This second English edition of Grivet's book provides a full account of the principles of electron optics, and of many of their applications. The underlying theory is set out in physical terms rather than abstruse mathematics, and no previous knowledge of the subject is required. Much new material has been added since the first edition, including a completely new chapter entitled *Prism Optics*.

The new edition has been improved for the benefit of two categories of reader. Firstly, the student who will profit from the introductory sections describing selected but nevertheless classical methods of calculating electric and magnetic fields. These are intended as a guide for the beginner who often feels lost in the abundant specialized literature. Secondly, the text keeps the research worker informed of the important recent progress in this field; an essential prerequisite for a book which aims at helping scientists to form their scientific taste and to choose their own path in research.
<table>
<thead>
<tr>
<th>Date</th>
<th>Date</th>
<th>Date</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 Jun 1982</td>
<td>1 April 1984</td>
<td>22 May 1993</td>
<td>3 Dec 1993</td>
</tr>
</tbody>
</table>

All Books must be Returned to the College Library or Renewed not later than the last date shown below.
P. GRIVET
Professor at the University of Paris

With the Collaboration of
M.-Y. BERNARD, F. BERGEIN, R. CASTAING,
M. GAUZIT AND A. SEPTIER

ELECTRON OPTICS
SECOND ENGLISH EDITION

Translated by P. W. HAWKES
Revised afresh for this Edition by A. SEPTIER

Part 1 OPTICS
Part 2 INSTRUMENTS

PERGAMON PRESS
OXFORD · NEW YORK
TORONTO · SYDNEY · BRAUNSCHWEIG
This book is a translation, with much revision and addition, of Optique Électronique, originally published in French by Bordas Éditeur, Paris.
CONTENTS

Preface to the Second English Edition xi
Preface to the First English Edition xiii

PART 1. OPTICS

1. Fundamental Features and Basic Techniques
   1.1 The Historical Background 3
   1.2 Outline of a Demonstration in Electron Optics 4
   1.3 The Object 5
   1.4 The Fluorescent Screen 11
   1.5 Photography 14
   1.6 The Optical Bench 15

2. The Distribution of Potential and Field in Electrostatic Lenses
   2.1 Solution of the Laplace Equation 20
   2.2 The Electrolytic Tank 45
   2.3 The Resistance Network 54

3. The Field Distribution in Magnetic Lenses
   3.1 Direct Calculation of the Magnetic Field 58
   3.2 Direct Methods of Measuring the Induction B and Its Gradient 72
   3.3 The Use of Electrical Analogues 77

4. The Optical Properties of Electrostatic Lenses
   4.1 Image Formation 85
   4.2 The Gaussian Approximation 94
   4.3 Optical Formalism 101
   4.4 Relativistic Lenses 110
   4.5 Practical Determination of Electron Trajectories 111
   4.6 Experimental Determination of the Standard Cardinal Elements 119

5. The Optical Properties of Magnetic Lenses
   5.1 The Form of the Trajectories in an Axially Symmetric Field 123
   5.2 The Gaussian Approximation 126
   5.3 Properties of Magnetic Lenses 130
   5.4 Relativistic Lenses 132
   5.5 Methods of Determining the Fundamental Rays 133
   5.6 Calculation of the Trajectories and Optical Elements 135

6. Similarities and Differences Between Glass Optics and Electron Optics 139
   6.1 Electrostatic Lenses 139
   6.2 Magnetic Lenses 144
   6.3 The Refractive Index in Interference Phenomena and the "Wave" Mechanics of the Electron 149
7. Lens Defects
   7.1 The Limits of the Gaussian Approximation 155
   7.2 Third Order Aberrations 156
   7.3 Calculation of the Aberration Coefficients 164
   7.4 Chromatic Aberration 176
   7.5 Mechanical Defects 182
   7.6 Measurement of the Aberration Coefficients, and Their Orders of
      Magnitude 185
   7.7 Correction of the Aberrations 198

8. Some Electrostatic Lenses and Triode Guns 207
   8.1 Two Coaxial Cylinders 207
   8.2 Lenses Formed from Three Diaphragms 211
   8.3 “Mesh” or “Grid” Lenses 224
   8.4 The Immersion Objective 228
   8.5 The Triode Electron Gun 240

9. Magnetic Lens Models 259
   9.1 The Long Solenoid 259
   9.2 Models Representing Short Systems 260
   9.3 The Unshielded Short Coil 273
   9.4 Shielded Coils 276
   9.5 Superconducting Coils 283

10. Strong Focusing Lenses 287
    10.1 The Strong Focusing Principle 287
    10.2 The Field Distribution 291
    10.3 The First Order Optical Properties 296
    10.4 Combinations of Lenses 301
    10.5 A Chain of \( N \) Identical Crossed Quadrupole Lenses 307
    10.6 Applications of Quadrupole Lenses 312
    10.7 The Aberrations of Quadrupole Lenses 313

10A. Prism Optics 323
    10A.1 Magnetic Prisms 324
    10A.2 Electrostatic Prisms 354

**Table 1. Some Useful Constants** xv
**Table 2. Some Useful Formulae** xvi
**Table 3. Fundamental Characteristics of Free Electrons in Motion** xvii
**Table 4. The Legendre Functions of Fractional Order** xx

**PART 2. INSTRUMENTS**

11. Electron Optical Instruments 375
    11.1 Light Radiation and Corpuscular Radiation; the Simple Character
        of Photons 375
    11.2 Mass and Velocity Spectrographs; Diffraction Devices 375
    11.3 Cathode Ray Tubes; Oscillographs 376
    11.4 Microscopes 377
    11.5 Telescopes and Image-converters 378
<table>
<thead>
<tr>
<th>SECTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. THE CATHODE RAY TUBE</td>
<td></td>
</tr>
<tr>
<td>12.1 Historical Introduction; the Importance of the Cathode Ray Tube</td>
<td>380</td>
</tr>
<tr>
<td>12.2 Formation of the Spot</td>
<td>383</td>
</tr>
<tr>
<td>12.3 Deflexion of the Beam</td>
<td>407</td>
</tr>
<tr>
<td>12.4 The Post-acceleration</td>
<td>423</td>
</tr>
<tr>
<td>13. THE EMISSION MICROSCOPE</td>
<td></td>
</tr>
<tr>
<td>13.1 Introduction</td>
<td>427</td>
</tr>
<tr>
<td>13.2 Optical Properties</td>
<td>427</td>
</tr>
<tr>
<td>13.3 The Various Types of Microscope</td>
<td>434</td>
</tr>
<tr>
<td>13.4 Some Actual Instruments</td>
<td>452</td>
</tr>
<tr>
<td>13.5 The Filament Microscope and the Spherical (or Point-projection)</td>
<td>460</td>
</tr>
<tr>
<td>Microscope</td>
<td>463</td>
</tr>
<tr>
<td>13.6 The Point-projection Microscope</td>
<td></td>
</tr>
<tr>
<td>14. THE IMAGE-CONVERTER AND THE IMAGE-INTENSIFIER</td>
<td></td>
</tr>
<tr>
<td>14.1 The Basic Principle</td>
<td>474</td>
</tr>
<tr>
<td>14.2 The Main Types of Instrument</td>
<td>474</td>
</tr>
<tr>
<td>14.3 Principal Characteristics</td>
<td>483</td>
</tr>
<tr>
<td>14.4 The Development and the Future of These Instruments</td>
<td>489</td>
</tr>
<tr>
<td>14.5 Amplification of the Brilliance of Radioscopic Images</td>
<td>496</td>
</tr>
<tr>
<td>15. THE ELECTRON MICROSCOPE; STRUCTURE AND SPECIAL FEATURES</td>
<td></td>
</tr>
<tr>
<td>15.1 The Notions of Resolving Power and Useful Magnification</td>
<td>498</td>
</tr>
<tr>
<td>15.2 A Comparison of the Electron and Optical Microscopes</td>
<td>500</td>
</tr>
<tr>
<td>15.3 The General Properties of the Instrument</td>
<td>503</td>
</tr>
<tr>
<td>15.4 Examining and Recording the Image</td>
<td>509</td>
</tr>
<tr>
<td>15.5 The Importance of the Projective Lens</td>
<td>517</td>
</tr>
<tr>
<td>15.6 Measurement of the Characteristics of a Microscope</td>
<td>522</td>
</tr>
<tr>
<td>16. THE FACTORS WHICH GOVERN THE RESOLVING POWER IN ELECTRON MICROSCOPY</td>
<td></td>
</tr>
<tr>
<td>16.1 The Final Resolving Power; Diffraction Alone</td>
<td>529</td>
</tr>
<tr>
<td>16.2 The Optimum Resolving Power; Diffraction and Spherical Aberration</td>
<td>531</td>
</tr>
<tr>
<td>16.3 The Resolving Power in Practice; the Influence of the Chromatic Aberration</td>
<td>538</td>
</tr>
<tr>
<td>16.4 The Practical Resolving Power; the Influence of Defects in the Mechanical Construction</td>
<td>544</td>
</tr>
<tr>
<td>16.5 “Revolutionary” Attempts to Improve the Resolving Power</td>
<td>557</td>
</tr>
<tr>
<td>16.6 New Tendencies in Electron Microscopy</td>
<td>560</td>
</tr>
<tr>
<td>17. THE OPTICAL ELEMENTS OF A MICROSCOPE</td>
<td></td>
</tr>
<tr>
<td>17.1 The Object Illumination</td>
<td>563</td>
</tr>
<tr>
<td>17.2 General Characteristics of Microscope Lenses</td>
<td>573</td>
</tr>
<tr>
<td>17.3 Electrostatic Lenses</td>
<td>573</td>
</tr>
<tr>
<td>17.4 Magnetic Lenses</td>
<td>579</td>
</tr>
<tr>
<td>18. THE ANCILLARY PARTS OF AN ELECTRON MICROSCOPE</td>
<td></td>
</tr>
<tr>
<td>18.1 The Electrical Supply</td>
<td>602</td>
</tr>
<tr>
<td>18.2 High Tension Effects</td>
<td>610</td>
</tr>
<tr>
<td>18.3 The Vacuum</td>
<td>619</td>
</tr>
</tbody>
</table>
19. The Action of the Object in a Microscope on the Electron Beam; Contrast in the Image; Velocity Analysers
   19.1 The Two Effects to which Electrons May Be Subjected on Passing through a Thin Film
   19.2 Contrast Formation in an Amorphous Object
   19.3 Velocity Analysers
   19.4 Velocity Filters

20. Specimen Preparation
   20.1 The Problem of Preparation
   20.2 Supporting the Object
   20.3 Direct Observation of the Object
   20.4 Indirect Observation of the Object
   20.5 The Increase in the Temperature of the Object, and the Disturbance of the Latter by the Impact of the Electron Beam

21. The Diffraction Camera
   21.1 Electron Diffraction
   21.2 The Properties of a Diffraction Camera
   21.3 Some Practical Examples
   21.4 Some Applications

22. Electron Probe Devices
   22.1 Electron Probes and Micro-probes
   22.2 Micro-probe Analysers
   22.3 The Scanning Electron Microscope
   22.4 The X-ray Projection Microscope
   22.5 Machining by Electron Bombardment

23. The Electron (Ion) Prism
   23.1 The History of the Mass Spectroscope
   23.2 The Principle of Mass Spectroscopy; the Spectrograph and the Spectrometer
   23.3 Elementary Examples of Analysing Fields: Uniform Fields
   23.4 Ion Optics in Fields with a Symmetry Plane
   23.5 Sector Fields; Angular Focusing; Their Application in Spectrometers
   23.6 Sector Fields; Mass and Energy Dispersion; Double Focusing; Spectrographs
   23.7 The Various Perturbing Effects
   23.8 The Resolving Power of Spectrometers and Spectrographs

24. The Individual Parts of a Mass Spectroscope
   24.1 The Ion Source
   24.2 Collection of the Ions
   24.3 The Electrical Circuitry of a Spectroscope
   24.4 The Vacuum and the Gas Circulation
   24.5 The Various Types of Measurement in Mass Spectroscopy
   24.6 An Overall View of Present-day Instruments, and the Fields in which They Are Employed
   24.7 Mass Spectroscopes with a High Frequency Field
CONTENTS

25. β-ray Spectrographs with Electron Lenses

25.1 Introduction
25.2 The General Principle
25.3 Energy Selection by an Aperture
25.4 Clarity, the Size of the Instrument, and the Radius of the Source
25.5 The Annular Selective Diaphragm (Witcher)
25.6 Modern Trends

Table 1. Some Useful Constants
Table 2. Some Useful Formulae
Table 3. Fundamental Characteristics of Free Electrons in Motion
Table 4. The Legendre Functions of Fractional Order

BIBLIOGRAPHY

INDEX
PREFACE TO THE SECOND ENGLISH EDITION

We have tried to improve this new edition of Electron Optics for the benefit of two categories of reader. First of all, we have endeavored to make it more useful to the graduating student. With this in mind we have, for example, added some introductory sections describing selected but nevertheless classical methods of calculating electric and magnetic fields; these indications are intended as a guide for the beginner who sometimes feels lost in the abundant specialized literature.

On the other hand, we thought it essential to keep the research worker (even the beginner) informed of the important recent progress in the field: this seems to us an essential prerequisite for a book which aims at helping scientists to form their scientific taste and choose their own path in research.

This proved to be a fascinating but difficult task because activity in the field of electron optics has been both important and successful during recent years. The familiar magnetic objective has been significantly perfected, mainly by its discoverer, Professor Ruska, while the practical introduction of superconducting lenses gives some realistic hopes of a breakthrough in the domains of high resolution and high energy: today, several careful measurements prove that a resolving power of 4 Å is already achieved and that one can distinguish lines spaced 1 or 2 Å apart: the traces of Bragg planes are clearly distinguishable and these results give good reason to hope that in a not too distant future the individual atoms of most solids will be made visible by high-voltage electron optics.

On the theoretical side, the phenomenon of interference between two electron beams has gained in interest, as it may be usefully compared to the superconducting interferometer, a more mysterious but also more practical instrument making use of the stability of Cooper electron pairs and producing beats in the time domain instead of fringes in space. For this reason, greater care has been taken to bridge the gap between refractive index theory and the theory of the phase shift for electron waves.

The new appearance of this second edition is again mainly due to the efforts of Dr. Septier and the new text was also very ably translated into English by the English specialist of electron optics, Dr. Hawkes. I am glad to thank them for their renewed efforts on behalf of this book, in the name of all our team. The success of the first edition induces me
to express our gratitude to the founder of this team during the hard times of the forties, Academician Dr. Ponte. We hope that the qualities of this book reflect in no negligible part the precious lessons of his two teachers during the twenties, Professor L. de Broglie and Sir John (J. J.) Thomson: their creative enthusiasm still had a beneficial impact on us 40 years later.

Pierre Grivet
PREFACE TO THE FIRST ENGLISH EDITION

With the appearance of this new edition in English it is a pleasure to look back at the original French version of the book; the latter is not so very old, and yet how important have been the advances which have been made since its publication, in 1955 and 1958. The most fundamental sphere of application of electron optics, electron lenses, seemed at that time to have been fully explored, but nevertheless, the subject has a fresh appeal to young research workers. The methods of calculation have developed beyond recognition, for nowadays no-one hesitates to invoke the aid of an electronic computer, so that the atmosphere of slightly timorous respect in which the differential equations to be solved used to be enveloped has practically disappeared. The importance of these electronic computers has been shown by the major discoveries which have resulted from their use, the behaviour of the quadrupole lens, for example. The need to consider this new type of lens was imposed by the requirements of nuclear physicists; none of the ordinary methods was suitable for studying these lenses, but even so, their properties and their surprising qualities are now well known after less than five years' study—ordinary lenses, with their much simpler structure, had to wait almost two decades before being thoroughly understood.

Other pleasant surprises await us in the instrumental field. To begin with, the electron microscope is the most perfected instrument, having bettered the performance which the theoreticians were reasonably hoping for around 1949. In their caution, these latter were doubtless too severe in their estimates of the limiting resolving power which could be attained, for all over the world today, industrial instruments of various nationalities—English, German, Japanese, French and Russian—are in current use and reach and even slightly exceed the most optimistic of the estimates of the theoreticians of ten years ago.

Still more astonishing are the results which have been obtained with this microscope. Ten years ago, it was still thought that the electron microscope was capable of giving only a rather frozen view of reality, unnaturally stiff like the view which we are given of the life of prehistoric organisms by paleontology. But at Toulouse, Professor Dupouy has already begun to study living biological cells with the aid of very high voltages; and even in more ordinary laboratories, we have no longer to be satisfied with an examination of a replica or desiccated skeleton of reality. Proper sections can now be cut which show the real structure of the dead cell. The success is still more striking where solids are concerned; with an electron microscope
we can fasten on to the very life of a metal, by filming, for example, the
evolution of the dislocations when the metal is heated; this can be
performed directly on a very thin layer cut from a block of the metal
being studied.

Metallurgy provides us with yet another example of a great success. The
metallurgists have now whole-heartedly adopted the electron probe; with
its aid, they are able to perform extremely detailed and accurate metallurgical
analyses.

This analyser was invented in France in 1949, and since 1954 Castaing
has made a highly perfected form available to the metallurgists. It has taken
six years for metallurgists all over the world to become convinced of
the excellence of this method, and 1960 marks the beginnings of world-
wide employment of this micro-analyser. This example demonstrates how
useful it always is to struggle for the propagation of scientific ideas and
methods, even when they have received the most brilliant experimental
confirmation in the laboratory.

In these present times it is always necessary to make a very great effort
towards the diffusion of scientific thought and the renewal of experimental
methods and to disseminate them among a wide public, as, in the last
resort, it is this public which reaps all the benefit. It is for this reason that
the senior author of the French edition feels justified in his delight at seeing
that this book is going to present the ideas of the French school of electron
optics to an English-speaking public, to a public whose language is that
of the country in which the "electron" was born.

It is a pleasure to thank the many colleagues, both at home and abroad,
who have graciously contributed to this work by helping us to clarify some
difficult points, or by providing us with illustrations taken from their own
personal work. From among them, we should like to single out Professors
Castaing, Dupouy, Fert, Lallement, Möllenstedt, Müller, Noziéres and
Pernoux, and Doctors Duchesne, Haine, Mulvey, Stohr and Wlerick.

We are also grateful to various industrial firms who furnished us with
certain documents, and in particular to C.S.F. and O.P.L. (Paris) and
Zeiss (Oberkochen).

This book, in its present form, owes much to the painstaking work of
Dr. Septier, who brought it up to date. Dr. Hawkes has been more than
a translator, and his profound knowledge of Electron Optics is present, if
not apparent, in many parts of this book. Last, but not least, I should like
to express my personal gratitude, and that of most of my colleagues in this
enterprise, to Professor Marton; without his continual encouragement
during the past fifteen years, much of the research and teaching included
within these pages would perhaps never have been performed.

Pierre Grivet
PART 1

OPTICS
CHAPTER 1

FUNDAMENTAL FEATURES AND BASIC TECHNIQUES

1.1 THE HISTORICAL BACKGROUND

Towards the end of the nineteenth century, physicists directed a large proportion of their efforts towards elucidating the properties of cathode rays which, by virtue of their rectilinear propagation, exhibited a close affinity with light rays. An example is furnished by the celebrated experiment in which Crookes introduced a metal cross into the path of the rays and threw its shadow onto a fluorescent screen. That there was, however, a fundamental difference between light rays and these cathode rays composed of electrons in motion, was demonstrated by Pluecker and Goldstein who showed that the latter are deflected by magnets and by charged bodies.

Later, with the revolutionary new ideas which followed the discovery of the wave nature of the electron by L. de Broglie, it became apparent that this deviation could be regarded as an optical property; further, the exploration of the vast new sphere of phenomena, now classed generically as “electron diffraction”, could begin.

The creation of this “physical optics” of electrons led to a more profound examination of particle mechanics, and the analogy between the dynamics of a material point and geometrical optics, which had been formulated in theory by Hamilton in 1827, soon provided the foundations for the new subject of “electron optics”. This “corpuscular” optics—valid for charged particles other than electrons: ions for example—consists in studying the laws which govern the behaviour of electrons from their source, through the “object”, to the image which they trace out upon either a fluorescent screen or a photographic plate.

After the first lenses had been built, by H. Busch (1927) and by Davisson and Calbick (1931), the new optics developed rapidly, considerably aided as it was by the important advances in “radio valve” techniques which were taking place.

In fact, many of the difficulties were becoming less important: in particular, to obtain a vacuum within a capacious envelope containing large metal surfaces had become a straightforward procedure and fluorescent powders were now standard industrial products. Technology, too, provided fascinating scientific problems to be solved, and from the beginning, electron
optics has been used successfully both to unravel the mysteries of oxide-coated cathodes and secondary emission, and to satisfy the requirements of television engineers.

To-day, electron optics is a branch of science which has achieved maturity, a branch represented by a rich diversity of instruments—cathode ray tubes, electron microscopes and electronic image converters for example. Throughout the whole of this vast field of applied electron optics, however, it remains true that there are two alternative ways of controlling a beam of cathode rays. The beam passes through the aperture either of an "electrostatic lens", in which the rays are deviated by the electric field produced by a system of annular electrodes, or of a "magnetic lens", where annular magnets or current-carrying coils surround the beam.

Two quite different methods of exploring the theory of electrons in electric and magnetic fields are possible.

If, from the beginning, we examine the analogy between the trajectory of an electron and the track of a light ray, it is possible to define a refractive index, a function of position $n(r, z)$, in such a way that the path of the particle is identical with that which a light ray, obeying the usual rules of refraction, would trace out in a medium with refractive index $n$. Once $n$ has been established, it becomes a simple matter to tabulate the properties of electron lenses by a direct transposition from those of their optical models; this is not, however, physically particularly meaningful in the case of the magnetic lens. In his book *Optique Electronique et Corpusculaire* (1950), de Broglie has displayed both the power of this method, and the symmetry which can be established between optics and electromagnetic theory.

Alternatively, the elementary classical formalism of the laws of mechanics and electromagnetism can be used, but this involves rejection of the powerful range of optical methods. However, although the investigation is made less concise, it does relate the optical properties more directly to the concrete features of the electrical experiment. Even if we deliberately restrict the extent of our vision, we do nevertheless avoid the abstract mathematical reasoning which is necessary to relate elementary geometrical optics to Hamilton's theory. For this reason, we shall—to begin with—employ the trajectory method.

1.2 Outline of a Demonstration in Electron Optics

The formation of an image by an electron lens can be demonstrated with the apparatus shown schematically in Fig. 1. An electron gun, which comprises an oxide-coated cathode and an accelerating grid $C_1$ held at a poten-
ential $\Phi_1$, is placed within an evacuated envelope. The cathode is held at a temperature of about 800°C by an incandescent heating element and electrons, in consequence of their thermal motion, are ejected with a small velocity which corresponds for the majority to an energy of less than 0.1 eV.

![Diagram](image)

**Fig. 1.** Diagram of the basic demonstration.

The velocity $v_1$ at which they pass through the grid is given by the conservation of energy relation

$$\frac{1}{2} m v_1^2 = e \Phi_1.$$ 

The initial velocity is negligible in comparison with the velocity produced by $\Phi_1$, which is usually of the order of thousands of volts. The beam then illuminates the object, which is a grating cut in the diaphragm $D_1$, passes through the lens $L$, and finally forms an image on the plate $D_2$, which is coated with some fluorescent substance. The potential of the lens $L$ is negative with respect to the potentials of $D_1$ and $D_2$, which are held at the same potential as the outer electrodes $H$ and $H'$, the purpose of which is to localize the effect of the lens. The extent of the lens field is of the order of magnitude of the diameter of the holes in the grids; outside (between the object and the grid $H$ and between the grid $H'$ and the screen), the field is zero and the potential has the constant value $\Phi_1$.

