from (14.13)

$$a = \frac{2\omega_C}{\hbar}, \quad T_{4r} = \frac{m_e c^2}{2}. \quad (14.15)$$

We construct the covariant interaction energy V_4 for the charged particle in external fields by considering terms that involve tensors up to rank 2 inclusively:

$$V_4 = V_0 + V_v + V_s. \quad (14.16)$$

The first term $V_0 = V_0(x_v, \tau)$ describes a scalar coupling and the second term a vector coupling, given by

$$V_v = e \sum_{\nu=1}^4 A_\nu \dot{x}_\nu. \quad (14.17)$$

This interaction term results from the coupling of the 4-current $\vec{J} = (\vec{j}, e\dot{x}_4)$ with the external electromagnetic 4-vector potential. The third term accounts for the coupling of the spin tensor with the electromagnetic field tensor:

$$V_s = \frac{e}{2m_e} \sum_{\mu, \nu} F_{\mu\nu} S_{\mu\nu}(\tau). \quad (14.18)$$

The factor $e/2m_e c$ represents the classical gyromagnetic ratio relating the magnetic moment of the electron with its angular momentum. Within the frame of our four-dimensional approach, we do not need to introduce the Lande factor $g = 2$ because we have taken it implicitly into account by the double summation in (14.18). We can understand the origin of the Lande factor by considering that each component $S_{\mu\nu}$ of the spin tensor describes a rotation in the (x_μ, x_ν) plane. In the four-dimensional space, the surface of a cube consists of 12 subsurfaces, twice as many as in the case of a three-dimensional cube. The introduction of the spin term seems to be somewhat artificial because it is widely believed that the spin interaction is a quantum-mechanical effect which results from the Dirac equation. However, we shall show later that it is also possible to derive the Dirac equation from a four-dimensional Hamilton–Jacobi equation in a similar way as the nonrelativistic Schrodinger equation is obtained from the three-dimensional Hamilton–Jacobi equation of classical mechanics.

The potentials V_c and V_s are of entirely electromagnetic nature, whereas the potential V_0 is not. This potential describes the coupling of two scalar quantities. If we suppose that this potential energy accounts for the gravitation, it has the form

$$V_0 = -m_e G. \quad (14.19)$$

In this case, the scalar function $G = G(x_\nu, \tau)$ is the gravitation potential. The potential energy V_s is gauge invariant, whereas the potential energies V_0 and V_c are not. This ambiguity is of no importance because the equations of motion of the particle are invariant under gauge transformations of the potentials G and $\vec{A}$, respectively. Considering the relations (14.4), (14.17), and (14.18), the four-dimensional Lagrangian (14.2) adopts the form

$$L_4 = \sum_{\mu=1}^4 \left[\frac{m_e}{2} \dot{x}_\mu^2 + m_e \tau_r \int_{\dot{x}_{\mu i}}^{\dot{x}_\mu} \ddot{x}_\mu d\dot{x}_\mu - e \dot{x}_\mu A_\mu \right] - \frac{m_e}{2} c^2 - \frac{e}{2m_e} \sum_{\mu,\nu} F_{\mu\nu} S_{\mu\nu} - V_0. \quad (14.20)$$

By partial differentiation with respect to the velocity components, we define the components of the canonical momentum 4-vector as

$$p_\mu = \frac{\partial L_4}{\partial \dot{x}_\mu} = m_e (\dot{x}_\mu + \tau_r \ddot{x}_\mu) - e A_\mu, \quad \mu = 1, 2, 3, 4. \quad (14.21)$$

These components differ from the components of the classical canonical momentum by the term $m_e \tau_r \ddot{x}_\mu$ which we can conceive as the repulsion of the emitted radiation.

14.2 Propagation in Minkowski Space

In order to account most conveniently for relativistic effects, it is advantageous to describe the motion of particle in the four-dimensional Minkowski space as a function of the universal time. As a result, the time-like velocity component $\dot{x}_4$ of the particle can change its direction in Minkowski space apart from the three spatial components. According to Feynman, such a change is related to the creation and annihilation of particles.

14.2.1 Path Equation

We derive the path equations most conveniently from the action integral (14.1) by employing Hamilton's principle of least action $\delta W_4 = 0$. This condition states that among all possible paths along which the particle may move from one point to another in four-dimensional space within a certain time interval $\tau - \tau_0$, the actual path yields an extremum for the action. In order to obtain the equations for the particle motion, we cannot directly apply the standard calculation of variation because the radiation energy is an integral expression whose integrand depends on the acceleration of the particle. Accordingly, the equations resulting from the variation will differ from the standard Euler–Lagrange equations. Considering the representation (14.20) of the Lagrangian L_4 and using the last expression of the relations (14.5) for the radiation term, the variation of the action (14.1) for fixed boundaries ($\delta x_\mu(\tau) = \delta x_\mu(\tau_0) = 0$) gives

$$\delta W_4 = \int_{\tau_0}^{\tau} \sum_{\mu=1}^4 \left(\frac{L_4}{\partial x_\mu} - \frac{d}{d\tau} \frac{\partial L_4}{\partial \dot{x}_\mu} \right) \delta x_\mu d\tilde{\tau} = 0. \quad (14.22)$$

Because we can perform the small variation $\delta x(\tilde{\tau})$ of the path at any time within the interval $\tau < \tilde{\tau} < \tau_0$, the action vanishes only if

$$\frac{d}{d\tau} \frac{\partial L_4}{\partial \dot{x}_\mu} - \frac{\partial L_4}{\partial x_\mu} = 0. \quad (14.23)$$

Inserting the expression (14.20) into (14.23) and taking into account (14.21), we eventually obtain the path equations

$$m_e \ddot{x}_\mu + m_e \tau_r \ddot{x}_\mu - e \sum_v \dot{x}_v F_{\mu v} = + \frac{e}{2m_e} \frac{\partial}{\partial x_\mu} \sum_{\lambda, v} S_{\lambda v} F_{\lambda v} - \frac{\partial V_0}{\partial x_\mu} + e \frac{\partial A_\mu}{\partial \tau}, \mu = 1, 2, 3, 4. \quad (14.24)$$

For deriving the last term on the left-hand side of (14.24), we employ the relation

$$\frac{dA_\mu}{dt} = \frac{\partial A_\mu}{\partial t} + \sum_{v=1}^4 \frac{\partial A_\mu}{\partial x_v} \dot{x}_v. \quad (14.25)$$

With (14.25) we obtain

$$\frac{dA_\mu}{dt} - \sum_{v=1}^4 \dot{x}_\mu \frac{\partial A_v}{\partial x_\mu} = \frac{\partial A_\mu}{\partial t} + \sum_{v=1}^4 \left(\frac{\partial A_\mu}{\partial x_v} - \frac{\partial A_v}{\partial x_\mu} \right) \dot{x}_v = \frac{\partial A_\mu}{\partial t} + \sum_{v=1}^4 F_{\mu v} \dot{x}_v. \quad (14.26)$$

The equations of motion (14.24) are invariant under gauge transformations of the potentials G and $\vec{A}$, respectively, although the Lagrangian (14.20) and hence the action (14.1) will change. By introducing new gauges $G \rightarrow G + G_0$ and $\vec{A} \rightarrow \vec{A} + \text{Grad } \Lambda$, the action (14.1) transforms to

$$\tilde{W}_4 = W_4 + m_e [G_0(\tau_0) - G_0(\tau_i)] + e [\Lambda(x_v(\tau_0)) - \Lambda(x_v(\tau_i))]. \quad (14.27)$$

The result shows that the change of the gauge of the potentials adds only a constant value to the original action W_4 . The variation of this contribution is zero because the boundaries are kept fixed ($\delta\tau = \delta\tau_0 = 0$).

14.2.2 Hamiltonian

We obtain the conservation of the total energy in Minkowski space by multiplying the path equation (14.25) with $\dot{x}_\mu$ and by summing subsequently over the index μ , giving

$$\begin{aligned} \frac{d}{d\tau} \left\{ \frac{m_e}{2} \left[\sum_{\mu} \dot{x}_\mu^2 - c^2 \right] + V_0 + V_t + \frac{e}{2m_e} \sum_{\mu, v} S_{\mu v} F_{\mu v} \right\} \\ = +m_e \tau_0 \sum_{\mu} \ddot{x}_\mu \dot{x}_\mu - e \sum_{\mu} \dot{x}_\mu \frac{\partial A_\mu}{\partial \tau} + \frac{\partial(V_t + V_0)}{\partial \tau}. \end{aligned} \quad (14.28)$$

In order that the dynamics of the particle is stationary in Minkowski space, the terms on the right-hand side of (14.28) must vanish. This condition is satisfied if we assume that the external potentials V_0 and A_μ do not depend explicitly on the universal time and if

$$\frac{\partial V_t}{\partial \tau} = \frac{e}{2m_e} \sum_{\mu, v} \dot{S}_{\mu v} F_{\mu v} = 0. \quad (14.29)$$

$$\frac{\partial A_\mu}{\partial \tau} = -\frac{m_e \tau_r}{e} \ddot{x}_\mu. \quad (14.30)$$

Condition (14.29) holds always true, as will be shown in Sect. 14.6. Condition (14.30) implies that the external field must compensate for the radiation emitted by the electron, as it is the case in electron synchrotrons. We may also encounter this situation in atoms realized by virtual emission and absorption of photons between the electron and the nucleus. The situation differs if $\vec{A}$ and $A_4 = i\varphi/c$ do not depend on the universal time ($\partial A_\mu/\partial\tau = 0$). In this case, the motion of the electron is affected by the radiation emitted if $\ddot{x}_\mu \neq 0$. Hence, we can readily perform the integration in Eq. (14.30). The result defines the 4-Hamiltonian.

In order to cast the 4-Hamiltonian in the appropriate form, we partially integrate the radiation term (14.5), giving

$$\begin{aligned} m_e \tau_r \sum_\mu \int_{\tau_0}^\tau \ddot{x}_\mu \dot{x}_\mu d\tau &= m_e \tau_r \sum_\mu \ddot{x}_\mu \dot{x}_\mu - m_e \tau_r \sum_\mu \int_{\tau_0}^\tau \ddot{x}_\mu^2 d\tau \\ &= \tau_r \frac{d}{d\tau} \sum_{\mu=1}^4 \frac{m_e}{2} \dot{x}_\mu^2 - m_e \tau_r \sum_\mu \int_{\tau_0}^\tau \ddot{x}_\mu^2 d\tau. \end{aligned} \quad (14.31)$$

The first term of the second row represents the radiant repulsion energy which is transferred to the electron by the emitted radiation. This term is proportional to the temporal rate at which the radiation transfers kinetic energy to the particle. Because this term is a total differential, it adds to the action (14.1) terms which depend only on the boundaries. Since the variation of the action is performed for fixed boundaries, the first term of the second row of (14.31) does not affect the path equation (14.24). The negative second term accounts for the total energy carried away by the radiation during the time interval $\tau - \tau_0$. Both terms cancel out each other in the case of constant linear acceleration in Minkowski space. The result can be cast in the form

$$\begin{aligned} \frac{d}{d\tau} \left\{ \frac{m_e}{2} \left[\sum_\mu \left(\dot{x}_\mu^2 + 2\tau_r \ddot{x}_\mu \dot{x}_\mu \right) - c^2 \right] + V_0 - \frac{e}{2m_e} \sum_{\mu,\nu} S_{\mu\nu} F_{\mu\nu} \right\} - m_e \tau_r \sum_\mu \ddot{x}_\mu^2 \\ = -e \sum_\mu \dot{x}_\mu \frac{\partial A_\mu}{\partial\tau} + \frac{\partial(V_s + V_0)}{\partial\tau}. \end{aligned} \quad (14.32)$$

The term $m_e \tau_r \sum_\mu \ddot{x}_\mu \dot{x}_\mu$ in the expression vanishes if the acceleration 4-vector is perpendicular to the velocity 4-vector. This behavior may be the reason why radiation losses are larger in circular accelerators than in linear accelerators.

In the following, we assume that the external potentials $\vec{A}$, $A_4 = i\varphi/c$ and V_0 do not depend explicitly on the universal time τ :

$$\frac{\partial A_\mu}{\partial\tau} = 0, \quad \frac{\partial V_0}{\partial\tau} = 0. \quad (14.33)$$

In this case, the terms on the right-hand side of Eq. (14.32) vanish provided that

$$\frac{\partial V_s}{\partial \tau} = \frac{e}{2m_e} \sum_{\mu, \nu} \dot{S}_{\mu\nu} F_{\mu\nu} = 0. \quad (14.34)$$

We shall prove in subsequent considerations for the spin precession that the condition (14.34) holds always true. By considering this behavior, we can readily integrate Eq. (14.32). The result defines the 4-Hamiltonian:

$$\begin{aligned} H_4 &= H_{4e} + H_{\text{rad}} = H_{40}, \\ H_{4e} &= \frac{m_e}{2} \sum_{\mu} (\dot{x}_{\mu} + \tau_r \ddot{x}_{\mu})^2 - \frac{m_e}{2} c^2 - \frac{m_e \tau_r^2}{2} \ddot{x}_{\mu}^2 \\ &\quad + \frac{e}{2m_e} \sum_{\mu, \nu} S_{\mu\nu} F_{\mu\nu} + V_0, \\ H_{\text{rad}} &= T_{4\text{rad}} = -m_e \tau_r \sum_{\mu} \int_{\tau_0}^{\tau} \ddot{x}_{\mu}^2 d\tau. \end{aligned} \quad (14.35)$$

We have split up the Hamiltonian $H_4 = H_{4e} + H_{\text{rad}}$ in the term $H_{4e} = H_{4\text{mot}}$ which governs the motion of the electron and the radiation term H_{rad} which accounts for the electromagnetic radiation emitted by the particle. By applying this separation and assuming that the fields in Minkowski space are conservative ($\partial A_{\nu}/\partial \tau = 0$, $\partial V_0/\partial \tau = 0$), we obtain from (14.20) and (14.32) the relations

$$\begin{aligned} \frac{dH_{4e}}{d\tau} + \frac{dH_{\text{rad}}}{d\tau} &= 0, \\ \frac{dH_{\text{rad}}}{d\tau} &= -m_e \tau_r \sum_{\mu} \ddot{x}_{\mu}^2 = \frac{\partial L_4}{\partial \tau}. \end{aligned} \quad (14.36)$$

Conservative fields in Minkowski space imply that particles are neither created nor annihilated. In this stationary case, the total four-dimensional Hamiltonian $H_4 = H_{4\text{mot}} + H_{\text{rad}}$ is conserved. As a consequence, the action $W_4 = W_4(x_{\mu}, \tau; x_{0\mu}, \tau_0) = W_4(x_{\mu}, x_{0\mu}; \tau - \tau_0)$ depends only on the time interval $\tau - \tau_0$. An explicit dependence of the fields on the universal time τ will enable one to describe the creation and annihilation of particles within the frame of the dynamics in the four-dimensional Minkowski space.

If the electron propagates in field-free Minkowski space ($V_0 = 0$, $F_{\mu\nu} = 0$, $\ddot{x}_{\mu} = \ddot{x}_{\mu} = 0$), its time-like velocity component $\dot{x}_{40} = ic$ equals the velocity of light apart from the imaginary factor i . By assuming this situation, we find that the constant energy $H_{40} = -E_0 = -m_e c^2$ of the 4-Hamiltonian equals the negative rest energy of the particle:

$$H_4 = \frac{1}{2} m_e (\dot{x}_{40}^2 - c^2) = H_{40} = -m_e c^2. \quad (14.37)$$

This Lorentz-invariant total energy should not be confused with the fourth component $m_e \dot{x}_4$ of the momentum 4-vector. Because E_0 does not depend on the sign of $\dot{x}_4$, this energy must be the same for the particle and its antiparticle.

It should be noted that covariant Hamiltonians suggested so far in the literature only consider the electromagnetic vector potential [39]. Within the frame of these approaches, the energy is either zero or $-mc^2/2$, respectively. Both energies differ from the rest energy of the particle in the absence of radiation. Our gauge of the Lagrangian is in accordance with that in classical mechanics. In the absence of radiation, the four-dimensional Hamiltonian is identical with the negative rest energy of the particle in any laboratory frame. We can conceive the rest energy $m_e c^2 = E_0$ as the canonic conjugate “variable” of the universal time τ in the same way as the energy in three-dimensional space is the conjugate variable to the laboratory time t . This conjecture implies that the universal time becomes obsolete or meaningless, respectively, for massless particles. Therefore, we can conclude that τ is not some meaningless Lorentz-invariant parameter but may have a realistic physical meaning in the context of creation and annihilation of particles. The relation (14.36) represents a true dynamical constraint that confines the motion of the particle to a particular three-dimensional hypersurface in the four-dimensional space. Our supposition differs from the conventional assumption that the absolute value of the velocity 4-vector is a constant of motion and equal to c . This condition is used as a definition of the parameter τ , which is considered as the proper time of the particle [37, 39].

The radiation term vanishes in the expression (14.35) of the 4-Hamiltonian if the components $\ddot{x}_\mu$ of the 4-acceleration are constant. This surprising result contradicts classical electrodynamics according to which an accelerated charge radiates. Although the radiation term of the 4-Hamiltonian (14.35) vanishes for constant 4-acceleration, the radiation term (14.6) of the 4-Lagrangian does not. In this case, however, the radiation term becomes a total differential with respect to the independent variable τ of the 4-action (4.1). Because we can integrate this total differential, the result depends only on the fixed boundary values which do not contribute to the variation of S_4 . We can conceive this behavior if we consider that the phase $\Omega = S_4/\hbar$ of the electron wave is a nonmeasurable quantity which can only be determined up to an arbitrary constant value.

By employing the relations (14.20) and (14.21), we can rewrite the 4-Hamiltonian (14.35) in the standard form of classical mechanics:

$$\begin{aligned}
 H_{4e} = H_{4\text{mot}} &= \sum_{\mu=1}^4 \dot{x}_\mu \frac{\partial L_4}{\partial \dot{x}_\mu} - L_4 - H_{4\text{rad}} \\
 &= \frac{1}{2m_e} \sum_{\mu} (p_\mu + eA_\mu)^2 - \frac{m_e}{2} c^2 - \frac{e}{2m_e} \sum_{\mu, \nu} S_{\mu\nu} F_{\mu\nu} + V_0 - \frac{m_e \tau_r^2}{2} \sum_{\mu} \ddot{x}_\mu^2.
 \end{aligned}
 \tag{14.38}$$

This four-dimensional Hamiltonian should not be mixed up with the classical three-dimensional Hamiltonian $H = H_3$. Although our approach is not based on quantum-mechanical considerations, the Hamiltonian (14.38) shows its relation with quantum electrodynamics because it contains a radiation term. However, because τ_r does not involve the Planck constant $\hbar$, the radiation is emitted continuously in our semiclassical four-dimensional approach. The classical relativistic three-dimensional Hamiltonian $H_3 = H_3(x_\nu, \dot{x}_\nu, t); \nu = 1, 2, 3$ is based on the condition

$$\sum_{\mu=1}^4 \dot{x}_\mu^2 = -c^2. \quad (14.39)$$

Relation (14.38) shows that this assumption holds true only if we neglect spin, radiation, and the scalar interaction V_0 . Generally, τ is considered as the so-called eigenzeit of the particle.

14.2.3 Propagation of Free Particles in Minkowski Space

To illustrate the propagation of a point-like particle in Minkowski space, it is advantageous to choose ct as the fourth spatial coordinate instead of the imaginary coordinate $x_4 = ict$. For reasons of simplicity, we consider the motion of a free particle in the x, ct -plane ($y = 0, z = 0$) of the four-dimensional Minkowski space as a function of the universal time τ . The propagation of the particle with rest mass m_0 starts at the origin $x = x_0 = 0, ct = ct_0 = 0$ at universal time $\tau = \tau_0 = 0$. In the absence of external forces, this point is located throughout its entire motion on the action surface

$$W_4 = S_4 - m_0 c^2 \tau = m_0 c \left\{ \sqrt{c^2 t^2 - x^2} - c\tau \right\} = \text{const} = 0. \quad (14.40)$$

The surface of reduced action $S_4 = m_0 c^2 \tau$ propagates in Minkowski space with the velocity of light in the direction of the time-like coordinate ct . It follows from the relation (14.40) that the free particle is located on a hyperbola. To survey the propagation of the wave surface $S_4 = 0$, we rewrite (14.40) in the form

$$L^2 = c^2(t - t_0)^2 - (\vec{r} - \vec{r}_0)^2 = c^2(\tau - \tau_0)^2. \quad (14.41)$$

This equation describes two hyperboloids that propagate from the light cone $|\vec{r} - \vec{r}_0| = c|t - t_0|$ in opposite ct -directions, as illustrated in Fig. 14.1.