### 1.3 The Object

Electronic devices have been adapted to enable a great variety of objects to be examined; however, despite outward appearances, the optical properties of these instruments can always be traced back to those of the apparatus shown in Fig. 1, an electron gun in conjunction with a grating, the properties of which we shall now examine in more detail.
1.3.1 Opaque Object

A grating is a characteristic example of an object which is genuinely opaque to fast electrons; the bars of a grid are at least some dozens of microns thick, and even though they be made of some light metal, aluminium or beryllium for example, the electrons will be stopped and their energy is absorbed. The result is that the light transmitted is so greatly reduced as to be almost negligible in intensity.

Fig. 2. The electron image of an opaque object (zinc oxide); magnification 20,000.

Fig. 3. The image of an opaque object (zinc oxide) "illuminated" by Li⁺ ions (Gauzit, 1951).
converted into heat, as Lenard showed towards the end of the last century, by a series of very careful measurements (see the article by Bothe in Geiger and Scheel's *Handbuch der Physik* 24). Such objects—powders, fibres, colloids, for example—are thus genuinely opaque in the usual sense of the word, and it is natural that they should give "black and white" images. They are, in fact, the only types of object which can at present be observed in ion microscopy, as such heavy particles are halted by films of matter only a few millimicrons thick (see Figs. 2 and 3).

1.3.2 Thin Objects for Transmission Microscopy

This category, to-day extremely wide—microscopic observation of such objects is an everyday occurrence—produces images of which the black regions represent only an apparent opacity. These objects have in fact the form of very thin films, only a few millimicrons thick (between 10 and 50 µm for electrons accelerated to 60 kV), and such thin layers are effectively transparent to electrons, which pass through them without any appreciable loss of speed. Each electron does undergo, however, a slight deviation—it is "elastically scattered" through an angle θ. A very fine incident pencil is therefore scattered into a cone of a few degrees at the apex, on leaving the thin film; the scattering angles are in fact those which are observed directly in electron diffraction patterns, and lie between a few hundredths and a few tenths of a radian, whereas the initial angular aperture of a beam from a typical electron gun scarcely exceeds a thousandth of a radian. For a given substance, the mean scattering angle θ is proportional to the film thickness; for a heterogeneous specimen, θ is essentially proportional to the mass traversed by the electrons. The effect of the microscope is to convert the variation of the mean scattering angle between different regions of the specimen into an apparent variation of opacity. A diaphragm with a small opening (a few hundredths of a millimetre for a focal length f of a few millimetres) is placed in the focal plane of the objective lens. Rays which have left the object at an angle θ to the optic axis intersect the focal plane at a distance r = θf from the axis; the radius ρ of the hole in the diaphragm is chosen so that only those rays which were scattered through an angle α or less can pass through (ρ = f·α), while those scattered through angles θ greater than α are intercepted (θ > α implies θf > ρ). The presence of a "contrast diaphragm" subtracts a certain number of electrons from each pencil, an action which becomes progressively more important as the scattering becomes more marked, which occurs when the thickness—or, more generally, the quantity of matter traversed—is increased (or again, for a given thickness, as the atomic weight of the constituent elements is made greater). The apparent opacity which appears in the image is therefore proportional to the quantity of matter through which the electrons have passed, and is thus a direct visual representation of variations of thickness.
Fig. 4. The principle behind the formation of contrast by a transparent object (for clarity, the rays are drawn as rectilinear).

Fig. 5a. Contrast in the image of a transparent object (molybdenum oxide); \( \Delta = 20,000 \) (Gauzit).

or of structure of the object. Figure 4 shows schematically this imaging mechanism; for simplicity's sake, the contrast diaphragm is represented as lying between the object and the objective lens, while the electrons follow rectilinear paths, an arrangement which has been successfully used experimentally (Hall, 1948).

Figure 5a is a photograph of the image of a transparent object. The contrast of particular regions of the object can be accentuated by covering them with a thin layer of "heavy" metal (gold or uranium), by the shadow
technique, or by a method similar to the phase contrast techniques of classical optics (Locquin, 1953).

The complex action of the object can, therefore, be summarized by ascribing to it a fictitious transparency, characteristic of the structure at each point. Although clearly only a first approximation, this description proves to be quite adequate as a foundation for the theory of the instruments customarily used to examine such objects. It is, too, extremely difficult to acquire more detailed knowledge about the interaction between a thin object and an electron beam; recent work is only just beginning to yield a few definite results, which will be mentioned later (see also the publications of Marton, 1946; Gabor, 1945; and Boersch, 1947).

1.3.3 Self-luminous Objects

These objects behave as if each point of their surfaces were emitting a fine pencil of electrons ($x = 10^{-4}$ radians, for example); as geometrical optics alone is necessary to describe the mechanism which leads to this conclusion, we shall postpone its description to the paragraph devoted to immersion objectives. Meanwhile, we simply summarize the principle underlying the phenomena which produce in the image features characteristic of the structure of the object, that is, which produce variations in the rate of emission of electrons at different points on the source. Self-luminous objects can be classified according to the type of emission which they produce.

Thermionic electron emission: the object is heated to a temperature at which it emits electrons spontaneously (at least 1300°C for pure metals).

Photoelectric emission: the object is illuminated by an auxiliary light source (usually a source of ultra-violet radiation).

Secondary emission: the object is bombarded with an auxiliary beam of electrons or ions.

Thermionic emission is the most familiar of these, and we shall take it as an example, but experimental evidence shows that in the two other cases the phenomena are qualitatively similar. Within a heated metal, the electrons acquire a thermal excitation, distributed at random among the individual particles, which is ranged about a mean value according to the Maxwellian probability distribution. The electrons are held within the solid by a potential barrier, known as the "work function", whose height depends not only upon the nature of the substance but also upon the condition of its surface. The current density which is emitted depends, therefore, upon three factors, the natural conditions, specimen preparation, and the temperature. By natural conditions, we means such properties as the nature of the metal and the nature of the crystal face at the point being considered, since a slight variation in work function produces a large variation in the emission intensity. The manner in which the specimen has been prepared is reflected
in the cleanliness and relief of the surface, the nature of such impurities as may accidentally have been absorbed at the surface, and by the nature of layers known as "activation layers" with which the surface may become covered as a result of some particular preparation to which the specimen has been subjected. Finally, increasing the specimen temperature increases the energy of the thermal excitation.

![Image of a self-emissive object](image)

Fig. 5b. The image of a self-emissive object (the surface of recrystallized niobium); G = 300 (Septier, 1951).

The factors which help to provide contrast in the image are relief and variations in crystal orientation. We usually try to work under conditions of controlled cleanliness, choosing whichever are most favourable for revealing the surface structure of the crystal; this structure can be seen clearly with the refractory metals, for example, when they have been cleaned by prolonged heating in a vacuum at high temperature (1400 to 2000°C); with the less refractory metals, the surface structure is more elusive—it is necessary to hold them at a low temperature (800°C), after first cleaning the surface, and then covering it with an activation layer of one of the alkaline earths (usually barium or strontium).

These processes are described in detail in the literature of emission electron microscopy (Gauzit and Septier, 1950, 1951). Figure 5b shows the image of such a cathode.
Objects which emit positive ions can be studied equally easily, and again provide images with contrast, under certain conditions; in the case of ions, however, the origin of this contrast is still unknown.

1.4 THE FLUORESCENT SCREEN

As electrons are invisible, a fluorescent screen is necessary to make the image perceptible; the screen fulfils the double function of transforming into light as large a part of the energy of the incident particles as possible, and of ensuring that the flow of electrons is towards the positive pole of the source of the accelerating potential.

1.4.1 Fluorescence

The first function of the screen, that of fluorescing under electron bombardment, is satisfied nowadays by a whole range of substances which differ in the colour of the light emitted, in the duration of the emission of light, in resistance to electron or ion bombardment or in the efficiency with which they convert electron energy into light. So many phosphors are now available that whatever may be the combination of properties necessary for some particular application, it is usually possible to obtain it. Here, we shall simply mention the range of substances which are usually employed in conjunction with an electron optical bench: calcium tungstate, which gives a blue light and can support extreme bombardment and ionic discharge, but is rather insensitive; zinc silicate, or willemite, which fluoresces green, and zinc and cadmium sulphide, which give a white or a yellow light, depending upon which activating element is used, are considerably less robust although much more sensitive. High quality specimens, distributed over a metal support and used at high tension (60,000 volts) will transform up to 10 per cent of the incident electron energy into light, and an efficiency of three or four per cent is standard. For further details, the reader is referred to the books by Kröger (1948), Garlick (1949), and Leverenz (1950).

The light intensity emitted by the screen obeys a law of the form:

\[ B = A f(i)(V - V_0)^n \]

Under normal conditions (the current density \( i \) not larger than a few \( \mu A \text{ cm}^{-2} \)), \( B \) is proportional to \( i \), but not to the accelerating potential which affects \( B \) according to \( (V - V_0)^n \), where the index \( n \) varies between one and three according to the substance used, and \( V_0 \), depending upon
the condition of the screen, is of the order of hundreds of volts. For will-
emite, \( n \sim 2 \). With larger current densities, \( B \) approaches a constant
value; for zinc silicate, this saturation appears beyond about 10 A cm\(^{-2}\),
and for zinc sulphide activated with silver or with copper, beyond
100 A cm\(^{-2}\). These saturation conditions occur whenever the electrons
are concentrated into a spot as nearly as possible a geometrical point,
which is the usual situation in television sets and cathode ray tubes. It is
worth mentioning that fluorescent substances provide an extremely sensitive
detector of radiation. On the final screen of an electron microscope, for
example, with a direct magnification of 50,000, the current density rarely
exceeds 10\(^{-10}\) A cm\(^{-2}\), but nevertheless, the screen is sufficiently luminous
for the eye to be capable of distinguishing clearly the detail in the image,
once it has become accustomed to the darkness, and for the focus to be
adjusted with precision (the situation known as "scotopic" vision). In
other applications, television for example, the spot emits a blinding light.
At such a high brilliance, however, the temperature of the fluorescent
screen rises, and its quality deteriorates correspondingly. It is necessary
to provide a cooling mechanism for the screen, and this is done by using
a copper support, cooled by a current of air or of water.

Nevertheless, with very intense beams of electrons, the sensitive layer
is destroyed gradually by the impact of negative ions which are produced
at the cathode and by ionization of residual gas in the tube.

The high sensitivity of fluorescent substances explains why it has been
for a long time possible to dispense with electron lenses when studying
the properties of electrons. "Electron rays", stopped down by small holes,
a tenth of a millimetre in diameter, cut in a lead screen, still possess
sufficient energy to produce very bright spots, and by such simple means
as these, quite brilliant diffraction patterns can be obtained.

For the last few years, the question of the sensitivity and the robustness
of fluorescent screens to the arrival of powerfully accelerated ions has
arisen, the ions now being considered no longer as stray destructive particles,
but as the "illuminating" particles in ion microscopy. We now know how
to produce sources of monoenergetic ions sufficiently intense to give visible
images, although the image quality improves markedly as the intensity is
reduced, as this renders the numerous parasitic effects less harmful.

All the fluorescent substances commercially available have been system-
atically investigated, and classified according to their sensitivity and their
resistance to bombardment (Gauzit, 1954). Certain of the zinc sulphides,
and willemite, have proved to be superior to calcium tungstate which
nevertheless remains the most robust, with a lifetime of several hours at
a current density of about 10\(^{-12}\) A cm\(^{-2}\).

The object, however, is strongly affected by an ion beam, even though
the intensity of the latter be comparable with the values of beam intensity
which are usual in electron microscopy.
1.4.2 The Screen as an Electron Sink

A fluorescent screen comprises an active layer formed on a passive support. The active layer is sufficiently thin and is composed of a substance sufficiently translucent for the layer to be very transparent—the image, therefore, can be observed from either side of the screen, either by “reflection” looking at the side on which the electrons actually impinge, or by transmitted light, provided that the support is transparent.

The first method is used in electron microscopes, and the support is then a metal plate whose high optical reflecting power enables some of the light which is emitted in directions away from the observer to be used. The majority of the incident charge is conducted away across the fluorescent film to the conducting support; secondary emission effects play a minor part, a phenomenon which will be described in more detail later.

If, on the other hand, we choose to observe the rear face of the fluorescent layer through a transparent support, which is usual in cathode ray oscillographs, the incident current circuit is closed by secondary electron emission. The screen is effectively insulated, as the conductivity of the fluorescent layer and the glass is too poor for there to be an appreciable current towards the conducting ring at their edge. Fortunately, however, provided the accelerating potential lies between 100 and about 10,000 V, each incident electron ejects at least one secondary electron as it arrives at the screen, and frequently many (four or five, say, or even as many as nine if the nature of the layer and the magnitude of the potential are favourable); these secondary electrons are ejected at low speeds, in the direction opposite to the incident electrons, and are collected either on a metal electrode, or on a band of colloidal graphite smeared on the walls of the vacuum chamber (a layer of “aquadag” or carbon black serves the same purpose in oscillographs). These secondary electrons close the “circuit” of the primary electrons, since the collector electrode is connected to the positive pole of the source of high tension; this electrode receives a current equal to the primary current, as any superfluous secondary electrons return to the screen, the potential of which adjusts itself automatically to the value for which the primary and secondary currents exactly balance.

The secondary emission phenomenon is highly stable for potentials of some thousands of volts, but at very low or very high accelerating potentials, it becomes erratic and the multiplication factor falls below unity. With very high potentials, however, we can dispense with the assistance which secondary emission provides at lower potentials, by designing screens which are both transparent and conducting. Such screens consist of the usual thin film of fluorescent material, one surface of which is attached to a glass support, but whose other surface, on which the beam impinges, is covered with a film of aluminium, evaporated onto the fluorescent film in vacuo. The aluminium is sufficiently thin for high energy electrons to pass through
it with no appreciable retardation. This film of high conductivity metal closes the primary current circuit, and in addition, provides a mirror which returns towards the observer that part of the emitted light which would otherwise have been wasted. Further, this layer of aluminium is a most effective shield for the fluorescent material against the destructive impacts of the ions which are inevitable products of the residual gas in the tube. Screens of this kind, brilliant, stable, and robust, are most convenient for work on an electron optical bench.

It is far more difficult to conduct away the superfluous charge at the screen in the ion microscope. The fluorescent film, resting on a metal support, must be extremely thin, but very rich in fluorescent centres, and must be prepared with as little binding material as possible, perhaps with no cellulose binder at all (Gauzit, 1953). There is no question of lining the film with a layer of conductor as this would be completely opaque to an ion beam.

1.5 PHOTOGRAPHY

The photographic plate is even more sensitive than the fluorescent screen as a detector of electrons, provided that the energy of the electrons exceeds a few thousand volts. Instruments using electrons are, however, considerably less stable than their optical equivalents, and the faculty of the photographic plate of "integrating" automatically the incident energy is rarely exploited. In the electron microscope and in electron diffraction, the exposure time usually lies between 0.1 and 10 sec, and only in a few special situations has it been possible to produce lenses and sources of sufficient stability for long exposures—some hours for ions, a day for electrons—to be usefully employed. Examples of these exceptional cases are provided by the photographs taken with \( \beta \)-rays from gadolinium (Marton and Abelson, 1947) or from Th B (Barker, Richardson and Feather, 1950); by astronomy using the electron "telescope" (Lallemand and Duchesne, 1951); and by the first attempts at ion microscopy (Boersch, 1942c, 1948). Generally speaking, any advance which leads to a reduction in exposure time produces higher quality images.

The response characteristics of photographic emulsions remain qualitatively unchanged, but the classification of the various emulsions according to their optical sensitivity has to be completely re-evaluated for electron "illumination". Emulsions which are "slow" for light photographs often prove to be "fast" for electrons, and the film which is used commercially for recording sound is extensively used among electron microscopists; this particular emulsion is extremely sensitive to electrons, and is also very fine-grained.

A great deal has been published on the subject of electron sensitometry; the reader is referred to the books and bibliographies of electron microscopy
mentioned at the end of this book. Here, we shall do no more than mention
the orders of magnitude which are habitually used in electron microscopy,
following a very careful examination by von Borries (1942, 1944) of
German emulsions.

For Agfa “Normal”, the photographic density \( D \), defined by

\[
D = \log_{10} \frac{I_{\text{Incident}}}{I_{\text{Transmitted}}},
\]

is 1.5 for a charge density of \( 10^{-10} \text{ C cm}^{-2} \) and an accelerating potential \( V \)
of 24 kV; for \( V = 75 \text{ kV} \), \( D \) increases to 2.6, but decreases again to 1.5
for \( V = 200 \text{ kV} \). This form of variation with potential is common to all
emulsions; only the optimum potential varies. For \( V = 200 \text{ kV} \), \( D = 1.85 \)
for Agfa “Mikro” and 1.6 for Agfa “Kontrast”; in addition \( D \) is dependent
upon the charge density.

The problem of recording ion images photographically is very difficult
and largely unsolved since ions have only a very feeble penetrating power,
and a highly idiosyncratic sensitometry.

Precise data are scant among the literature. The most numerous studies
have been made with slow ions and mass spectrographs (Aston).

Some information about the relative sensitivity of a few different emulsions
can be found among works on emission ion microscopy (Couchet, Gauzit
and Septier, 1951) or transmission ion microscopy (Magnan and Chanson,
1951; Gauzit, 1951). Emulsions which are suitable for electrons are too rich
in gelatine for ions. An emulsion with large surface density of sensitive grains
is required; Kodak B 10, Ilford Q, and nuclear emulsion plates give tolerable
results.

Yet another problem appears in emission microscopy, where the ion
intensity is stronger: the flow of charge across the insulating support.
For currents of the order of microamperes, large potential differences
build up at certain points, with the consequence that surface currents flow,
sparking over to the metal wall of the camera occurs, and it becomes quite
impossible to focus intense beams. Only with low intensities can the images
be recorded.

1.6 THE OPTICAL BENCH

Varied though the applications of electron microscopy are, they can be
divided, from the technological point of view, into two major categories.
On the one hand, there are those in which monomolecular layers provide
the electron source either photoelectrically or by secondary emission. These
layers are excessively fragile and can only be produced within glass envelopes,
sealed off after very careful evacuation (down to a pressure of \( 10^{-7} \text{ mm Hg} \))
and baked at a high temperature to outgas the objects inside the envelope.
It is not, therefore, possible to make a preliminary study with an arrangement which can easily be dismantled and prototypes have in fact to be very similar to the final model, sealed off in the same way. There is, nevertheless, the possibility of moving some of the internal electrodes from outside, either through flexible membranes sealed to the glass, or with magnets which act on small pieces of soft iron which are free to move inside the evacuated envelope. Demountable joints are quite out of the question, and experimental work is time-consuming, exacting and expensive.

Quite distinct from these systems are those in which only robust sources of electrons are used, refractory metals for example, or else sources which can easily be replaced, like the oxide-coated cathode, and which have in addition comparatively long lifetimes even at pressures of the order of $10^{-5}$ mm Hg. Demountable metal devices can now be built, with neoprene joints and a number of interchangeable elements which can be varied at will, and a whole range of measurements and experiments is feasible with a single piece of apparatus, which is evacuated and is known as an electron optical bench by analogy with its light optical counterpart.

The arrangement of these benches is usually based on the layout of a standard electron microscope, which is the most highly-developed of present day electron devices; often, the bench is nothing more than a microscope which has been modified to be very flexible, but which at the same time remains quite unrefined. This kind of bench is well described in the literature, by Boersch (1951), by Marton et al. (1951) and by Fert (1954), for example.

The most satisfying form of electron optical bench consists of an assembly of separate interchangeable metal components which are assembled one above the other on the end-plate of a pump with a high pumping rate. Each element plays an individual role; there might well be, for example, an electron or ion gun, with a control for each of the variable parameters, electrostatic or magnetic lenses, magnetic screens, observation windows, a photographic chamber, and connecting cylinders. Each unit, of the most robust construction, is centred automatically so that the axes of all the units always coincide; to ensure that the junction between each pair of units is leak-proof, each joint consists of an annular ring of neoprene which is tightly clamped into an annular recess (which also provides accurate centring) cut into the end-plate of the adjacent unit.

Another version, which has been described by Marton et al. (1951), is particularly suitable for magnetic lenses, and requires only a single potential lead for the gun. The whole assembly is ranged within a metal cylinder coupled to a high output vacuum pump. The gun is fixed permanently to the cylinder, from which it is insulated to withstand up to 100 kV. The optical system is mounted on a trolley which can either be within the evacuated chamber, or outside it; the lenses and diaphragms can be shifted from the outside in three mutually perpendicular directions, by means of arms, set in Wilson joints distributed around the fluorescent
screen, which are rotated or moved longitudinally. The screen is set in the middle of a thick plate which seals off the end of the evacuated cylinder, and is observed from outside through an eyepiece.

In conclusion, we describe a recent model due to Septier (1957). This is a special type of optical bench, which has been designed with a view to the exploration of lenses and prisms (optical systems with a symmetry plane) which may be either electrostatic or magnetic, but which are of large dimensions like those employed in particle accelerators. The whole system (see Fig. 6a) consists of the following components: a source of particles, \( S \), (which may be either electrons emitted by a tungsten filament or alkali metal ions obtained by heating a sphere of the aluminosilicate of the alkali metal in question); an accelerating tube containing several lenses \( E_1, E_2 \), giving a beam of particles of some tens of microamperes and one or two millimetres in diameter for potentials between 50 and 200 kV; and a deflexion system. This latter is a rotating electric field \( D \), produced by six electrodes which are supplied with three equal accurately sinusoidal potentials of amplitude \( v \), at a frequency of 50 c/s, and out of phase by 120°; \( v \) can be varied between zero and 1,500 V. On emerging from the deflector, the beam is initially tangential to a cone of semi-aperture \( \alpha_0 \) which is variable between zero and 15°; the apex of the cone lies on the axis at the point \( M \) which is situated at the centre of the deflector. The rotating beam behaves like a hollow divergent beam emerging from the object point \( M \). The cross-section of the beam which would be seen on a fluorescent screen is therefore a circle of variable radius, centred on the axis. There is also an electrostatic lens with large openings, \( L \), which makes the beam parallel to the axis—we thus have at our disposal a hollow cylindrical beam of radius \( R_0 \) which can be varied between zero and \( R_{\text{max}} = 4 \text{ cm} \) simply by varying the amplitude of the potential applied to the deflector. A fluorescent screen, \( E \), supported by a plate of special conductive glass, closes off the vacuum chamber; it is the outside face of this

\[ \text{Fig. 6a. An optical bench for lenses with large dimensions (Septier); } S: \text{particle source; } E_1, E_2: \text{accelerating lenses; } D: \text{electrostatic deflector; } L: \text{electrostatic lens; } Q_1, Q_2: \text{the system to be studied; } E: \text{fluorescent screen.} \]
plate which is examined. The fluorescent substance is either ordinary highly luminous zinc sulphide, if electrons are being used, or a very thin layer of zinc orthosilicate free of binding material, if an ion image is to be formed.

Fig. 6b. The appearance of a hollow emergent beam \((2R_0 = 45\text{mm})\) in the neighbourhood of a line focus, after passing through a highly astigmatic system; the cross-section has been deformed by spherical aberration: (c) should ideally be a straight line.

A grid, which consists of parallel wires, is placed at the exit of \(L\); it serves to check the parallelism of the beam which emerges from \(L\), simply by examining the shadow of the wires cast on the final screen.