In the standard three-dimensional space, the wave surfaces of the propagator are concentric spheres, and the trajectories start from the center $\vec{r}_0$ at laboratory time t_0 in radial directions. In the four-dimensional space, the rays start at universal time $\tau = \tau_0$ from the light cone which represents the source of the four-dimensional elementary wave. The apex of the cone is located at the point $\vec{r} = \vec{r}_0, t = t_0$.

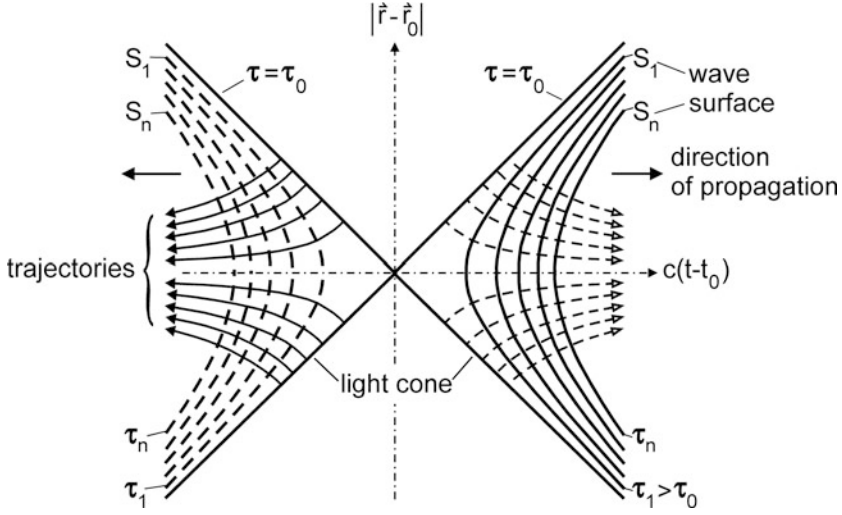
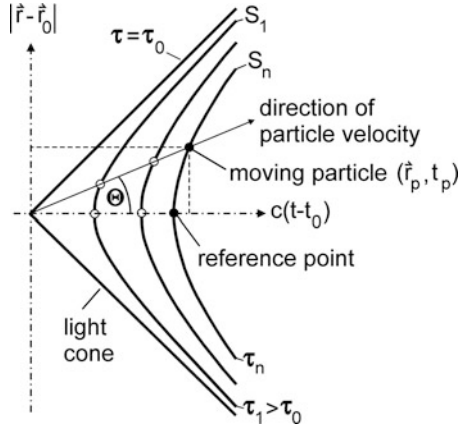


Fig. 14.1 Propagation of the elementary wave in Minkowski space. The wave surfaces $S_n = S_{4n}$, $n = 1, 2, \dots$, are surfaces of constant optical path length L (14.41) taken at universal times $\tau_n > \tau_{n-1}$, $n = 1, 2, \dots$

Fig. 14.2 Trajectory and surfaces $S_n = S_{4n} = m_0 c^2 (\tau_n - \tau_0)$ of constant reduced action of a free particle propagating along a straight line in the x, ct -plane of the four-dimensional Minkowski space

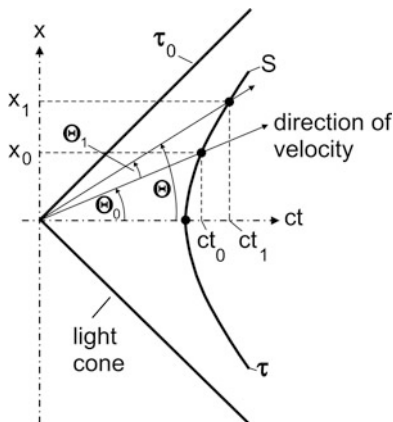


The trajectories form hyperbolas which are orthogonal to the set of hyperbolic surfaces $W_4(\tau_n) = S_{4n} \pm m c^2 (\tau_n - \tau) = 0$ of constant phase taken at times $\tau_n > \tau_{n-1}$, $n = 1, 2, 3, \dots$. Note that the path of a free particle in Minkowski space is a straight line, as illustrated in Figs. 14.2 and 14.3.

Using a parameter representation, the coordinates of the point x, t can be expressed as

$$\begin{aligned} t &= \tau \cosh \Omega, \\ x &= c \tau \sinh \Omega. \end{aligned} \quad (14.42)$$

Fig. 14.3 Motion of the points P_0 and P_1 on the action surface $S = S_4 = c\sqrt{c^2t^2 - x^2} = c\tau$ which propagates in Minkowski space with velocity c in the ct -direction



The components of the velocity in Minkowski space are given by

$$\begin{aligned} c \frac{dt}{d\tau} &= ci = c \cosh \Omega, \\ \frac{dx}{d\tau} &= \dot{x} = c \sinh \Omega. \end{aligned} \quad (14.43)$$

For an observer moving with the action surface S_4 along the axis $x = 0, ct = c\tau$, the point x, ct moves in x -direction with velocity

$$v = \frac{dx}{dt} = \frac{\dot{x}}{\dot{t}} = c \tanh \Omega. \quad (14.44)$$

The parameter Ω is related with the relative velocity $\beta = v/c$ by the relation

$$\Omega = \operatorname{arctanh} \beta = \frac{1}{2} \ln \frac{1 + \beta}{1 - \beta}. \quad (14.45)$$

For a fixed relative velocity β , the free particle moves in Minkowski space along the straight line:

$$\begin{aligned} x &= ct \tanh \Omega = ct\beta = ct \tan \Theta, \\ \Theta &= \arctan \beta. \end{aligned} \quad (14.46)$$

The direction of flight in Minkowski space is determined by the angle $\Theta = \arctan \beta$. The motion of the particle and that of its action surface is illustrated in Fig. 14.2.

It follows from the relations (14.42) that the absolute value of the velocity of a free particle in Minkowski space equals the velocity of light regardless of its velocity v :

$$\sqrt{c^2 \dot{t}^2 - \dot{x}^2} = c \sqrt{\cosh^2 \Omega - \sinh^2 \Omega} = c. \quad (14.47)$$

In the following, we consider two points $P_0(x_0, y_0)$ and $P_1(x_1, y_1)$ which move on the same action surface (14.40). The velocity of the point P_1 measured in the system which moves with this point along the ct -axis is given by

$$\beta_1 = \frac{dx_1}{c dt_1} = \tanh \Omega_1 = \tanh(\Omega_0 + \Omega) = \tan \Theta. \quad (14.48)$$

The parameter Ω determines the velocity difference $v = v_x = c \tanh \Omega = c \tan \Theta_1$ between the two points P_0 and P_1 . They propagate in Minkowski space along straight lines which originate at the apex $x_1 = x_0 = 0, ct_1 = ct_0 = 0$ of the action surface, as shown in Fig. 14.3. By considering the relation (14.44) together with $v = \beta c \tanh \Theta$, we find that the velocity v_1 in the laboratory system is connected with the velocities v_0 and v in this system by the relation

$$\begin{aligned} v_1 &= c \tanh \Omega_1 = c \tanh(\Omega_0 + \Omega) \\ &= c \frac{\tanh \Omega_0 + \tanh \Omega}{1 + \tanh \Omega_0 \tanh \Omega} = \frac{v_0 + v}{1 + v_0 v / c^2} \leq c. \end{aligned} \quad (14.49)$$

The result shows that the velocity measured in any laboratory system cannot exceed the velocity of light.

The coordinates $x_1(\tau), ct_1(\tau)$ of the moving point P_1 in Minkowski space are given by

$$\begin{aligned} ct_1 &= c\tau \cosh(\Omega_0 + \Omega) = c\tau \{\cosh \Omega_0 \cosh \Omega + \sinh \Omega_0 \sinh \Omega\}, \\ x_1 &= c\tau \sinh(\Omega_0 + \Omega) = c\tau \{\sinh \Omega_0 \cosh \Omega + \cosh \Omega_0 \sinh \Omega\}. \end{aligned} \quad (14.50)$$

To obtain the relations existing between the coordinates and the velocity difference v of the two moving points, we consider the motion of the particle x_1, ct_1 measured in the system of an observer who moves with the particle x_0, t_0 in Minkowski space. In addition, we employ the relation

$$\gamma = \cosh \Omega = \frac{1}{\sqrt{1 - \tanh^2 \Omega}} = \frac{1}{\sqrt{1 - \beta^2}}. \quad (14.51)$$

By inserting the relations (14.42) and (14.51) into (14.50) and writing x', t' for x_1, t_1 and x, t for x_0, t_0 , we readily obtain the standard Lorentz transformations:

$$\begin{aligned} x' &= c\tau \cosh \Omega [\sinh \Omega_0 + \tanh \Omega \cosh \Omega_0] \\ &= \gamma (x + vt), \\ t' &= \tau \cosh \Omega [\cosh \Omega_0 + \tanh \Omega \sinh \Omega_0] \\ &= \gamma (t + \beta x / c). \end{aligned} \quad (14.52)$$

According to our considerations, all inertial systems propagate with the velocity of light in the direction of the time-like coordinate in the four-dimensional Minkowski space regardless of their velocity in the three-dimensional space.

14.3 Four-Dimensional Hamilton–Jacobi Equation

To obtain the equivalence of the conventional Hamilton–Jacobi (HJ) equation for the four-dimensional space, we vary the four-dimensional action function (14.2) with respect to the universal time τ at the endpoint. Considering that $\delta W_4 = 0$ for fixed boundaries, the variation of (14.1) at the point of observation by the infinitesimal time element $\delta\tau$ gives

$$\delta W_4 = L_4 \delta\tau + \sum_{\mu=1}^4 \frac{\partial L_4}{\partial \dot{x}_\mu} \delta x_\mu(\tau). \quad (14.53)$$

To derive the differential displacement at time $\tau + \delta\tau$, we utilize the relation

$$\delta x_\mu(\tau + \delta\tau) \approx \delta x_\mu(\tau) + \dot{x}_\mu(\tau) \delta\tau. \quad (14.54)$$

Substituting $\delta x_\mu(\tau)$ in (14.53) by means of the relation (14.54), we obtain

$$\delta W_4 = \left[L_4 - \sum_{\mu} \dot{x}_\mu \frac{\partial L_4}{\partial \dot{x}_\mu} \right] \delta\tau + \sum_{\mu} \frac{\partial L_4}{\partial \dot{x}_\mu} \delta x_\mu(\tau + \delta\tau). \quad (14.55)$$

We can perform the spatial and temporal variations of the endpoint arbitrarily. Performing the variation with respect to each differential quantity gives

$$\begin{aligned} \frac{\partial W_4}{\partial x_\mu} &= \frac{\partial L_4}{\partial \dot{x}_\mu} = p_\mu = m_e(\dot{x}_\mu + \tau_{r0} \ddot{x}_\mu) - e A_\mu, \\ \frac{\partial W_4}{\partial \tau} &= L_4 - \sum_{\mu=1}^4 \dot{x}_\mu \frac{\partial L_4}{\partial \dot{x}_\mu} = -H_{4e} - H_{4r}, \\ \frac{\partial^2 W_4}{\partial \tau^2} &= -\frac{\partial H_{4r}}{\partial \tau} = m_e \tau_r \sum_{\mu} \ddot{x}_\mu^2. \end{aligned} \quad (14.56)$$

The explicit dependence of the Hamiltonian on $\ddot{x}_\mu$ and τ is caused by the radiation terms in the expression (14.38) for H_4 . Partial differentiation of the second relation of (14.56) with respect to τ gives

$$\frac{\partial^2 W_4}{\partial \tau^2} = -\frac{\partial H_4}{\partial \tau} = -\frac{dH_{4\text{rad}}}{d\tau} = m_e \tau_r \sum_{\mu} \ddot{x}_\mu^2. \quad (14.57)$$

By employing this relation, (14.18) and (14.38), we obtain from (14.56) the equation

$$\frac{\partial W_4}{\partial \tau} + \frac{1}{2m_e} \sum_{\mu=1}^4 \left(\frac{\partial W_4}{\partial x_\mu} + eA_\mu \right)^2 - \frac{m_e}{2} c^2 + V_s + V_0 = \frac{\tau_r}{2} \frac{\partial^2 W_4}{\partial \tau^2} + \int_{\tau_0}^{\tau} \frac{\partial^2 W_4}{\partial \tau^2} d\tau. \quad (14.58)$$

To obtain the dynamics of the electron, we separate the total action $W_4 = W_{4e} + W_{4r}$ into the term $W_{4e} = W_{4\text{mot}}$ which accounts for the particle motion and the term W_{4r} which accounts for the radiation. We further simplify Eq. (14.58) by making use of the second equation of (14.56), giving

$$\int_{\tau_0}^{\tau} \frac{\partial^2 W_4}{\partial \tau^2} d\tau = \frac{\partial W_{4r}}{\partial \tau} - \frac{\partial W_{4r}}{\partial \tau} \Big|_{\tau_0} = -m_e \tau_r \sum_{\mu} \int_{\tau_0}^{\tau} \ddot{x}_\mu^2 d\tau. \quad (14.59)$$

In the stationary case, we can put the constant on the right-hand side of (14.59) equal to zero without loss of generality. By inserting the relation (14.59) into (14.58), we obtain the relativistic Hamilton–Jacobi (RHJ) equation for stationary fields as

$$\frac{\partial W_{4e}}{\partial \tau} - \frac{\tau_r}{2} \frac{\partial^2 W_{4e}}{\partial \tau^2} + \frac{1}{2m_e} \sum_{\mu=1}^4 \left(\frac{\partial W_{4e}}{\partial x_\mu} + eA_\mu \right)^2 + \frac{m_e}{2} c^2 + V_s + V_0 = 0. \quad (14.60)$$

This equation is quite universal because it considers electromagnetic interaction, gravitation, and radiation in a correct relativistic way. Moreover, Eq. (14.60) shows that the radiation of the particle results solely from the spin interaction $V_s = V_s(\tau)$ because in the stationary case (14.33) the external fields, A_μ and V_0 , do not depend explicitly on the universal time. The relation (14.100) shows that the electromagnetic field causes a precession of the spin. The moving magnetic dipole produces an electric dipole in the laboratory system. The temporal change of the dipole moments is the source of the electromagnetic radiation.

If we neglect the effect of radiation ($\tau_r = 0$), the Hamiltonian (14.35) reduces to

$$H_4(x_\mu, \dot{x}_\mu, \tau) = \frac{m_e}{2} \sum_{\mu=1}^4 \dot{x}_\mu^2 - \frac{m_e c^2}{2} + V_s + V_0. \quad (14.61)$$

Moreover, the components (14.56) of the canonical 4-momentum adopt the standard form

$$p_\mu = m_e \dot{x}_\mu - eA_\mu. \quad (14.62)$$

By substituting $\partial W_4 / \partial x_\mu + eA_\mu$ for $m_e \dot{x}_\mu$ in the Hamiltonian (14.61), we obtain from the second equation of (14.56) the standard Hamilton–Jacobi equation extended to the Minkowski space:

$$\frac{\partial W_4}{\partial \tau} + H_4 \left(x_\mu, \frac{\partial W_4}{\partial x_\mu}, \tau \right) = 0. \quad (14.63)$$

According to the concepts of quantum mechanics, we may conceive this equation as the short-wavelength limit of a four-dimensional wave equation describing the dynamics of the electron wave in Minkowski space. Because the Hamiltonian (14.61) does not depend explicitly on the universal time τ , we can separate the time variable τ from the four spatial variables x_μ by means of the ansatz:

$$W_4 = S_4 + E_0\tau. \quad (14.64)$$

Inserting this expression into Eq. (14.63) results in the four-dimensional HJ equation for the reduced action $S_4 = S_4(x_\mu, x_{\mu 0}; E_0)$:

$$H_4\left(x_\mu, \frac{\partial S_4}{\partial x_\mu}\right) = -E_0 = -m_e c^2. \quad (14.65)$$

This equation no longer involves the universal time τ and represents the eikonal equation of the quantum-mechanical Klein–Gordon equation.

A constant reduced action $S_4 = 0$ represents a continuous set of surfaces in the four-dimensional Minkowski space:

$$S_4(x_\mu, x_{\mu 0}; E_0) = -E_0(\tau - \tau_1). \quad (14.66)$$

The relation

$$\frac{\partial S_4}{\partial x_\mu} = \frac{\partial W_4}{\partial x_\mu} = p_\mu = m_e \dot{x}_\mu - e A_\mu \quad (14.67)$$

demonstrates that in field-free space ($\vec{A} = 0$), the trajectories of all identical particles emanating from the singular initial point form the orthogonal trajectories to the surfaces of constant reduced action $S_4(x_\mu, x_{\mu 0}, E_0)$.

Note that the action functions S_4 and W_4 do not contain the initial velocity coordinates as variables explicitly. This behavior resembles the uncertainty principle of quantum mechanics, according to which it is not possible to precisely determine the position and the momentum of a given state. We obtain the initial canonical momentum by differentiating $S_4(x_\mu, x_{\mu 0})$ with respect to the components $x_{\mu 0}$ of the initial position vector, giving

$$p_{\mu 0} = -\frac{\partial S_4}{\partial x_{\mu 0}} = -\frac{\partial W_4}{\partial x_{\mu 0}}, \quad \mu = 1, 2, 3, 4. \quad (14.68)$$

Owing to the existence of the action surface, the trajectories of identical particles emanating from a common point in the four-dimensional space are correlated because their canonical momenta are orthogonal to this surface. It changes its shape when it propagates through the field region, yet it will never be torn apart. However, the surface can degenerate in sheets that intersect each other, forming a caustic. The caustic represents the loci of the intersections of neighboring *rays*, which start in

different directions from the point source, as discussed extensively in Sect. 6.4. If we take into account the wave nature of the electron, the surfaces of constant action also represent wave surfaces of constant phase. The action is a minimum for all points located in front of the caustic and may become a maximum if the end points are located behind the caustic. Note that the particle description breaks down in the region of the caustic because of pronounced interference effects.

14.4 Relativistic Maupertuis Principle

The relation (14.38) between the Hamiltonian and the Lagrangian L_4 differs from the corresponding relation of classical mechanics by the radiation ad spin terms. The four-dimensional Hamiltonian (14.35) is a constant of motion if the fields do not depend explicitly on the universal time. We can conceive this relation as a constraint for the motion of the particle in Minkowski space. Moreover, we can use the constraint to replace the independent variable τ by the time $t = -ix_4/c$. Neglecting the radiation effects, we readily derive from the Hamiltonian (14.61) that the differential time elements $d\tau$ and $d\tau$ are related by

$$\frac{d\tau}{dt} = \pm \frac{1}{\gamma} = \sqrt{\frac{1 - \beta^2}{1 + 2V_{0\text{ret}}/m_e c^2 + e \sum_{\mu, \nu} S_{\mu\nu} F_{\mu\nu})/m_e^2 c^2}}, \quad \beta^2 = \vec{\beta}^2 = \left(\frac{d\vec{r}}{c dt} \right)^2. \quad (14.69)$$

The plus sign is attributed to the electron, the minus sign to the positron. Note that the substitution of the laboratory time for the universal time affects the potential energy V_0 which we must replace by the retarded potential $V_{0\text{ret}} = V_{0\text{ret}}(\vec{r}, t; \vec{\beta})$ in order to satisfy causality. The dependence of $V_{0\text{ret}}$ on the velocity $\vec{v} = \dot{\vec{r}} = \vec{\beta}c$ accounts for the change of the potential if its source is moving relative to the system of rest of the observer. If we attribute the scalar potential to the gravitation, the retarded potential energy of two masses m_1 and m_2 whose centers of gravity are moving with velocity $\vec{v} = \dot{\vec{r}} = \dot{\vec{r}}_2 - \dot{\vec{r}}_1 = \vec{\beta}c$ with respect to each other is given by

$$V_{0\text{ret}} = -\gamma_G \frac{m_1 m_2}{r(1 - \vec{n} \vec{\beta})}, \quad \vec{n} = \frac{\vec{r}}{r} = \frac{\vec{r}_2 - \vec{r}_1}{|\vec{r}_2 - \vec{r}_1|}. \quad (14.70)$$

The unit vector $\vec{n}$ is directed from the center of the mass m_1 to the center of the mass m_2 ; γ_G is the gravitational constant. The derivation of the retarded gravitational potential is equivalent to that of the electric potential of a moving charge outlined in Chap. 15. By calculating the motion of a planet with mass $m_1 = m$ about the central star with mass $m_2 = M$ by means of the retarded gravitational potential $G_{\text{ret}} = -V_{0\text{ret}}/m_1$, we obtain the so-called perihelion rotation predicted by general relativity. It readily follows from (14.70) that the relativistic effect is the larger the more the orbit differs from a circle because in this case $\vec{n} \vec{\beta} \neq 0$.