The vacuum chamber consists of a number of sections of various lengths; the system to be studied is placed between \(L\) and the final screen, and by using a thin hollow beam, we are able to avoid, even for large apertures, the undesirable influence which the spherical aberration of \(L\) would have on a solid beam. The cross-section of the beam and the forms of the aberration figures can be determined, either for known values of \(R_0\) (when the
incident beam is parallel) or of $\alpha_0$ (when $L$ is not excited and the incident beam diverges from $M$). Magnetic lenses are placed outside the vacuum chamber, and electrostatic lenses within special sections of the chamber which are inserted between $L$ and $E$. This bench has been used to study strong-focusing lenses in particular (see Chapter 10). Figure 6b illustrates examples of the figures obtained on the screen with a hollow incident beam, circular in cross-section, after the beam has passed through a highly astigmatic optical system. The convergence of the latter is slightly varied in the neighbourhood of the value which corresponds to an image of $M$ in focus on the screen for a given value of $\alpha_0$. Instead of a straight line, we obtain a curve which corresponds to a straight line deformed by aberrations.
CHAPTER 2

THE DISTRIBUTION OF POTENTIAL
AND FIELD IN ELECTROSTATIC LENSES

2.1 SOLUTION OF THE LAPLACE EQUATION

The electric field distribution created by a system of electrodes in a
region from which free electric charge is absent can, in theory, be fully
described in terms of a "potential" function, $\Phi$, which everywhere satis-
ifies Laplace's equation:

$$\nabla^2 \Phi = 0.$$  

We recall that $\nabla^2 \Phi$, in which $\nabla^2$ is known as the Laplace operator, takes
the following form in Cartesian coordinates

$$\nabla^2 \Phi(x, y, z) = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2},$$  \hspace{1cm}  (2.1)

and in cylindrical polar coordinates

$$\nabla^2 \Phi(r, \theta, z) = \frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r} \frac{\partial \Phi}{\partial r} \frac{1}{r^2} \frac{\partial^2 \Phi}{\partial \theta^2} + \frac{\partial^2 \Phi}{\partial z^2}.$$  \hspace{1cm}  (2.2)

The electric field $E$ at a point $M$ is obtained by differentiating $\Phi$:

$$E = -\text{grad} \Phi = -\nabla \Phi;$$

$E$ satisfies two fundamental conditions:

$$\text{div} \ E = 0 \quad \text{and} \quad \text{curl} \ E = 0.$$ 

Over the surface of a metal electrode, $\Phi$ is constant and the electric
field $E$ is orthogonal to the electrode surface. Whether or not $\Phi$ is difficult
to calculate depends wholly on the degree of complexity of the boundary
conditions that have to be satisfied.

If the potential depends only on two variables, a solution can be obtained
in many cases from conformal representation theory, but the corresponding
lens would have to possess "translation symmetry"—that is, there would have
to be a direction for which the potential was unaffected by movements
of the lens along it. Such a situation is relatively rare in electron optics.
Usually, however, the lens potential varies in all three directions, and
we are restricted to electrodes whose shapes are those of the coordinate
surfaces of a triply orthogonal system, as it is then possible to find a solution
by separating the variables, that is, by reducing the partial differential
equation to three ordinary differential equations.

Alternatively, it is possible to make an approximate calculation of the
potential distribution along the axis of electron optical systems with a
symmetry axis, by making simplifying assumptions about the boundary
conditions (Bertram, 1940, 1942). Nevertheless, in the majority of cases,
there is no option but to obtain a numerical solution of \( \nabla^2 \Phi = 0 \), not at
all an easy task despite recent efforts (Shortley et al., 1947, 1948; Motz
and Klanfer, 1946; and Hesse, 1949). However complex be the system,
with axial or translation symmetry, a knowledge of the potential can
always be obtained by the so-called “relaxation method”; in this method,
the partial derivatives in the Laplace equation are replaced by finite differ-
ences, and we estimate the value of the potential at the intersections of a
square lattice. A series of successive approximations produces an ever
closer estimate of the true potential at each lattice-point. A detailed analy-
sis and extensive bibliography are to be found in the books by Durand, by
Southwell (1946) and by Shaw (1953).

We now give a brief account of each of these methods of solving the
Laplace equation, after which we consider in greater detail analogue
methods of solution involving the electrolytic tank or the resistance
network.

2.1.1 The Method of Conformal Transformation

*Principle of the method.* This method can be used only in the case of
systems in two variables, that is, systems with translation symmetry; it
depends upon the properties of analytic functions of a complex variable.

If we set \( j = \sqrt{-1} \), the variable \( Z = x + jy \) (in which \( x \) and \( y \) are two real
variables) is known as a complex variable and may be represented as a
point having coordinates \((x,y)\) in the complex plane \(xOy\) (the Argand
diagram, Fig. 7a). In polar coordinates, we write

\[
r = \sqrt{x^2 + y^2} \quad \text{and} \quad \theta = \tan^{-1}(y/x)
\]

so that

\[
Z = x + jy = r(\cos \theta + j \sin \theta) = r \exp j\theta.
\]

We now consider another complex variable,

\[
W = u(x, y) + jv(x, y).
\]

To every value of \( Z \) will correspond a value of \( W \), and we can write

\[
W = f(Z).
\]
If $Z$ is continuously varied, the corresponding point will describe a curve $(C)$ in the plane $xOy$ and the point corresponding to $W$ will trace out a curve $(C')$ in the complex plane $uOv$ (Fig. 7a). The function $W$ is said to be "analytic" if the derivative

$$\frac{dW}{dZ} = \lim_{\Delta Z \to 0} \frac{f(Z + \Delta Z) - f(Z)}{\Delta Z}$$

exists and is unique (irrespective of the direction of $\Delta Z$ in the plane $xOy$). It is easy to show that this is the case when the real functions $u(x, y)$ and $v(x, y)$ satisfy the Cauchy–Riemann relations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \quad (2.3)$$

All functions that can be expressed as a series expansion in the neighbourhood of the origin ($W = \sum_n c_n Z^n$) belong to this family (examples are $e^z$, $\sin Z$, ...).

On differentiating relations (2.3), we obtain the two equations

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \quad (2.4)$$

or

$$\nabla^2 u(x, y) = 0, \quad \nabla^2 v(x, y) = 0.$$

Thus each of the two component functions of $W$ satisfies the Laplace equation and either may be regarded as the potential function in a two-dimensional system.

Since the function $W$ is analytic, the derivative $dW/dZ$ at a point is independent of the direction of $\Delta Z$ at this point. We can therefore write

$$\frac{dW}{dZ} = A e^{i\phi} \quad \text{or} \quad dW = A e^{i\phi} dZ.$$

The segment $dW$ in the $W(u, v)$-plane is obtained from $dZ$ by a change
of scale $A$ and a rotation $\phi$. Hence, if two curves in the $Z$-plane intersect at an angle $\alpha$, the transformed curves will intersect at the same angle in the $W$-plane; such a transformation is known as a conformal transformation. In particular, the curves $u = \text{const}$ and $v = \text{const}$ in the $W$-plane intersect at right angles; their transformations $u(x, y)$ and $v(x, y)$ in the $Z$-plane will therefore be orthogonal (Fig. 7b). If, then, $u(x, y)$ is selected to be the potential function, the family of curves $u(x, y) = \text{const}$ in the $Z$-plane will form a map of the equipotentials and if we plot the curves $v(x, y) = \text{const}$, we obtain the electric field lines immediately (and vice versa).

A conformal transformation will therefore relate a region of uniform field in the $W$-plane to a more complex situation in the $Z$-plane, in which certain curves $u(x, y) = \text{const}$ (or $v(x, y) = \text{const}$) will coincide with the real electrodes considered.

![Fig. 7b. The curves $u = \text{const}$ and $v = \text{const}$ in the $Z$- and $W$-planes.](image)

**Two simple examples**

(a) Let $W = Z^n$. We can then write $W$ in the form

$$W = r^n e^{in\theta}$$

so that

$$u = r^n \cos n\theta,$$

$$v = r^n \sin n\theta.$$  

We select $v$ as the potential function. Then $v = 0$ at $\theta = 0$ and $\theta = \pi/n$. For $n = 2$, the corresponding map is shown in Fig. 7c; we can therefore obtain a potential

$$\Phi = Ar^2 \sin 2\theta$$

by means of two plane electrodes coinciding with the axes $OX$ and $OY$ and held at zero potential together with a hyperbolic conducting electrode.
at a distance \( a \) from \( O \) and held at potential \( +\Phi_0 \). We then have

\[
\Phi(r, \theta) = \Phi_0 \frac{r^2}{a^2} \sin 2\theta
\]
or

\[
\Phi(x, y) = \frac{2\Phi_0}{a^2} xy.
\]  \hfill (2.5)

This distribution can also be obtained with the aid of four identical electrodes held at potentials \( \pm \Phi_0 \) (Fig. 7c), placed symmetrically with respect to the axes \( Ox \) and \( Oy \): this transformation has thus provided us with a quadrupole lens (see Chapter 10). With \( n = 3 \), we should obtain a sextupole lens and with \( n = 4 \), an octupole, and so on; the case \( n = 1 \) corresponds to the parallel plate condenser.

The general solution of Laplace’s equation will therefore be of the form

\[
\Phi(r, \theta) = \sum_n A_n r^n \sin n\theta.
\]

![Fig. 7c. Electrodes creating the distribution \( \Phi(r, \theta) = Ar^2 \sin 2\theta \) \((W = Z^2)\).](image)

(b) Let us now consider the function \( W = A \log Z \). We have

\[
u(r, \theta) + jv(r, \theta) = A \log r + Aj(\theta - \theta_0).
\]

Taking \( \Phi(r, \theta) = v(r, \theta) \), the curves \( v = \text{const} \) will be straight lines emerging from the origin \( O \), and the curves \( u = \text{const} \) will be circles centred on \( O \).

For the special case of the electrodes illustrated in Fig. 7d, consisting of two planes separated at \( O \) by an infinitesimally small gap and held at potentials \( V_1 \) and \( V_2 \), we find

\[
\begin{align*}
v(r, \theta) &= \frac{V_2 - V_1}{\theta_2 - \theta_1} \theta + \frac{V_1 \theta_2 - V_2 \theta_1}{\theta_2 - \theta_1}, \\
u(r, \theta) &= \frac{V_2 - V_1}{\theta_2 - \theta_1} \log r.
\end{align*}
\]  \hfill (2.6)
When $\theta_2 - \theta_1 = \pi$, $V_1 = 0$ and $V_2 = V_0$ so that by choosing $\theta_1 = 0$, we obtain

$$v(\theta) = \frac{V_0}{\pi} \theta,$$

$$u(r) = \frac{V_0}{\pi} \log r.$$

If, on the contrary, we take $\Phi(r, \theta) = u(r, \theta)$ we shall obtain the solution of the Laplace equation between two coaxial metal cylinders of radii $r_1$ and $r_2$ held at potentials $V_1$ and $V_2$ respectively (Fig. 7d):

$$u(r) = \frac{V_2 - V_1}{\log (r_2/r_1)} \log \frac{r}{r_1} + V_1,$$

$$v(\theta) = \frac{V_2 - V_1}{\log (r_2/r_1)} \theta. \tag{2.7}$$

Other examples are to be found in the books by Ramo and Whinnery and Durand. An electrode system can be associated with every analytic function $W$ in this way, and a list of these solutions may thus be prepared.

The Schwarz transformation. It would be convenient if one could (by means of a conformal transformation) transform a given set of electrodes into a simpler set, in which the equipotentials were expressible in terms of a known function, belonging to the preceding category. A few solutions of this difficult problem are available.

One of the most useful of these is the Schwarz transformation; this is applied when the electrodes are planes, intersecting around a polygonal boundary (Fig. 8a), which may indeed have some of its apices at infinity.

This transformation is written

$$\frac{dZ}{d\zeta} = A \prod_{n} (\zeta - \zeta_n)^{(a_0/\pi - 1)}. \tag{2.8}$$
in which

\[ Z = A \int_0^\zeta (\zeta - \zeta_1)^{(\alpha_1/\pi - 1)} (\zeta - \zeta_2)^{(\alpha_2/\pi - 1)} \cdots (\zeta - \zeta_n)^{(\alpha_n/\pi - 1)} \, d\zeta + B. \]

The points \( \zeta_i = \pm \infty \) do not enter into this formula.

Consider a simple example: the outline of the parallel plate condenser shown in Fig. 8b. From symmetry considerations, we need only calculate the function \( u(x, y) \) in a simpler system consisting of two electrodes: one a plane, held at potential \( \Phi = V_0 \), and the other a half-plane, distant \( y = h \) from the first for \( x > 0 \), and held at \( \Phi = 0 \). In this special case, we have \( \alpha_1 = 0, \alpha_2 = 2\pi \) and \( \alpha_3 = \pi \). In the complex \( \zeta \)-plane, the apices \( A, B, C, D, E \) of the polygon (only \( D \) is not at infinity) all lie on the real axis \( Ox' \). We may seek a particular transformation such that the first apex (\( CB \)) is situated at \( \zeta = 0 \), the second (\( D \)) at \( \zeta = 1 \) and the third (\( EA \)) at \( \zeta = \pm \infty \) (Fig. 8c). We then have

\[ Z = A \int_0^\zeta (\zeta - 0)^{(\alpha/\pi - 1)} (\zeta - 1)^{(2\pi/\pi - 1)} d\zeta + B \]

or

\[ Z = A \int_0^\zeta (1 - 1/\zeta) \, d\zeta + B = A (\zeta - \log \zeta) + B. \]
For $Z = 0 + jh$, $\zeta = 1$ so that $jh = A + B$:

$$Z = A(\zeta - 1 - \log \zeta) + jh.$$ 

In the vicinity of $\zeta = 0$, we have

$$\frac{dZ}{d\zeta} = -\frac{A}{\zeta}.$$ 

Integrating over a circle of radius $r'$ about $\zeta = 0$, from $\alpha' = \pi$ to $\alpha' = 0$ (that is, from $C$ to $B$), and using the relation

$$d\zeta = d(r' e^{j\alpha'}) = j r' e^{j\alpha'} d\alpha' = j\zeta d\alpha',$$

we have

$$\int_B^C dZ = jh = \int_\pi^0 \left(\frac{A}{\zeta}\right) j\zeta d\alpha' = jA\pi,$$

so that

$$A = \frac{h}{\pi}.$$

The required transformation is therefore

$$Z = \frac{h}{\pi}(\zeta - 1 - \log \zeta + j\pi).$$

The potential function of the system shown in Fig. 8c, consisting of two half planes held at $\Phi = 0$ and $\Phi = V_0$, is given (see above) by the $v$ term of the analytic function

$$W = \frac{V_0}{\pi} \log \zeta = u + jv$$

or

$$\zeta = \exp \left(\pi W/V_0\right) = \exp \frac{\pi(u + jv)}{V_0}.$$ 

Hence

$$Z = \frac{h}{\pi} \left( e^{\pi W/V_0} - 1 \right) - \frac{\pi W}{V_0} + j\pi. \quad (2.9)$$
After extracting $u$ and $v$ from this formula, the curves $V = \text{const}$ can be plotted in the $Z$-plane; this provides us with the potential distribution in the fringe-field region of the parallel plate condenser. This distribution is of importance in particle deflectors; the fringe-field at the edges of the pole-pieces of an electromagnet may be treated in a similar way.

2.1.2 General Solutions. The Separation of Variables Method

This is a general method, and can be used to solve all partial differential equations; it is, however, only applicable for certain simple types of boundary conditions, and hence only for certain simple electrode shapes.

In essence, the solution is written as a product of functions, each of which contains only one of the variables of the coordinate system employed.

*Cartesian coordinates*

(a) *Two-dimensional systems*

$$\nabla^2 \Phi(x, y) = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} = 0.$$  

We write

$$\Phi(x, y) = X(x) \cdot Y(y).$$

Substituting into Laplace’s equation and writing $\frac{\partial^2 X}{\partial x^2} = X''$ and $\frac{\partial^2 Y}{\partial y^2} = Y''$, we find

$$\frac{X''}{X} = \frac{Y''}{Y}.$$  

The term $X''/X$ is a function of $x$ only and the term $Y''/Y$ of $y$ only. The equation will be satisfied everywhere only if these two terms are constant:

$$\frac{X''}{X} = k^2, \quad \frac{Y''}{Y} = -k^2,$$

giving general solutions

$$X(x) = A_1 \cosh kx + B_1 \sinh kx,$$
$$Y(y) = A_2 \cos ky + B_2 \sin ky,$$

so that $\Phi(x, y)$ can be written

$$\Phi(x, y) = \begin{bmatrix} A_1 \cosh kx \\ B_1 \sinh kx \end{bmatrix} \begin{bmatrix} A_2 \cos ky \\ B_2 \sin ky \end{bmatrix}. \quad (2.10)$$
The constants $A_1, A_2, B_1, B_2$ and $k$ are determined for a specific system by the boundary conditions at the electrodes. If $k$ is imaginary, the circular functions become hyperbolic functions, and conversely.

(b) Three-dimensional systems. We write

$$\Phi(x, y, z) = X(x) \cdot Y(y) \cdot Z(z)$$

and following the same procedure, we find

$$\frac{X''}{X} = \alpha^2, \quad \frac{Y''}{Y} = \beta^2, \quad \frac{Z''}{Z} = -\gamma^2,$$

with

$$\alpha^2 + \beta^2 = \gamma^2$$

and hence

$$\Phi(x, y, z) = \left\{ A_1 \cosh \alpha x \right\} \left\{ A_2 \cosh \beta y \right\} \left\{ A_3 \cos \gamma z \right\}$$

$$\left\{ B_1 \sinh \alpha x \right\} \left\{ B_2 \sinh \beta y \right\} \left\{ B_3 \sin \gamma z \right\}.\quad (2.11)$$

Other solutions are found by giving $\alpha, \beta, \gamma$ imaginary values, but the relation $\alpha^2 + \beta^2 = \gamma^2$ must always be satisfied; this means that at least one of the partial solutions $X, Y$ or $Z$ will contain hyperbolic functions.

(c) The idea of spatial harmonics. The term $\Phi(x, y) = A \sinh kx \sin ky$ is an elementary solution of (2.10). We now show that it is impossible to create such a potential distribution exactly, with electrodes of finite size; with finite electrodes, therefore, only an approximation to the theoretical field can be created.

For $x = 0, y = 0$ and $y = \pi/k = b$, we have $\Phi = 0$. We assume that the line $y = b/2$ intersects an electrode held at potential $\Phi_0$ at $x = l$.

We then have

$$\Phi_0 = A \sinh kl, \quad \text{so that} \quad A = \Phi_0 \cosech (\pi l/b)$$

and

$$\Phi(x, y) = \frac{\Phi_0}{\sinh \pi l/b} \sinh \frac{\pi x}{b} \sin \frac{\pi y}{b}.\quad (2.12)$$

The cross-section of the electrode at potential $\Phi = \Phi_0$ will be the curve defined by $\Phi(x, y) = \Phi_0$, or

$$\sinh \frac{\pi x}{b} \sin \frac{\pi y}{b} = \sinh \frac{\pi l}{b}.$$
This curve tends asymptotically to the straight lines $y = 0$ and $y = b$, and extends from $x = l$ to $x = +\infty$ (Fig. 8d); such an electrode cannot be built in practice. (The same can be shown to be true of all the elementary solutions of the Laplace equation.) If we discontinue the electrode beyond a finite distance, it is only possible to satisfy the boundary conditions by superimposing an infinite number of these elementary solutions. Consider, for example, the system shown in Fig. 8e, which can be made in practice, and which consists of planes $x = 0$, $y = 0$ and $y = b$ held at potential $\Phi = 0$ together with a planar region at $x = l$ and held at $\Phi = \Phi_0$. The gap between the latter and the other electrodes is assumed to be infinitesimally narrow.

At $x = l$, the elementary solution

$$\Phi(x, y) = A \sinh kx \sin ky \quad \text{(with } k = \pi/b)$$

cannot on its own satisfy the condition

$$\Phi = \Phi_0 \quad \text{for} \quad 0 < y < b.$$

If, however, we take the higher order solutions, with $k_n = n\pi/b$, we obtain a general solution of $\nabla^2 \Phi = 0$ in the form

$$\Phi(x, y) = \sum_n A_n \sinh k_n x \sin k_n y.$$

This sum of elementary solutions can now be made to satisfy the condition:

$$\Phi_0 = \sum_n A_n \sinh \frac{n\pi l}{b} \sin \frac{n\pi y}{b} \quad 0 < y < b. \quad (2.13)$$

This expression is in fact the Fourier series expansion of the function
\[ f(y) = \Phi_0. \] Here, this series contains only sine terms:

\[ f(y) = \Phi_0 = \sum_{n=1}^{\infty} a_n \sin \frac{n\pi y}{b} \]  

(2.14)

and it is easy to show that if \( n \) is odd, \( a_n = 4\Phi_0/n\pi \), and if \( n \) is even, \( a_n = 0 \). Comparing the physical solution (2.13) and the series (2.14), we see that

\[ a_n = A_n \sinh \frac{\pi l}{b}, \]

so that finally,

\[ \Phi(x, y) = \sum_{n \text{ odd}} \frac{4\Phi_0}{n\pi} \frac{\sinh (n\pi x/b)}{\sinh (n\pi l/b)} \sin \frac{n\pi y}{b}. \]  

(2.15)

This function, which converges rapidly (except in the immediate vicinity of \( \pi = l \)), can be used to determine the potential everywhere within the electrode system.

The term \( n = 1 \) is the fundamental, and the terms \( n > 1 \) are known as "spatial harmonics". They are necessary in order that the boundary conditions at the electrodes determining the field can be satisfied.

In practice, the preceding system, with no gaps between the electrodes, could not be built. If the electrode at \( x = l \) is in the form of a band of width \( d < b \), and is symmetrical about \( y = b/2 \), we need only know the function \( \Phi(l, y) \) for \( 0 < y < b \) and calculate the coefficients \( a_n \) of its Fourier series in order to find the coefficients \( A_n \) and hence the expression for the general solution, \( \Phi(x, y) \); in this special case, it is a very good approximation to replace \( \Phi(l, y) \), as a function of \( y \), by an isosceles trapezium.

Cylindrical polar coordinates

(a) The general solution. We now have

\[ \nabla^2 \Phi(r, \theta, z) = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \Phi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial \Phi}{\partial \theta^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0, \]

into which we substitute

\[ \Phi(r, \theta, z) = R(r)\Theta(\theta)Z(z). \]

Differentiating, we obtain

\[ \frac{1}{rR} \frac{d}{dr} \left( rR' \right) + \frac{1}{r^2} \Theta'' + \frac{Z'}{Z} = 0. \]

We can then write \( \Theta''/\Theta = -n^2 \) and \( Z''/Z = k^2 \); \( n \) is necessarily a real integer, since the physical solution can only contain \( \cos n\theta \) or \( \sin n\theta \),
because the potential must take the same value when we advance through $\theta = 2\pi$ around a circle $r = \text{constant}$. $k$ may be real or imaginary.

We thus obtain

$$
\frac{1}{r} \frac{d}{dr} \left( r R' \right) + \left( k^2 \frac{n^2}{r^2} \right) R = 0.
$$

Introducing a new variable $v = kr$, this $R$-equation is transformed into a Bessel equation:

$$
\frac{1}{v} \frac{d}{dv} \left( v \frac{dR}{dv} \right) + \left( 1 - \frac{n^2}{v^2} \right) R = 0,
$$

of which the general solution for real $k$ is expressed as

$$
R_n(kr) = AJ_n(kr) + BN_n(kr).
$$

($J_n(kr)$ is a Bessel function of the first kind of order $n$ in the variable $kr$ and $N_n(kr)$ is a Bessel function of the second kind.)

If $k$ is imaginary, the solution is of the form

$$
R_n(kr) = AI_n(kr) + BK_n(kr),
$$
in which $I_n$ and $K_n$ are modified Bessel functions of the first and second kinds. Since the functions $N_n(kr)$ and $K_n(kr)$ become infinite on the axis $Oz$ ($r = 0$), $B$ is necessarily zero when there are no electrodes on the axis; this is the case for all electrostatic lenses used in practice.

The general solution of the Laplace equation is therefore of the following form:

$$
\text{real } k: \quad \Phi(r, \theta, z) = A \begin{bmatrix} J_n(kr) \\ N_n(kr) \end{bmatrix} \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \begin{bmatrix} \cosh kz \\ \sinh kz \end{bmatrix},
$$

$$
\text{imaginary } k: \quad \Phi(r, \theta, z) = A \begin{bmatrix} I_n(kr) \\ K_n(kr) \end{bmatrix} \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \begin{bmatrix} \cos kz \\ \sin kz \end{bmatrix}.
$$

(2.16)

There are two important cases to be considered.

**Systems with translation symmetry ($k = 0$):** The differential equation for $R$ is of the form

$$
r \frac{d}{dr} \left( r \frac{dR}{dr} \right) - n^2 R = 0,
$$

which has the general solution

$$
R(r) = a_n r^n + b_n r^{-n}
$$

so that

$$
\Phi(r, \theta) = \sum_n (a_n r^n + b_n r^{-n}) \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix}.
$$
Notice that if \( k = 0 \), we may have \( Z(z) = Az + B \), but the system no longer has translation symmetry. In a lens with no electrode at the axis, we have \( b_n = 0 \) (so that the potential remains finite at the axis). We recover the general solution

\[
\Phi(r, \theta) = \sum_n A_n r^n \left\{ \begin{array}{c} \cos n\theta \\ \sin n\theta \end{array} \right\}
\]

encountered in the preceding section.

**Systems with rotational symmetry** \( \left( \frac{\partial}{\partial \theta} = 0, n = 0 \right) \): The Bessel functions are of zero order. If the axis is free of electrodes, we have

\[
\Phi(r, z) = AI_0(kr) \left\{ \begin{array}{c} \cos kz \\ \sin kz \end{array} \right\}
\]

or

\[
\Phi(r, z) = AJ_0(kr) \left\{ \begin{array}{c} \cosh kz \\ \sinh kz \end{array} \right\} = BJ_0(kr) \left\{ \begin{array}{c} \exp kz \\ \exp (-kz) \end{array} \right\}.
\]

**(b) Spatial harmonics.** As before, these expressions are only elementary solutions of the Laplace equation, and cannot be exactly reproduced in practice. The boundary conditions can be satisfied in a real system only by introducing an infinite number of spatial harmonics.

Consider, for example, the electrodes shown in Fig. 8f, consisting of

\[\begin{array}{c}
\Phi = \Phi_0 \\
0 \\
\end{array}\]

\[\begin{array}{c}
\Phi = 0 \\
z \\
\end{array}\]

\[\begin{array}{c}
r = a \\
0 \\
\end{array}\]

**Fig. 8f.** A cylindrical box (two planes held at \( \Phi = 0 \) and a cylinder at \( \Phi = \Phi_0 \)).

two planes \( z = 0 \) and \( z = l \), held at potential \( \Phi = 0 \), together with a tubular portion of radius \( r = a \); the function \( Z(z) \) can be represented by a series of terms of the form

\[Z(z) = A_m \sin k_m z \quad \text{with} \quad k_m = m\pi/l\]

so that

\[\Phi(r, z) = \sum_m A_m l_0(k_m r) \sin k_m z.\]
The coefficients $A_m$ are determined by expanding the function $\Phi(a, z)$ for $0 < z < l$ as a Fourier series:

$$\Phi(a, z) = \sum_m a_m \sin\frac{m\pi z}{l}.$$  

In the simple case where $\Phi(a, z) = \Phi_0 = \text{const}$, for $0 < z < l$, it is easy to show that

$$\Phi(r, z) = \sum_{m \text{ odd}} 4\Phi_0 I_0(m\pi r/l) \frac{m\pi z}{l}.$$  \hspace{1cm} (2.17)

In this way, the majority of practical problems involving simple electrode shapes may be solved. With the aid of an electronic computer, the potential distribution can be plotted rapidly from solutions in series form. This method requires a knowledge of the potential distribution over a cylinder $r = a$. In the case of the system of Fig. 8g, we assume that, for

$$0 < z < z_0 \text{ and } l - z_0 < z < l,$$ 

the potential varies linearly with $z$ so that $\Phi(a, z)$ is represented by an isosceles trapezium; we accept a good approximation (often perfectly adequate), for want of a rigorously exact expression.

(c) The use of Fourier integrals. In a system in which $\Phi_0(z)$ is not bounded in the direction $Oz$ (as is the case in lenses formed from cylindrical tubes or diaphragms with openings in them), the function $\Phi(a, z)$ extends in theory from $z = -\infty$ to $z = +\infty$. It can then be represented not by a Fourier series but by a Fourier integral.

If $\Phi(r, z)$ can be put into the form of an infinite series, thus:

$$\Phi(r, z) = \sum_k \{A(k)I_0(kr) \sin kz + B(k)I_0(kr) \cos kz\},$$  \hspace{1cm} (2.18)
we may write

$$\Phi(r, z) = \int_0^\infty I_0(kr)\{A(k) \sin kz + B(k) \cos kz\} \, dk$$

and

$$\Phi(a, z) = \int_0^\infty I_0(ka)\{A(k) \sin kz + B(k) \cos kz\} \, dk.$$

The coefficients $A_k I_0(ka)$ and $B_k I_0(ka)$ can be calculated from the relations

$$A(k) I_0(ka) = \frac{2}{\pi} \int_0^\infty \Phi(a, z) \sin kz \, dz,$$

$$B(k) I_0(ka) = \frac{2}{\pi} \int_0^\infty \Phi(a, z) \cos kz \, dz.$$  \hspace{1cm} (2.19)

We now consider a few examples.

A lens, consisting of two cylinders of equal radius, and separated by an infinitesimally narrow gap (Fig. 8h). At $z = 0$, we take as origin of poten-

![Fig. 8h. Lens consisting of two identical cylinders in contact.](image)

Fig. 8h. Lens consisting of two identical cylinders in contact.

tial the value $\Phi = (V_1 + V_2)/2$. The function $\Phi(a, z)$ becomes $-(V_2 - V_1)/2$ for $z < 0$ and $(V_2 - V_1)/2$ for $z > 0$; this odd function can be expressed in terms of $\sin kz$ alone:

$$\Phi(r, z) - \frac{V_1 + V_2}{2} = \int_0^\infty A(k) I_0(kr) \sin kz \, dk.$$

For $\Phi(a, z) = \frac{V_2 - V_1}{2}$ ($z > 0$)

$$A(k) = \frac{2}{\pi I_0(ka)} \int_0^\infty \frac{V_2 - V_1}{2} \sin kz \, dz = \frac{V_2 - V_1}{\pi k I_0(ka)}$$
and
\[ \Phi(r, z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi} \int_0^\infty \frac{I_0(kr)}{k I_0(ka)} \sin kz \, dk ; \]
on the axis,
\[ \Phi(0, z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi} \int_0^\infty \frac{\sin kz}{k I_0(ka)} \, dk = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi} F(k, z). \]

Bertram (1940) has pointed out that the integral \( F(k, z) \) can be represented to a very good approximation by the function
\[ F(k, z) \approx \tanh \left( \frac{wz}{a} \right), \quad w = 1.318, \]
so that
\[ \Phi(0, z) \approx \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi} \tanh \left( \frac{wz}{a} \right). \tag{2.20} \]

*If the two cylinders are separated by a distance \( 2d \) (Fig. 8i), we have*

\[
\Phi(a, z) = \begin{cases} 
V_1 & (z < d), \\
\frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{2d} z & (-d < z < +d), \\
V_2 & (z > d).
\end{cases}
\]

\[ A(k) = \frac{2}{\pi I_0(ka)} \left( \int_0^d \frac{V_2 - V_1}{2d} \sin kz \, dz + \int_d^\infty \frac{V_2 - V_1}{2} \sin kz \, dz \right) \]
so that
\[ A(k) = \frac{\sin kd}{\pi dk^2 I_0(ka)} \]

![Fig. 8i](image_url)

Fig. 8i. Lens consisting of two cylinders of the same radius.
and
\[ \Phi(r, z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi d} \int_0^\infty \frac{I_0(kr)}{k^2 I_0(ka)} \sin kd \sin kz \, dk. \]  
\( (2.21) \)

On the axis,
\[ \Phi(0, z) = \frac{V_1 + V_2}{2} + \frac{V_2 - V_1}{\pi d} \int_0^\infty \frac{\sin kd \sin kz}{k^2 I_0(ka)} \, dk. \]

The integral on the right-hand side can, however, be written in the form (Bertram, 1951a)
\[ \int_0^\infty \frac{\sin kd \sin kz}{k^2 I_0(ka)} \, dk = \frac{1}{2} \int_{(z-d)/a}^{(z+d)/a} \, dz \int_0^\infty \frac{\sin k \cdot \cdot \cdot}{k I_0(ka)} \, dk, \]
\[ \frac{1}{2} \int_{(z-d)/a}^{(z+d)/a} F(k, z) \, dz. \]

Replacing \( F(k, z) \) by its approximate expression, we have
\[ \Phi_0(z) = \Phi(0, z) \]
\[ = \frac{V_1 + V_2}{2} + \frac{(V_2 - V_1)a}{4wd} \log \left( \cosh \frac{w(z + d)}{a} / \cosh \frac{w(z - d)}{a} \right). \]  
\( (2.22) \)

Using this method, the reader can recover the potential distribution on the axis of the lens of Fig. 87, given by formula (8.39), and hence derive the function \( \Phi_0(z) \) for a three-diaphragm lens.

**Remark.** Instead of the solution (2.18), we can equally well use a series of the form
\[ \Phi(r, z) = \sum_k A(k) J_0(kr) \exp (-kz). \]

(The solution in \( \exp (+kz) \) must be discarded, as \( \Phi(r, z) \) must not become infinite at \( z = \infty \).)

Taking \( J_0(ka) = 0 \), we can write the function \( \Phi(r, z) \) as
\[ \Phi(r, z) = V_2 + \sum_{n=1}^\infty A_n J_0(\mu_n r / a) \exp (-\mu_n z / a), \]
in which \( \mu_n \) denotes the \( n \)-th root of \( J_0(x) \).

At \( z = 0 \), we have \( \Phi(r, 0) = (V_1 + V_2)/2; \) we can regard this series as an expansion in Bessel functions in the variable \( (\mu_n r / a) \), and following a procedure similar to that used for circular functions, the coefficients \( A_n \) may be calculated in terms of the known function \( \Phi(a, z) = V_2 \).
\[ A_n = \left( \frac{V_1 + V_2}{2} - V_2 \right) \int_0^a r J_0(\mu_n r/a) \, dr \]

and

\[ \Phi(r, z) = V_2 + (V_1 - V_2) \sum_{n=1}^{\infty} \frac{J_0(\mu_n r/a)}{\mu_n J_1(\mu_n)} \exp \left( -\frac{\mu_n z}{a} \right). \quad (2.23) \]

This expression is equivalent to (2.21) but is better adapted to calculations with a computer.

Superposition methods. By using elliptical coordinates (see Durand), it is possible to calculate rigorously the potential distribution, \( \Phi(r, z) \), created by an infinitesimally thin conducting plane \( P \) held at a potential \( \Phi_0 \), containing a circular opening of radius \( R \):

\[ \Phi(r, z) = \Phi_0 - \frac{1}{2} \left( E_1 - E_2 \right) z + \frac{E_1 - E_2}{\pi} \left| z \right| \left( \frac{R}{\alpha} \tan^{-1} \frac{\alpha}{R} \right), \quad (2.24) \]

where

\[ \alpha = \left[ \frac{1}{2} \left( r^2 + z^2 - R^2 \right) + \frac{1}{2} \left( r^2 + z^2 - R^2 \right)^2 + 4z^2 R^2 \right]^{1/2}. \]

On the axis of the hole \( (r = 0) \), \( \alpha = \left| z \right| \), and

\[ \Phi_0(z) = \Phi_0 - \frac{1}{2} \left( E_1 + E_2 \right) z - \frac{E_1 - E_2}{\pi} R \left( 1 + \frac{z}{R} \tan^{-1} \frac{z}{R} \right) \]

or

\[ \Phi_0(z) = a + bz + cz \tan^{-1} \frac{z}{R}. \quad (2.25) \]

For a family of parallel conducting planes, \( P_i \), with abscissae \( z_i \), held at potentials \( \Phi_i \), and containing holes of radii \( R_i \) all having the same axis \( Oz \), Regenstreif (1951a) and Bertein (1952a) have shown that \( \Phi_0(z) \) can be obtained by superimposing solutions of the foregoing type (see § 8.2). The constants are found from the boundary conditions: far from the axis, the electric fields between the planes are uniform and given by

\[ E_i = -\frac{\Phi_{i+1} - \Phi_i}{z_{i+1} - z_i}. \]

2.1.3 Numerical Solution. Relaxation

The preceding methods yield rigorous or approximate solutions in analytic form; they make it possible to obtain the potential function or the flux function at every point \( M(x, y, z) \). There are other methods, however, which yield the values of these functions only at a finite number of points situated at the intersections of a mesh of step-length \( h \), which
may be as small as required. These methods, which are called “relaxation methods”, are particularly well-suited to calculations with a computer and can be applied to systems of any shape. It is only necessary to know the boundary conditions at the electrodes, and also between the electrodes a reasonable distance from the region of interest.

These methods are described in detail in the books by Durand and Southwell, who also discuss their accuracy and the methods of obtaining rapid convergence towards the final solution. By using finite difference equations (instead of partial differential equations), we can determine the value of the function at some given point in terms of its values at neighbouring points. Calculation of a potential “map” now becomes a process of successive approximations (or relaxation), each stage of which is more accurate than its predecessor. When two successive sequences of calculations have given results differing by less than some fraction \( \Delta \Phi/\Phi \) decided beforehand, the calculation is halted.

We shall give only the principle of these methods.

**Systems with translation symmetry.** The function \( V(x, y) \) can be expanded as a Taylor series in the neighbourhood of a point \( M(x_0, y_0) \):

\[
V = V_0 + \frac{1}{1!}\left\{(x-x_0)\left(\frac{\partial V}{\partial x}\right)_0 + (y-y_0)\left(\frac{\partial V}{\partial y}\right)_0\right\} \\
+ \frac{1}{2!}\left\{(x-x_0)^2\left(\frac{\partial^2 V}{\partial x^2}\right)_0 + 2(x-x_0)(y-y_0)\left(\frac{\partial^2 V}{\partial x\partial y}\right)_0 + (y-y_0)^2\left(\frac{\partial^2 V}{\partial y^2}\right)_0\right\} + \cdots
\]

(2.26)

If we draw a rectangular grid of mesh-length \( h \), we need only know the values of the function \( V \) at five neighbouring points in order to calculate the five partial derivatives at \( M \) that appear in this formula.

Let us apply this relation to the points 1, 2, 3 and 4 of Fig. 8j \((x-x_0 = \pm h; y-y_0 = \pm h)\). We obtain:

\[
\left(\frac{\partial V}{\partial x}\right)_0 = \frac{1}{2h}(V_1 - V_3), \quad \left(\frac{\partial V}{\partial y}\right)_0 = \frac{1}{2h}(V_4 - V_2), \\
\left(\frac{\partial^2 V}{\partial x^2}\right)_0 = \frac{1}{h^2}\{(V_1 - V_0) - (V_0 - V_3)\}, \quad \left(\frac{\partial^2 V}{\partial y^2}\right)_0 = \frac{1}{h^2}\{(V_4 - V_0) - (V_0 - V_2)\}.
\]

![Fig. 8j. Plane network with square openings.](image)
Hence,

\[ (\nabla^2 V)_0 = \left( \frac{\partial^2 V}{\partial x^2} \right)_0 + \left( \frac{\partial^2 V}{\partial y^2} \right)_0 = \frac{1}{\hbar^2} (V_1 + V_2 + V_3 + V_4 - 4V_0), \]  

(2.27)

and the condition that

\[ (\nabla^2 V)_0 = 0 \]

leads to

\[ V_1 + V_2 + V_3 + V_4 = 4V_0. \]  

(2.28)

Using the points 5, 6, 7 and 8, we should find

\[ \left( \frac{\partial^2 V}{\partial x \partial y} \right)_0 = \frac{1}{4\hbar^2} (V_5 - V_6 + V_7 - V_8). \]

By this method, therefore, we can solve a partial differential equation containing first and second derivatives. If the first derivatives are discontinuous at the point \( M(x_0, y_0) \), we calculate them using points all on the same side of the point \( M: x - x_0 = h \) and \( 2h \), for example.

The calculation can be made more accurate by retaining higher degree terms in \( (x - x_0) \) and \( (y - y_0) \) in the expansion of \( V(x, y) \). It is then possible to show, taking advantage of the fact that \( \nabla(\nabla^2 V) = 0 \) and \( \nabla^2(\nabla^2 V) = 0 \), that the relation

\[ 4(V_1 + V_2 + V_3 + V_4) + (V_5 + V_6 + V_7 + V_8) = 20V_0 \]

is correct up to eighth order.

Relations of a similar kind involving the dielectric constant can be established for the surface of an insulator.

To solve a given problem in practice, we proceed as follows. The electrodes in question are drawn to a large scale (Fig. 8k), and a rectangular grid is ruled in such a way that the electrodes coincide with the

![Fig. 8k. Solution of a planar problem.](image-url)
lines of the mesh. (Sloping lines or curves will be represented by "steps" and the smaller the mesh-length, the greater the accuracy.)

The values of the potential at the outer points are inserted: over the electrodes, the potential is known and between the electrodes, it obeys a known law (along the segments $AB$ and $A'B'$ of Fig. 8k, for example). To all the other points of the grid arbitrary values of the potential are then ascribed, but these should be as reasonable and self-consistent as possible. (Thus on Fig. 8k, the values should fall between $V_1$ and $V_2$, and from electrode 1 to electrode 2 they should vary from $V_1$ to $V_2$.)

Using this first set of values, $\Phi^{(1)}_{ij}$, a second set, $\Phi^{(2)}_{ij}$, is calculated, using the foregoing formulae, from which a third set $\Phi^{(3)}_{ij}$ is obtained, and so on until $\Phi^{(n)}_{ij}$ and $\Phi^{(n-1)}_{ij}$ scarcely differ. The length of the calculation obviously depends upon the difference between the arbitrary distribution $\Phi^{(1)}_{ij}$ and the real solution; with a lot of practice, this difference can be reduced to as little as a few percent, and the solution is then obtained after a few "relaxations" to the degree of accuracy, $\Delta \Phi / \Phi$, selected in advance. After interpolation between the values found at the mesh-points, the equipotentials can be inserted.

Greater accuracy can also be obtained, over a restricted region, by repeating the calculation with a much smaller step length and using the values $\Phi^{(n)}_{ij}$ as boundary conditions.

These methods were originally employed without the assistance of electronic computers, and a great deal of mathematical work went into determining rules for increasing the speed of "convergence" towards the final solution.

*Rotationally symmetric systems.* Here, only the $r$ and $z$ coordinates are involved. The expansion of $\Phi(r, z)$ in the neighbourhood of $M(r_0, z_0)$ is given by (2.26), when $y$ and $x$ are replaced by $r$ and $z$ respectively.

We first consider a point $M$ such that $r_0 = ph$ (Fig. 8l). Laplace's

![Diagram](image)

**Fig. 8l.** Rotationally symmetric network. (a) A point 0 not on the axis; (b) a point 0 on the axis.
equation takes the form
\[ \frac{\partial^2 \Phi}{\partial z^2} + \frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r} \frac{\partial \Phi}{\partial r} = 0, \]
from which it is easy to derive the relation
\[ (\nabla^2 \Phi)_0 = \frac{1}{2 \rho h^2} \left\{ 2p(\Phi_1 - \Phi_0) + 2p(\Phi_3 - \Phi_0) + (2p - 1)(\Phi_2 - \Phi_0) + (2p + 1)(\Phi_4 - \Phi_0) \right\} \]

(2.29)

so that at \( M \),
\[ 8p \Phi_0 = 2p(\Phi_1 + \Phi_3) + (2p - 1)\Phi_2 + (2p + 1)\Phi_4. \]

(2.30)

When \( M \) lies on the axis, \( r = 0 \), this formula no longer applies. On symmetry grounds, however, \( (\partial \Phi / \partial r)_0 = 0 \), so that
\[ \frac{\partial \Phi}{\partial r} = 0 + r \left( \frac{\partial^2 \Phi}{\partial r^2} \right)_0 \]

and
\[ \Phi_4 = \Phi_0 + 0 + \frac{h^2}{2} \left( \frac{\partial^2 \Phi}{\partial r^2} \right)_0. \]

Hence
\[ \left( \frac{1}{r} \frac{\partial \Phi}{\partial r} \right)_0 = \left( \frac{\partial^2 \Phi}{\partial r^2} \right)_0 = \frac{2(\Phi_4 - \Phi_0)}{h^2}. \]

The relation
\[ (\nabla^2 \Phi)_0 = 0 \]
leads to
\[ 6\Phi_0 = \Phi_1 + \Phi_3 + 4\Phi_4. \]

(2.31)

Just as in the preceding case in the two variables \( x, y \), improved formulae may be derived.

2.1.4 Expression for the Potential and Fields in the Vicinity of a Symmetry Plane or an Axis of Symmetry

*Systems with translation symmetry, possessing a plane of symmetry \( Ox \).* The potential in the neighbourhood of \( Ox \) can be written in the form
\[ \Phi(x, y) = \sum_n A_n(x)y^{2n}. \]

The term \( A_0(x) \) denotes the potential distribution in the plane of symmetry, along \( Ox \):
\[ A_0(x) = \Phi_0(x). \]

Substituting this expansion into the Laplace equation and writing
\[ \nabla^2 \Phi(x, y) = 0, \] we obtain the recurrence relation
\[ A_{n-1}'' + 2n(2n-1)A_n = 0 \]
or
\[ A_n = -\frac{A_{n-1}''}{2n(2n-1)}. \]

This gives
\[ A_1 = -\frac{1}{2!} A_0'', \]
\[ A_2 = \frac{1}{4!} A_0^{(4)}, \]
\[ \cdots \]

Hence
\[ \Phi(x, y) = \sum_n \frac{(-1)^n}{(2n)!} y^{2n} \Phi_0^{(2n)}(x) \]
so that near Ox, we have
\[ \Phi(x, y) = \Phi_0(x) - \frac{y^2}{2} \Phi_0''(x) + \frac{y^4}{24} \Phi_0^{(4)}(x) - \cdots \quad (2.32) \]

With the aid of this expression, we can show that two equipotentials that intersect at Ox are at right angles to one another (Fig. 8m).

![Diagram](image)

Fig. 8m. The intersection of two equipotentials (a two-dimensional system): \( \alpha = 45^\circ \).

If two equipotentials intersect at \( x = x_0 \), then necessarily,
\[ \Phi_0'(x_0) = 0, \]
and hence
\[ \Phi(x, 0) = \Phi(x_0, 0) + \frac{1}{2} (x - x_0)^2 \Phi''(x_0, 0), \quad (2.33) \]
\[ \Phi''(x, 0) = \Phi''(x_0, 0). \]
Substituting into equation (2.32) and recalling that at all points along one of the equipotentials $\Phi(x, y) = \Phi(x_0, 0)$ we obtain
\[ y = \pm x - x_0. \]

The equipotentials intersect the axis again at an angle of $\pm 45^\circ$.