If we neglect both the spin and the scalar interaction, we obtain the standard relation:

$$d\tau = \sqrt{1 - \beta^2} dt. \quad (14.71)$$

Unfortunately, we cannot directly use (14.69) to substitute the laboratory time for the universal time τ because the spin components $S_{\mu\nu} = S_{\mu\nu}(\tau)$ are functions of τ . In order to reduce the number of dependent variables, it would be preferable to substitute the three-component spin vector $\vec{s} = \vec{s}(t)$ for the six-component antisymmetric spin tensor $\vec{\vec{S}}$. Within the frame of our non-quantum-mechanical calculations, this procedure corresponds to the transformation of the spin tensor from the laboratory frame to the particle rest frame. This transformation corresponds to the Foldy–Wouthuysen transformation [173] of the Dirac equation which reduces the four-component Dirac spinor to the two-component Pauli spinor. By going from the Minkowski space to the laboratory frame, we must consider that the rest frame of the electron rotates with respect to the laboratory frame if the direction of the particle velocity $\vec{v} = \vec{\beta}c$ changes. The corresponding angular velocity is known as Thomas precession [174]:

$$\vec{\omega} = \vec{\omega}_T = \frac{\gamma^2}{\gamma + 1} \vec{\beta} \times \left(\frac{d\vec{\beta}}{dt} \right). \quad (14.72)$$

Using this expression, the rotational energy referred to the laboratory frame is found as

$$T_{4r} = \vec{s}(\vec{\omega}_C + \vec{\omega}_T) = \frac{m_e c^2}{2} + \frac{\gamma^2}{\gamma + 1} \vec{s} \left(\vec{\beta} \times \frac{d\vec{\beta}}{dt} \right). \quad (14.73)$$

The second term is generally small compared with the first term. Moreover, the spin forces are negligibly small compared with the Coulomb forces. Therefore, we can replace the acceleration with a sufficient degree of accuracy by the relativistic Lorentz equation approximation:

$$m_e \gamma \frac{d\vec{\beta}}{dt} \approx -m_e \frac{d\gamma}{dt} \vec{\beta} - e \left(\frac{\vec{E}}{c} + \vec{\beta} \times \vec{B} \right). \quad (14.74)$$

The result of the substitution is given by

$$T_{4r} \approx \frac{m_e c^2}{2} + \frac{e\gamma}{m_e c(1 + \gamma)} \vec{s} \left\{ \vec{\beta} \times \left(\vec{E} + c\vec{\beta} \times \vec{B} \right) \right\}. \quad (14.75)$$

By assuming the validity of special relativity, we can express the components of the spin tensor $\vec{\vec{S}}$ by those of the spin vector $\vec{s}$ by transforming the spin from the particle's rest frame to the laboratory system. Because the spin is described by an antisymmetric second-rank tensor, it transforms from the system at rest to a system moving with velocity $\vec{v} = d\vec{r}/dt = \vec{\beta}c$ by Lorentz transformations in the same way

as the electric and magnetic field strengths. Considering that $\vec{S}_e = 0$ and $\vec{S}_m = \vec{s}$ in the system at rest, we readily obtain in the laboratory system

$$\vec{S}_e = \gamma \vec{s} \times \vec{\beta}, \quad \vec{S}_m = \gamma \vec{s} - \frac{\gamma^2}{\gamma + 1} \vec{\beta}(\vec{\beta} \cdot \vec{s}), \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \quad (14.76)$$

Here, $\vec{s} = \vec{s}(t)$ and $\vec{\beta} = \vec{\beta}(t)$ are functions of the laboratory time t . The relations (14.72) demonstrate that in the laboratory system the components of the spin tensor depend on the velocity of the electron and on the orientation of the spin vector $\vec{s}$. Employing these relations, we readily obtain

$$\begin{aligned} \frac{m_e}{e} V_s(t, \vec{r}) &= \frac{1}{2} \sum_{\mu, \nu} S_{\mu\nu} F_{\mu\nu} = \vec{S}_m \vec{B} - \vec{S}_e \vec{E}/c \\ &= \gamma \vec{s} \vec{B} - \frac{\gamma^2}{1 + \gamma} (\vec{\beta} \vec{B})(\vec{s} \vec{\beta}) + \gamma \vec{\beta}(\vec{s} \times \vec{E}/c) \\ &= \vec{s} \vec{B} - \frac{\gamma}{1 + \gamma} (\vec{s} \times \vec{\beta})(\vec{\beta} \times \vec{B}) + \gamma \vec{\beta}(\vec{s} \times \vec{E}/c). \end{aligned} \quad (14.77)$$

We define the variational principle for constant total energy $H_4 = -E_0$ in Minkowski space as the *generalized Maupertuis principle* in accordance with the conventional three-dimensional case. In order to obtain the reduced Lagrangian, we neglect radiation effects and replace the universal time τ by the laboratory time t as independent variable. However, we cannot directly use the relation (14.69) for this purpose because the spin term is a function of τ . In principle, we perform the substitution most conveniently by considering only variations $\tilde{\delta} = \delta_{H_4 = -E_0}$ of W_4 with the constraint $H_4 = -E_0 = -m_e c^2$, so that the variational principle becomes

$$\begin{aligned} \tilde{\delta} W_4 &= \tilde{\delta} \int L_4 d\tau = \tilde{\delta} \int \left[\sum (m_e \dot{x}_\mu - e A_\mu) \dot{x}_\mu - H_4 \right] d\tau \\ &= \tilde{\delta} \int \sum p_\mu \dot{x}_\mu d\tau = \delta \int L_3 dt = 0. \end{aligned} \quad (14.78)$$

The variation $\tilde{\delta}$ does not allow for arbitrary variations of the four coordinates, whereas the variation δ allows arbitrary variation of the three position coordinates $x_1 = x(t)$, $x_2 = y(t)$, and $x_3 = z(t)$. In order to substitute t for τ , we put $\tau_r = 0$ in the Hamiltonian (14.57) and express the remaining expression in the form

$$\begin{aligned} H_4(x_\mu, \dot{x}_\mu) &= \frac{m_e}{2} (\beta^2 - 1) \left(\frac{dt}{d\tau} \right)^2 - \frac{m_e c^2}{2} + V_s(t) \\ &\quad + \frac{e\gamma}{m_e c (1 + \gamma)} \vec{s} \left(\vec{\beta} \times (\vec{E} + c \vec{\beta} \times \vec{B}) \right) + V_{0\text{ret}}. \end{aligned} \quad (14.79)$$

Because now all quantities are functions of $\vec{r}$ and t , we can utilize (14.74) for replacing in (14.60) $d\tau$ by dt , giving the reduced Lagrangian L_3 in the form

$$L_3 = -ec\vec{A}\vec{\beta} + e\varphi - m_e c^2 \sqrt{(1 - \beta^2) \left(1 - 2 \frac{V_{0\text{ret}} + V_s(t)}{m_e c^2}\right)}. \quad (14.80)$$

By incorporating the conservation of energy, we have transformed the restricted variational principle for the four variables $x_\mu(\tau)$, $\mu = 1, 2, 3, 4$, into a variational principle for the three position coordinates $x_1 = x(t)$, $x_2 = y(t)$, and $x_3 = z(t)$. The transformation into the three-dimensional laboratory coordinate system conserves the total energy $H_4 = -E_0$ in Minkowski space.

Substituting in the Lagrangian (14.80) the relation (14.77) for $V_s(t)$, we obtain

$$L_3 \approx -ec\vec{\beta}\vec{A} + e\varphi - \frac{c}{\gamma} \sqrt{m_e^2 c^2 - 2m_e V_{0\text{ret}} - 2e \left(\vec{s}\vec{B} + \frac{\gamma^2 \vec{\beta}(\vec{s} \times \vec{E})}{c(1 + \gamma)} \right)}. \quad (14.81)$$

By considering only electromagnetic forces, we disregard the scalar interaction ($V_{0\text{ret}} = 0$). Moreover, the spin terms are small compared with $m_e^2 c^2$. Accordingly, it suffices to only consider the linear terms of the Taylor expansion of the square root giving

$$L_3 \approx -\frac{m_e c^2}{\gamma} - ec\vec{\beta}\vec{A} + e\varphi + \frac{e}{\gamma m_e} \left(\vec{s}\vec{B} + \frac{\gamma^2 \vec{\beta}(\vec{s} \times \vec{E})}{c(1 + \gamma)} \right). \quad (14.82)$$

If we neglect the spin terms, we obtain the standard relativistic Lagrangian for electron motion:

$$L = L_3(\vec{s} = 0) = -m_e c^2 \sqrt{1 - \beta^2} - ec\vec{A}\vec{\beta} + e\varphi. \quad (14.83)$$

We obtain the path equations in the laboratory space including spin effects by employing the form (14.77) for the Lagrangian L_3 instead of the standard form (14.83) in the Euler–Lagrange equations:

$$\frac{d}{dt} \frac{\partial L_3}{\partial \dot{x}_\mu} - \frac{\partial L_3}{\partial x_\mu} = 0, \quad \mu = 1, 2, 3. \quad (14.84)$$

Note that the dot indicates in this equation derivatives with respect to the laboratory time t .

14.5 Approximate Relativistic Canonical Momentum and Hamiltonian in the Laboratory System

The phase ϕ of the electron wave is directly related with the action function S_3 in the laboratory system:

$$\phi = \frac{S_3}{\hbar} = \frac{1}{\hbar} \int L_3 dt = \frac{c}{\hbar} \int (\vec{p}\vec{\beta} - H_3/c) dt. \quad (14.85)$$

The Hamiltonian $H_3 = \hbar\omega_e$ is a constant of motion for stationary fields. In this case, it is advantageous to substitute the arc length z of the central trajectory of a confined electron beam for the time t . In most electron optical systems, one chooses the symmetry axis as the z -axis. In this case, the phase adopts the standard form

$$\phi = \int \vec{k} d\vec{r} - \omega_e t = \frac{1}{\hbar} \int \vec{p} \frac{d\vec{r}}{dz} dz - \omega_e t. \quad (14.86)$$

We must perform the first integration over the canonical momentum $\vec{p} = \hbar\vec{k}$ along the true path of the electron.

Employing the standard procedure of classical mechanics, we derive from (14.82) the canonical momentum:

$$\begin{aligned} \vec{p} &= \frac{1}{c} \text{grad}_{\vec{\beta}} L_3 \approx m_e c \gamma \vec{\beta} - e \vec{A} + \frac{e \vec{s} \vec{B}}{m_e c} \gamma \vec{\beta} \\ &\quad - \frac{e \gamma}{m_e c^2} \frac{\vec{s} \times \vec{E}}{(1 + \gamma)} - \frac{e \vec{\beta}}{m_e c^2} \frac{\gamma^3}{(1 + \gamma)^2} (\vec{\beta} (\vec{s} \times \vec{E})). \end{aligned} \quad (14.87)$$

We can cast this rather involved expression in a more suitable form by means of the relation

$$\vec{\beta} (\vec{\beta} (\vec{s} \times \vec{E})) = \vec{\beta} \times (\vec{\beta} \times (\vec{s} \times \vec{E})) + (\vec{s} \times \vec{E}) \beta^2. \quad (14.88)$$

By inserting this relation into the last term on the right-hand side of (14.87), we eventually derive at

$$\begin{aligned} \vec{p} &\approx m_e c \gamma \left(1 + \frac{e \vec{s} \vec{B}}{m_e^2 c^2} \right) \vec{\beta} - e \vec{A} - \frac{e}{m_e c^2} \frac{\gamma}{(1 + \gamma)} (\vec{s} \times \vec{E}) \\ &\quad - \frac{e}{m_e c^2} \frac{\gamma^3}{(1 + \gamma)^2} \vec{\beta} \times (\vec{\beta} \times (\vec{s} \times \vec{E})). \end{aligned} \quad (14.89)$$

Note that the interaction of the spin with the magnetic field affects the relativistic mass of the electron, which we define as

$$m = m_e \gamma \left(1 + \frac{e \vec{s} \vec{B}}{m_e^2 c^2} \right). \quad (14.90)$$

If we disregard spin effects ($\vec{s} = 0$), the canonical momentum adopts the standard relativistic form

$$\vec{p} = m_e c \gamma \vec{\beta} - e \vec{A}. \quad (14.91)$$

The expression (14.89) demonstrates that the interaction of the spin with the electromagnetic field affects the canonical momentum by small terms which are proportional to $\hbar$. Although these terms are small with respect to the kinetic momentum $m_e c \gamma \beta$, they may cause an appreciable phase shift of the electron wave, especially if the electron interacts with atomic fields.

So far, we have not made any restrictions to any of the electromagnetic fields. If these fields are stationary ($\partial A_\nu / \partial t = 0$), the three-dimensional Hamiltonian is a constant quantity of motion. Using the expressions (14.89) and (14.82), we find for the Hamiltonian the approximate relation

$$\begin{aligned} H_3 &= \vec{\beta} \text{grad}_{\vec{\beta}} L_3 - L_3 \\ &\approx m_e c^2 \gamma \left(1 + \frac{e \vec{s} \vec{B}}{m_e^2 c^2} \right) - e \varphi = m_e c^2. \end{aligned} \quad (14.92)$$

We have chosen the gauge of the electric potential in such a way that φ is zero at the emitting surface placed in the region $\vec{B} = 0$.

We utilize the relation (14.92) to express the relativistic factor γ and the relative velocity β as functions of φ and $\vec{s} \vec{B}$, giving

$$\gamma = \frac{m_e c^2 + e \varphi}{m_e c^2 + e \vec{s} \vec{B} / m_e}, \quad \gamma \beta = \sqrt{\gamma^2 - 1} \approx c \frac{\sqrt{2e(m_e \varphi^* - \vec{s} \vec{B})}}{m_e^2 c^2 + e \vec{s} \vec{B}}. \quad (14.93)$$

The relativistic modified potential is given by

$$\varphi^* = \varphi \left(1 + \frac{e \varphi}{2 E_0} \right), \quad E_0 = m_e c^2. \quad (14.94)$$

The conservation of the Hamiltonian H_3 allows us to construct the reduced eikonal:

$$S_2 = \int_{z_0}^z L_2(x, y, x', y'; \vec{s}; z) dz, \quad L_2 = \vec{\beta} \text{grad}_{\vec{\beta}} L_3. \quad (14.95)$$

We derive the Lagrangian L_2 by substituting the relations (14.93) for γ and $\gamma\beta$ in the expression (14.89) and by choosing the z -coordinate as the independent variable. By considering that the rest energy is large compared with the potential energy of the electron magnetic momentum ($m_e c^2 \gg e\vec{s}\vec{B}/m_e$), we obtain

$$\begin{aligned} L_2 &= L_2(x, y, x', y'; z) = \vec{p} \frac{d\vec{r}}{dz} \\ &\approx \left| \frac{d\vec{r}}{dz} \right| \sqrt{2em_e\varphi^* \left(1 - \frac{\vec{s}\vec{B}}{m_e\varphi^*} \right)} - e\vec{A} \frac{d\vec{r}}{dz} - \frac{e}{E_0} \frac{E_0 + e\varphi}{2E_0 + e\varphi} \frac{d\vec{r}}{dz} (\vec{s} \times \vec{E}). \end{aligned} \quad (14.96)$$

The unit vector along the trajectory is given by

$$\frac{d\vec{r}}{dz} = x'\vec{e}_x + y'\vec{e}_y + g_3\vec{e}_z, \quad \left| \frac{d\vec{r}}{dz} \right| = \sqrt{g_3^2 + x'^2 + y'^2}. \quad (14.97)$$

Dashes denote derivatives with respect to the z -coordinate measured along the optic axis, which may be straight ($g_3 = 1$) or curved ($g_3 = \text{Re}(\Gamma\bar{w}) = 1 + \kappa_x x + \kappa_y y$).

We can vary the reduced eikonal S_2 arbitrarily with respect to the two off-axial coordinates $x(z)$ and $y(z)$ because expression (14.89) of the canonical momentum keeps the conventional energy (time-like component of the momentum 4-vector) fixed for each position of the electron along the trajectory. Hence, we obtain the trajectory equations for fixed boundaries from the condition

$$\delta S_2 = \delta \int_{z_0}^z L_2 dz = \sqrt{2em_e\Phi_0^*} \delta \int_{z_0}^z \mu dz = 0. \quad (14.98)$$

The variation of this expression leads to the Euler–Lagrange equations

$$\frac{d}{dz} \frac{\partial L_2}{\partial x'} = \frac{\partial L_2}{\partial x}, \quad \frac{d}{dz} \frac{\partial L_2}{\partial y'} = \frac{\partial L_2}{\partial y}. \quad (14.99)$$

The resulting path equations depend on the orientation of the spin $\vec{s} = \vec{s}(z)$ along the trajectory. Therefore, we can only solve the path equation if we know the precession of the spin along the optic axis.

14.6 Spin Precession

We cannot derive the equations for the spin precession from a proper Lagrangian by employing Hamilton's principle. Therefore, we must try to construct these equations in such a way that they satisfy the constraints (14.11) and (14.34). In addition, we

require that the equations reduce to the standard nonrelativistic form at the limit $c \rightarrow \infty$:

$$\frac{d\vec{s}}{dt} = -\frac{e}{m_e} \vec{s} \times \vec{B}. \quad (14.100)$$

We construct a suitable set of equations for the dynamics of the four-dimensional spin in Minkowski space by assuming that the precession corresponds to that of classical mechanics for a spinning top affected by an external angular momentum. As a proper set of equations, we propose

$$\dot{S}_{\mu\nu} = -\frac{e}{m_e} \sum_{\lambda} (S_{\mu\lambda} F_{\lambda\nu} - S_{\nu\lambda} F_{\lambda\mu}), \quad \mu, \nu = 1, 2, 3, 4. \quad (14.101)$$

These equations do not alter if we exchange the subscripts μ and ν and consider the antisymmetric property $S_{\mu\nu} = -S_{\nu\mu}$ of the components of the spin tensor. The driving term on the right-hand side of (14.101) represents the external angular momentum, which determines the precession of the electron along its path.

The spin tensor has the same structure as the electromagnetic field tensor because both tensors are Lorentz-invariant antisymmetric tensors. We construct from the spatial components of the spin tensor (14.8) a three-dimensional axial vector $\vec{S}_m$ and from the imaginary time-like components a real electric axial vector $\vec{S}_e$ with components $S_{ex} = iS_{41}$, $S_{ey} = iS_{42}$, $S_{ez} = iS_{43}$. The vector $\vec{S}_e$ accounts for the electric dipole moment of the electron induced by the motion of the magnetic dipole. If we also express the components of the electromagnetic field tensor by the components of the electric and magnetic field strengths and consider the relation $\omega_{\mu\nu} = S_{\mu\nu}/\hbar$, we can rewrite equations (14.100) as two coupled three-dimensional vector equations:

$$\dot{\vec{S}}_m = -\frac{e}{m_e} \{ \vec{S}_m \times \vec{B} + \vec{S}_e \times \vec{E}/c \}, \quad \dot{\vec{S}}_e = \frac{e}{m_e} \{ \vec{S}_m \times \vec{E}/c - \vec{S}_e \times \vec{B} \}. \quad (14.102)$$

The electric and magnetic field vectors $\vec{E} = \vec{E}(x_\mu)$, $\vec{B} = \vec{B}(x_\mu)$ depend implicitly on τ because we must substitute the position coordinates $x_\mu = x_\mu(\tau)$ of the particle for the coordinates of the electromagnetic field vectors. The position of the electron is determined by the four path equations (14.25), which also depend on the orientation of the spin. Therefore, the path equations (14.25) and the six equations (14.102) of the spin precession form a coupled system of ten differential equations, which define the position of the particle and the orientation of its spin in Minkowski space as functions of the universal time τ .