*Rotationally symmetric systems.* For symmetry reasons, the function $\Phi(r, z)$ takes the form
\[ \Phi(r, z) = \sum_n A_n(z) r^{2n} \]
and Laplace's equation imposes the condition
\[ A_n = -\frac{A_{n-1}''}{4n^2}, \]
so that
\[ \Phi(r, z) = \sum_n \frac{(-1)^n}{(n!)^2 \left(\frac{1}{2}\right)^n} \Phi_0^{(2n)}(z) \]
or
\[ \Phi(r, z) = \Phi_0(z) - \frac{r^2}{4} \Phi''_0(z) + \frac{r^4}{64} \Phi_0^{(4)}(z) - \cdots. \] (2.34)

Reasoning similar to that of the preceding sections shows that two equipotentials that intersect on the axis at $z = z_0$ are given by the equation
\[ r = \pm \sqrt{2}(z - z_0) \]
(to first order in $r$). They intersect the axis at an angle of $54^\circ 44'$.

The curvature of an equipotential in the neighbourhood of the axis (Fig. 8n) can be determined simply with the aid of equation (2.34). We have
\[ \Phi(r, z) = \Phi_0(z) - \frac{r^2}{4} \Phi''_0(z) \]
and a Taylor expansion gives
\[ \Phi'_0(z) = \Phi''_0(z_0), \]
\[ \Phi_0(z) = \Phi_0(z_0) + (z - z_0) \Phi'_0(z_0), \]
so that
\[ \Phi(r, z) = \Phi_0(z_0) + (z - z_0) \Phi'_0(z_0) - \frac{r^2}{4} \Phi''_0(z_0). \]

![Fig. 8n. The curvature of an equipotential near the axis.](image)
Since the points \( P(r, z) \) and \( A(0, z_0) \) lie on the same equipotential, we have
\[
\Phi(r, z) = \Phi_0(z_0)
\]
and so
\[
r^2 = 4(z - z_0) \frac{\Phi'_0(z_0)}{\Phi''_0(z_0)}.
\]
The radius of curvature \( R \) is thus given by
\[
r^2 = 2(z - z_0)R \quad \text{or} \quad R = \frac{2\Phi'_0(z_0)}{\Phi''_0(z_0)}.
\] (2.35)

Finally, relation (2.34) can be used to determine the shape of the electrode system needed to produce a given distribution \( \Phi_0(z) \). Plass (1942) has calculated lenses such that, on the axis, we have
\[
\Phi_0(z) = \Phi_0(1 - \frac{1}{2} e^{-2z/a})
\]
and
\[
\Phi_0(z) = \Phi_0\{1 - \frac{1}{2} \tanh(z/a)\}
\]
(a denotes the radius of the aperture in one of the electrodes).

2.2 THE ELECTROLYTIC TANK

2.2.1 The Underlying Principle

Consider a homogeneous isotropic conductor, conductivity \( \sigma \), carrying a current density \( i \); let the potential distribution be \( V \). The continuity equation gives
\[
\text{div} i = 0.
\] (2.36)

Combining Ohm's law
\[
i = \sigma E
\]
with
\[
E = - \text{grad} V,
\]
we find
\[
i = - \sigma \text{grad} V.
\]

On writing equation (2.36) in Cartesian coordinates:
\[
\frac{\partial}{\partial x} \left( \sigma \frac{\partial V}{\partial x} \right) + \frac{\partial}{\partial y} \left( \sigma \frac{\partial V}{\partial y} \right) + \frac{\partial}{\partial z} \left( \sigma \frac{\partial V}{\partial z} \right) = 0,
\] (2.37)
and substituting $\sigma = \text{constant}$, we obtain
\[
\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0.
\]
(2.38)

$V$ therefore satisfies the Laplace equation: there is an exact analogy between the distribution of electrostatic potential in a lens and the distribution of electrical potential in a homogeneous, isotropic conductor, provided that the boundary conditions are identical.

In electrostatics as in dynamic electricity, the boundary conditions between two media 1 and 2 relate the tangential components of the field $E$ and the normal components of the current density (either displacement or conduction current). Each pair is continuous across the boundary, thus:
\[
\left(\frac{\partial V}{\partial t}\right)_1 = \left(\frac{\partial V}{\partial t}\right)_2,
\]
(2.39)
\[
a_1\left(\frac{\partial V}{\partial n}\right)_1 = a_2\left(\frac{\partial V}{\partial n}\right)_2
\]
(2.40)

(in 2.40 $a$ represents dielectric constant and conductivity in electrostatics and current electricity respectively). The electrolytic tank is in fact a very general tool, since any phenomenon which can be described by an equation of the form (2.37) or (2.38), or by an equation which can be transformed into one of these forms, is amenable to analysis with its aid. The metal electrodes, the conductivity of which is supposed infinite, are immersed in an electrolyte with a very low conductivity, $\sigma_1$, and are then raised to potentials proportional to those at which they are to be held in the real experiment. A dielectric body is represented in the tank by a substance with conductivity $\sigma_2$, such that $\sigma_2/\sigma_1$ is equal to the relative dielectric constant of the body.

In most cases, the electrolyte is simply tap-water with a resistivity of about 2500 $\Omega$ cm; the impedances in the model are then of the order of thousands of ohms, and are thus comparable with those in the supply. The conductivity can be increased by adding a small amount of hydrochloric acid; it is also advantageous to add a few drops of some wetting agent (0.05 per cent Teepol, for example) to avoid the formation of a troublesome meniscus in the zones where the water is in contact with the electrodes or the walls of the tank.

For quick measurements, the electrodes are made of sheet-copper or brass, which can easily be bent into the required shape. For more precise results, however, it is necessary to use massive electrodes, as thin sheets are too easily deformed; the electrode surfaces must be cleaned with extreme care as surface tension would otherwise be liable to produce local distortion of the surface of the electrolyte. The cleaning may be performed
electrolytically—a steady current of the order of 10 A cm$^{-2}$ is passed, with the electrode acting as the anode. The electrodes must be positioned with the utmost precision.

2.2.2 The Various Types of Tank

Practical electron systems fall into one of two categories, those possessing axial symmetry, and those with planar or translational symmetry.

Axially symmetric systems

A priori it would seem necessary to build a model of the whole system which is to be studied, but in fact a much simpler model suffices, as we can demonstrate very simply—if an insulated plate is placed in the tank, the equipotential surfaces are all normal to it because the normal component of the electric field is zero; in the presence of such a plate, therefore, the potential distribution is identical to the distribution which would be produced by a combination of the original electrodes and their mirror images with respect to the plate. The surface of the liquid is a particular case of this effect, and by arranging that the axis of the system lies in the surface, only one half of the system, bisected along a meridian plane, needs to be constructed.

It is, however, possible to simplify matters still further, since the liquid in a tank with an insulated sloping bottom forms a wedge whose edge represents the axis of symmetry of the system. The model electrodes are reduced to a small fraction of their originals, and provided the wedge angle is small (5 to 15°), cylinders, with the meridian of the system under consideration as base, can be used (see Fig. 9a).

![Diagram](image.png)

**Fig. 9a.** An example of the actual form of the model electrodes in a tank with a sloping bottom.
Unfortunately, owing to the shallowness of the electrolyte and the locally disturbing effects of the meniscus, measurements are least easily made in the most interesting region, close to the axis. To obtain the potential actually on the axis, however, we can use an extrapolation technique. This method is particularly tempting, in any case, as the model electrodes are of so simple a shape.

Introducing insulated walls into the tank, we can make use of such further symmetry as the actual system may have, to make one further simplification of the model—the symmetrical unipotential lens is such a case; only the quadrant \( BAA' \) need be built, the half of the central electrode being attached to an insulating sheet which is normal to the symmetry axis (see Fig. 9a).

*Systems with translation symmetry*

To represent the electrodes of the slit lenses which are used in mass-spectrometry, we have only to place portions of right cylinders on the bottom of a tank with a flat horizontal bottom; the corresponding electrodes would be infinitely long in the vertical direction.

### 2.2.3 The Measurement of Potential and Field

In order to obviate polarization phenomena, it is better in fact to supply the electrodes with a low frequency sinusoidal current, between 50 c/s and one kc/s, from an oscillator of some tens of watts, which provides a potential of a few tens of volts. The potentials between the electrodes should not exceed a few volts so that strong currents which would decompose the electrolyte are avoided.

The potential \( \Phi_s \) at any point of the liquid surface is measured by inserting a fine metal probe (a platinum wire of 0·05 mm or 0·1 mm diameter) at the point, and using a null method to compare its potential with that provided by an accurate potentiometer across the terminals of the oscillator. By annulling the difference between \( \Phi_s \) and \( \Phi_A \) (the potentiometer potential), there is no risk of the introduction of a probe disturbing the potential in the electrolyte. The most sensitive way of detecting the null point is by applying the oscillator output to the horizontal deflector plates of a cathode ray oscillograph and a signal proportional to \( (\Phi_s - \Phi_A) \) to the vertical plates; this signal is obtained by amplifying \( (\Phi_s - \Phi_A) \) with the aid of a selective amplifier which eliminates any parasitic potentials which have frequencies different from that of the useful signal. In theory, the Lissajous figure which appears on the screen is an inclined straight line, and to obtain \( \Phi_s = \Phi_A \), either the probe or the slider of the potentiometer is adjusted until the line is horizontal.

In practice, stray interelectrode capacities and the stray capacities which are produced by the meniscus wherever there is a liquid–solid contact
transform the straight line into an ellipse. In equilibrium, we should obtain an ellipse with its major axis horizontal; the minor axis must be reduced to the minimum to increase the accuracy with which the zero can be detected, and in particular, the precautions mentioned above must be taken in the cleaning. Alternatively, a variable capacitance can be introduced into one of the other arms of the measuring bridge.

The electrodes are supplied with the aid of two potentiometers which are placed in parallel across the main generator. Their potential is adjusted while they are connected to the input of the amplifier attached to the probe; as a result of the currents which are drawn, however, the potentials

![Diagram](image)

**Fig. 9b.** An electrolytic tank with its accessories.


*L:* A.C. source.

![Photograph](image)

**Fig. 9c.** Photograph of the tank and its accessories.
of the electrodes are not independent of one another—we are forced to proceed by a series of successive adjustments. In addition to an amplitude control, it is useful to provide a phase control, which entails inserting either a variable resistance or a phase-change circuit in series into the electrode supply lead.

If we select values of $\Phi_A$ between 0 and 100, say (in steps of 10, for example), we can follow equipotentials $\Phi_S = \Phi_A$ with the probe. With the aid of a pantograph coupled to the system which moves the probe, a direct point by point plot of these equipotentials can be made on a drawing board. Great care is necessary in the construction of the mechanical parts if recording errors are to be avoided.

The tank provides, in this way, a map of the potential distribution $\Phi$, in the form of equipotentials numbered from 0 to 100, numbers which represent the two extremes of potential which are applied to the system. In Fig. 9b, the arrangement is shown schematically, while Fig. 9c is a photograph of a tank actually in use.

Since the early work of Fortescue and Farnsworth (1913) and Peres and Malavard (1938), the apparatus has undergone numerous improvements. Ingenious devices for ensuring that the recording-pen or marker follows the probe faithfully have been introduced: for example, the pen may be fixed directly above the probe, where it traces out the equipotentials on a sheet of paper which rests against a glass plate parallel to the surface of the electrolyte.

In the best tanks, $\Phi$ can be determined to within a few parts in $10^4$. Devices have been built to trace the equipotentials automatically, by mounting the probe on a trolley which is driven parallel to one side of the tank by means of a motor $M_1$, and perpendicular to this direction by a second motor $M_2$; this latter stops, however, immediately $(\Phi_S - \Phi_A)$ becomes zero. Given $\Phi_A$, therefore, the probe is forced to follow the equipotential $\Phi = \Phi_A$. An installation of this type is fully described in an article by Picquendar and Cahen (1960).

Direct measurement of the field

One could calculate the field at any point, given the distribution of potential $\Phi$, from the relation $E = - \text{grad} \, \Phi$. Alternatively, we can conceive of a method of measuring the field electrically, by using two separate probes maintained a known distance apart ($\Delta S$), and measuring the potential difference $\Delta \Phi$ between their tips, after amplification. The field would then be given by

$$E = \frac{\Delta \Phi}{\Delta S}.$$
The presence of the meniscus in the liquid between the two probes makes such a measurement extremely delicate, however. If accurate and reproducible results are to be obtained, special probes must be employed, consisting of rigid wires embedded in glass, only the tips of which are exposed; the meniscus is then formed around the insulated parts. The electronic equipment is here somewhat more complicated than that which is required for the measurement of potential, as the two extremities of the signal (ΔΦ) which is to be amplified are “floating”. A great deal of information about devices which have been designed specially for field measurement is to be found in the articles by Hollway (1955) and Sanders and Yates (1953, 1956). The two latter authors have made a thorough examination of the influence of the meniscus, and have been able to eliminate it completely by constructing a tank to a very special design; the pair of probes is fixed in the bottom of the tank, and the model is displaced mechanically with respect to the double probe. In this way, an accuracy of a few parts in 10^3 can be reached.

2.2.4 Space Charge

Several different methods are at present available for determining the potential distribution of a beam in the presence of space charge. These consist in solving an equation of the form

\[ \nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} = -K \frac{i}{\sqrt{\Phi(x, y)}} \]

in the region through which the beam passes (the system considered is two-dimensional). The principle of the earliest solution, suggested by Musson-Genon (1947), is as follows.

The local conductivity (σ) in the tank is proportional to the depth h(x, y): σ = a h(x, y). The relation div σE = 0 thus takes the form

\[ -\nabla^2 \Phi = \text{div } E = -E \cdot \frac{\text{grad } h}{h} = -E \cdot \text{grad } (\log h). \]

We obtain an analogy between the two equations, and hence between \( \Phi(x, y) \) in the beam and \( V(x, y) \) in the tank, if we write

\[ E \cdot \text{grad } (\log h) = -K \frac{i}{\sqrt{\Phi}}. \]

In a tank the bottom of which is shaped in such a way that h varies and this relation is satisfied (see Fig. 10a), the potential distribution across the surface of the electrolyte (h = 0) will provide a chart of the potential \( \Phi(x, y) \).
Fig. 10a. Cross-section of the tank with a shaped bottom, as used by Musson-Genon.

Fig. 10b. The simulation of space charge in the electrolytic tank (after van Duzen and Brewer, 1959): the probes for injecting currents. (a) A single probe, and the corresponding cylinder of charge. $A$: water surface; $B$: cylinder of charge; $C$: probe; $D$: tank floor. (b) The arrangement of the probes, seen in plan. (c) Wire probe current source—details. $A$: surface of the electrolyte; $B$: tank floor; $C$: conductor; $D$: insulating varnish; $E$: O-ring water-seal; $F$: electrical connection.
The equation for \( h \), however, contains functions which are \textit{a priori} unknown, namely \( \Phi \) and \( E = -\nabla \Phi \). We have, therefore, to proceed by successive approximations and a tank of constant depth is first of all used to solve the equation without a space charge term, \( \nabla^2 \Phi = 0 \). This gives the distributions \( \Phi_0(x, y) \) and \( E_0(x, y) \) with which a first approximation to the expression for the variation of \( h \) is obtained:

\[
h_1(x, y) = h_0 \exp \left[ -K \int_{s_0}^{s} \frac{i}{E_0 \sqrt{\Phi_0}} \, ds \right].
\]

A bottom of this shape is then built, \( \Phi \) and \( E \) again measured, and second approximations \( \Phi_1(x, y) \) and \( E_1(x, y) \) obtained. A more precise expression for the depth function, \( h_2(x, y) \), can then be calculated; this procedure is repeated until a self-consistent solution is found, which requires no further modification to be made to the bottom of the tank.

In practice, however, this is a long process, and a solution which is simpler is nowadays employed (Hollway, 1955). The space charge in the beam is simulated in the tank by introducing a current density with the same volume distribution. This current is, in fact, led along a finite number of very fine electrodes, each of which simulates a cylinder of charge which is homogeneous in density. The juxtaposition of a number of such cylinders must create the same field distribution at the surface of the electrolyte in the tank as the distribution which prevails across a plane cross-section of the real beam.

Hollway, and later Brewer (1957) and van Duzen and Brewer (1959) have studied the problem of choosing the shape and size of electrodes which best simulate reality, both for flat-bottomed tanks and tanks with sloping bottoms. Each electrode should consist of a thin insulated rod which passes through the insulating bottom of the tank in a direction normal to the liquid surface; the end of the rod is stripped of insulation over a length \( l \), and lies at a depth \( a \) beneath the surface of the liquid. If the depth of the liquid is \( h \) at this point (see Fig. 10b), and if we obey the rule

\[
a + l = \frac{h}{2},
\]

experiment shows that this localized charge is practically equivalent to a cylinder of current, its axis along the wire, its height \( h \) and its radius \( a \). We have, thus, simply to provide the bottom of the tank with a regular array of electrodes, distance \( a \) apart, and to supply them with the intensities which are appropriate to the zone through which the beam passes and which will be proportional to the charge density in the real beam. Here again, a series of successive approximations is necessary, but we pass from one to the next simply by adjusting the intensities of the various currents.
2.3 THE RESISTANCE NETWORK

In many laboratories, the elaborate installation represented by a tank is replaced with advantage by a resistance network, comparable in accuracy and simpler to use.

The possibility of using such networks to solve partial differential equations had been suggested in 1943 by Hogan, but it was not until considerably later (de Packh, 1947; Hennequin, 1948; Charles 1949; Liebmann, 1949, 1950a) that the idea of actually using them to solve the Laplace equation—or rather, the corresponding finite difference equation—was exploited. Provided the mesh-length of the network is short enough, the solution which is obtained can be regarded as very close to the theoretical solution.

Fig. 11. Resistance networks. (a) Calculation of $\Phi$ at a particular point $M$. (b) The values of the resistances in a network designed for axially symmetrical systems (after Francken, 1959/60).
We shall consider first the case in which the system possesses translational symmetry. At any point $M$, the electrostatic potential satisfies the equation

$$\left(\nabla^2 \Phi\right)_M = \left(\frac{\partial^2 \Phi}{\partial x^2}\right)_M + \left(\frac{\partial^2 \Phi}{\partial y^2}\right)_M = 0.$$  \hspace{1cm} (2.41)

In the plane $Oxy$ of the system, we draw an imaginary network of mesh-length $h$, which consists of two sets of straight lines, parallel to $Ox$ and $Oy$ respectively; $M$ is situated at a point of intersection or node of the network. We denote by $P$, $Q$, $R$, $S$ the four nodes which surround $M$ (see Fig. 11a). Given the potentials at these points, we can calculate the second derivatives which appear in equation (2.41), see §2.2.3, and we obtain

$$\left(\nabla^2 \Phi\right)_M = \frac{1}{h^2} \left[ (\Phi_p - \Phi_M) + (\Phi_Q - \Phi_M) + (\Phi_R - \Phi_M) + (\Phi_S - \Phi_M) \right] = 0.$$ \hspace{1cm} (2.42)

We construct a resistance network with the same interval $h$ as the fictitious network above. Each point $M$ is linked to the neighbouring nodes by four resistances $R_1$, $R_2$, $R_3$ and $R_4$. Applying Kirchhoff’s law to the node $M$, we have

$$\sum i_M = 0$$

or

$$\frac{V_p - V_M}{R_1} + \frac{V_Q - V_M}{R_2} + \frac{V_R - V_M}{R_3} + \frac{V_S - V_M}{R_4} = 0.$$ \hspace{1cm} (2.43)

Equation (2.43)—as regards the value of the potential $V(x, y)$ at the point $M$ in the resistance network—is analogous to (2.42)—as regards the electrostatic potential $\Phi(x, y)$—if we set

$$R_1 = R_2 = R_3 = R_4.$$  

All the resistances in the network are, for this case, equal. Electrodes of infinite conductivity are represented by short-circuiting the nodes of the network which lie along the curve which corresponds to a magnified version of the electrostatic system, and applying potentials proportional to those which are actually present in this system.

This same technique can equally well be used to study systems with axial symmetry, provided a network in which the resistances vary between the axis and the periphery is used; the network corresponds to one of the half-planes of radial symmetry of the system.

The Laplace equation now has the form:

$$\frac{\partial^2 \Phi}{\partial r^2} + \frac{1}{r} \frac{\partial \Phi}{\partial r} + \frac{\partial^2 \Phi}{\partial z^2} = 0,$$ \hspace{1cm} (2.44)
and the corresponding finite difference equation for the point $M$ is (see § 2.2.3)

$$(\nabla^2 \Phi)_M = \frac{1}{2ph^2} \left\{ 2p(\Phi_p - \Phi_M) + 2p(\Phi_q - \Phi_M) + (2p - 1)(\Phi_R - \Phi_M) + (2p + 1)(\Phi_S - \Phi_M) \right\} = 0 \quad (2.45)$$

when $M$ is at a distance $r_M = ph$ from the axis and

$$(\nabla^2 \Phi)_M = \frac{1}{h^2} \left\{ (\Phi_p - \Phi_M) + (\Phi_q - \Phi_M) + 4(\Phi_S - \Phi_M) \right\} = 0 \quad (2.46)$$

if $M$ is located on the axis.

In the corresponding resistance network, we have the relation

$$\frac{V_p - V_M}{R_1} + \frac{V_q - V_M}{R_2} + \frac{V_R - V_M}{R_3} + \frac{V_S - V_M}{R_4} = 0 \quad (2.47)$$

when $M$ is not on the axis. This relation will be analogous to (2.45) if the resistances are given the following values:

$$R_1 = R_2 = R_0/2p, \quad R_3 = R_0/(2p - 1), \quad R_4 = R_0/(2p + 1). \quad (2.48)$$

For $p = 1$ (so that $M$ is a distance $r_M = h$ from the axis), we shall have $R_3 = R_0$. The resistances that are connected at one end to the axis are all equal to $R_0$, the value of which is chosen arbitrarily.

When $M$ is on the axis, Kirchhoff’s law gives us

$$\frac{V_p - V_M}{R_1} + \frac{V_q - V_M}{R_2} + \frac{V_S - V_M}{R_4} = 0, \quad (2.49)$$

and this is the analogue of (2.46) provided we set

$$R_1 = R_2 = 4R_4 = R_0. \quad (2.50)$$

In conclusion, therefore, we have

$$R(Oz) = 4R_0$$

on the axis. Along rows parallel to Oz,

$$R_{p=1} = R_0/2, \quad R_{p=2} = R_0/4, \ldots$$

and along columns parallel to Or (see Fig. 11(b))

$$R_{0,1} = R_0, \quad R_{1,2} = R_0/3, \quad R_{2,3} = R_0/5, \quad \ldots$$
The network can be constructed from a sheet of insulating bakelite, punched at regular intervals of length $h$. Each node consists of a conducting rod attached to the board, and the resistances are soldered to these rods on one side of the board; the connexions which represent electrodes, and the measurements are made on the other side.