We have shown in Sect. 14.4 that we can substitute the laboratory time t for the universal time τ if we neglect radiation. In this case, we obtain the components $S_{\mu\nu} = S_{\mu\nu}(t)$ of the spin tensor from Eqs. (14.102) and (14.69) as follows:

$$\begin{aligned}\frac{d\vec{S}_m}{dt} &= \dot{\vec{S}}_m \frac{d\tau}{dt} = -\frac{e}{m_e\gamma} \{\vec{S}_m \times \vec{B} + \vec{S}_e \times \vec{E}/c\}, \\ \frac{d\vec{S}_e}{dt} &= \dot{\vec{S}}_e \frac{d\tau}{dt} = \frac{e}{m_e\gamma} \{\vec{S}_m \times \vec{E}/c - \vec{S}_e \times \vec{B}\}.\end{aligned}\quad (14.103)$$

The electric and magnetic field strengths and the relativistic factor depend implicitly on the time t because these quantities are functions of the position $\vec{r} = \vec{r}(t)$ and of the velocity $\vec{v} = \dot{\vec{r}} = \vec{v}(t)$ of the electron in the laboratory system.

Our approach incorporates correctly the spin of the particle in the equations of motion without the need of a phenomenological g -factor and/or quantum-mechanical considerations, even in the relativistic case. This result contradicts the general belief that a microscopic consideration of the spin is beyond the scope of classical electrodynamics.

In order to show the validity of Eqs. (14.101) and (14.102), we first prove that they satisfy the constraint (14.34). To this purpose, we multiply (14.97) with $F_{\mu\nu}$ and sum over the indices μ and ν , giving

$$\begin{aligned}\frac{m_e}{2e} \sum_{\mu,\nu} \dot{S}_{\mu\nu} F_{\mu\nu} &= \frac{m_e}{2e} (\dot{\vec{S}}_m \vec{B} - \dot{\vec{S}}_e \vec{E}/c) \\ &= \sum_{\mu,\nu,\lambda} F_{\mu\lambda} F_{\mu\nu} S_{\lambda\nu} = \sum_{\mu,\nu,\lambda} F_{\mu\nu} F_{\mu\lambda} S_{\nu\lambda} = - \sum_{\mu,\nu,\lambda} F_{\mu\lambda} F_{\mu\nu} S_{\lambda\nu} = 0.\end{aligned}\quad (14.104)$$

We obtain the second and third relations by exchanging two indices. Because this procedure is merely a change of notation, it does not affect the value of the summation. Subsequently, we multiply Eq. (14.101) with $\omega_{\mu\nu}$ and perform the same procedure. The result demonstrates that the constraint (14.32) is automatically fulfilled for the equations (14.75). In order to prove the conservation of the absolute value of the spin (14.11), we multiply Eq. (14.101) with $S_{\mu\nu}$ and employ the same method as in (14.104), giving

$$\sum_{\mu,\nu} \dot{S}_{\mu\nu} S_{\mu\nu} = \frac{1}{2} \frac{d}{d\tau} \sum_{\mu,\nu} S_{\mu\nu}^2 = 0. \quad (14.105)$$

The result demonstrates that the absolute value of the spin is conserved as postulated by the expression (14.11).

We derive an invariant of the spin precession by scalar multiplying the first equation in (14.105) with $\vec{S}_e$ and the second equation with $\vec{S}_m$. Addition of the resulting equations gives

$$\dot{\vec{S}}_m \vec{S}_e + \vec{S}_m \dot{\vec{S}}_e = \frac{d(\vec{S}_m \vec{S}_e)}{d\tau} = 0. \quad (14.106)$$

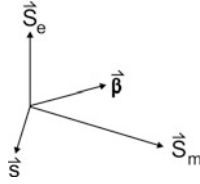


Fig. 14.4 Location of the spin vectors $\vec{S}_m$ and $\vec{S}_e$ in the laboratory system and of the spin $\vec{S}$ in the system of rest moving with constant velocity $\vec{\beta}c$ with respect to the laboratory system

Hence, the scalar product of the two vectors must be a constant. Because the vector $\vec{S}_e$ is zero in the system at rest, it follows that the two vectors are orthogonal:

$$\vec{S}_m \vec{S}_e = 0. \quad (14.107)$$

The validity of the Lorentz transformations (14.76) for the spin at rest $\vec{S}$ is readily demonstrated by substituting the transformations for the vectors $\vec{S}_e$ and $\vec{S}_m$ in (14.107). We readily obtain from (14.76) the additional relations

$$\vec{S}_e \vec{\beta} = 0, \quad \vec{S}_e \vec{S} = 0 \quad \vec{S} \times \vec{S}_m = \frac{\gamma}{\gamma + 1} (\vec{\beta} \vec{S}) \vec{S}_e, \quad \vec{S}_m \times \vec{\beta} = \vec{S}_e. \quad (14.108)$$

Note that the Lorentz transformations are valid for inertial systems which move relative to each other with constant velocity $\vec{\beta}c$. In this case, the vector $\vec{S}_e$ is orthogonal to the vectors $\vec{S}_m, \vec{S}$ and $\vec{\beta}$ which are embedded in a mutual plane, as depicted in Fig. 14.4.

We conceive (14.102) as extensions of the so-called BMT equation [38], which is only valid for homogeneous electromagnetic fields. We derive the BMT equation by constructing the spin 4-vector:

$$S_\lambda = (1/2c) \varepsilon^{\lambda\kappa\mu\nu} \dot{x}_\kappa S_{\mu\nu} = \frac{1}{c} (-)^{\lambda+1} \sum_p \dot{x}_\kappa S_{\mu\nu}. \quad (14.109)$$

Here, $\varepsilon^{\lambda\kappa\mu\nu}$ is the totally antisymmetric fourth-rank unit tensor; p denotes the cyclic permutation of the indices κ, μ , and ν . These indices and λ differ from each other and each defines one of the four numbers 1, 2, 3, 4. It readily follows from the relation (14.109) that the scalar product $\sum S_\lambda \dot{x}_\lambda$ vanishes identically. To derive the equation for the precession of the spin, we take the derivative of Eq. (14.109) with respect to the universal time τ and replace $\ddot{x}_\kappa$ by means of the path equations (14.25). By considering the assumption $\partial A_\mu / \partial \tau = 0$, we eventually obtain

$$\dot{S}_\lambda = (-)^{\lambda+1} \sum_p \{ \ddot{x}_\kappa S_{\mu\nu} + \dot{x}_\kappa \dot{S}_{\mu\nu} \} / c = -\frac{e}{m_e} \sum_\mu F_{\lambda\mu} S_\mu + \frac{1}{m_e c} (-)^{\lambda+1} \sum_p \frac{\partial V_2}{\partial x_\kappa} S_{\mu\nu}. \quad (14.110)$$

The last term accounts for the gradient forces, which vanish for homogeneous electromagnetic fields. Neglecting this term, we obtain the BMT equation for a charged particle with Lande factor $g = 2$ [38].

In order to reduce the number of variables, we consider the motion of the rest-frame spin $\vec{s} = \vec{s}(t)$ as a function of the laboratory time t . The corresponding equation for the spin precession of the electron in the laboratory system is

$$\frac{d\vec{s}}{dt} = -\frac{e}{m_e} \vec{s} \times \left(\frac{\vec{B}}{\gamma} - \frac{\gamma}{\gamma+1} \vec{\beta} \times (\vec{\beta} \times \vec{B}) - \vec{\beta} \times \vec{E}/c \right) + \frac{\gamma^2}{\gamma+1} \vec{s} \times \left(\vec{\beta} \times \frac{d\vec{\beta}}{dt} \right). \quad (14.111)$$

If we replace the acceleration $\vec{a} = c d\vec{\beta}/dt$ in the laboratory frame by the relation (14.74), we readily find

$$\frac{d\vec{s}}{dt} = -\frac{e}{m_e} \vec{s} \times \left(\frac{\vec{B}}{\gamma} - \frac{\vec{\beta} \times \vec{E}}{(1+\gamma)c} \right) - \frac{e}{m_e^2 c^2 (1+\gamma)} \vec{s} \times (\vec{\beta} \times \text{grad}(\vec{s} \vec{B})). \quad (14.112)$$

This equation is valid for arbitrary macroscopic electromagnetic fields. The last term on the right-hand side is zero for homogeneous magnetic fields. In this case, we derive at Thomas's equation for the spin precession of the electron in a uniform electromagnetic field [174]. At the limit $c \rightarrow \infty$, Eq. (14.112) adopts the nonrelativistic form (14.100).

To solve the path equations for stationary fields (14.99), we need the spin precession as a function of the z -coordinate. Therefore, we must substitute this coordinate for the time t in Eq. (14.105) by means of the relation

$$\frac{d}{dt} = \frac{\beta c}{\sqrt{g_3^2 + x'^2 + y'^2}} \frac{d}{dz}. \quad (14.113)$$

This substitution gives

$$\frac{\beta c}{\sqrt{g_3^2 + x'^2 + y'^2}} \frac{d\vec{s}}{dz} = -\frac{e}{m_e} \vec{s} \times \left(\frac{\vec{B}}{\gamma} - \frac{\vec{\beta} \times \vec{E}}{(1+\gamma)c} \right) - \frac{e}{m_e^2 c^2 (1+\gamma)} \vec{s} \times (\vec{\beta} \times \text{grad}(\vec{s} \vec{B})). \quad (14.114)$$

This vector equation defines three coupled differential equations for the three spin components. Fortunately, we can reduce these equations to two equations by utilizing the fact that the absolute value of the spin is a constant of motion (14.11). We readily verify this behavior by scalar multiplication of (14.114) with the spin vector $\vec{s}$. Hence, if we neglect radiation effects, we must solve in the stationary case only four coupled differential equations compared with ten in the most general case.

14.7 Stern–Gerlach Effect Revisited

In 1921, Stern and Gerlach measured the magnetic moment of single silver atoms by transmitting the atom beam through an extremely inhomogeneous magnetic field [175]. Monovalent atoms like silver (Ag) have all their electrons arranged in closed shells except of the valence electron. Because the shells are completed, their electrons cannot contribute to the magnetic moment. The valence electron occupies an s-state which is symmetric and has no angular moment. Therefore, only the spin of the valence electron can contribute to the atomic magnetic moment of the Ag atom. In the presence of an external magnetic field $\vec{B}$, the magnetic moment of the bound valence electron can adopt only two orientations, namely, either parallel or antiparallel to $\vec{B}$. Accordingly, the energy shift can only take the values $\pm\mu_B B$. This behavior corresponds to the quantization of the energy of the atomic electrons. On the other hand, free electrons can have any positive kinetic energy. We have shown in the preceding section that the spin of free electrons precesses when they are passing through a magnetic field. The precession depends on the initial orientation of the spin. In addition, the interaction of the magnetic moment attributed to the spin with the inhomogeneous magnetic field produces a force which deflects the electron.

The spin axes in an ordinary non-polarized beam are oriented entirely at random. It is widely assumed that the orientation of the spin of the electrons becomes quantized if they propagate through a magnetic field as it is the case for the spin of bound atomic electrons. If this conjecture holds true, one should observe a splitting of the electron beam in a Stern–Gerlach experiment. However, the experiments did not show any splitting for non-polarized electron beams. Bohr has tried to explain this negative result by the effect of the Lorentz force which acts like a beam limiting aperture. He showed that the diffraction at this aperture will entirely swamp the splitting of the beam. However, this argument does not hold true if we substitute an inhomogeneous Wien filter for the inhomogeneous magnet of the original Stern–Gerlach experiment. The inhomogeneous Wien filter consists of crossed electric and magnetic dipole and quadrupole fields which are excited in such a way that the Lorentz force is nullified for monochromatic electrons moving in axial direction. In this case, the beam splitting would not be appreciably affected by diffraction at a narrow aperture. A negative experimental result will then prove that the spin orientation of *free* electrons is not quantized in an external magnetic field. Evidence for this behavior is the fact that the spin of relativistic electrons in synchrotrons points in the direction of propagation and not in the direction of the beam-guiding magnetic field. If the system satisfies the Wien condition $\vec{E} + c\vec{\beta} \times \vec{B} = 0$, the contribution of the Thomas precession to the rotational energy (14.75) vanishes, and the potential energy of the spin (14.77) adopts the form

$$V_s = -\frac{e}{m_e} \left\{ \frac{\vec{s} \cdot \vec{B}}{\gamma} + \frac{\gamma}{1 + \gamma} (\vec{\beta} \cdot \vec{B})(\vec{s} \cdot \vec{\beta}) \right\}. \quad (14.115)$$

We assume a long inhomogeneous Wien filter whose scalar magnetic potential has the form

$$\psi = \Psi_{1c}x + \Psi_{2c}(x^2 - y^2). \quad (14.116)$$

The electric potential is given by

$$\varphi = \Phi_{1s}y + 2\Phi_{2s}xy. \quad (14.117)$$

The electric and magnetic multipole strengths are related with each other by the Wien condition, giving

$$\Phi_{1s} = -v_0\Psi_{1s}, \quad \Phi_{2s} = -v_0\Psi_{2s}. \quad (14.118)$$

The multipole strengths Ψ_{1c} , Ψ_{2c} , Φ_{1s} , and Φ_{2s} are constant within the Wien filter and diminish to zero in the region of the fringing fields. Therefore, the magnetic flux density $\vec{B}$ in the interior of the Wien filter has only lateral components

$$B_x = -\Psi_{1c} - 2\Psi_{2c}x, \quad B_y = -2\Psi_{2c}y, \quad B_z = 0. \quad (14.119)$$

Hence, the spin of electrons traveling with relative velocity $\vec{\beta} = \beta\vec{e}_z$ along the optic axis ($x = 0, y = 0$) is only affected by the dipole component $B_x(x = 0, y = 0) = \Psi_{1c}$. If the orientation of the spin of the axial electrons is quantized, the spin of these electrons must point in the positive or negative x -direction so that $\vec{\beta}\vec{s} = \vec{v}_0\vec{s}/c = 0$. By considering this result and (14.117), we readily obtain from (14.115) the relations

$$V_s = \mp \frac{e\hbar}{2m_e\gamma} B_x, \quad \vec{F}_s = \text{grad } V_s = \frac{\partial V_s}{\partial x} \vec{e}_x = \pm \frac{e\hbar}{m_e\gamma} \Psi_{2c} \vec{e}_x. \quad (14.120)$$

Since $\vec{F}_s$ is the only force affecting the path of the x -component of the axial electrons, the paraxial path equation for the x -coordinate adopts the form

$$x'' = \pm \frac{\hbar\Psi_{2c}}{2m_e\Phi^*}. \quad (14.121)$$

We obtain the equation for the y -coordinate from (4.13) as

$$y'' + \frac{\Phi_{1s}^2}{4\Phi^{*2}} y = 0. \quad (14.122)$$

Equations (14.121) and (14.122) demonstrate that the electron is deflected in x -direction and focused in the y -direction within the inhomogeneous Wien filter. We can readily integrate the differential equation (14.121) giving

$$x' = x'_0 \pm \frac{\hbar\Psi_{2c}}{2m_e\Phi^*} l. \quad (14.123)$$

The deflection angle increases linearly with the length ℓ of the Wien filter. The initial slope x'_0 is in the range $-\Theta_{il} \leq x'_0 \leq \Theta_{il}$. The characteristic illumination angle Θ_{il} is determined by the size of the effective source and the illumination optics. In the case of parallel illumination (point source), diffraction imposes a constraint. We assume that the lateral extension of the incident parallel beam is limited at the entrance plane $z = z_0$ of the filter by a circular aperture with opening radius $\rho_a \ll \ell = z_e - z_0$. The diffraction at the aperture introduces an angular deflection

$$\frac{\lambda}{\rho_a} = \frac{2\pi\hbar}{\sqrt{2em_e\Phi^*}} \frac{l}{\rho_a}. \quad (14.124)$$

Therefore, we can observe the spin splitting only if

$$\frac{\lambda}{\rho_a} < \frac{\hbar |\Psi_{2c}|}{2m_e\Phi^*} l > \Theta_{il}. \quad (14.125)$$

Because we can choose the length ℓ of the filter arbitrarily, at least in principle, we should be able to realize the beam splitting. This statement holds also true for a finite energy width of the beam because the Wien filter deflects axial electrons with non-average energy in the positive or negative y -direction depending on the sign of the energy deviation. However, according to our considerations, the spin axes of the electrons of an ordinary beam are oriented at random regardless if the electrons travel in field-free region or within a macroscopic electromagnetic field. Therefore, we expect a negative result also for our modified Stern–Gerlach experiment even if the condition (14.122) is satisfied. Summarizing, we conclude that spin-induced beam splitting by an inhomogeneous external magnetic field shows solely up for atoms because only two different energy states exist for the bound atomic electron. This situation differs from that of the free electron whose continuous energy states are not quantized.

Chapter 15

Electron Self-Action

Abstract A covariant expression is given for the self-action of the electron. By postulating that the self-action adopts an extremum, the equation for the derivative of the four-dimensional vector potential with respect to the universal time is obtained by applying the calculus of variation to the action integral. By considering the electron as a moving point-like particle and by employing the four-dimensional Green formalism, we express the four-dimensional vector potential of the moving electron as a loop integral which can be solved analytically. The resulting electromagnetic field consists of the accompanying field attached to the electron and the radiation field resulting from the acceleration.

So far, we have considered the action W of a charged particle in the presence of external fields. However, the electromagnetic field of the particle also contributes to the total action $W_t = W + W_{sa}$. To guarantee that the motion of the particle is not affected by its own field, we must postulate that the self-action W_{sa} is a constant of motion forming an extremum. This constant can be set to zero because the action is defined only up to an arbitrary constant. The charge and the magnetic moment of the electron are the sources of its accompanying electromagnetic field. Any change of the velocity of the electron results in an additional radiation field. We assume that the external electromagnetic field does not depend on the universal time τ ($\dot{F}_{\mu\nu, ex} = 0$). However, this constraint does not hold true for the components $F_{\mu\nu}$ of the electromagnetic field of the electron. Its field depends on the universal time τ , because the position coordinates $x_\mu = x_\mu(\tau)$ of the electron are functions of τ . In order to completely describe the self-action of the electron, we must take into account the interaction of its charge and its spin with its own electromagnetic field $F_{\mu\nu}$ together with the temporal change of its field energy [28]. Because the electromagnetic field is continuously distributed in the four-dimensional space we must introduce a proper four-dimensional Lagrange density $l_{sa} = l_{sa}(x_\mu, \tau)$, given by

$$l_{\text{sa}} = - \sum_{\mu} e \rho_e \dot{x}_{\mu} A_{\mu} - \sum_{\mu, \nu} \frac{e}{2m_e} \rho_e S_{\mu\nu} F_{\mu\nu} - \frac{i}{2c\mu_0} \sum_{\mu, \nu} F_{\mu\nu} \dot{F}_{\mu\nu}. \quad (15.1)$$

The normalized four-dimensional mass density $\rho_e = \rho_e(x_{\mu}, \tau)$ of the particle is imaginary and a function of its four coordinates x_{μ} in Minkowski space and the universal time τ . The first and the second term on the right of (15.1) represent the self-interaction of the charge and the spin of the electron with its own electromagnetic field, whereas the third term comprises the contribution of this field to the self-energy density; μ_0 is the permeability of the vacuum.