An alternating supply may be used, although if the resistances are of high precision, a direct current supply to the network is preferable to avoid self-induction effects. With resistances of ordinary quality which have been chosen to be equal to the calculated values to within 1 per cent, the accuracy which can be attained in the measurements is considerably better than 1 per cent (Liebmann, 1950a). If manganin resistances are used, chosen to be accurate to within two parts in $10^3$, an accuracy of one part in $10^4$ is to be hoped for, or even better with a larger network. Any errors will then originate in the fact that electrodes which have in fact a curved meridional cross-section must be represented by a series of straight lines in the network; the interval $h$ must in consequence be small in comparison with the dimensions of the system. The accuracy of the measurements can also be increased by “enlarging” a region of special interest. The complete system is first represented on a small scale on the network, and subsequently the interesting region is produced on a larger scale; the first experiment suggests the appropriate potentials which should be applied at the boundaries.
CHAPTER 3

THE FIELD DISTRIBUTION
IN MAGNETIC LENSES

3.1 DIRECT CALCULATION OF THE MAGNETIC FIELD

In the various systems that are used in electron optics, the magnetic fields can be produced in a number of different ways.

We may use current-carrying coils, with or without a high permeability magnetic circuit; the latter enables us to concentrate the magnetic flux within a limited region, and to obtain higher fields with a given coil. Unfortunately, these circuits, made of some ferromagnetic material, saturate when the magnetic induction \( B \) reaches values from 20 to 22 kG (2 to 2.2 tesla) for soft iron, 22 to 23 kG for pure iron and 23 to 25 kG for certain cobalt steels. Unshielded coils can be used to obtain fields considerably higher than these values but the electrical power dissipated in the windings then becomes prohibitive and in practice only pulsed operation is feasible. The recent discovery of superconducting materials with high critical fields has, however, brought iron-free coils back into favour, for with them steady fields greater than 100 kG can be achieved (see Chapter 9).

Far rarer is the use of permanent magnets, in conjunction with circuits of some ferromagnetic material. The limitations are the same as those of ordinary electromagnets. Furthermore, the field variations in the useful zone can only be of limited amplitude; they are produced by moving a soft iron shunt in the neighbourhood of the magnetized blocks (see Chapter 9).

3.1.1 General Equations

It is extremely difficult to calculate the magnetic induction everywhere in a system containing magnetic materials. In electron optics, this problem need only be solved \textit{in vacuo} (or in air), that is to say, in a medium of magnetic permeability \( \mu \) equal to \( \mu_0 = 4\pi \times 10^{-7} \) (R.M.K.S. units), in which the induction \( B \) and the magnetic field \( H \) are connected by the simple relation

\[
B = \mu_0 H.
\]  

(3.A)
At a point in a region in which no currents are flowing, $B$ and $H$ satisfy the relations
\begin{align*}
\text{div } B &= 0. \quad (3.\, B) \\
\text{curl } H &= 0. \quad (3.\, C)
\end{align*}

The latter can also be written
\begin{equation} \quad \text{curl } B = 0. \quad (3.\, C') 
\end{equation}

Within a conductor carrying current density $i$, we should have
\begin{align*}
\text{curl } H &= i, \quad (3.\, D) \\
\text{curl } B &= \mu_0 i. \quad (3.\, D')
\end{align*}

By analogy with electrostatics, we may assume that, if $i = 0$ in the region in question, $B$ (or $H$) is derived from a function $\Phi_m$ called the "scalar magnetic potential".

\begin{equation} \quad B = -\text{grad } \Phi_m. \quad (3.\, E)
\end{equation}

and this potential $\Phi_m$ satisfies Laplace's equation

\begin{equation} \quad \nabla^2 \Phi_m = 0. \quad (3.\, F)
\end{equation}

All the methods described in the preceding chapter of solving equation (3.\, F) will therefore be applicable, but now we must always take the boundary conditions into account. In particular, the surfaces of ferromagnetic substances of very high permeability can be considered, with the approximation $\mu_r = \infty$, as equipotential surfaces. The usual shapes of the pole-pieces are different from those of the electrodes encountered in electrostatics, however, and new problems arise.

In addition, a magnetic "vector potential" function is defined at every point by the relations

\begin{align*}
B &= \text{curl } A, \\
\text{div } A &= 0. \quad (3.\, G)
\end{align*}

At points where $i = 0$, we shall have

\begin{equation} \quad \nabla^2 A = 0.
\end{equation}

Over an arbitrary closed curve $(C)$, the circulation of $A$, given by $\oint_C A \cdot ds$, is equal to the flux $\Psi$ of the vector $B$ through any surface $\Sigma$ bounded by this curve. Thus

\begin{equation} \quad \oint_C A \cdot ds = \iint_{\Sigma} \text{curl } A \cdot ds = \iint_{\Sigma} B \cdot ds = \Psi_{(\Sigma)}. \quad (3.\, H)
\end{equation}

The vector potential is particularly useful in calculating the fields of iron-free coils.
3.1.2 The Special Case of a System Containing no Magnetic Material (Iron-free Coils)

If $B$ is produced by a number of conductors of volume $V$, through which currents of density $i$ flow, $A$ is given at a point $M(x,y,z)$ by

$$A = \frac{\mu_0}{4\pi} \int \int \int \frac{i \, dV}{r},$$

(3.1)

in which $r$ denotes the distance of the point $M$ from the volume element $dV$. For a filamental conductor of cross-section $S$, we have (Fig. 12a)

$$i \, dV = |i| \, S \, ds = J \, ds$$

so that

$$A = \frac{\mu_0 J}{4\pi} \int \frac{ds}{r},$$

(3.J)

This gives

$$B = \frac{\mu_0 J}{4\pi} \text{curl} \left( \int \frac{ds}{r} \right).$$

Fig. 12a. Field of a conducting wire.

For every point $M$, the integration must be performed along the whole length of the conductor. From the relation

$$\text{curl} \left( \frac{ds}{r} \right) = \frac{1}{r} \text{curl} \, s + \left( \text{grad} \frac{1}{r} \right) \times ds$$

and using the fact that $s$ does not depend upon the position of $M$, we obtain

$$\text{curl} \left( \frac{ds}{r} \right) = \left( \text{grad} \frac{1}{r} \right) \times ds = -\frac{r \times ds}{r^3},$$

with $r = MP$. Using instead the convention $r = PM$, we have

$$B = \frac{\mu_0 J}{4\pi} \int \frac{r \times ds}{r^3}.$$

(3.K)

This expression is the law of Biot and Savart. To calculate the com-
ponents of $B$, we can use (3.K) directly or, alternatively, we may use (3.J) together with (3.G).

In Cartesian coordinates, we have

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z},$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x},$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y},$$

and in cylindrical polars,

$$B_r = \frac{1}{r} \frac{\partial A_z}{\partial \theta} - \frac{\partial A_\theta}{\partial z},$$

$$B_\theta = \frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r},$$

$$B_z = \frac{1}{r} \left( \frac{\partial (r A_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta} \right).$$

(3.L')

In the two simple cases which follow, these relations simplify considerably.

(a) **Straight coils, infinitely long in the Oz direction (translation symmetry).** The only non-zero component of $i$ is $i_z$, and we have

$$A_x = A_y = 0, \quad A = A_z,$$

$$B_z = 0, \quad B_x = \frac{\partial A_z}{\partial y}, \quad B_y = -\frac{\partial A_z}{\partial x}.$$  

(3.M)

(b) **Circular coils having a common axis Oz (rotational symmetry).**

We now have $i = i_\theta$, so that

$$A_r = A_z = 0, \quad A = A_\theta,$$

and

$$B_\theta = 0, \quad B_r = \frac{\partial A_\theta}{\partial z}, \quad B_z = \frac{1}{r} \frac{\partial (r A_\theta)}{\partial r}.$$  

(3.N)

In the case of a closed current loop (C) carrying current $J$, we can show that the scalar magnetic potential $\Phi_m$ at a point $M$ is given (Fig. 12b) by

$$\Phi_m = \frac{\mu_0 J}{4\pi} \Omega.$$  

(3.O)

in which $\Omega$ denotes the solid angle subtended at $M$ by the curve (C). We then find

$$B = -\frac{\mu_0 J}{4\pi} \text{grad} \Omega.$$  

(3.P)
EXAMPLE 1. The field produced by a single circular turn of negligible cross-section and radius \( a \) (Fig. 12c).

On symmetry grounds, \( B_r = 0 \). Let \( z \) be the abscissa of \( M \) and \( z_0 \) that of the turn \( (S) \); \( x = a \cos \theta \) and \( y = a \sin \theta \) are the coordinates of a point \( P \) on the turn. The projections of \( \overrightarrow{PM} \) on the axes are equal to \( x, y \) and \( (z - z_0) \).

Thus

\[
dx = -a \sin \theta \, d\theta, \quad dy = a \cos \theta \, d\theta,
\]

and (3.K) gives

\[
B_z = \frac{\mu_0 J}{4\pi} \int \frac{x \, dy - y \, dx}{r^3} = \frac{\mu_0 J}{4\pi} \int \frac{a^2 \, d\theta}{r^3} = \frac{\mu_0 Ja^2}{2r^3}
\]

and

\[
B_z = \frac{\mu_0 Ja^2}{2\left((z - z_0)^2 + a^2\right)^{3/2}} \quad \text{(3.Q)}
\]

With the aid of (3.Q), we can calculate the magnetic induction \( B_z \) on the axis of a thin solenoid of radius \( a \) and length \( 2l \) centred on \( z = 0 \); by considering the induction \( dB_z \) produced by an element of the solenoid
of length \( d\zeta \) situated at \( z = \zeta \) and containing \( n \) identical turns per unit length, and integrating from \(-l\) to \(+l\), we find

\[
B_z = \frac{\mu_0 n J a^2}{2} \int_{-l}^{l} \frac{d\zeta}{\{(z - \zeta)^2 + a^2\}^{3/2}}.
\]

Setting \((z - \zeta)/a = \cot \phi\), we obtain

\[
B_z = \frac{\mu_0 n J}{2} (\cos \phi_1 - \cos \phi_2)
\]

or

\[
B_z = \frac{\mu_0 l}{4l} \left[ \frac{z + l}{\{a^2 + (z + l)^2\}^{1/2}} - \frac{z - l}{\{a^2 + (z - l)^2\}^{1/2}} \right],
\]

in which

\[
I = 2\ln J.
\]

If \( I \) tends to infinity, we have \( \phi_1 = 0 \) and \( \phi_2 = \pi \), and so

\[
B_z = \mu_0 n J.
\]

For a coil of appreciable thickness, we must integrate in the two directions \( Oz \) and \( Or \). From (3.R) we know the contribution from a layer of thickness \( dr \), radius \( r \) and length \( 2l \), carrying a total current

\[
I = 2\ln J = 2l dr \cdot i.
\]

\((i\) is the current density and is independent of \( z \).)

Writing (see Fig. 12d)

\[
\phi_1 = \alpha(r), \quad \phi_2 = \beta(r),
\]

we integrate (3.R) from \( r_1 \) to \( r_2 \),

\[
B_z = \frac{\mu_0 l}{2} \int_{r_1}^{r_2} \{\cos \alpha(r) - \cos \beta(r)\} \, dr.
\]

**Fig. 12d.** A thick solenoid.
We have
\[ \tan \alpha = \frac{r}{z+l}, \quad \tan \beta = \frac{r}{z-l}, \]
\[ \frac{d\alpha}{\cos^2 \alpha} = \frac{dr}{z+l}, \quad \frac{d\beta}{\cos^2 \beta} = \frac{dr}{z-l}, \]
so that
\[ B_z = \frac{\mu_0 i}{2} \left[ (z+l) \int_{\alpha_1}^{\alpha_2} \frac{d\alpha}{\cos \alpha} - (z-l) \int_{\beta_1}^{\beta_2} \frac{d\beta}{\cos \beta} \right] \]
or
\[ B_z = \frac{\mu_0 i}{2} \left[ (z+l) \log \frac{\tan \left( \frac{\alpha_2 + \pi}{4} \right)}{\tan \left( \frac{\alpha_1 + \pi}{4} \right)} - (z-l) \log \frac{\tan \left( \frac{\beta_2 + \pi}{4} \right)}{\tan \left( \frac{\beta_1 + \pi}{4} \right)} \right]. \tag{3.S} \]
If \( I \) denotes the total current in the thick solenoid, we have
\[ i = \frac{I}{2l(r_2-r_1)}. \tag{3.S'} \]

Knowing the field distributions along the axis \( OZ \), we shall see later that it is possible to calculate \( B_r \) and \( B_z \) in the neighbourhood of the axis.

**REMARKS.** (i) We may calculate \( B_z \) on the axis \( OZ \), starting from equations (3.O) and (3.P). For a single turn, we should have
\[ \Omega = 2\pi (1 - \cos \phi) \]
(see Fig. 12c), and so
\[ \Phi_m = \frac{\mu_0 J}{2} (1 - \cos \phi) = \frac{\mu_0 J}{2} \left\{ 1 - \frac{z-z_0}{\sqrt{a^2+(z-z_0)^2}} \right\}. \]
Hence
\[ B_z = -\frac{\partial \Phi_m}{\partial z} = -\frac{\mu_0 J}{2} \left\{ -\frac{1}{r} + \frac{(z-z_0)^2}{r^3} \right\} = \frac{\mu_0 I a^2}{r^3} \]
with
\[ r = \sqrt{a^2+(z-z_0)^2}. \]
Alternatively,
\[ B_z = \frac{\mu_0 J}{2a} \sin^2 \phi. \]
For a thin solenoid,
\[ B_z = \frac{\mu_0 n J}{2a} \int_{\phi_1}^{\phi_2} \sin^3 \phi \, d\phi \]
and we recover the formula
\[ B_z = \frac{\mu_0 n j}{2} (\cos \phi_1 - \cos \phi_2). \]

(ii) If the field created throughout the whole space is required, it is preferable first to calculate \( A_\theta (r, z) \) (cf. Durand, 1952) and to deduce \( B \) from it. The expression for \( A_\theta \) (and hence, those for \( B_z \) and \( B_r \)) contains the elliptic integrals of Legendre; these are well tabulated.

**Example 2.** The expression for the field produced at a point \( M(x, y) \) by a straight wire of cross-section \( dS = d\xi \, d\eta \), carrying current \( I = j d\xi \, d\eta \) (Fig. 12e(a)), is of the form
\[
\begin{align*}
\frac{dB_x}{2\pi} & = -\frac{\mu_0 j}{2\pi} \frac{y - \eta}{(x - \xi)^2 + (y - \eta)^2} \, d\xi \, d\eta, \\
\frac{dB_y}{2\pi} & = +\frac{\mu_0 j}{2\pi} \frac{x - \xi}{(x - \xi)^2 + (y - \eta)^2} \, d\xi \, d\eta.
\end{align*}
\]

From this we can calculate the magnetic induction produced by a narrow band of thickness \( d\xi \) and width \( 2b \) (Fig. 12e(b)):
\[
\begin{align*}
\Delta B_x & = \frac{\mu_0 j}{2\pi} \frac{d\xi}{d\eta} \log \frac{(x - \xi)^2 + (y - b)^2}{(x - \xi)^2 + (y + b)^2}, \\
\Delta B_y & = -\frac{\mu_0 j}{2\pi} \frac{d\xi}{d\eta} \left( \tan^{-1} \frac{x - \xi}{y + b} - \tan^{-1} \frac{x - \xi}{y - b} \right),
\end{align*}
\]
or
\[
\begin{align*}
\Delta B_x & = \frac{\mu_0 I}{4\pi b} \log \frac{R_2}{R_1}, \\
\Delta B_y & = \frac{\mu_0 I}{4\pi b} (\alpha_2 - \alpha_1),
\end{align*}
\]
in which \( I = 2bj \, d\xi \); \( R_1, R_2, \alpha_1 \) and \( \alpha_2 \) are defined in Fig. 12e(b).

![Fig. 12e. A bar of rectangular cross-section. (a) A wire of square cross-section; (b) a thin ribbon; (c) a thick bar.](image-url)
Finally, integrating with respect to $x$, we obtain the induction produced by a bar of rectangular cross-section $2b \times 2a$ (Fig. 12e(c)), carrying a current $I = 4abj$:

\[
B_y = \frac{\mu_0 I}{8\pi ab} \left\{ (x+a) \log \frac{R_2}{R_3} + (x-a) \log \frac{R_4}{R_1} \\
+ (y+b)(\alpha_3 - \alpha_4) + (y-b)(\alpha_1 - \alpha_2) \right\},
\]

\[
B_y = \frac{\mu_0 I}{8\pi ab} \left\{ (y+b) \log \frac{R_3}{R_4} + (y-b) \log \frac{R_1}{R_2} \\
+ (x+a)(\alpha_3 - \alpha_2) + (x-a)(\alpha_1 - \alpha_4) \right\}.
\]

With four thin sheets or four rectangular bars (Septier, 1960), we can produce field distributions with quadrupolar symmetry, having a transverse field gradient $G = \partial B/\partial r$ that is virtually constant in all directions; iron-free quadrupole lenses can thus be designed in this simple way (see Chapter 10).

**Example 3.** The use of complex functions, and the creation of a multipole field in a system with translation symmetry (Beth, 1966). From the relations

\[
B_x = -\frac{\partial \Phi_m}{\partial x} = \frac{\partial A_z}{\partial y},
\]

\[
B_y = -\frac{\partial \Phi_m}{\partial y} = -\frac{\partial A_z}{\partial x}
\]

we have

\[
\nabla^2 \Phi_m = 0 \quad \text{and} \quad \nabla^2 A_z = 0,
\]

so that the function

\[
W(z) = -(A + j\Phi_m)
\]

is an analytic function in the complex variable \(z = x + jy\).

In consequence, the "complex potential" \(W(z)\) may be written in the form

\[
W(z) = -(A + j\Phi_m) = W_0 + W_1z + W_2z^2 + \cdots W_nz^n.
\]

We then have

\[
B(z) = \frac{dW}{dz}
\]

or

\[
B(z) = W_1 + 2W_2z + 3W_3z^2 + \cdots nW_nz^{n-1}.
\]
FIELD DISTRIBUTION IN MAGNETIC LENSES

Successive terms in this expression correspond to constant fields, dipole fields, quadrupole fields and so on. Thus for \( n = 2 \), for example,

\[
W(z) = z^2 = r^2 \exp 2j\theta = u(x, y) + jv(x, y),
\]

where

\[
\begin{align*}
    u &= x^2 - y^2 = r^2 \cos 2\theta, \\
v &= 2xy = r^2 \sin 2\theta.
\end{align*}
\]

The function \( u(x, y) \) represents \( A_2(x, y) \) and \( v(x, y) \) gives \( B(x, y) \).

For an infinitely long wire carrying current \( I \) situated at \( z_0 = x_0 + jy_0 \), we should have

\[
W(z) = -(A + j\Phi_m) = \frac{\mu_0 I}{2\pi} \log(z - z_0),
\]

\[
B(z) = B_y + jB_x = \frac{\mu_0 I}{2\pi(z - z_0)}.
\]

Along a curve \((C)\) surrounding the wire,

\[
\oint_{(C)} B(z) \, dz = \oint_{(C)} \frac{\mu_0 I}{2\pi(z - z_0)} \, dz = j\mu_0 I.
\]

Consider now an infinitesimally thin ribbon of width \( \Delta z \) carrying a total current \( \Delta I \). The values of \( B(z) \) at two infinitesimally close points \( A_1 \) and \( A_2 \) on either side of the ribbon are not the same, and letting \( \Delta z \) tend to zero, we have

\[
B_1(z) - B_2(z) = j\mu_0 \frac{dI}{dz},
\]

in which \( dI/dz \) represents the current density per unit length in the ribbon. Integrating, we find

\[
W_1(z) - W_2(z) = j\mu_0 I + \text{const}.
\]

Let us apply this relation to the following specific problem: we consider a circular cylinder of radius \( r_0 \) centred on \( z = 0 \). The equation of this cylinder is \( z = r_0 e^{i\theta} \). The current at each point is denoted by \( I(\theta) \). Let us assume that the total current is zero and that the dependence of \( I \) upon \( \theta \) is such that \( I(\theta) = I(\theta + 2\pi) \). \( I(\theta) \) can be written

\[
I(\theta) = \sum_{n=1}^{\infty} C_n \sin(n\theta + \theta_n)
\]

and

\[
\sin(n\theta + \theta_n) = \frac{1}{2j} \{ e^{j(n\theta + \theta_n)} - e^{-j(n\theta + \theta_n)} \}.
\]
Replacing $e^{in\theta}$ by $(z/r_0)^n$, we have

$$\sin (n\theta + \theta_n) = \frac{1}{2j} \left\{ e^{j\theta_n} \left( \frac{z}{r_0} \right)^n - e^{-j\theta_n} \left( \frac{r_0}{z} \right)^n \right\}.$$ 

At every point $z$ on the (outer and inner) surface of the curve,

$$W_{\text{ext}} - W_{\text{int}} = \frac{\mu_0}{2} \sum_{n}^{\infty} C_n \left\{ e^{j\theta_n} \left( \frac{z}{r_0} \right)^n - e^{-j\theta_n} \left( \frac{r_0}{z} \right)^n \right\} + \text{const.}$$

At a point $M(z)$, $W(z)$ must be finite when $|z| \to 0$ and when $|z| \to \infty$. We must therefore have

$$W_{\text{ext}}(z) = -\frac{\mu_0}{2} \sum_{n} C_n e^{-j\theta_n} \left( \frac{r_0}{z} \right)^n + \text{const.}$$

$$W_{\text{int}}(z) = \frac{\mu_0}{2} \sum_{n} C_n e^{j\theta_n} \left( \frac{z}{r_0} \right)^n + \text{const.}$$

Inside the cylinder, the field will therefore be of the form

$$B_{\text{int}}(z) = -\frac{\mu_0}{2} \sum_{n} n C_n r_0^n e^{j\theta_n} z^{n-1}$$

and outside,

$$B_{\text{ext}}(z) = \frac{\mu_0}{2} \sum_{n} n C_n r_0^n e^{j\theta_n} z^{-(n+1)}.$$

**Example.** We wish to create inside the cylinder a field with constant gradient:

$$B_x = G y, \quad B_y = G x.$$ 

Thus

$$B(z) = B_y + jB_x = G(x + jy) = Gz - \frac{1}{2} n \mu_0 C_n r_0^n e^{j\theta_n} = 0 \quad \text{for } n = 1,$$

$$= G \quad \text{for } n = 2,$$

$$= 0 \quad \text{for } n > 2.$$ 

Hence

$$C_2 = -\frac{G}{\mu_0} r_0^2 \sin 2\theta,$$

and therefore

$$I_\theta = -\frac{G}{\mu_0} r_0^2 \sin 2\theta.$$

The current density per unit length (in $A/m$) on the cylinder is

$$j(\theta) = \frac{dI}{r_0 d\theta} = -\frac{2G}{\mu_0} r_0 \cos 2\theta = -j(0) r_0 \cos 2\theta.$$ 

The field gradient is given by

$$G = \frac{\mu_0 j(0)}{2}.$$
and the maximum induction at the layer is

\[ B_0 = \frac{\mu_0 j(0)r_0}{2} \]

If, for example, we wish to obtain \( G = 1 \) tesla/m, then over a circle of radius \( r = 0.10 \) m we must satisfy the relation

\[ j(\theta) = -\frac{0.2}{\mu_0} \cos 2\theta = -j(0) \cos 2\theta \]

or

\[ j(0) = 160.000 \text{ A/m.} \]

Current densities as high as this can be attained by using superconducting windings (see Chapter 7).