The field term differs from that of standard field theory, because we have extended the space from three to four dimensions. Because ρ_e , $\dot{x}_4$, and A_4 are imaginary, we must introduce the imaginary unit in the third term in order that all terms of the Lagrange density l_{sa} are imaginary. This measure guarantees that the self-action W_{sa} is real:

$$\begin{aligned} W_{\text{sa}} &= Ex \int_{\tau_1}^{\tau_2} \left\{ \int l_{\text{sa}} d^4 \vec{R} \right\} d\tau \\ &= -Ex \int_{\tau_1}^{\tau_2} \left\{ \int \sum_{\mu} \left[e \rho_e \dot{x}_{\mu} A_{\mu} \right. \right. \\ &\quad \left. \left. + \sum_{\nu} \left(\frac{e}{2m_e} \rho_e S_{\mu\nu} F_{\mu\nu} + \frac{i}{2c\mu_0} F_{\mu\nu} \dot{F}_{\mu\nu} \right) \right] d^4 \vec{R} \right\} d\tau. \end{aligned} \quad (15.2)$$

15.1 Electromagnetic Field of the Moving Electron

We derive the equations for the electromagnetic field associated with the electron by employing Hamilton's principle $\delta W_{\text{sa}} = 0$. The variation has to be performed with respect to the four components $A_{\mu} = A_{\mu}(x_{\nu}, \tau)$ of the four-vector potential of the electron. Derivatives of the differential δA_{μ} with respect to x_{ν} are removed by partial integrations. For fixed boundaries, the deviations are zero at the limits τ_1 and τ_2 . In this case, the variation gives

$$\begin{aligned} \delta W_{\text{sa}} &= i \int_{\tau_1}^{\tau_2} \int \sum_{\mu} \left[j_{\mu} + \sum_{\nu} \left(\frac{e}{im_e} \frac{\partial \rho_e}{\partial x_{\nu}} S_{\mu\nu} + \frac{i}{c\mu_0} \frac{\partial \dot{F}_{\mu\nu}}{\partial x_{\nu}} \right) \right] \delta A_{\mu} d^4 \vec{R} d\tau = 0, \\ j_{\mu} &= e \rho_e \dot{x}_{\mu}. \end{aligned} \quad (15.3)$$

Here, $j_{\mu} = j_{\mu}(\vec{R}, \tau)$ is the μ th component of the four-dimensional current density $\vec{J}$. Since the variation δA_{μ} can be taken at an arbitrary time $\tau_1 \leq \tau \leq \tau_2$, the

expression in the brackets must vanish identically. Imposing in addition the Lorentz gauge $\vec{\nabla} \cdot \vec{A} + \partial A_4 / \partial x_4 = 0$, we find that the time derivative $\dot{A}_\mu$ of each component of the four-dimensional vector potential (A_1, A_2, A_3, A_4) of the electron must satisfy the inhomogeneous equation

$$\sum_{\nu=1}^4 \frac{\partial^2 \dot{A}_\mu}{\partial x_\nu^2} = i c \mu_0 \left[j_\mu + \frac{e}{m_e} \sum_{\nu} \frac{\partial \rho_e}{\partial x_\nu} S_{\mu\nu} \right], \quad \mu = 1, 2, 3, 4. \quad (15.4)$$

We obtain the solution of this differential equation most conveniently by means of the Green method employing the four-dimensional Green function:

$$G = G(\vec{R}, \vec{R}') = \frac{1}{4\pi^2(\vec{R} - \vec{R}')^2}. \quad (15.5)$$

This function is the inhomogeneous solution of the differential equation

$$\sum_{\mu} \frac{\partial^2 G}{\partial x_\mu^2} = -\delta^4(\vec{R} - \vec{R}'). \quad (15.6)$$

Here, $\delta^4(\vec{R} - \vec{R}')$ is the four-dimensional delta function, and $\vec{R}'$ denotes the position of the point source. By employing Green's method to the equation (15.4), we obtain the solution:

$$\begin{aligned} \dot{A}_\mu &= \frac{dA_\mu}{d\tau} = \frac{c\mu_0}{4i\pi^2} \int \frac{1}{(\vec{R} - \vec{R}')^2} \left\{ j_\mu(\vec{R}') - \frac{e}{m_e} \sum_{\nu} S_{\mu\nu}(\tau) \frac{\partial \rho_e}{\partial x'_\nu} \right\} d^4 \vec{R}' \\ &= \frac{c\mu_0}{4i\pi^2} \int \left\{ \frac{j_\mu(\vec{R}')}{(\vec{R} - \vec{R}')^2} - \frac{e}{m_e} \sum_{\nu} \rho_e S_{\mu\nu}(\tau) \frac{x_\nu - x'_\nu}{(\vec{R} - \vec{R}')^4} \right\} d^4 \vec{R}'. \end{aligned} \quad (15.7)$$

We have removed the derivatives of ρ_e with respect to x'_ν , $\nu = 1, 2, 3, 4$, by partial integration and by considering the relation

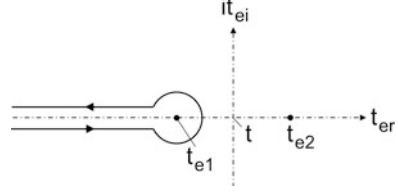
$$\frac{\partial \left[1/(\vec{R} - \vec{R}')^2 \right]}{\partial x'_\nu} = - \frac{\partial \left[1/(\vec{R} - \vec{R}')^2 \right]}{\partial x_\nu}. \quad (15.8)$$

In order to evaluate the integral analytically, we consider the electron as a point-like particle with four-dimensional density:

$$\rho_e = \delta^4(\vec{R}' - \vec{R}_e(\tau)). \quad (15.9)$$

By employing this density and assuming that the particle originates at the initial time $\tau_i = -\infty$, we obtain

Fig. 15.1 Path of integration within the complex t_e -plane; only the residues at the pole $t_{e1} < t$ contribute to the integral (15.13)



$$A_\mu = \frac{c\mu_0}{4i\pi^2} \int_{-\infty}^{\tau_0} \left\{ \frac{e\dot{x}_{e\mu}(\tau)}{[\vec{R} - \vec{R}_e(\tau)]^2} + \frac{e}{m_e} \sum_v S_{\mu v}(\tau) \frac{\partial}{\partial x_v} \frac{1}{[\vec{R} - \vec{R}_e(\tau)]^2} \right\} d\tau. \quad (15.10)$$

The integration must be taken over the entire lifetime of the particle starting at the initial universal time τ_i up to the time of observation τ_0 . To evaluate the integral (15.10), it is advantageous to substitute the time t_e for the universal time τ . However, in order to obtain the correct results, we must consider that

$$\frac{dt_e}{d\tau} = \frac{1}{ic} \frac{dx_4}{d\tau} = \pm\gamma \quad (15.11)$$

may be positive and negative. According to Feynman, we suppose that the negative sign considers the positron contribution to the electron. Moreover, to evaluate the integral (15.10) analytically, we consider the time variable $t_e = t_{er} + it_{ei}$ as a complex variable. The path of integration within the complex t_e -plane is shown in Fig. 15.1 and consists of the loop:

$$t_e = -\infty \rightarrow t_{e1} + \varepsilon e^{i\phi} \rightarrow -\infty, \quad -\pi \leq \phi \leq \pi. \quad (15.12)$$

With these assumptions, we write (15.10) in the form

$$A_\mu = \frac{c\mu_0}{4i\pi^2} \oint \left\{ \frac{e\dot{x}_{e\mu}(t_e)}{[(\vec{r} - \vec{r}_e)^2 - c^2(t - t_e)^2]} - \frac{2e}{m_e\gamma} \sum_v S_{\mu v}(t_e) \frac{\partial}{\partial x_v} \frac{1}{(\vec{r} - \vec{r}_e)^2 - c^2(t - t_e)^2} \right\} dt_e \quad (15.13)$$

The residues of the integrand at the poles determine the integral. These poles are the zeros $t_{e1} = t_r(\vec{r}, t) < t$ and $t_{e2} = t_a(\vec{r}, t) > t$ of the implicit equation

$$[\vec{r} - \vec{r}_e(t_e)]^2 - c^2(t - t_e)^2 = 0. \quad (15.14)$$

The zeros t_{e1} and t_{e2} are functions of the coordinates x, y, z of the point of observation and of the laboratory time t . However, only the pole $t_{e1} = t_r < t$ located within the integration loop contributes to the integral. Because the components

$\dot{x}_{ev}(t_e)$ of the velocity vector $\dot{\vec{r}}(t_e)$ and the components $S_{\mu\nu}(t_e)$ of the spin tensor either vanish or stay finite at infinity, we can perform the integrations in (15.13) analytically by determining the residues at the pole $t_{e1} = t_r < t$.

15.1.1 Field of an Electron Moving with Constant Velocity

To demonstrate the validity of the integral (15.13), we consider an electron which propagates with constant velocity $\dot{\vec{r}}_e = \vec{v} = \vec{\beta}c$, $\dot{x}_{e4} = ic$ and constant spin with components $S_{\mu\nu}$ in field-free space. For simplicity, we assume without loss of generality that at time $t_e = 0$ the electron is located at position $\vec{r}_e(0) = 0$. In this case, (15.14) adopts the simple form

$$(\vec{r} - \vec{v}t_e)^2 - c^2(t - t_e)^2 = 0. \quad (15.15)$$

This quadratic equation has the two solutions:

$$\begin{aligned} ct_{e1} &= \gamma^2 \left(tc - \vec{\beta}\vec{r} - \sqrt{(t\vec{\beta}c - \vec{r})^2 - \vec{\beta}^2\vec{r}^2} \right) < ct, \\ ct_{e2} &= \gamma^2 \left(tc - \vec{\beta}\vec{r} + \sqrt{(t\vec{\beta}c - \vec{r})^2 - \vec{\beta}^2\vec{r}^2} \right) > ct, \\ \gamma &= 1/\sqrt{1 - \beta^2}. \end{aligned} \quad (15.16)$$

The solution t_{e2} violates causality in the laboratory frame. Fortunately, it does not contribute to the integral (15.13) because it is located outside the integration loop. By taking the residue at the pole $t_e = t_{e1}$ and considering $x_4 = ict$, we eventually obtain

$$\begin{aligned} A_\mu &= \frac{e\mu_0}{4\pi\gamma^2} \left\{ \frac{\dot{x}_\mu}{\sqrt{(\vec{r} - \vec{\beta}tc)^2 - \beta^2\vec{r}^2}} + \frac{1}{m_e\gamma} \sum_v \frac{\partial}{\partial x_v} \frac{S_{\mu\nu}}{\sqrt{(\vec{r} + i\vec{\beta}x_4)^2 - \beta^2\vec{r}^2}} \right\} \\ &= \frac{e\mu_0}{4\pi\gamma^2} \left\{ \frac{\dot{x}_\mu}{\sqrt{(\vec{r} - \vec{\beta}tc)^2 - \beta^2\vec{r}^2}} \right. \\ &\quad \left. - \frac{1}{m_e\gamma} \frac{\sum_{v=1}^3 (x_v - v_v t \gamma^2) S_{\mu\nu} + i(\vec{\beta}\vec{r} - \beta^2 ct) S_{\mu 4}}{((\vec{r} - \vec{\beta}tc)^2 - \beta^2\vec{r}^2)^{3/2}} \right\}. \end{aligned} \quad (15.17)$$

The components of the spin vectors $\vec{S}_e$ and $\vec{S}_m$ are related with the components $S_{\mu\nu}$ of the spin tensor via $S_{23} = S_{mx}$, $S_{31} = S_{my}$, $S_{12} = S_{mz}$; $S_{14} = iS_{ex}$, $S_{24} = iS_{ey}$, $S_{34} = iS_{ez}$. In the system at rest, we have $\vec{S}_e = 0$, $\vec{S}_m = \vec{s}$. The components of the spin tensor transform in the same way as the components of the electromagnetic field tensor. Accordingly, if the particle moves with constant velocity $\vec{v}$ with respect to the observer, the moving spin has components:

$$\vec{S}_m = \vec{s} + \frac{\gamma^2}{1 + \gamma} \vec{\beta} \times (\vec{s} \times \vec{\beta}), \quad \vec{S}_e = \gamma \vec{s} \times \vec{\beta}. \quad (15.18)$$

By considering in addition the relations $\mu_0 \varepsilon_0 = 1/c^2$, we eventually find

$$\varphi = icA_4 = -\frac{e}{4\pi\varepsilon_0\gamma^2\sqrt{(\vec{r} - \vec{v}t)^2 - \beta^2\vec{r}^2}} - \frac{e\mu_0}{4\pi m_e\gamma^2} \frac{(\vec{r} \times \vec{\beta}) \cdot \vec{s}}{((\vec{r} - \vec{v}t)^2 - \beta^2\vec{r}^2)^{3/2}}, \quad (15.19)$$

$$\vec{A} = \frac{e\mu_0}{4m_e\pi\gamma^2} \frac{\vec{v}}{\sqrt{(\vec{r} - \vec{v}t)^2 - \beta^2\vec{r}^2}} - \frac{e\mu_0}{4m_e\pi\gamma^3} \frac{(\vec{r} - \vec{v}t\gamma^2) \times \vec{S}_m + (\vec{\beta}\vec{r} - \beta^2 ct)\vec{S}_e}{((\vec{r} - \vec{v}t)^2 - \beta^2\vec{r}^2)^{3/2}}. \quad (15.20)$$

By neglecting the spin terms ($\vec{S}_m = \vec{S}_e = \vec{s} = 0$), we obtain the components of the so-called *Lienard–Wiechert potentials* [37]. In the special case $\vec{\beta}c = \vec{v} = 0$, $\gamma = 1$, the spin tensor degenerates to the three-dimensional axial vector $\vec{S}_m = \vec{s} = s_x\vec{e}_x + s_y\vec{e}_y + s_z\vec{e}_z$. As a result, (15.19) and (15.20) adopt the simple forms

$$\varphi = \frac{-e}{4\pi\varepsilon_0 r}, \quad \vec{A} = \frac{e\mu_0}{m_e} \frac{\vec{s} \times \vec{r}}{4\pi r^3}. \quad (15.21)$$

These familiar relations represent the electrostatic potential φ and the magnetic vector potential of an electron with charge $q = -e$ and intrinsic magnetic moment $\vec{\mu} = e\mu_0\vec{s}/m_e$ in the system at rest. Due to our four-dimensional relativistic treatment of electron motion, we obtain the correct result for the magnetic vector potential of the electron without the need to introduce a g -factor.

15.2 Radiation Field Emitted by the Accelerated Electron

Any unbound charged particle emits electromagnetic radiation if its velocity changes. One utilizes this behavior for producing Roentgen radiation of variable energy in synchrotrons. We divide the total field of the electron into its accompanying field and its radiation field caused by acceleration. The accompanying field is stationary and attached to the particle, whereas the radiation field is emitted by the

particle if it is accelerated. To obtain the components of the four-dimensional vector potential of the arbitrarily moving electron, we must evaluate the integral (15.13) for an arbitrary motion $\vec{r}_e(t_e)$ of the electron. We tackle this task by considering that in the vicinity of the zero $t_{e1} = t_r(\vec{r}, t) < t$, the nominator (15.14) of the first integrand of the integrals (15.13) has the expansion

$$\begin{aligned} [\vec{r} - \vec{r}_e(t_e)]^2 - c^2(t - t_e)^2 &\approx \frac{d}{dt_e} \left[[\vec{r} - \vec{r}_e(t_e)]^2 - c^2(t - t_e)^2 \right]_{t_e=t_r} (t_e - t_r) \\ &= 2 \left[c^2(t - t_r) - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r) \right] (t_e - t_r) = 2c(t_e - t_r) \left[|\vec{r} - \vec{r}_e| (1 - \vec{n} \cdot \vec{\beta}) \right]_{\text{ret}}. \end{aligned} \quad (15.22)$$

Here, $c\vec{\beta} = \dot{\vec{r}}_e(t_r)$ is the instantaneous velocity of the particle. The unit vector $\vec{n} = (\vec{r} - \vec{r}_e(t_r)) / |\vec{r} - \vec{r}_e(t_r)|$ points from the position $\vec{r}_e(t_r)$ of the charge e to the observation point $\vec{r}$. We perform the integrations in (15.13) by expanding the integrands in a Laurent series about the pole $t_{e1} = t_r(\vec{r}, t) < t$. By considering (15.22), we can evaluate the residues which give the components of the 4-vector potential as

$$\begin{aligned} A_\mu &= \frac{ec\mu_0}{4\pi} \frac{\dot{x}_{e\mu}(t_r)}{c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r)} \\ &\quad - \frac{ec\mu_0}{4\pi m_e \gamma_r} \sum_{\nu=1}^4 S_{\mu\nu}(t_r) \frac{\partial}{\partial x_\nu} \frac{1}{c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r)}. \end{aligned} \quad (15.23)$$

The velocity $\vec{v}_e = c\vec{\beta} = \vec{v}_e(t_r)$, the spin vectors $\vec{S}_m = \vec{S}_m(t_r)$ and $\vec{S}_e = \vec{S}_e(t_r)$, the electron position vector $\vec{r}_e = \vec{r}_e(t_r)$, and the relativistic factor $\gamma_r = \gamma_r(t_r)$ are functions of the retarded time $t_r = t_r(t, \vec{r})$. Therefore, these quantities depend implicitly on the position $\vec{r}$ and time t of observation.

For performing the differentiations with respect to x_ν , it is advantageous to utilize the relations

$$\frac{\partial t_r}{\partial t} = \frac{1}{1 - \vec{\beta} \cdot \vec{n}}, \quad \frac{\partial t_r}{\partial x_\nu} = \frac{x_{e\nu} - x_\nu}{c |\vec{r} - \vec{r}_e|} \frac{1}{1 - \vec{\beta} \cdot \vec{n}}, \quad \nu = 1, 2, 3. \quad (15.24)$$

By employing these expressions, we obtain

$$\frac{\partial |\vec{r} - \vec{r}_e|}{\partial t} = -\frac{\vec{n} \cdot \dot{\vec{r}}_e}{1 - \vec{n} \cdot \vec{\beta}}, \quad \frac{\partial (\vec{r} - \vec{r}_e) \dot{\vec{r}}_e}{\partial t} = \frac{(\vec{r} - \vec{r}_e) \ddot{\vec{r}}_e - \dot{\vec{r}}_e^2}{1 - \vec{n} \cdot \vec{\beta}}, \quad (15.25)$$

$$\frac{\partial |\vec{r} - \vec{r}_e|}{\partial x_\nu} = \frac{x_\nu - x_{e\nu}}{|\vec{r} - \vec{r}_e|} \frac{1}{1 - \vec{n} \cdot \vec{\beta}}, \quad \frac{\partial (\vec{r} - \vec{r}_e) \dot{\vec{r}}_e}{\partial x_\nu} = \dot{x}_{e\nu} - \frac{(\vec{r} - \vec{r}_e) \ddot{\vec{r}}_e - \dot{\vec{r}}_e^2}{1 - \vec{n} \cdot \vec{\beta}} \frac{x_\nu - x_{e\nu}}{c |\vec{r} - \vec{r}_e|}, \quad (15.26)$$

$$\frac{\partial(1/\gamma_r)}{\partial t} = \frac{d(1/\gamma_r)}{dt_r} \frac{\partial t_r}{\partial t} = -\frac{\gamma_r \vec{\beta} \cdot \dot{\vec{\beta}}}{1 - \vec{\beta} \cdot \vec{n}}. \quad (15.27)$$

Using the results (15.25) and (15.27), we find

$$\begin{aligned} \frac{\partial}{\partial x_v} \frac{1}{c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r)} &= - \frac{c \partial |\vec{r} - \vec{r}_e| / \partial x_v - \partial \left[(\vec{r} - \vec{r}_e) \dot{\vec{r}}_e \right] / \partial x_v}{\left[c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r) \right]^2} \\ &= \frac{\dot{x}_v}{c^2 (1 - \vec{n} \vec{\beta})^2 |\vec{r} - \vec{r}_e|^2} - \frac{(x_v - x_{ev}) (c^2 + (\vec{r} - \vec{r}_e) \ddot{\vec{r}}_e - \dot{\vec{r}}_e^2)}{c^3 (1 - \vec{n} \vec{\beta})^3 |\vec{r} - \vec{r}_e|^3}, \end{aligned} \quad (15.28)$$

$$\begin{aligned} \frac{\partial}{\partial t} \frac{1}{c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r)} &= - \frac{c \partial |\vec{r} - \vec{r}_e| / \partial t - \partial \left[(\vec{r} - \vec{r}_e) \dot{\vec{r}}_e \right] / \partial t}{\left[c |\vec{r} - \vec{r}_e(t_r)| - (\vec{r} - \vec{r}_e(t_r)) \dot{\vec{r}}_e(t_r) \right]^2} \\ &= \frac{c^2 \vec{n} \vec{\beta} + (\vec{r} - \vec{r}_e) \ddot{\vec{r}}_e - \dot{\vec{r}}_e^2}{c^2 (1 - \vec{n} \vec{\beta})^3 |\vec{r} - \vec{r}_e|^2}. \end{aligned} \quad (15.29)$$

By taking into account the expressions (15.23)–(15.29), the components of the four-dimensional vector potential (15.23) adopt the form

$$\begin{aligned} \varphi &= -icA_4 = \frac{e}{4\pi\epsilon_0 |\vec{r} - \vec{r}_e| (1 - \vec{n} \vec{\beta})_{\text{ret}}} \\ &\quad + \frac{e\mu_0 c}{4\pi m_e \gamma_r} \sum_{v=1}^3 i S_{4v} \frac{\partial}{\partial x_v} \frac{1}{c |\vec{r} - \vec{r}_e| - (\vec{r} - \vec{r}_e) \dot{\vec{r}}_e} \\ &= \frac{e}{4\pi\epsilon_0 |\vec{r} - \vec{r}_e| (1 - \vec{n} \vec{\beta})_{\text{ret}}} \\ &\quad + \frac{e\mu_0}{4\pi m_e \gamma_r} \frac{\vec{\beta} \cdot \vec{S}_e (1 - \vec{n} \vec{\beta}) |\vec{r} - \vec{r}_e| - (\vec{r} - \vec{r}_e) \cdot \vec{S}_e (1 + (\vec{r} - \vec{r}_e) \cdot \ddot{\vec{r}}_e / c^2 - \vec{\beta}^2)}{(1 - \vec{n} \vec{\beta})^3 |\vec{r} - \vec{r}_e|^3} \\ \vec{A} &\approx \frac{e\mu_0}{4\pi} \frac{\vec{v}_e}{|\vec{r} - \vec{r}_e| (1 - \vec{n} \vec{\beta})_{\text{ret}}} + \frac{e\mu_0}{4\pi m_e \gamma_r c} \frac{\dot{\vec{S}}_m \times (\vec{r} - \vec{r}_e) - \vec{S}_m \times \dot{\vec{r}}_e}{\left[|\vec{r} - \vec{r}_e| (1 - \vec{n} \vec{\beta})_{\text{ret}} \right]^2}. \end{aligned} \quad (15.30)$$

These relations adopt the forms (15.19) and (15.20) for an electron moving with constant velocity $\ddot{\vec{r}}_e = 0$ and constant spin $\dot{\vec{S}}_m = 0$.

We must also consider the relations (15.23)–(15.29) for calculating the electromagnetic field strengths of the moving charge given by

$$\vec{E} = -\vec{\nabla} \varphi - \frac{\partial \vec{A}}{\partial t}, \quad \vec{B} = \vec{\nabla} \times \vec{A}. \quad (15.31)$$

The energy flow of the radiation (emitted energy per area and time) is given by the Poynting vector:

$$\vec{S}_P = \vec{E}_{\text{rad}} \times \vec{H}_{\text{rad}} = \frac{1}{\mu_0} \vec{E}_{\text{rad}} \times \vec{B}_{\text{rad}}. \quad (15.32)$$

The radiation field strengths $\vec{E}_{\text{rad}}$ and $\vec{H}_{\text{rad}} = \vec{B}_{\text{rad}}/\mu_0$ comprise all terms of $\vec{E}$ and $\vec{B}$ which are inversely proportional to the distance $\vec{r} - \vec{r}_e$ of the observer from the particle because the radiation energy E_r , propagating per unit time through the surface of a sphere surrounding the particle, must be the same for any distance $|\vec{r} - \vec{r}_e|$ from the center of the particle:

$$\dot{E}_r = \frac{dE_r}{dt} = \oint \vec{S}_P \cdot d\vec{\sigma} = \frac{(\vec{r} - \vec{r}_e)^2}{\mu_0} \oint (\vec{E}_{\text{rad}} \times \vec{B}_{\text{rad}}) \cdot d\vec{\Omega} > 0. \quad (15.33)$$

We have derived the expression on the right-hand side of (15.33) by substituting (15.32) for $\vec{S}_P$ and by replacing the differential surface element of the sphere $d\vec{\sigma}$ by the differential solid angle $d\vec{\Omega}$:

$$d\vec{\sigma} = (\vec{r} - \vec{r}_e)^2 d\vec{\Omega}. \quad (15.34)$$

The calculation readily shows that the radiation terms of the electromagnetic field strengths are linear in $\ddot{x}_\mu$ and $\ddot{S}_{\mu\nu}$ and inversely proportional to the distance $|\vec{r} - \vec{r}_e|$. In deriving these terms from (15.30), we consider the relations

$$\frac{\partial \dot{x}_{e\mu}(t_{e\lambda})}{\partial x_\nu} = \ddot{x}_{e\mu} \frac{\partial t_{e\lambda}}{\partial x_\nu}, \quad \frac{\partial \dot{S}_{\mu\nu}(t_{e\lambda})}{\partial x_\nu} = \ddot{S}_{\mu\nu} \frac{\partial t_{e\lambda}}{\partial x_\nu}. \quad (15.35)$$

Our results demonstrate that the radiation of a charged particle originates from the acceleration of its motion and its spin.

15.2.1 Dipole Approximation

In many cases, radiation is emitted by charged particles oscillating or rotating about a fixed position $\vec{r}_0$ such that the distance $\vec{a} = \vec{a}(t_e)$ of the particle from the center of motion $\vec{r}_0$ is small compared with the distance $\vec{r} - \vec{r}_0$ of the observer. Within the frame of the dipole approximation, we assume that $\vec{a} = \vec{r}_0 - \vec{r}_e(t)$ is small compared with $\vec{r} - \vec{r}_0$:

$$\vec{r} - \vec{r}_e(t_e) = \vec{r} - \vec{r}_0 + \vec{a}(t) \approx \vec{r} - \vec{r}_0. \quad (15.36)$$

By substituting $\vec{r} - \vec{r}$ for $\vec{r} - \vec{r}_e(t_e)$ in (15.14), we readily find for the contributing pole $t_e = t_{e1} < t$ the familiar relation

$$t_{e1} = t_{\text{red}} = t - |\vec{r} - \vec{r}_0|/c. \quad (15.37)$$

Using this result together with (15.36) in (15.13), we obtain for the components of the 4-vector potential the dipole approximation:

$$\begin{aligned}
 A_\mu &\approx -\frac{e\mu_0}{4\pi} \frac{\dot{x}_{e\mu}(t_{\text{ret}})}{|\vec{r} - \vec{r}_e(t_{\text{ret}})|} + \frac{e\mu_0}{4\pi m_e \gamma} \sum_\nu S_{\mu\nu}(t_{\text{ret}}) \frac{\partial}{\partial x_\nu} \frac{1}{|\vec{r} - \vec{r}_e(t_{\text{ret}})|} \\
 &\approx -\frac{e\mu_0}{4\pi} \frac{\dot{x}_{e\mu}(t_{\text{ret}})}{|\vec{r} - \vec{r}_0|} - \frac{e\mu_0 \dot{x}_{e\mu}(t_{\text{ret}})}{4\pi} \frac{\vec{a} \cdot (\vec{r} - \vec{r}_0)}{|\vec{r} - \vec{r}_0|^3} \\
 &\quad + \frac{e\mu_0}{4\pi m_e \gamma} \sum_\nu S_{\mu\nu}(t_{\text{ret}}) \frac{\partial}{\partial x_\nu} \frac{1}{|\vec{r} - \vec{r}_0|}. \tag{15.38}
 \end{aligned}$$

By considering that $\vec{S}_m = \vec{S}_m(t_{\text{ret}})$, $\vec{S}_e = \vec{S}_e(t_{\text{ret}})$, $\vec{v}_e = \vec{v}_e(t_{\text{ret}})$, and $\vec{a} = \vec{a}(t_{\text{ret}})$ are functions of t and $\vec{r}$, we rewrite (15.38) in the form

$$\begin{aligned}
 \varphi &= -icA_4 \approx \frac{e}{4\pi\epsilon_0 |\vec{r} - \vec{r}_0|} + \frac{e\vec{a} \cdot (\vec{r} - \vec{r}_0)}{4\pi\epsilon_0 |\vec{r} - \vec{r}_0|^3} + \frac{e\mu_0}{4\pi m_e \gamma} \frac{c\vec{S}_e \cdot (\vec{r} - \vec{r}_0)}{|\vec{r} - \vec{r}_0|^3}, \\
 \vec{A} &\approx \frac{e\mu_0}{4\pi} \frac{\vec{v}_e}{|\vec{r} - \vec{r}_0|} + \frac{e\mu_0 \vec{v}_e}{4\pi} \frac{\vec{a} \cdot (\vec{r} - \vec{r}_0)}{|\vec{r} - \vec{r}_0|^3} + \frac{e\mu_0}{4\pi m_e \gamma} \frac{\vec{S}_m \times (\vec{r} - \vec{r}_0)}{|\vec{r} - \vec{r}_0|^3}. \tag{15.39}
 \end{aligned}$$

The spin terms account for the electromagnetic potentials resulting from the intrinsic magnetic moment of the electron. Because these terms are small compared with those produced by the electric charge, they are generally disregarded.

15.2.2 Radiation Field of the Moving Electron

The electromagnetic field strengths $\vec{E}_{\text{rad}}$ and $\vec{B}_{\text{rad}}$ of the radiation field are the far-field terms of (15.31) which are inversely proportional to the distance $\vec{r} - \vec{r}_e$ of the observer from the particle. We obtain these terms from (15.39) after some analytical manipulations as

$$\begin{aligned}
 \vec{E}_{\text{rad}} &= -\frac{e\mu_0}{4\pi} \frac{\dot{\vec{v}}}{|\vec{r} - \vec{r}_e|} - \frac{e\mu_0(\vec{r} - \vec{r}_0)}{4\pi\gamma cm_e |\vec{r} - \vec{r}_e|^2} \times \left\{ \frac{\ddot{\vec{S}}_e \times (\vec{r} - \vec{r}_e)}{|\vec{r} - \vec{r}_0|} - \ddot{\vec{S}}_m \right\}, \\
 \vec{H}_{\text{rad}} &= \frac{\vec{B}_{\text{rad}}}{\mu_0} = \frac{e}{4\pi c} \frac{\dot{\vec{v}} \times (\vec{r} - \vec{r}_e)}{|\vec{r} - \vec{r}_e|^2} + \frac{e(\vec{r} - \vec{r}_e)}{4\pi\gamma m_e c^2 |\vec{r} - \vec{r}_e|^2} \times \left\{ \frac{\ddot{\vec{S}}_m \times (\vec{r} - \vec{r}_e)}{|\vec{r} - \vec{r}_e|} - \ddot{\vec{S}}_e \right\}. \tag{15.40}
 \end{aligned}$$

If we neglect the radiation originating from the acceleration of the magnetic and electric spin components ($\ddot{\vec{S}}_m = \ddot{\vec{S}}_e = 0$), we derive the standard equations for

the radiation field of an accelerated electron or charged particle, respectively. By employing the relations (15.40), we verify after a lengthy calculation that the electric field strength $\vec{E}_{\text{rad}}$ and the magnetic field strengths $\vec{B}_{\text{rad}}$ of the radiation field are orthogonal to each other, as it must be the case for any electromagnetic wave propagating in vacuum.

References

1. E. Abbe, in *Die optischen Hilfsmittel der Mikroskopie*, ed. by A.W. Hofmann. Report on Scientific Instruments at the London International Exhibition 1876 (Vieweg, Braunschweig, 1878), pp. 383–420
2. L. De Broglie, *Ann. de Phys.* **3**, 22 (1925)
3. H. Busch, *Ann. Phys.* **4**, 974 (1927)
4. M. Knoll and E. Ruska, *Ann. Phys.* **12**, 607 (1931)
5. E. Brueche, O. Scherzer, *Geometrische Elektronenoptik* (Springer, Berlin, 1934)
6. O. Scherzer, in *Beitraege zur Elektronenoptik*, ed. by H. Busch, E. Brueche (Barth, Leipzig, 1937), pp. 33–41
7. W. Glaser, *Z. Physik.* **80**, 451 (1933)
8. O. Scherzer, *Z. Physik.* **101**, 593 (1936)
9. W. Glaser, *Grundlagen der Elektronenoptik* (Springer, Wien, 1952)
10. P.A. Sturrock, *Proc. R. Soc. A (London)*. **210**, 269 (1951)
11. O. Klemperer, *Electron Optics* (Cambridge University Press, Cambridge, 1953)
12. V.E. Cosslett, *Introduction to Electron Optics* (Oxford University Press, Oxford, 1946)
13. V.K. Zworykin, G.A. Morton, E.G. Ramberg, J. Hillier, A.W. Vance, *Electron Optics and the Electron Microscope* (Wiley, New York, 1945)
14. P. Grivet, *Electron Optics* (Pergamon, London, 1972)
15. P.W. Hawkes, E. Kasper, *Principles of Electron Optics*, vols. 1–3 (Academic, London, 1995)
16. O. Scherzer, *Optik* **2**, 114 (1947)
17. W. Bernhard, *Optik* **57**, 73 (1980)
18. D.F. Hardy, Dissertation, University of Cambridge (1967)
19. V. Beck, *Proc. Annu. Meet. EMSA.* **35**, 90 (1977)
20. E. Plies, *Microelectron. Eng.* **12**, 189 (1989)
21. E. Plies, in *Advances in Optical and Electron Microscopy*, vol. 13, ed. by T. Mulvey, C.J.R. Sheppard (Academic, London, 1994), pp. 123–242
22. H. Rose, *Optik* **85**, 19 (1990)
23. L. Reimer (ed.), *Energy-Filtering Electron Microscopy, Springer Series in Optical Sciences*, vol. 71 (Springer, Berlin, 1995)
24. M. Haider, H. Rose, S. Uhlemann, B. Kabius, K. Urban, *J. Electron Microsc.* **47**, 395 (1998)
25. O.L. Krivanek, N. Dellby, A.R. Lupini, *Ultramicroscopy* **78**, 1 (1999)
26. S. Uhlemann, H. Rose, *Optik* **96**, 163 (1994)
27. F. Kahl, H. Rose, *Proceedings of the 14th International Conference on Electron Microscopy, Cancun*, vol. 1 (Institute of Physics, Bristol, 1998), pp. 71–72
28. H. Rose, in *Advances in Imaging and Electron Physics*, vol. 132, ed. by P.W. Hawkes (Academic, London, 2004), pp. 247–285
29. R.K. Luneburg, *Mathematical Theory of Optics* (University California Press, Berkeley, 1966)

30. E. Ruska, *The Early Development of Electron Lenses and Electron Microscopy* (Hirzel, Stuttgart, 1980)
31. A. Septier (ed.), *Focusing of Charged Particles*, vols. I, II (Academic, New York, 1967)
32. H. Wollnik, *Optics of Charged Particles* (Academic, London, 1987)
33. O. Scherzer, *Optik* **22**, 314 (1965)
34. H. Wiedemann, *Particle Accelerator Physics: Basic Principles and Linear Beam Dynamics* (Springer, New York, 1995)
35. M.E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961)
36. J.J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, New York, 1967)
37. J. D. Jackson, *Classical Electrodynamics*, 2nd edn. (Wiley, New York, 1975)
38. V. Bargmann, L. Michel, V.L. Telegedi, *Phys. Rev. Lett.* **2**, 435, (1959)
39. H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980)
40. M. Born, E. Wolf, *Principles of Optics*, 7th edn. (Cambridge University Press, Cambridge, 1999)
41. Y. Aharonov, D. Bohm, *Phys. Rev.* **115**, 485, (1959)
42. G. Moellenstedt, H. Dueker, *Z. Phys.* **145**, 377, (1956)
43. H. Lichte, in *Handbook of Microscopy*, ed. by I.S. Amelinckx, D. van Dyk, J. van Landuyt, G. van Tendeloo (VCH, Weinheim, 1997)
44. A. Tonomura, *Electron Holography* (Springer, Heidelberg, 1999)
45. G. Kirchhoff, *Ann. Phys.* **18**, 663 (1882)
46. A. Sommerfeld, *Optics* (Academic, New York, 1954)
47. G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, London, 1966)
48. M.R. Spiegel, *Vector Analysis* (McGraw-Hill, New York, 1974)
49. F. Ollendorff, *Potentialfelder der Elektrotechnik* (Springer, Berlin, 1932)
50. E. Munro, in *Image Processing and Computer-Aided Design in Electron Optics*, ed. by P.W. Hawkes (Academic, London, 1978), pp. 284–323
51. E. Plies, H. Rose, *Optik* **34**, 171, (1971)
52. E. Plies, D. Typke, *Z. Naturforsch. A* **33**, 1361 (1978)
53. P.A. Sturrock, *Static and Dynamic Electron Optics* (Cambridge University Press, Cambridge, 1955)
54. H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **258**, 374 (1987)
55. G.H. Hoffstaetter, H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **328**, 398 (1993)
56. H. Hoch, E. Kasper, D. Kern, *Optik* **50**, 413 (1978)
57. G. Schoenecker, R. Spehr, H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **299**, 360 (1990)
58. B. Lencova, *Nucl. Instrum. Meth. Phys. Res. A* **427**, 329 (1999)
59. E. Kasper, *Optik* **46**, 271 (1976)
60. W. Glaser, P. Schiske, *Optik* **11**, 422, (1954)
61. W. Glaser, in *Handbuch der Physik*, ed. by S. Fluegge, vol. 33 (Springer, Berlin, 1956), pp. 123–395
62. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965)
63. G. Schoenecker, H. Rose, R. Spehr, *Ultramicroscopy* **20**, 203 (1986)
64. D. Preikszas, H. Rose, *Optik* **100**, 179 (1995)
65. A.A. Rusterholz, *Elektronenoptik*, vol. 1. Grundzuege der theoretischen Elektronenoptik (Birkhaeuser, Basel, 1950)
66. J. Picht, *Einführung in die Theorie der Elektronenoptik* (Barth, Leipzig, 1939)
67. A. Melkich, *Sitzungsber. Akad. Wiss. Wien, Math-Nat. Kl. Abt. IIa*, **155**, 393 (1947)
68. E.D. Courant, H. Snyder, *Ann. Phys.* **3**, 1 (1958)
69. O. Rang, *Optik* **5**, 518 (1949)
70. H. Kawakatsu, K.G. Vosburgh, B. Siegel, *J. Appl. Phys.* **30**, 245 (1968)
71. H.D. Bauer, *Optik* **23**, 596 (1965/1966)
72. G.D. Archard, *Br. J. Phys.* **5**, 294 (1954)
73. J.M.H. Deltrap, in *Proceedings of the 3rd European Regional Conference on Electron Microscopy, Prague*, vol. A, ed. by M. Titlbach (Czechoslovak Academy of Sciences, Prague, 1964), pp. 45–46