### 3.1.3 Magnets with Magnetic Material

The general relations given above are still valid. To calculate \( B \) in the gap of an unsaturated electromagnet (or a permanent magnet), we assume that the permeability \( \mu_r \) is infinite. The surfaces of the pole-pieces are then equipotentials, and the Laplace equation

\[ \nabla^2 \Phi_m = 0 \]

is solved in whichever system is employed (Hesse, 1949; Lenz, 1950). Thus the field \( B_z \) of a rotationally symmetric lens similar to the one shown schematically in Fig. 13b will effectively be given by the expression for \( E_z \) in an electrostatic lens consisting of two equal tubes. It is far more difficult to determine the scalar potentials at the poles, however: we merely determine the potential difference between the two poles. The distribution of \( B_z \) on the axis is such that

\[ \Delta \Phi_m = \int_{-\infty}^{\infty} B_z dz = \mu_0 n I, \]

where \( nI \) denotes the total current through the magnetizing coils; when \( \mu = \infty \), this potential difference is wholly between the poles.

It is, in fact, frequently better to measure this potential difference; for this, a small long coil is wound onto a flexible support, each end is placed in contact with one of the pole-pieces, and then withdrawn. The two ends are then brought into contact with each other, and it is easy to show that the difference of flux across the coil, (which is measured with a fluxmeter) is proportional to \( \Delta \Phi \), the magnetic potential difference between those points on the two pole-pieces to which the ends of the winding were applied. If the magnetic circuit saturates, \( \mu \) becomes small (\( \mu / \mu_0 \) tends to unity); as soon as \( \mu < 100 \), the "potential drop" in the magnetic material can
no longer be neglected, and this drop will depend upon the path followed in determining the loop-integral of $H$ in the iron: the pole-pieces will no longer be "equipotentials".

If, nevertheless, the distribution of $B$ is the same as that in iron, $\mu$ can be deduced from the characteristic curve of the material employed. We can then proceed as follows. Consider an electromagnet (see Fig. 12f) consisting of a large yoke and two poles surrounded by coils each carrying total current $nI/2$. Over the path $ABCDEA$ we have

$$\int \vec{H} \cdot d\vec{l} = nI$$

and for $\mu \approx \infty$ we have $\int H \cdot dl = 0$ except between $B$ and $C$; hence $nI = \int \vec{H} \cdot d\vec{l}$ and we can regard the potential difference $\Phi_B - \Phi_C$ as equal to $\mu_0 nI$, and the pole-pieces as equipotentials.

If, on the contrary, $\mu$ is small, we have $\int H \cdot dl \neq 0$ with $H = B/\mu$. We write

$$\int_{A}^{B} H \cdot dl = \alpha_B nI, \quad \int_{C}^{D} H \cdot dl = \alpha_D nI.$$

We can always arrange that $\int_{D}^{A} H \cdot dl = 0$ by making the cross-section of the

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{fig12f.png}
\caption{Cross-section of an electromagnet with a soft iron yoke.}
\end{figure}
yoke big enough. We now have

$$\Phi_B - \Phi_C = \mu_0 \left\{ 1 - (\alpha_1 + \alpha_2) nI \right\}.$$  

If the value of $B$ is known everywhere in the cross-section, the potential difference between various pairs of points ($B'C', B''C''$) can be calculated and thus we can find the potential distribution over the surfaces of the pole-pieces. We then proceed by successive approximations; we start from the distribution corresponding to the unsaturated case. For a given value of $B_0$ in the gap, we can calculate the flux in the poles, and hence obtain a first approximation to the distribution of $B$ in the iron. We then carry out the calculation above and again solve $\nabla^2 \Phi_m = 0$ with the new boundary conditions; this gives a second, more exact approximation to $B$. After two or three "relaxations", $B$ is known very accurately.

This procedure can only be used when a computer or a resistance network (see below) is available.

3.1.4 The Expressions for the Field Components and the Vector Potential in the Neighbourhood of the Axis (or of a Symmetry Plane)

Just as in electrostatics, we can expand in series near the axis of a rotationally symmetric lens, since $\Phi_m(r, z)$ is a solution of the Laplace equation.

We obtain the expression

$$\Phi_m(r, z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} \left( \frac{r}{2} \right)^{2n} \Phi_0^{(2n)}(z),$$

in which

$$\Phi_0(z) = \Phi(0, z).$$

Hence

$$B_z(r, z) = -\frac{\partial \Phi(r, z)}{\partial z},$$

$$B_r(r, z) = -\frac{\partial \Phi(r, z)}{\partial r}.$$  

We know, however, that

$$B_r(r, z) = -\frac{\partial A(r, z)}{\partial z},$$

so that

$$\frac{\partial A}{\partial z} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!(n-1)!} \left( \frac{r}{2} \right)^{2n-1} \Phi_0^{(2n)}(z),$$

and integrating with respect to $z$,

$$A(r, z) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!(n-1)!} \left( \frac{r}{2} \right)^{2n-1} \Phi_0^{(2n-1)}(z).$$
In the neighbourhood of the axis, therefore, we have

\[ B_z(r, z) = -\frac{\partial \Phi(r, z)}{\partial z} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} \left( \frac{r}{2} \right)^{2n} B_0^{2n}(z), \]

\[ B_x(r, z) = B_0(z) - \frac{r^2}{4} B_0'(z) + \frac{r^4}{64} B_0''(z) - \cdots, \]

\[ B_y(r, z) = -\frac{r}{2} B_0'(z) + \frac{r^3}{16} B_0''(z) - \cdots, \]

\[ A_\theta(r, z) = \frac{r}{2} B_0(z) - \frac{r^3}{16} B_0''(z) + \cdots. \]

If we consider a circle \((C)\) of radius \(r\), centred on the axis \(Oz\), we have

\[ \oint_C A_\theta \, r \, d\theta = \int \int B_z \, dS = \Psi(r, z) = 2\pi r A_\theta, \]

where \(\Psi\) denotes the flux of \(B_z\) through this circle.

Hence

\[ \Psi(r, z) = \pi r^2 \left\{ B_0(z) - \frac{r^2}{8} B_0'(z) + \cdots \right\}. \]

In a similar way we could obtain \(B_x\), \(B_y\) and \(A_z\) in the vicinity of a symmetry plane in a system possessing translation symmetry.

### 3.2 DIRECT METHODS OF MEASURING THE
### INDUCTION \(B\) AND ITS GRADIENT

A knowledge of the distribution of magnetic induction along the axis of a lens, \(B(z)\), is essential, and the magnitude of \(B\) may vary from zero to several kilogauss over a distance of the order of millimetres; another difficulty is that we are very often obliged to use extremely tiny measuring probes, as the dimensions of the lenses are themselves so small. For both of these reasons, high precision measuring techniques which depend upon nuclear resonance are not applicable here. From among the numerous methods available for measuring strong magnetic fields (see Symonds, 1955, for example), we shall therefore describe only those which can usefully be employed to measure the fields in lenses.

#### 3.2.1 Methods Based upon Induction Phenomena

When the lenses are free of iron, the simplest method is to use a small coil, very carefully positioned, and to supply the lens with alternating current. The potential difference which is induced by the alternating current
across the ends of the coil is measured after amplification in a linear amplifier or in an amplifier which has been calibrated beforehand to give values of the field directly.

If the lenses are shielded, however, alternating current can no longer be used, because of the hysteresis effects which would be produced; the lenses are supplied with a highly stabilized direct current, and the distribution of $B$ is then measured by one of the various methods described below.

(i) *Direct measurement of the flux through a small coil.* The coil is moved slowly in the zone to be explored—along the $Oz$ axis of a lens, for example—and the potential difference induced across its terminals is transmitted to a highly sensitive fluxmeter (Sauzade, 1958); the signal, $V$, at the output may reach one volt for $\psi = 100$ maxwells. By recording $V$ directly onto a sheet of paper which unrolls in synchronism with the motion of the search coil, we obtain automatically a plot of the distribution of flux $\psi(z)$. We can obtain the absolute value of $B(z)$ by calibrating the coil beforehand in a known homogeneous field, always provided the dimensions of the coil are considerably smaller than those of the lens, in which the field is inhomogeneous. Laudet (1957) has, however, developed a method of calculation with the aid of which we can derive the values of $B(z)$, $\frac{dB}{dz}$ and $\frac{d^2B}{dz^2}$ quite accurately from the curve produced by a relatively large search coil.

We can equally well plot the value of $B$, point by point, with a calibrated search coil connected to a fluxmeter or to a ballistic galvanometer. The coil is placed in position in the field which is to be charted, with its turns normal to $B$; it is then suddenly withdrawn from the field, or alternatively, the coil is left untouched but the sense of the current in the coil is reversed (Dosse, 1941. See Fig. 12g).

![Diagram](image)

**Fig. 12g.** The principle behind Dosse's measurements (1941). The coil $M$ is shifted along the axis, and in each position, the change of flux which is produced by reversing the magnetizing current is measured.
A alternating e.m.f. is created across the terminals of a small rotating coil which is placed in the field; this e.m.f., which is proportional to $B$ and to the frequency of rotation, is then measured. The coil is supported on the non-magnetic axle of a synchronous motor; for preference, the motor should be a synchronous three-phase motor, virtually free of mechanical vibration (see Fig. 13a). For the measurements to be accurate, the coil must be extremely small and must be placed at its successive positions within the gap between the pole-pieces with great care (see Langer and Scott, 1950, for example). This method can be made more accurate by using two coils driven by the same motor; one coil, $B_1$, rotates within the lens being studied, and gives a signal $v_1$, whilst the other, $B_2$, rotates within a homogeneous magnetic field which has been very accurately standardized (with a nuclear resonance probe, for example). The signal $v_1$ is then counteracted by a known fraction $k$ of the signal $v_2$ given by $B_2$; by adjusting the relative orientations of the symmetry axes of $B_1$ and $B_2$, the phase difference between $v_1$ and $v_2$ can be exactly annulled. Finally, we suppress as far as possible any vibration of the coils by using special vibration-proof bearings, and thus avoid parasitic signals being superimposed onto the fundamental (Germain, 1955a). We can thus arrange that the signals $v_1$ and $kv_2$ are equal, and hence deduce the value of the unknown induction from the quantity $k$ which is given by an accurate potentiometric method. In these conditions, we can easily reach an accuracy of 0.1 per cent.

**Fig. 13a.** Sketch of a rotating coil.


**Fig. 13b.** Fert and Gautier’s arrangement (1954), with an oscillating coil.

(iii) A third possibility is to set a long coil in vibration along the direction of the axis; as before, the induced e.m.f. is proportional to $B$ at the end face of the solenoid. However, if we have a short fixed coil of small diameter, $dB/dz$ can be obtained, and with two identical coils mounted in opposition, $d^2B/dz^2$ (Fert and Gautier, 1951; Gautier, 1952. See Fig. 13b).

A description of a large number of devices of this type for measuring the induction and its gradient in magnetic lenses with large dimensions is to be found in an article by Grivet and Septier (1960). In particular, the gradient of $B$ in a direction, $OX$, normal to $B$ can be measured directly by using a short coil which has its axis parallel to $B$ and is attached to a Perspex rod which vibrates transversely along $OX$; the oscillation frequency of the rod is selected to be equal to one of its resonant frequencies, and a large amplitude, $\Delta x$, can thus be obtained with only a small expenditure of power.

### 3.2.2 Method Using the Force which Is Exerted by the Induction $B$ on a Current-carrying Conductor

In this method (Ments and Le Poole, 1947) a long slender solenoid, $S$, wound on a non-magnetic arm and carrying a current $i$, is suspended horizontally along the lens axis. The end-face which lies within the field experiences a force proportional to the value of the induction $B$ at the point. The

![Diagram](image)

Fig. 13c. Durandeau's design (1953) in which the field is measured by the force which it exerts on a coil.


displacement of the whole mobile assembly is measured in a microscope. An improvement on this design has been made by Durandeau (1953), using two solenoids, $S_1$ and $S_2$, identical save in that they are wound in opposite senses (Fig. 13c), each carrying a current $i$. Either the force exerted by the induction on the common face of the two coils is measured, or else the
displacement of the mobile part is annulled with the aid of an auxiliary coil $B_1$, which also carries a current $i$, and which is placed in the known field of a system of Helmholtz coils; the Helmholtz coil current is then proportional to the magnitude of $B$.

### 3.2.3 The Hall Effect

The probe is now a small thin rectangular plate $ABCD$, thickness $a$ (see Fig. 14). $AB$ and $CD$ are the short sides which are covered with a conductive layer, and between these a steady current $I$ flows. If we label the midpoints of the longer sides, $BC$ and $DA$, $M$ and $N$ respectively, and if we place the plate in a region where the magnetic induction is $B$ and is normal to the plane $ABCD$, we shall find a potential difference between electrodes placed at $M$ and $N$ which is proportional to the magnitude of $B$ and to the current $I$:

$$V_M - V_N = K_H \frac{IB}{a},$$

in which $K_H$ is the “Hall constant” of the substance employed, normally a semi-conductor such as indium arsenide or indium antimonide. When $I \sim 0.1$ amps and $B \sim 10$ kG, a plate of a few square millimetres can produce a potential difference of a few tens of millivolts.

![Fig. 14. A Hall plate for measuring an induction $B$ perpendicular to the face $ABCD$; $AB$ and $CD$ are metallized.](image)

In fact, the terminals at $M$ and $N$ are connected through an external resistance, $R$, and the potential which is measured, $(V_M - V_N)$, is a “load” potential; we must, therefore, make allowance for the internal resistance of the plate between $M$ and $N$, $K_{MN}$. Due to the magneto-resistance effect however, the latter is a function of $B$, and hence $(V_M - V_N)$ is never perfectly proportional to $B$, especially when the induction reaches high values. For
accurate measurements (between one part in $10^8$ and $10^9$), $I$ has to be stabilized to better than one part in $10^4$, and the temperature of the plate has to be held constant to within less than 0.1°C. ($K_H$ varies with temperature—from this point of view, indium arsenide is superior to every other semi-conductor, as its temperature coefficient is very slight at room temperature).

The potential difference between the edges of the plate is measured by a standard method in which the potential to be measured is opposed by a potential which has been calibrated to within about one part in $10^4$. The plate itself is accurately calibrated by using nuclear resonance, and a known electrical circuit which provides a curve $B = f(I)$ to which we subsequently refer.

By juxtaposing two identical plates, we can measure the gradient of $B$ directly. This technique has as yet only been used in electromagnets and lenses of large dimensions even though we now know how to construct plates on a miniature scale with which point by point measurements become possible.

### 3.3 THE USE OF ELECTRICAL ANALOGUES

#### 3.3.1 The Standard Transposition

The results of §2.3 suggested that both the electrolytic tank and the resistance network are convenient tools for studying fields; provided that the hypothesis that the pole-piece surfaces are equipotentials proves to be valid, there is a formal identity between electrostatic and scalar magnetic potentials.

![Diagram (a)](image1)

**FIG. 15a.** The direct analogy. The pole-pieces and the magnetic circuit are represented by metal electrodes. No allowance is made for the windings of the electromagnet and the lens. A chart of the "scalar equipotential surfaces", $\varphi = \text{constant}$, is traced out.

![Diagram (b)](image2)

**FIG. 15b.** The conjugate analogy. The iron is represented by a block of paraffin; the coils $B_1$, $B_2$ are represented schematically by injecting currents. We obtain "vector equipotentials", $A = \text{constant}$. 

The pole-piece profiles are equivalent to "perfect" conductors, held at
electrostatic potentials proportional to their magnetic counterparts. Coils
are represented, very roughly, by a break in the magnetic circuit across
whose ends an overall magnetic potential difference $NI$ is applied, to
represent the effect of the magnetizing current (see Fig. 15a). In the tank,
such a break is represented by an electrical gap across whose ends an elec-
tric potential proportional to $NI$ is applied.

Even when certain portions of the magnetic circuit are saturated, or in
the process of becoming saturated, it is still possible to obtain the distri-
bution of magnetic induction, $B$ (Liebmann, 1953). Once the characteristics
of the magnetic material in question are available, we are in a position
to know the value of the permeability $\mu$ for each local value of $B$. If the
value of the resistance at a point at which $\mu = \mu_0$ (air) is $R_n$, then a perme-
ability $\mu_1$ at the same point will be represented by a resistance $R'_n$ such that

$$R'_n = \left(\frac{\mu_0}{\mu_1}\right) R_n.$$

If $\mu_1 = \infty$ then clearly $R'_n = 0$ (the nodes are short-circuited). In practice,
we proceed by successive approximation; a chart of the scalar equipotentials
is drawn up for the infinite permeability case, and the induction $B$ at the
surface of the steel is calculated starting from a value $B_0$ at some point on
the axis, $B_0 = 20$ kG for example. The resistances which represent the
boundaries of the pole-pieces and the interior of the pole-pieces themselves
can then be modified, a new chart of the potential can be drawn and fresh
values of $B$ extracted. Continuing in this way, we tend gradually to-
wards a self-consistent solution, for which the network has no further
need of modification.

The method depending upon a direct analogy can, however, only be used
to determine the distribution of $B$ within the narrow gaps in magnetic
lenses—the windings lie far enough away from the useful zone to produce
no appreciable perturbation. This is by no means true of an electromagnet
with a wide opening and for this case, the "conjugate" analogy which we
are about to describe must be invoked; this analogue is also capable of
handling the case of systems free of iron.

3.3.2 The Conjugate Tank

We now pass on to a more recent method, again using an electrical
analogy, in which it is possible to make allowance for the magnetizing
coils. The method is due to Peierls and Skyrme (1947, 1949), and has been
tried in practice by Bracher (1950)—it is known as "the conjugate tank".

Consider the equations which govern the ordinary tank, valid for the whole
system of electrodes, current sources, and insulators which is immersed
in the electrolyte. If $i$ be current density at some point, and $J$ the intensity
emitted per unit volume, then
\[
\text{div} i = J.
\]
Elsewhere than at point sources, \( \text{div} \, i = 0 \).
Further,
\[
\text{curl} \, E = 0,
\]
and
\[
i = \sigma \, E.
\]
The electrical potential \( V \) satisfies the equations
\[
E = -\text{grad} \, V,
\]
\[
\nabla^2 V = -J/\sigma,
\]
in which \( \sigma \) is the conductivity of the medium.
On the boundary between two media, free of surface sources, \( \text{div} \, i = 0 \)
provides the boundary conditions for the normal components of current
density:
\[
i_{2n} = i_{1n},
\]
while \( \text{curl} \, E = 0 \) implies
\[
E_{2t} = E_{1t}
\]
(that is, that the tangential component of the electric field is continuous).
As indicated in § 2.2.1, these two relations can be put in the form
\[
\sigma_1 \left( \frac{\partial V}{\partial n} \right)_1 = \sigma_2 \left( \frac{\partial V}{\partial n} \right)_2
\]
and
\[
\left( \frac{\partial V}{\partial t} \right)_1 = \left( \frac{\partial V}{\partial t} \right)_2.
\]
Consider now the magnetic field of an electromagnet. The magnetic flux \( B \)
is conservative, that is.
\[
\text{div} \, B = 0.
\]
If \( j \) is the current density in the coils, Ampère’s theorem relates the magnetic
field \( H \) and \( j \) by
\[
\text{curl} \, H = j.
\]
Defining vector potential \( A \) by
\[
B = \text{curl} \, A,
\]
\[
\text{div} \, A = 0,
\]
substituting \( B = \mu \mu_0 \, H \), and using the standard identity
\[
\text{curl} \, \text{curl} \, u = \text{grad} \, \text{div} \, u - \nabla^2 u,
\]
we find
\[
\nabla^2 A = -\mu \mu_0 j.
\] (3.5)
If we suppose that the magnetizing current and the boundary between the
two media are parallel to the axis $Oz$ of a system of Cartesian coordinates, this implies

$$j_x = j_y = 0, \quad j = (0, 0, j_z).$$

Since $\mathbf{B}$ lies in a plane normal to $j$, we have

$$A_x = A_y = 0, \quad \mathbf{A} = (0, 0, A_z),$$

so that

$$\nabla^2 A_z = -\mu \mu_0 j_z. \quad (3.6)$$

Let $\mathbf{n}$ be the unit vector, normal to some separating surface, and $\mathbf{b}$ the unit vector perpendicular both to $\mathbf{n}$ and to $Oz$; the boundary conditions at the surface become:

$$\left( \frac{\partial A_z}{\partial t} \right)_1 = \left( \frac{\partial A_z}{\partial t} \right)_2, \quad (3.7)$$

and

$$\frac{1}{\mu_1 \mu_0} \left( \frac{\partial A_z}{\partial n} \right)_1 = \frac{1}{\mu_2 \mu_0} \left( \frac{\partial A_z}{\partial n} \right)_2. \quad (3.8)$$

Comparing equations (3.2) and (3.6), (3.3) and (3.8), and (3.4) and (3.7), we find the correspondence:

$$\begin{align*}
\text{Electrolytic tank} & \quad \left\{ \begin{array}{c}
V \longrightarrow A_z \\
J \longrightarrow j_z \\
\sigma \longrightarrow 1 \\
\mu \mu_0 \end{array} \right. \\
\text{Electromagnet.} & \quad \end{align*}$$

In practice, this correspondence enables us to interpret the equipotential curves which the electrolytic tank has provided as vector equipotentials of $A_z$, provided that pole-pieces (assumed in this case to be of infinite permeability) are represented by insulators ($\sigma = 0$), and coils by a line of current-sources of current density proportional to $j_z$ lying parallel to $Oz$ (see Fig. 15b).

From the $A_z$ curves, it is simple to deduce $\mathbf{B}$, which is tangential to the equipotentials at every point, and whose magnitude is proportional to grad $A_z$. The triad: grad $A_z$, $\mathbf{B}$ and $Oz$ is right-handed.

One further result is of interest—if the values of the equipotentials are equally spaced on the potential scale, the same flux of magnetic induction lies in each of the tubes defined by each pair of equipotential surfaces.

This method is not suitable for direct application to systems with axial symmetry. Nevertheless, it has been applied successfully to the study of very large magnets whose curvature can be neglected (those of synchrotrons, for example). The results provided by the tank accord well with those obtained by direct measurement (Germain, 1955a).

Attempts have been made to apply this technique to systems with axial symmetry, always assuming infinite permeability (Germain, 1955b). The only components of the current density $j$ and the vector potential $\mathbf{A}$ of the
induction $B$ in this case are of the form $j_0$ and $A_0$. The relation
\[ \text{curl } B = \text{curl } \text{curl } A = \mu_0 \mu_1 j \]
reduces to a single scalar expression
\[ \frac{\partial}{\partial z} \left[ \frac{1}{r} \frac{\partial}{\partial z} (r A_0) \right] + \frac{\partial}{\partial r} \left[ \frac{1}{r} \frac{\partial}{\partial r} (r A_0) \right] = -\mu_0 \mu_1 j_0 \]  
(3.9)
in which $\mu_1$ represents the relative permeability of the medium.

Apart from a factor of $2\pi$, the quantity $(r A_0)$ represents the magnetic flux $\psi$ which threads a turn of radius $r$:
\[ \psi = 2\pi r A_0, \]  
(3.10)
so that (3.9) is equivalent to
\[ \frac{\partial^2 \psi}{\partial z^2} - \frac{1}{r} \frac{\partial \psi}{\partial r} + \frac{d^2 \psi}{dr^2} = -2\pi r \mu_0 \mu_1 j_0. \]

At the frontier between two media of relative permeabilities $\mu_1$ and $\mu_2$, Ampère's theorem states that
\[ \text{curl } H = 0, \]
and Gauss's theorem that
\[ \text{div } B = 0, \]
which in this case leads to the relations:
\[ \left[ \frac{1}{\mu_0 \mu_1} \frac{\partial}{\partial n} (r A_0) \right]_1 = \left[ \frac{1}{\mu_0 \mu_1} \frac{\partial}{\partial n} (r A_0) \right]_2 \]  
(3.11a)
between the normal components, and
\[ \left[ \frac{\partial}{\partial t} (r A_0) \right]_1 = \left[ \frac{\partial}{\partial t} (r A_0) \right]_2 \]  
(3.11b)
between the tangential components.