74. H. Rose, *Optik* **33**, 1 (1971)
75. P.W. Hawkes, *Philos. Trans. R. Soc. A* **257**, 479 (1965)
76. P.W. Hawkes, *Quadrupole Optics, Springer Tracts in Modern Physics*, vol. 42 (Springer, Berlin, 1966)
77. H. Rose, *Optik* **84**, 91 (1990)
78. H. Rose, E. Plies, *Optik* **40**, 336 (1974)
79. H. Rose, D. Krahl, in *Energy-Filtering Transmission Electron Microscopy*, ed. by L. Reimer (Springer, Berlin, 1995), pp. 43–149
80. R. Degenhardt, H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **298**, 171 (1990)
81. H. Mueller, D. Preikszas, H. Rose, *J. Electron Microsc. A* **48**, 191 (1999)
82. H. Rose, *Optik* **85**, 95 (1990)
83. D.W.O. Heddle, *Electrostatic Lens Systems* (Adam Hilhger, Breistol, 1991)
84. E. Harting, F.H. Read, *Electrostatic Lenses* (Elsevier, Amsterdam, 1976)
85. L.A. Baranova, S.Y. Yavor, *Sov. Phys. Tech. Phys.* **29**, 827 (1984)
86. W. Glaser, *Z. Phys.* **117**, 285 (1941)
87. W.D. Riecke, in *Magnetic Electron Lenses*, ed. by P.W. Hawkes (Springer, Berlin, 1982), pp. 163–357
88. A.D. Dymnikov, S.Y. Yavor, *Sov. Phys. Tech. Phys.* **8**, 639 (1963)
89. H. Rose, *Ultramicroscopy* **78**, 13 (1999)
90. E. Regenstreif, in *Focusing of Charged Particles*, ed. by A. Septier, vol. 1 (Academic, New York, 1967), pp. 353–410
91. H. Rose, *Optik* **36**, 19 (1972)
92. C. Caratheodory, *Geometrische Optik, Ergeb. Math. Grenzgeb.*, vol. 5 (Springer, Berlin, 1937)
93. H. Rose, *Optik* **24**, 36 (1966/1967)
94. G. Wuestefeld, *Proceedings of the Workshop on Polarized Protons at High Energies* (DESY, Hamburg, 1999). A. Luccio, T. Roser, *Proceedings of the 3rd Workshop on Siberian Snakes and Spin Rotators* (Brookhaven National Laboratory, Upton, NY, 1994)
95. M. Cotte, *Ann. Phys. (Paris)*. **10**, 333 (1938)
96. H. Rose, *Optik* **51**, 15 (1978)
97. P. Schmid, H. Rose, *J. Vac. Sci. Technol. B* **19**, 2555 (2011)
98. J.W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968)
99. H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **187**, 187 (1981)
100. V.M. Kelman, S.Y. Yavor, *Zh. Tekh. Fiz.* **31**, 1439 (1961)
101. H. Rose, *Optik* **34**, 285 (1971)
102. J. Zach, M. Haider, *Optik* **98**, 112 (1995)
103. M. Haider, W. Bernhardt, H. Rose, *Optik* **68**, 9 (1982)
104. M. Berz, *Modern Map Methods in Particle Beam Physics* (Academic, San Diego, 1999)
105. W. Glaser, in *Beiträge zur Elektronenoptik*, ed. by H. Busch, E. Brüche (Barth, Leipzig, 1937), pp. 24–32
106. H. Rose, U. Petri, *Optik* **33**, 151 (1971)
107. H. Rose, in *High-Resolution Imaging and Spectrometry of Materials*, ed. by F. Ernst, M. Ruehle (Springer, Berlin, 2002), pp. 189–270
108. H. Bruns, *Abh. K. Saechs. Ges. Wiss., Math.-Phys. Kl.* **21**, 321 (1895)
109. E.H. Linfoot, *Recent Advances in Optics* (Oxford University Press, Oxford, 1955)
110. E. Zeitler, *Nucl. Instrum. Meth. Phys. Res. A* **298**, 234 (1990)
111. H. Boersch, *Z. Phys.* **139**, 115 (1954)
112. H. Rose, R. Spehr, in *Applied Charged Particle Optics*, vol. C, ed. by A. Septier (Academic, New York, 1983), pp. 475–530
113. A. Weidenhausen, R. Spehr, H. Rose, *Optik* **69**, 126 (1985)
114. P. Kruit, G. Jansen, in *Handbook of Charged Particle Optics*, ed. by J. Orloff (CRC Press, Boca Raton, 2009)
115. C. Lejeune, J. Aubert, in *Applied Charged Particle Optics*, Suppl. 13A to Adv. Electron. Electron Phys. vol. A, ed. by A. Septier (Academic, New York, 1980), pp. 159–259
116. S. Uhlemann, H. Rose, *Ultramicroscopy* **63**, 161 (1996)

117. P. T. Saunders, *Catastrophe Theory* (Cambridge University Press, London, 1980)
118. A.J. Dragt, *Physics of High-Energy Particle Accelerators*, *AIP Conference Proceedings* **87** (1982)
119. H. Rose, *Optik* **27**, 466 (1968)
120. R.L. Seliger, *J. Appl. Phys.* **43**, 2352 (1972)
121. H. Boersch, J. Geiger, W. Stickel, *Z. Phys.* **180**, 415 (1964)
122. P.E. Batson, *Rev. Sci. Instrum.* **57**, 43 (1986)
123. H.W. Mook, P. Kruit, *Ultramicroscopy* **78**, 43 (1999)
124. E. Plies, J. Baertle, *Microsc. Microanal.* **9** (Suppl. 3), 28 (2003)
125. K. Tsuno, M. Terauchi, M. Tanaka, *Instrum. Phys. Conf. Ser.* **98**, 71 (1989)
126. H. Rose, *Optik* **77**, 26 (1987)
127. S. Lanio, *Optik* **73**, 99 (1986)
128. S. Lanio, H. Rose, D. Krah, *Optik* **73**, 56 (1986)
129. O.L. Krivanek, A.J. Gubbens, N. Dellby, *Microsc. Microanal. Microstruct.* **2**, 315 (1991)
130. S. Uhlemann, H. Rose, *Optik* **96**, 163 (1994)
131. K. Tsuno, *Nucl. Instrum. Meth. Phys. Res. A* **519**, 286 (2004)
132. V.D. Beck, *Optik* **53**, 241 (1979)
133. A.V. Crewe, D. Kopf, *Optik* **56**, 301 (1980)
134. J.A. Rouse, in *Advances in Optical and Electron Microscopy*, vol. 13, ed. by T. Mulvey (Academic, London, 1994), pp. 1–121
135. E. Munro, in *Handbook of Charged Particle Optics*, ed. by J. Orloff (CRC, Baton Rouge, 1997), pp. 1–76
136. R. Fink, M.R. Weiss, E. Umbach, D. Preikszas, H. Rose, R. Spehr, P. Hartel, W. Engel, R. Degenhardt, R. Wichtendahl, H. Kühlenbeck, W. Erlebach, K. Ihrmann, R. Schloegel, H.J. Freund, A.M. Bradshaw, G. Lilienkamp, T. Schmidt, E. Bauer, G. Benner, *J. Electron Spectrosc. Relat. Phenom.* **84**, 231 (1997)
137. E. Plies, *Ultramicroscopy* **93**, 305 (2002)
138. J. Frosien, E. Plies, K. Anger, (1989) *J. Vac. Sci. Technol. B* **7**: 1874–1877
139. H. Rose, *Optik* **26**, 289 (1967/1968)
140. S. Uhlemann, M. Haider, *Ultramicroscopy* **72**, 109 (1998)
141. W. Tretner, *Optik* **11**, 312 (1954)
142. W. Tretner, *Optik* **16**, 155 (1959)
143. R.W. Moses, in *Image Processing and Computer-Aided Design in Electron Optics*, ed. by P.W. Hawkes (Academic, New York, 1973), pp. 250–272
144. O. Scherzer, *J. Appl. Phys.* **20**, 20 (1949)
145. J. Zach, M. Haider, *Nucl. Instrum. Meth. Phys. Res. A* **365**, 316 (1995)
146. D. Typke, *Optik* **34**, 573 (1972)
147. D. Typke, *Optik* **36**, 124 (1972)
148. G. Schoenhense, H. Spieker, *J. Vac. Sci. Technol. B* **20**, 2526 (2002)
149. A. Septier, in *Advances in Optical and Electron Microscopy*, vol. 1, ed. by R. Barer, V.E. Cosslett (Academic, New York, 1966), pp. 204–274
150. C. Weissbaecker, H. Rose, *J. Electron Microsc.* **50**, 383 (2001)
151. C. Weissbaecker, H. Rose, *J. Electron Microsc.* **51**, 45 (2001)
152. A. Huber, J. Baertle, E. Plies, *Nucl. Instrum. Meth. Phys. Res. A* **519**, 320 (2004)
153. D.C. Carey, *Nucl. Instrum. Meth. Phys. Res. A* **189**, 365 (1981)
154. P. Hartel, D. Preikszas, R. Spehr, H. Mueller, H. Rose, *Adv. Imaging Electron Phys.* **120**, 41 (2002)
155. M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, K. Urban, *Nature* **392**, 768 (1998)
156. C.L. Jia, M. Lentzen, K. Urban, *Science* **299**, 870 (2004)
157. H. Rose, *Nucl. Instrum. Meth. Phys. Res. A* **519**, 12 (2004)
158. A. Recknagel, *Z. Phys.* **104**, 381 (1936)
159. E.G. Ramberg, *J. Appl. Phys.* **20**, 183 (1949)
160. G.F. Rempfer, *J. Appl. Phys.* **67**, 6027 (1990)
161. D. Preikszas, H. Rose, *J. Electron Microsc.* **46**, 1 (1997)

- 162. W. Wan, J. Feng, H.A. Padmore, D. Robin, Nucl. Instrum. Meth. Phys. Res. A **519**, 222 (2004)
- 163. R. Castaing, L. Henry, J. Microsc. **3**, 133 (1964)
- 164. R.M. Henkelman, F.P. Ottensmeyer, J. Microsc. **102**, 79 (1979)
- 165. G.F. Rempfer, M.S. Mauck, Proc. Annu. Meet. EMSA. **43**, 132 (1985)
- 166. J. R. Pierce, *Theory and Design of Electron Beams* (Van Nostrand, Princeton, 1949)
- 167. E. Kasper, Adv. Opt. Electron Microsc. **8**, 207 (1982)
- 168. R. Lauer, Adv. Opt. Electron Microsc. **8**, 137 (1982)
- 169. E. Fischer, Z. Phys. **156**, 1 (1959)
- 170. F.M. Penning, Physica **3**, 873 (1936)
- 171. J.S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, London, 1987)
- 172. R.P. Feynman, Phys. Rev. **76**, 749 (1949)
- 173. L. Foldy, S.A. Wouthuysen, Phys. Rev. **78**, 29 (1950)
- 174. L.H. Thomas, Philos. Mag. **3**, 1 (1927)
- 175. W. Gerlach, O. Stern, Z. Phys. **9**, 353 (1922)

Nomenclature

Symbol	Quantity
e	Charge of electron
m_e	Rest mass of electron
c	Velocity of light
$E_0 = m_e c^2$	Rest energy of electron
v	Velocity
$\vec{p} = m\vec{v} - e\vec{A}$	Canonical momentum
$\vec{B}$	Magnetic induction
$\vec{E}$	Electric field strength
E	Total energy
$\vec{A}$	Magnetic vector potential
ψ	Scalar magnetic potential
$\Pi = \psi + i\Omega$	Complex magnetic potential
φ	Electric potential
Γ	Complex curvature
γ	Relativistic factor
τ	Generalized time (Chap. 10)
τ	Universal time
$\tau(z)$	Torsion of the optic axis
$T(z)$	Axial lens strength of round lenses
$G(z)$	Quadrupole lens strength
$\Phi_v(z)$	Electric multipole strength
$\Phi(z) = \Phi_0(z)$	Axial electric potential
$\Psi_v(z)$	Magnetic multipole strength
f	Focal length
$\chi(z)$	Angle of rotating coordinate system

M

$$1/\varepsilon = 2E_0/e$$

$$\Phi^* = \Phi(1 + \varepsilon\Phi)$$

$$\kappa = \Delta E/e\Phi_0$$

$$q = \sqrt{2m_e e \Phi^*}$$

$$W(x, y, z, t)$$

$$S(x, y, z)$$

 $\hbar$

$$\alpha_S \approx 1/137$$

 H L V U τ_r $\vec{F}$ λ

$$w = x + iy$$

$$\kappa^* = 2\gamma_0\kappa/(1 + \gamma_0)$$

Magnification

Normalization voltage

Relativistic modified axial electric potential

Relative energy deviation

Axial kinetic momentum

Action function

Reduced action function

Planck's constant

Sommerfeld fine-structure constant

Hamiltonian

Lagrangian

Potential energy

Voltage

Characteristic radiation time

Force

Wavelength

Complex off-axial distance

Relativistic modified chromatic parameter

Index

- Abbe formula, 43
- Abbe sine condition, 198
- Abbe's theory, 40
- Aberrations, 282
 - coefficients, 298, 299, 382
 - correctors, x
 - figure, 299
- Acceleration
 - lens, 148
 - voltage, 75, 135
- Acceptance domain, 232
- Achromatic
 - aperture, 315
 - aplanat, 385
 - electron optical aplanat, 385
- Achroplanator, 370
- Action, 8
- Action integral, 391
- Adaptor lenses, 373
- Ag atom, 474
- Aharanov–Bohm effect, 21
- Airy disc, 365
- Airy pattern, 43
- Analytical transmission electron microscope, 288
- Anamorphic
 - images, 181, 186, 371
 - lenses, 435
- Anamorphic, 131
- Anamorphic image, 176
- Anastigmatism condition, 269
- Angle
 - of acceptance, 411
 - eikonal, 209
 - of incidence, 16
 - of Larmor rotation, 29, 341
 - of reflection, 16
- Angular
 - dispersion, 268
 - frequency, 448
 - illumination, 111
 - radiation characteristics, 23
- Anisotropic coma, 307, 381
- Anisotropic distortion, 312
- Anisotropic medium of refraction, 20
- Apex, 78, 100, 309, 414
- Aplanat, 210
- Aplanatic
 - corrector, 374
 - imaging, 211
- Aplanatism, 211
- Aplanator, 381
- Apochromator, 385
- Aristotle, 444
- Astigmatic
 - difference, 312
 - domain, 320
 - image, 131, 155
 - image plane, 339
- Astigmatism, 187
- Astroid, 246
- Asymptotes, 26
- Axial
 - astigmatism, 243, 292
 - chromatic astigmatism, 291, 339
 - coma, 244, 323
 - deviation, 398
 - electron, 359
 - pseudo rays, 348
 - ray, 104, 112, 122
 - star aberration, 324, 368
- Azimuthal
 - chromatic distortion, 293
 - coma, 364

- distortion, 301
- Back-focal plane, 40, 108, 110
- Back nodal plane, 371
- Barrel distortion, 313
- Beam
 - chopper, 187
 - limiting aperture, 38
 - separator, 357, 410
- Beam-guiding systems, 1, 45, 135
- Beat, 21
- Bell parameter, 445
- Bending magnets, 94
- Bessel function, 37
- Bifurcation lines, 245
- Big Bang, 444
- Binomial coefficients, 57
- Binormal, 28, 49
- Biprism, 23
- BMT equation, 6, 472
- Boersch effect, x, 223
- Boundary
 - constraints, 216
 - surfaces, 71, 72
- Brightness, 33, 224
- Brightness function, 233
- Bundle of skew rays, 418
- Busch's theorem, 104, 106
- Canonical
 - form, 267
 - momentum vector, 226
 - representation, 256
 - variables, 199
- Canonical-momentum, 9, 190, 194, 208, 210, 216, 218, 266
- Canonical 4-momentum, 461
- Canonical momentum 4-vector, 450
- Cardinal elements, 112, 116, 128, 146, 172
- Cardinal planes, 114
- Catastrophe theory, 244
- Cathode, 110
- Causality, 481
- Caustic, 3, 31, 33, 205, 233, 235
 - curves, 235
 - equation, 238
 - of two sheets, 239
- CCD camera, 132, 305
- Center curvature, 395
- Center of curvature, 100
- Center of gravity, 117
- Central trajectory, 347, 387
- Chaotic behavior, 89
- Characteristics, 166, 223
 - function, 8
 - planes, 111
 - radiation time, 447
- Charged-particle beam, 223
- Charge elements, 84
- Charge-simulation
 - method, 84, 411
 - procedure, 73
- Chromatic
 - aberration, 342
 - aberration of magnification, 131
 - axial astigmatism, 348, 360
 - combination aberration, 354
 - correction, 335
 - defocus, 291, 339, 360
 - deviation, 272, 398
 - displacement, 351
 - distortion, 288, 291, 339, 354, 360
 - image rotation, 294
 - parameter, 99, 251, 262, 272, 295, 335, 397
 - round-lens distortion, 291, 314
- Chromaticity, 99, 136, 272, 291
- Chromaticity parameter, 91
- Circular accelerators, 94, 100
- Circulation, 27
- Coefficients
 - of axial chromatic aberration, 314
 - of the third-order aberrations, 377
- Coherent parasitic aberrations, 322
- Coma, 306
 - circle, 306
 - streak, 306
- Coma-free
 - aperture, 314
 - corrector, 332
 - magnetic lens, 384
 - matching, 378
 - plane, 308, 364, 383
 - point, 308, 364, 385
- Comb electrode, 177
- Combined trap, 427
- Comb lens, 178
- Complex
 - coordinates, 50
 - curvature, 48, 93
 - magnetic potential, 65
 - notation, 49, 94
 - potential, 64
 - slope, 113
 - variables, 47
- Compound lenses, 119
- Compton

- frequency, 449
- wavelength, 19
- Concave mirror, 400
- Conchoidal distortion, 325
- Condenser lens, 124
- Condenser-objective lens, 315
- Confocal hyperboloids, 73
- Conical sector magnets, 439
- Conjugate points, 8, 27, 205
- Conservation of energy, 7, 225, 391
- Continuation formula, 39
- Convergent section, 140
- Convex image field, 310
- Convex region of curvature, 409
- Correction
 - of coma, 380
 - conditions, 383
 - element, 344
- Coulomb
 - gauge, 64, 66, 68
 - interactions, 102, 179, 189
- Covariant Lagrangian, xi
- Cramer's rule, 162
- Cross-eye distortion, 325
- Crossover, 109, 418
- Crosstalk, 185
- Curl strength, 27
- Current
 - density, 31, 77, 224, 234
 - plane, 252
- Curvature, 27, 70
- Curved, 117
 - axis, 46, 64, 69, 98, 350
 - optic axis, 101
- Curvilinear coordinate system, 49, 271
- Cusp, 235
- Cylinder einzel lens, 178
- "Cylinder" lenses, 55
- Cylindrical coordinates, 72
- Deceleration lens, 148
- Decoupled correction procedure, 378
- Deflector, 187
- Defocus of least confusion, 306
- Degree, 252
- Delocalization, vii
- Deltroid, 244
- Dependent variables, 46
- Detection plane, 23, 132
- Diaphragm, 38
- Differential solid angle, 485
- Diffraction, vii
- Diffraction plane, 111, 112, 136
- Diode mirror, 408, 433
- Dipole
 - approximation, 485
 - field, 58
 - strength, 165
- Direction of flight, 387, 439
- Direct ray tracing, 281
- Disk of least confusion, 305
- Dispersion, 94, 99, 187
 - coefficient, 438
 - ray, 163, 349
- Dispersion-free monochromator, 350
- Distortion, 312
- Distribution function, 224, 226
- Divergent
 - lens, 97, 172
 - section, 140
- Dodecapole, 126, 165, 180
 - element, 269
 - Wien filter, 270
- Double fold, 245
- Double plane-section symmetry, 329
- Double symmetry, 357, 372
- Duality, 14
- Dynamical constraint, 455
- Dynamic transmission electron microscopy, 187
- Effective length, 127
- Effective source, 23, 109, 112, 166, 307, 415, 417
- Eigensolutions, 137
- Eigenvalue equation, 137
- Eikonal, 8, 151, 193, 203, 236, 242
 - approach, 392
 - coefficients, 276, 286
 - equation, 11, 16, 238
 - method, 208, 213
 - polynomials, 213, 258, 261, 269
- Einstein, 444
- Einstein relation, 18
- Einzel
 - cylinder lens, 148
 - lens, 75, 102, 108, 344
- Electric
 - dipole, 461
 - field strength, 132
 - potential, 7, 396
- Electric-field index, 102
- Electrode, 79
- Electromagnetic
 - field, 21
 - field tensor, 449

- Electron
 - gun, 413
 - holography, v, 38
 - lithography, ix, 74, 309
 - microscope, vii, 1, 108, 109
 - mirror, 388
- Electrostatic
 - corrector, 346
 - hexapole, 275
 - mirror, 360
- Elementary
 - catastrophes, 242
 - caustics, 242
 - wave, 32, 39
- Elliptical
 - chromatic distortion, 291, 339
 - distortion, 131
- Elliptic integral, 88
- Emergent asymptote, 115
- Emission
 - angle, 417
 - characteristics, 166, 223, 429
 - gun, 109
 - plane, 419
- Emittance, 224
- Emittance diagram, 232
- Energy
 - broadening, 188
 - deviation, 223
 - flow, 485
 - window, 282
- Energy-loss spectrum, 282, 349, 355
- Energy-selection plane, 282, 349, 432
- Entrance axis, 440
- Entrance plane, 174
- Euler-Lagrange equations, 10, 92, 254, 264, 392, 469
- Exchange symmetry, 371, 376
- Exit axis, 440
- Expansion parameter, 251
- Extended cylinder lens, 177
- Extraction field, 415

- Fermat's principle, 16, 62
- Field
 - aberrations, 325
 - astigmatism, 299, 310, 325, 376
 - curvature, 298, 376
 - of illumination, 110
 - index, 436
 - lens, 342
 - pseudo rays, 348
 - ray, 112
- Field-emission gun, 84
- Field of view, 433
- Fifth-order
 - aberrations of orthogonal systems, 328
 - aperture aberration, 329
 - coma, 369
 - combination aberrations, 328
 - path deviation, 261
 - spherical aberration, 383
- First-order Wien filter, 371
- Flux density, 67
- Focal
 - length, 90, 124, 133
 - plane, 147
 - point, 114
 - surfaces, 235
- Focusing power, 97
- Focusing strength, 116
- FODO cell, 139
- "Foundation stone" of geometrical electron optics, 119
- Four-dimensional
 - Green function, 479
 - Hamilton-Jacobi equation, 450, 460
 - Lagrange density, 477
 - Lagrangian, 446
 - mass density, 478
 - vector potential, 484
- Fourfold
 - astigmatism, 246
 - axial astigmatism, 318, 324, 346, 368, 380
 - umbilic, 246
- Fourth-order
 - perturbation eikonal, 316, 320, 367
 - variational polynomial, 316
- Fourth-rank
 - eikonal polynomial, 261
 - variational polynomial, 264
- Fraunhofer diffraction, 40, 43
- Freestanding quadrupole, 127
- Frenet-Serret trihedral, 27, 70
- Frequency, 18
- Fringe-field quadrupole, 101
- Fringe quadrupole, 358
- Fringing fields, 53, 101, 269
- Front nodal plane, 331
- Front principal plane, 371
- 4-f system, 180
- Fundamental
 - Poisson brackets, 221
 - pseudo rays, 294
 - rays, 112, 138, 164, 170
 - symplectic matrix, 200
 - trajectories, 104

- Gauge, 66
- Gauge invariant, 8, 18
- Gauss algorithm, 86
- Gaussian
 - approximation, 89, 91, 92, 107, 213
 - beam, 272
 - construction, 137
 - curvature, 238
 - distribution, 226
 - image, 41
 - image plane, 305, 308
 - image point, 111, 294, 308
 - optics, 100, 126, 195
 - path equations, 93, 271, 394
 - regime, 228
 - trajectory, 136
- Gauss theorem, 32, 34
- Generalized Helmholtz–Lagrange relations, 152
- Generalized Maupertuis principle, 465
- Generalized time, 390
- General relativity, 463
- Generating functions, 242
- Geometrical
 - aberrations, 212
 - electron optics, x
 - ray parameters, 251
- Glaser bell-shaped field, 120, 313
- Glaser lens, 314
- Glaser model field, v
- Gradient-index lenses, 2, 20, 89
- Gravitation, 461
- Gravitation potential, 450
- Green function, 34, 35
- Green theorem, 34
- Grid magnification, 158
- Group velocity, 21
- Gyromagnetic ratio, 450

- Half-space, 35
- Hamilton equations of classical mechanics, 191
- 4-Hamiltonian, 453
- Hamiltonian formalism, viii
- Hamilton–Jacobi equation, 10, 208
- Hamilton–Jacobi formalism, xi
- Hamilton’s principle, 8, 189, 208, 446, 451
- Harmonic component, 72
- Harmonic polynomial, 51
- Helmholtz–Lagrange, 153
 - invariants, 200
 - relations, 136, 137, 195, 258, 293
- Hemisphere, 34

- Hexapole
 - corrector, 180, 382
 - field, 381
 - function, 275
 - strength, 277, 359, 381
- Holography, 322
- Homocentric bundle of rays, 26, 155, 203
- Homogeneous Wien filter, 268
- Horizontal
 - astigmatic image, 368
 - emittance, 229
 - section, 224, 435
- Huyghens’s principle, 14
- Hyperbolic
 - fold, 243
 - mirror, 80
- Hyperboloids, 79
- Hyperemittance, 224
- Hyper-surface, 11, 455
- Hysteresis effects, 185

- Illumination aperture, 110
- Image
 - circle, 318
 - column, 439
 - curvature, 308
 - field, 308
 - focal plane, 114, 116, 128
 - formation, 108, 109
 - line, 158
 - plane, 104
 - principal plane, 114, 168, 175
 - principal ray, 122
 - section, 157
 - space, 26
 - tilt, 325
- Imaging energy filter, 282, 357
- Immersion
 - compound lens, 334
 - lens, 102, 117
- Incident asymptote, 115
- Independent variable, 46, 195, 402
- Index of refraction, ix, 2, 17, 107
- Information limit, ix, 322, 362
- Inhomogeneous Wien filter, 268, 474
- Initial plane, 194
- Inseparable systems, 151
- Instantaneous velocity, 483
- Instrumental resolution limit, 366
- Integral
 - equation, 146
 - representation, 319
- Interaction Hamiltonian, 443

- Interference
 - experiment, 20
 - pattern, 23, 24
- Intermediate
 - image, 342
 - magnification, 350
- Intrinsic angular velocity, 449
- Ion traps, 79, 425
- Isochromatic, 433
 - energy filtering, 289
 - imaging, 438
- Isoiduction lines, 100, 358
- Isoplanatic imaging, 41
- Isotropic
 - distortion, 312
 - index of refraction, 20
- Iteration
 - algorithm, 213
 - procedure, 86, 264, 401
- Jacobi determinant, 196, 227, 233
- Kinetic
 - energy, 8, 446
 - momentum, 6, 8, 193
- Klein–Gordon equation, 15, 30, 462
- Koehler illumination, 110
- Kronecker symbol, 57, 86, 256
- Laboratory system, 145, 447
- Lagrange
 - brackets, 152, 192, 222, 228
 - invariants, 106, 192, 197, 202
- Lagrange–Helmholtz relation, 107, 219, 299
- Lagrangian, 8, 189, 392
- Lamor rotation, 29
- Lande factor, 5, 450
- Laplace
 - equation, 50
 - operator, 16, 238
- Larmor rotation, 96, 123, 164
- Lateral
 - deviations, 401, 414
 - magnification, 111
- Laurent series, 483
- Legendre
 - differential equation, 76
 - functions, 76
 - polynomial, 76
 - transformation, 208
- Lie-algebraic expression, 260
- Lienard–Wiechert potentials, 482
- Line charge, 88
- Line focus, 125, 188
- Liouville’s theorem, 197, 226
- Longitudinal
 - deviations, 401
 - magnification, 111, 305
 - spherical aberration, 305
- Lorentz
 - force, 6, 31, 336
 - gauge, 16
 - microscopy, 385
 - scalar, 444
 - transformations, 459
- Low-energy electron microscope, 411
- Magnetic
 - cylinder lens, 69
 - dipole, 461
 - dipole strength, 358
 - flux, 22, 25
 - flux density, 24
 - lens, 121
 - model lens, 120
 - moment, 5, 447, 474, 477
 - saturation, 334
 - sheet, 77
 - vector potential, 11, 24, 25, 62, 69
- Magnification, 105, 107, 168
- MANDOLINE filter, 432, 435
- Mass spectroscopy, 425
- Mathieu equations, 426
- Mathieu’s differential equation, 79
- Matrix
 - method, 169, 213
 - representation, 202
- Maxwell equation, 47, 350
- Mean energy, 223
- Mechanical inaccuracies, 356
- Medium of refraction, 89
- Meissner effect, 24
- Meridional
 - line focus, 310
 - paraboloid, 311
- Method
 - of successive iteration, 147
 - of variation of coefficients, 161
 - of variation of parameters, 397
- Metric coefficients, 50
- Mid-plane symmetry, 357
- Midsection, 165
- Midsection symmetry, 98, 164, 267, 285
- Minkowski space, 9, 443

- Mirror, 56
 - approach, 35
 - charges, 72
 - corrector, 357, 411
 - electrode, 400
 - symmetry, 56, 68
- Mirror-prism filter, 433
- Misalignment, 323
- Mixed eikonal, 209
- Modified
 - eikonal polynomials, 257
 - emittance, 230
 - fourth-order eikonal, 295
 - path equation, 144
 - perturbation eikonal, 257
- Moellenstedt experiment, 22
- Momentary center of curvature, 28
- Momentum deviations, 258
- Monochromator, 165, 322, 429
- Monomials, 53, 60, 275, 323, 348, 405
- Moving objective lens, 176
- Multi-element systems, 359
- Multiplicity, 51, 53, 264, 276, 279, 317, 323, 329, 348
- Multipoles, 50, 267
 - component, 54
 - correctors, 242
 - element, 329
 - expansion, 45, 56
 - field, 55, 73
 - potential, 71, 72
 - representation, 66
 - strengths, 51, 56, 64, 71
- Nabla operator, 11
- Nanostructured devices, 432
- Neumann iteration procedure, 251
- Neumann series, 251
- “Newtonian” fields, 123
- Newton lens equation, 114
- Nodal planes, 115, 176, 277
- Nodal ray, 115, 184
- Nominal
 - energy, 161, 262, 350
 - velocity, 270
- Non-orthogonal systems, 157
- Non-polarized beam, 474
- Normalization potential, 265
- Normalized eikonal, 276
- Normal unit vector, 27
- column, 439
- focal length, 108
- focal plane, 128
- line, 158
- nodal point, 365
- plane, 104
- principal plane, 114, 168, 175
- principal ray, 121
- space, 26
- Objective
 - focal length, 279
 - lens, 40
- Oblate spheroidal coordinates, 73
- Octopole, 165, 346, 369
- Octopole strength, 319
- Odd chromatic distortion, 291
- Odd multiplicity, 296
- Off-axial (off-axis) coma, 199, 211, 242, 299, 376
- OMEGA filter, 433
- One-dimensional lattice, 350
- Optical
 - continuation, 36
 - medium, 242
 - path length, 62, 214, 236, 457
 - potential, 192
- Optic axis, 2, 46, 52, 58
- Orange spectrometer, 207
- Orthogonal
 - principal sections, 317
 - sections, 151
 - systems, 295
 - trajectories, 14
- Orthomorphic, 184
- Paraboloid of revolution, 309
- Parametric
 - representation, 236
 - representation of the caustic, 240
- Parasitic
 - aberrations, 322, 356
 - quadrupole fields, 359
- Paraxial
 - approximation, 90
 - emittance, 230
 - trajectory, 192
- Particle wave, 20
- Path
 - deviation, 252, 266
 - and momentum deviations, 220
- Paul trap, 425
- Pencil of rays, 38, 234, 405
- Penning trap, 425, 428
- Object

- Perihelion rotation, 463
- Periodicity length, 136
- Perturbation
 - eikonal, 214, 255
 - function, 264
- Petzval curvature, 302, 378
- Phase, 18
 - ellipse, 229
 - space, 195, 224, 227, 231
- Phase-contrast, 38
- Phase-space diagrams, 231
- Photo-emission electron microscope, 103, 411
- Pincushion distortion, 313
- Planar fields, 50
- Planar system, 376
- Planator, 377
- Planes, 2
 - double cusp, 246
 - of least confusion, 306
 - of observation, 193
- Poincaré invariant, 202
- Point eikonal, 207, 233
- Point-like particle, 446
- Point of observation, 38, 237
- Point spread function, 43
- Poisson bracket, 221, 259, 328
- Polarity, 130
- Pole, 480
- Pole pieces, 71, 120
- Polynomial, 348
- Positron, 447
- Post-column filter, 435
- Post-lens, 124
- Potential barrier, 414
- Potential energy, 7
- Power series expansion, 391
- Poynting vector, 485
- Precession, 6, 474
- Pre-lens, 124
- Primary
 - aberrations, 113
 - deviation, 266
 - effect, 275
 - field aberrations, 376
 - geometrical aberrations, 297
 - image, 111, 180
 - third-order aberrations, 295
- Principal
 - planes, 90, 147, 172
 - point, 114
 - rays, 104, 112, 121, 146
 - sections, 52, 125, 128, 264, 353
- Principle
 - of least action, 11
 - of Maupertuis, 11, 190
 - ray, 169
- Prism, 145
- Product ansatz, 82
- Projector
 - lens, 132
 - system, 176
- Propagator, 30, 33
- Proper time, 455
- Pseudo-Euclidian, 9
- Pseudo-Euclidian space, 443
- Pseudo rays, 159
- Pulsed beam, 335
- QO corrector, 367, 372
 - Quadrupole, 58, 82, 125
 - anastigmat, 131, 166
 - doublet, 130
 - quadruplet, 130, 166, 337
 - quintuplet, 167
 - septuplet, 184
 - stigmator, 292
 - strength, 128
 - system, 134, 138
 - Wien filter, 354
 - Quadrupole-octopole correctors, 362
 - Quantum mechanical tunneling, 414
 - Quasi-monochromatic electron source, 430
- Radial chromatic distortion, 293
- Radial coma, 306
- Radiation
 - characteristics, 233
 - damage, vii
 - energy, 446, 451
 - field, 486
- Radius of curvature, 439
- Rank, 99, 251, 252, 258
- Ray-defining parameters, 216
- Ray gradient, 388
- Rayleigh criterion, 43
- Rayleigh–Sommerfeld diffraction integral, 35
- Ray parameters, 212, 265
- Reciprocity
 - relation, 33
 - theorem, 35
- Recording plane, 245
- Rectilinear system, 66
- Recurrence
 - formula, 59, 60, 253, 258
 - relations, 402
- Reduced

- action, 11, 207, 456, 462
- axial brightness, 226
- brightness, 224, 226, 350
- eikonal, 468
- hamiltonian, 190
- lagrange function, 192
- lagrangian, 190, 465
- Reference
 - electron, 388
 - trajectory, 192
- Reflection, 388
- Refraction
 - matrices, 169
 - plane, 170
- Refractive power, 177
- Regular chromatic distortion, 291, 293
- Regular multipoles, 51
- Relativistic
 - factor, 468
 - limit, 135
 - Maupertuis principle, 463
 - modified axial electric potential, 396
 - modified axial potential, 91, 97
 - modified chromatic parameter, 100, 271, 272
- Repetitive symmetry, 350, 353
- Resolution limit, *v*, 334, 380
- Rest energy, 455
- Retarded gravitational potential, 463
- Roentgen radiation, 482
- Ronchigrams, 235
- Rosette, 318
- Rosette aberration, 331
- Rotating coordinate system, 95
- Rotational
 - cusp, 246
 - symmetry, 67, 74
- Rotation angle, 95
- Rotation-free imaging, 341
- Round lens, 75
- Runge–Kutta method, 400
- Russian quadruplet, 131, 166
- Saddle point, 78
- Sagittal line focus, 310
- Sagittal paraboloid concave, 311
- SALVE corrector, 375
- Scalar magnetic potential, 50, 63, 77, 359
- Scaling length, 171
- Scanning electron microscope (SEM), 368
- Scherzer
 - condition, 343
 - theorem, *ix*, 19, 199, 241, 263, 292, 303, 335, 388
- Schottky field emitters, 414, 430
- SCOFF approximation, 430, 436
- Secondary
 - dispersion ray, 355
 - fundamental rays, 279, 321
- Second-degree dispersion, 272, 288, 352, 354
- Second-order
 - axial aberrations, 288
 - distortions, 288
 - field astigmatism, 288
 - path deviations, 376
- Second-rank
 - achromats, 347, 353, 358
 - chromatic deviation, 264
 - deviation, 286, 355, 404
 - longitudinal deviation, 420
 - path deviation, 260, 267, 270, 274
- Second-zone mode, 124
- Sections, 2
- Seidel-order, 212, 213, 252, 259, 281, 295
- Self action, *v*
- Self action of the electron, 477
- Self-luminous object, 43
- SEM corrector, 369
- Semi-aplanat, 365
- Semi-classical propagator, 38
- Separation
 - condition, 149
 - distance, 118
- SESAM Microscope, 432
- Sextupole, 272
 - corrector, 362, 363
 - stigmator, 323
- Siberian Snakes, 161
- Sine condition, 211
- Single-field condenser-objective lens, 124
- Six-fold axial astigmatism, 331
- Sixth-order eikonal polynomial, 261
- Skew
 - axis, 58
 - multipoles, 51
 - octopoles, 374
- Skewness, 203
- Slit
 - apertures, 177
 - lenses, 50
- Slope angle, 198
- SMART microscope, 282, 411
- Solenoid, 2, 103
- Sommerfeld diffraction formula, 34
- Space charge, 47
- Space-like rotation, 445

- Spectral brightness, 429
- Speed of light, 5
- Spherical
 - aberration, vii, 185, 298, 367
 - cathode tip, 415
 - wave, 30, 32, 234
- Spin, 5, 15, 447
 - four-vector, 472
 - precession, xi, 469
 - tensor, 448, 482
- Spiral distortion, 312
- S-state, 474
- Stability
 - domain, 140
 - point, 142
 - region, 428
- Star aberration, 317, 331
- State
 - space, 195, 224, 228
- State of alignment, 316
- Static
 - lens defects, 322
- Stationary field, 33
- Stern–Gerlach effect, v
- Stern–Gerlach experiment, 474
- Stigmatic imaging, 41
- Stigmaticity, 177
- Stigmatic region, 279
- Stigmator, 102
- Stoke’s theorem, 22, 26, 203
- Straightened optic axis, 352
- Straight optic axis, 102
- Straight-vision
 - element, 350
 - prism, 268, 343
- Strong focusing, 79, 90, 125, 132
- Successive insertion, 54
- Symmetry requirement, 381
- Symplectic mapping, 196, 199, 222
- Systems with midsection symmetry, 271
- Tangential plane, 27, 417
- Tangential unit vector, 27
- Taylor series expansion, 56
- TEAM corrector, 331
- Telescopic
 - anastigmat, 181
 - corrector, 330
 - doublet, 373
 - filter, 439
 - limit, 115
 - magnification, 377
 - mode, 175, 181
 - octuplet, 370
 - quadrupole quadruplet, 279
 - quadrupole quintuplet, 182
 - round-lens doublet, 277
 - systems, 179
- Terminal planes, 209
- Tetrode mirror, 410, 411
- Theorem of alternating images, 104, 110
- Thermal electron emission, 415
- Thick lens, 113
- “Thin” lens, 98, 113
- Thin-lens approximation, 113, 169
- Third-degree dispersion, 356
- Third-order
 - aberration, 314
 - coma, 383
 - combination aberrations, 357
 - distortion, 299
 - focusing, 315, 324
 - perturbation eikonal, 324
- Third-rank
 - eikonal, 270, 273
 - eikonal polynomial, 274, 347
 - lateral deviation, 406
 - path deviations, 260, 263
 - perturbation eikonal, 284
 - perturbation function, 421
 - variational polynomials, 271
- Thomas precession, 464, 474
- Threefold
 - astigmatism, 243
 - axial astigmatism, 324
 - symmetric system, 320
 - umbilic, 243
- Time-dependent perturbation method, 405
- Time-dependent perturbations, 362
- Time-like angular rotation, 445
- Time of flight, 12
- Toroidal deflection electrodes, 431
- Toroidal electrodes, 102, 426
- Torsion, 27, 48, 70
- Torsion angle, 70
- Trajectory, 46, 102
 - density, 33, 233
 - displacement, 148
 - method, viii
- Transfer
 - doublet, 373
 - matrices, 170
- Transition time, 45
- Translational energy, 446
- Transmissivity, 232, 435
- Transverse emittance, 224
- Trihedral, 27, 49

Trilobe distortion, 325
Turning point, 387, 409
Turning time, 397
Twiss parameters, 229
Twist angle, 48, 96, 104
Twisted quadrupole, 150
Two-sheet hyperboloid, 79, 426

Ultracorrector, 185, 331
Unipotential lens, 75
Universal time, 443, 477

Valence electron, 474
Variable-axis anastigmat, 177
Variational
 function, 90, 91, 208, 262, 264
 polynomials, 254
 principle, 465
Vector potential, 32
Velocity of light, 135
Velocity 4-vector, 448
Vertical astigmatic image, 368
Vertical section, 227, 435
Vignetting, 307

Virtual
 image, 104
 source, 421

Wave
 surfaces, 17, 240
 vector, 19

Wavelength, 19

Weak lens, 133

Wehnelt
 electrode, 413
 potential, 413
 voltage, 187

W-filter, 433

Wien
 condition, 93, 337, 474
 filter, 94, 165, 263, 268

WKB-approximation, 30

WKP-approximation, 16

Wronskian, 106, 395, 416

Wronski determinant, 106, 153, 162

Zero-order Wien filter, 336

Zonal magnification, 198