We now consider a two-dimensional conducting medium, $M$, $(Ox,y)$, the conductivity, $\sigma$, of which is a function solely of $y$:
\[ \sigma = \frac{k}{y}. \]  
(3.12)
Current sources giving a current density $d$ are distributed across the conducting surface. The relations
\[ \text{div } i = d, \]
\[ i = -\sigma \text{ grad } V \]
give
\[ \frac{\partial}{\partial x} \left( \frac{1}{y} \frac{\partial V}{\partial x} \right) + \frac{\partial}{\partial y} \left( \frac{1}{y} \frac{\partial V}{\partial y} \right) = -\frac{d}{k}. \]  
(3.13)
The boundary conditions between two homogeneous media, of conductivities \( \sigma_1 = \frac{K_1}{y} \) and \( \sigma_2 = \frac{K_2}{y} \) respectively, take the form:

\[
\begin{align*}
(k_1 \frac{\partial V}{\partial n})_1 &= (k_2 \frac{\partial V}{\partial n})_2, \\
\left( \frac{\partial V}{\partial t} \right)_1 &= \left( \frac{\partial V}{\partial t} \right)_2
\end{align*}
\]  
(3.14a)  
(3.14b)

(using equations (3.3) and (3.4)).

An examination first of equations (3.10) and (3.11) and then of (3.13) and (3.14) reveals that they are identical if we make the following correspondences:

\[
\begin{align*}
 r A_0 &\to V, \\
 \mu &\to \frac{1}{k}, \\
 j_0 &\to d, \\
 z &\to x, \\
n &\to y.
\end{align*}
\]  
(3.15)

Once we possess the conducting medium \((M)\) we have simply to measure \(V\) by a standard method; the equipotential curves will represent the curves \(r A_0 = \text{constant}\)
or, what is equivalent, curves of equal flux, \(\psi = \text{constant}\). At a pair of corresponding points of the medium \((M)\) and the actual magnetic field, the electric field \(E\) and the induction \(B\) will be related in the following ways:

(i) \(V = k_r A_0\) with \(k > 0\),
(ii) \(E\) is normal to \(B\) (since \(E \cdot B = 0\)),
(iii) \(|E|\) is proportional to \(r|B|: |E| = k r |B|\),
(iv) The three vectors \(E, B\) and \(k\) form a right-handed triad \((k\) is a unit vector normal to the plane \(x0y\)).

The medium \((M)\) can be constructed in an approximate fashion by filling a tank of variable depth \(y\) with an electrolyte (more precisely, the bottom should be hyperbolic in form) provided that the depth does not change too rapidly; measurements which have been made in the region near the axis (where the liquid has to be very deep) cannot be regarded with any confidence (Germain, 1955a). Infinitely permeable pole-pieces are made of paraffin wax \((k = 0)\).

In conclusion, we should mention that pole-pieces with a permeability which is not infinite can be represented in the conjugate tank analogy by replacing the electrodes which have zero conductivity by others made of some substance (paraffin wax mixed with powdered graphite, for example)
which has a conductivity which is related to that of the electrolyte in the following way:

\[
\frac{\sigma_{\text{electrode}}}{\sigma_{\text{electrolyte}}} = \frac{\mu_{\text{air}}}{\mu_{\text{iron}}}.
\]

3.3.3 The Resistance Network

The resistance network is a far more flexible tool than the conjugate tank, and with its aid, the induction \( B \) can be obtained for systems with axial symmetry, with or without iron.

The simplest analogy is between a network and the medium \((M)\) defined above; the various quantities are thus related in the following way:

\[
V \rightarrow r A_\theta = \frac{\psi}{2\pi}.
\]

Networks of this kind have been described by Liebmann (1950b), and constructed by Wakefield (1958) and by Christensen (1959).

The resistances which are placed in rows parallel to the axis \( OX \) are of the form \( R_n = r R_0 \) (in which \( R_0 \) is a constant). Along the axis, \( R = 0 \); the nodes are short-circuited. The resistances which are placed in columns parallel to the straight line \( OY \) (perpendicular to \( OX \)) on either side of each node \( P_n \) which lies a distance \( r_n \) from the axis \( OX \) are given the values:

\[
R_n = \frac{(r_n^2 - r_{n-1}^2) R_0}{2}, \quad R_{n+1} = \frac{(r_{n+1}^2 - r_n^2) R_0}{2}.
\]

The analogy (3.15) between magnet and network cannot be used directly, as we have both to change the geometrical scale, writing

\[
\frac{y}{r} = \frac{x}{z} = s,
\]

and to scale the currents which are fed into the network. We suppose that \((NI)\) is the number of ampère-turns in one of the coils of the magnet, and \(i\) is the total current fed into the network in the zone which represents this coil, and we write

\[
p = \frac{NI}{i}.
\]

If we compare the current densities \( j \) in the real coil and \( d \) in the network, we find

\[
p = \frac{1}{s^2} \frac{j}{d}.
\]

If the analogy between equations (3.10) and (3.11) and equations (3.13)
and (3.14) is to be complete, therefore, we must write:

(a) \( \frac{\psi}{2\pi} = r A = \frac{k \mu p}{s} V \),

(b) \( j = p s^2 d \),

(c) \( \mu \mu_0 = \frac{1}{k} \),

(d) \( z = \frac{x}{s} \),

(e) \( r = \frac{y}{s} \).

(3.16)

Measurement of \( V \), therefore, gives the value of \( \psi \) directly from equation (3.16a); it will be expressed in webers. The magnitude of the induction, \( |B| \), can then be obtained as a function of the field \( E \) measured on the network; at a point associated with the value \( y \),

\[ |B| = \frac{k \mu p s}{y} |E| \text{ webers m}^{-2}. \]

The full calculation of the values which the resistances should be given both in the case when the scale is the same in the \( x \) and \( y \) directions and when an interesting region of the system is to be expanded are to be found in the article by Christensen.

Just as in the tank, coils are represented by currents which are fed into the network at the corresponding nodes. In complex systems, we take the sense of the currents in the coils with respect to the axis \( Ox \) into account by supplying the network from a current source with two polarities and earthing the axis of the network. An alternating current is usually employed; the detection device is commonly a valve voltmeter with a very high impedance, provided with a detector which indicates the polarity of the potential being measured with respect to earth (Wakefield, 1958).

Along the boundaries of the pole-pieces, of infinite permeability, \( R = \infty \); the circuit is open, since the condition \( \frac{\partial \psi}{\partial n} = 0 \) must be satisfied.

Another type of network of a more complicated design exists with which the curves \( A_\theta = \text{constant} \) can be obtained directly, just as they can in the conjugate tank when the system possesses a plane of symmetry (Liebmann, 1950b).
CHAPTER 4

THE OPTICAL PROPERTIES OF ELECTROSTATIC LENSES

4.1 IMAGE FORMATION

4.1.1 The Focusing Effect of the Radial Field

Having considered static electric and magnetic fields from a general point of view, we are now in a position to establish the properties which such fields must necessarily possess if they are to be capable of focusing into a point image all the electrons which emerge from a point object. In this chapter, only electrostatic lenses will be considered, and, as a simple case with which to begin, we select the three-electrode lens of Fig. 1. The

![Diagram of image formation in a thin lens](image)

**Fig. 16.** Image-formation in a thin lens.

two outer grids are held at the accelerating potential of the electrons, while the central electrode is held at a quite different value. In this way, the electric field is localized between the two grids, and outside the lens the electrons move in straight lines. If all the electrons emitted by an (axial) point $A$ are to pass through another (axial) point $A'$, each ray must be deflected towards the axis through some angle $\alpha$ as it passes through the lens (Fig. 16). Since an electric field $E$ produces a force $F = eE$ on each electron, the lens has to produce a radial field directed away from its centre. For a ray which reaches the lens at a distance $r$ from the axis, the angular deviation $\alpha$ has to be proportional to $r$, thus

$$\alpha = r \left( \frac{1}{a} + \frac{1}{a'} \right).$$

(4.1)
As the deviation is proportional to the force, the intensity of the radial field must increase with the distance $r$ from the axis according to a law of the form $E_r = K(z) r$, where $K(z)$ is either a constant or a function of $z$ alone. In every lens, the field can be described by an expression of this kind, as we shall subsequently see.

![Diagram](image_url)

Fig. 17. (a) Convergent lens—real image. (b) Convergent lens—virtual image. (c) Divergent lens—virtual image. (d) Convergent lens—the real image of a virtual object.

The consequence is that for every ray, the ratio of $\alpha$ to $r$ is a constant, $1/f$, such that the object and image distances satisfy the elementary lens relation

$$\frac{1}{a} + \frac{1}{a'} = \frac{1}{f}. \quad (4.1')$$

If, therefore, a system of electrodes produces a radial field, whose intensity at any point is proportional to the distance of the point from the axis, it will focus charged particles just as an optical lens focuses light rays. The rays
will converge onto a point such as \( A' \), if the radial force is directed towards the axis, and if the lens is powerful enough; if both of these conditions are satisfied simultaneously, we shall have a converging lens producing a real image, as in Fig. 17a. In such a lens, however, it would be possible to reduce the field to such an extent that it became no longer capable of turning the rays which have originated at \( A \) through an angle sufficiently large for them to intersect the axis again at \( A' \), however distant. Nevertheless, if we produce backwards the rays which diverge rectilinearly beyond \( H' \) (a dotted line represents the produced section of an actual ray), the "virtual rays" thus created will converge into a point \( A' \). The actual rays seem to have originated in \( A' \), which is called the "virtual" image of \( A \) (Fig. 17b). The same convention makes it possible to speak of "imaging" when the radial force repels the particles from the axis; the lens is called "divergent", in this case, and a real object always has a virtual image (see Fig. 17c). Figure 17d shows how this concept of virtuality can be extended to the object; a lens would have a virtual object if, for example, it were placed between another lens and the (real) image which this lens has produced.

For each field intensity, therefore, and for each particle velocity, an image is produced; these are the two factors which describe the action of the lens, or, optically speaking, its convergence. The slower the electrons and the stronger the radial field, the more convergent will be the lens—this must not be carried too far, however, as there is a point at which the lens turns into a mirror, as we shall see later.

4.1.2 The Impossibility of Separating the Longitudinal and Radial Field Components

The electrodes of such lenses as the three-electrode lens produce a satisfactory field configuration—at least in the axial region—but with a structure less simple than we have supposed in the previous section. In fact, each radial component is associated with a longitudinal component of electric field, parallel to the axis.

The map of lines of equal field strength shows that the radial component is appreciable only where the lines of force have a large curvature, and where, consequently, the longitudinal component of the field is changing rapidly over short distances. The laws which govern the electric field provide a relation between the radial and longitudinal field components in the neighbourhood of the axis; the flux across the surface of a small cylinder, whose axis coincides with that of the lens, is zero, so that for a field of axial symmetry,

\[
E_r = -\frac{1}{2} r \frac{\partial E_z}{\partial z} = \frac{1}{2} r \frac{d^2 \varphi}{dz^2}, \tag{4.2}
\]

provided that second and higher order terms are negligible within the cylinder. We see that the radial component is a function of the rate at which
the longitudinal component of the electric field is varying with \( z \), or of the axial potential \( \varphi(z) \). \( E_r \) depends upon \( r \) in the requisite fashion, and the factor \( K(z) \) which was introduced earlier is proportional to the gradient of the longitudinal field component. This suggests that the extent of the lens should be defined as the region over which the derivative of \( E_z \) is still appreciable.

The effect of the longitudinal component is most important, as it retards or accelerates the electrons and thus has an indirect influence over the deviation of the beam which the radial component can produce; in short, it affects the convergence.

4.1.3 Ordinary Lenses without Grids

The longitudinal component has a far more pronounced effect in lenses without grids, which are much more common in practice, as lenses with grids are unsuitable for a number of reasons. Firstly, it is difficult to construct a sufficiently transparent grid, combining extreme slimness of the meshes with mechanical rigidity; further, secondary electrons are produced at the wires, and their superfluous illumination at the final screen reduces the contrast of the image proper. In addition, the grid produces a slight haziness at the image.

So far, we have been assuming that the planes \( H \) and \( H' \) are effectively continuous conductors, at which lines of force will terminate normally (Fig. 18). In reality, when \( H \) and \( H' \) are grids, each line of force has a small
kink (Fig. 19) just before it is terminated; an estimate of the influence of this deformation can be made by comparing the variation of potential along a line of force with the potential difference along the bend near the grid. Small though this effect is, its effect is almost invariably harmful, however much we try to diminish it by reducing the mesh-length of the grid.

![Diagram of meshed grid with lines of force and equipotentials]

**Fig. 19.** The distortion of the field which the wires of the grids produce.

Subsequently, therefore, we shall consider only normal (grid-free) lenses, a typical example of which is the three-electrode lens, in which the grids are replaced by circular holes cut in plane electrodes ("Einzellinse" in German).

The longitudinal component now plays a considerably more important role. To make this more readily comprehensible, consider the symmetrical two-electrode lens of Fig. 20a.

In passing from the first half of the lens to the second, the sense of the radial field is reversed: if the field had a radial component only, it would have no overall effect. There is however, a longitudinal component which accelerates the electrons as they pass through the lens, which are as a result moving faster in the second half of the lens than in the first half. In the first half, therefore, the converging power is increased, and in the second, the diverging power diminished—the whole lens is convergent. In the three electrode lens, the action can be analysed similarly (see Fig. 20b) into an assembly consisting of a convergent unit between two divergent units.

Many of the features of ordinary electron lenses can be explained in terms of the influence of this longitudinal component. Lenses are, for example, always convergent. In order to explain this, we consider the shape of the lines of force of the electric field, which link the positive charges of one electrode with the negative charges of an adjacent electrode. The hole in the
middle of each electrode make the lines of force curved, convex towards the axis, a consequence of their mutual repulsion. A region of convergence must always, therefore, be accompanied by a region in which the lens action is divergent; but electrons move more slowly in a convergent region, which slows them down, than in a divergent region, which accelerates them, so that the convergent effect always exceeds the divergent. A three-electrode lens remains convergent even though the central electrode be at a positive

![Diagram](a)

![Diagram](b)

![Diagram](c)

Fig. 20. Some electron lenses, with their glass optical counterparts. (a) A two-cylinder lens, which corresponds to a doublet. (b) A three-electrode lens, equivalent to a symmetrical triplet. (c) The practical example of a lens with a central mesh electrode, of which the analogue is a pair of convergent lenses.
potential with respect to the outer diaphragms. In the special case of a plane cathode near an accelerating diaphragm, however, the rule is no longer applicable, as the convergent region has disappeared—it is the only grid-free lens which is divergent. No application of these lenses, studied by Davison and Calbick (1931, 1932) and MacNaughton (1952) has as yet been found.

Another feature of electron lenses which is explicable in terms of the effect of the longitudinal component is the fact that it is always found necessary to apply a potential difference across the two electrodes which is of the same order of magnitude as the beam accelerating potential, despite the differential retardation. This is quite different from the grid lens where the action of the field is everywhere uniform, and where only one tenth of the potential is necessary to produce the same convergence. Figure 20c shows a simple design for a lens with a single grid possessing the same optical properties as those of the version with two grids considered earlier. More details of these lenses, which are better than ordinary lenses in this respect, will be found in § 8.4.

4.1.4 The Electron Mirror; Minimum Focal Distance

We consider first a lens of the usual kind, with electrodes symmetrical about an axis \( z'0z \). The field in such a lens is then completely described by a set of equipotentials in a meridian plane, \( \Phi(r, z) = \text{constant} \). Our interest in \( \Phi \) is caused by the fact that the velocity of the electrons (each of mass \( m \) and carrying charge \(-e\)) is given (in the non-relativistic case) by

\[
v = \sqrt{\frac{2e\Phi}{m}}
\]  

(4.3)

provided the value of the potential at the cathode is chosen as origin \((V_c = 0)\) since the electrons are emitted with a negligible velocity. Then

\[
e/m = 1.759 \times 10^{11} \text{ C kg}^{-1},
\]

\[
\sqrt{\frac{2e}{m}} = 0.5932 \times 10^8 \text{ MKS units}.
\]

The basic relation (4.3) suggests immediately how an electron lens can be converted into an electron mirror—all that we have to do is to apply a potential to the central electrode of a three electrode lens which is not only negative with respect to the neighbouring electrodes, but also with respect to the cathode. The field which ensues will produce a lens action just so long as the minimum potential which the electrons encounter, and which is situated in the middle of the central electrode \((r = z = 0)\), is positive (if \( \Phi(0, 0) > 0 \), that is). A complete examination is to be found in § 8.2.

When the negative potential applied to the central electrode is numerically sufficiently large, however, the electrons are decelerated and begin to
return towards their source without having reached the centre of the lens. This situation is called an "electron mirror"; its optical properties are analogous to those of glass optical mirrors.

When \( \Phi(0,0) \) is negative and very large, the electrons are turned about before reaching the central region of the lens where the field is convergent, and the mirror is divergent; if, on the other hand, \( \Phi(0,0) \) is negative but only slightly so, the electrons penetrate deep into the lens, and before they are reflected, they will have experienced the converging action of the field—the mirror is convergent.

If \( \Phi(0,0) \) is negative but very close to zero (within a few tenths of a volt) oscillations appear in the electron trajectories. This phenomenon is intermediate between the lens and mirror regions. The potential zone which produces oscillating rays (which are known as "transgaussian" rays) is virtually useless for image formation, though it has a few rather specialized applications.

\( \Phi(0,0) \), therefore, cannot be reduced below a certain limit, and a very high convergence is not, as a result, attainable. In practice, the least focal length that one can obtain is about as large as the radius of the central electrode when the inter-electrode spacing is also of the order of this radius, which is frequently the case.

A similar phenomenon appears when \( \Phi(0,0) \) approaches zero from the positive side—the convergence of an electron lens too, therefore, is bounded.

Should the electron beam cross-sectional area not be small relative to the area of the holes, it is possible that peripheral rays will be reflected, while axial rays experience only lens action—the trajectories, in such a case, become most involved.

4.1.5 The Lens Considered as an Assembly of Prisms. Similitude Properties

Just as a glass lens can be decomposed into an assembly of prisms, so can an electron lens, and although the method is more artificial than those which we have so far used, the analogy provides suggestive results. The basic electron prism is the parallel plate condenser, the standard type of deflector in cathode ray oscillographs. This deviating property suggests a simple description of the action of the radial component \( E_r \) of the electric field. Each volume element of the field is to be thought of as a small condenser, with plates distance \( d \) apart, of length \( l \) parallel to the axis, supporting a potential difference \( U = E_r d \), and with a mean potential of \( \Phi \) (see Fig. 21). This elementary condenser deflects the ray through a small angle \( \alpha \) given by

\[
\tan \alpha = \frac{1}{2} \frac{UL}{\Phi d} = \alpha. \tag{4.4}
\]

We can draw a number of important conclusions from this law, which hold for both thick and thin lenses. Firstly, since neither the charge nor the mass
of the particles appears in the deviation law, the ray path is independent of \( e \) and \( m \), and the lens would focus no differently if an ion source replaced the electron source. This property provides the possibility of focusing ions, which is a most important operation in particle accelerators, and was brilliantly introduced into mass spectrography in France by Cartan (1937). Ion sources are also beginning to be introduced into microscopy (Gauzit, 1951, 1953, 1954; Magnan and Chanson, 1951).

![Figure 21. The electrostatic prism.](image)

It is worth mentioning in passing that electrostatic ion focusing can be hazardous in tubes which contain thermo- or photo-emissive cathodes, or luminous films, since an excessive bombardment by returning ions can deteriorate or destroy the delicate films (Broadway and Pierce, 1939; Schaeffler and Walcher, 1943). The positive ions which are produced all along the beam by the impacts of the electrons on the residual gas molecules retrace the electron paths and strike the cathode; negative ions, either produced by the cathode or originating in the residual gas, flow towards the fluorescent screen. In cathode ray tubes, these latter are caught in an ion trap, which consists of a pair of crossed deflecting fields, one magnetic, the other electrostatic, which act for a short distance on the beam. Their intensity is so adjusted that on electrons of mass \( m \), their effects just cancel out; on ions with a mass \( M \) very different from \( m \) (\( M \approx 1840 \) AM, where \( A \) is the atomic number of the ion), the two fields have a large residual deflecting effect, and the ions are carried away either to an electrode or towards the wall of the tube.

Another consequence of equation (4.4) is that the trajectory is independent of the absolute dimensions of the lens: only the ratio \( l/d \) matters. Similarly, only the ratio \( U/\Phi \), the ratio of electrode potential to accelerating potential, intervenes in (4.4) and not the absolute value of either potential. The action of an electrostatic lens will not be affected by small variations of \( U \) and \( \Phi \), which can even be oscillatory provided their ratio remains unaltered. It seems as though it should be possible to connect the electrodes to a potentiometer fed by an unstable or even oscillatory supply. In certain of the cruder applications of electron optics, high tension X-ray tubes, for example,
advantage is taken of this property by supplying the tube with the alternating voltage from a transformer. In more delicate applications, on the other hand, such as the electron microscope, only an extremely carefully stabilized high tension can be used, not varying by more than one per cent; so fine are the images which such an instrument produces that a defect due to relativity would otherwise appear. This is explained by the fact that for very fast electrons, equation (4.4) requires a relativistic correction, and becomes:

$$\alpha = \frac{Ul}{2\Phi d} \left( 1 + \frac{(e/m_0c^2)\Phi}{1 + (e/2m_0c^2)\Phi} \right)$$

(4.5)

(where \(c\) is the velocity of light and \(m_0\) is the rest mass of the electron). Equation (4.4) ceases to be linear as soon as \(\frac{e\Phi}{2m_0c^2}\), which is equal to \(0.977 \times 10^{-3} \Phi\) (\(\Phi\) measured in kV), is no longer negligible (Ramberg, 1942). For the optical properties to remain constant it is no longer sufficient that such ratios as \(\Phi_1/\Phi_0\), \(\Phi_2/\Phi_0\) should be invariant, but the actual potentials \(\Phi_1\), \(\Phi_2\) too must be narrowly stabilized; otherwise, the image will be slightly hazy as a result of the slight variations of these potentials about their mean values. In fact, however, this relativistic effect is most difficult to observe, masked as it is by more ordinary defects caused by imperfections in the source of high tension; for example, the capacities at the high tension entry and between the electrodes may be appreciable, and, by destroying the validity of the proportionality law in the presence of a varying potential, may well dwarf the relativistic effect.

4.2 THE GAUSSIAN APPROXIMATION

4.2.1 Gaussian Conditions

We have seen that the necessary conditions for a good image to be formed by an electron optical system obtain only within a tubular surface of revolution enclosing the axis of the system rather closely. This is the situation in which we can neglect terms in \(r\) and \(dr/dz\) of orders higher than the first in the differential equations of the electron trajectories. This first order approximation is known as "Gaussian optics", and the approximate trajectories which are obtained are called "Gaussian trajectories".

Equation (4.2) gives the lower order terms in the series expansion for \(E_r\) and \(E_z\), which we shall write for convenience

$$E_z = -\varphi'(z), \quad E_r = \frac{1}{2} r \varphi''(z);$$

(4.6)

\(\varphi(z), \varphi'(z)\) and \(\varphi''(z)\) represent the value, \(\Phi(0, z)\), of the potential on the axis, and its derivatives with respect to \(z\).
4.2.2 The Equations of Motion

Let us consider now the detailed calculation of trajectories to the Gaussian approximation for non-relativistic electrons. Applying Newton's law in the radial direction we obtain

$$f_r = m \frac{d^2 r}{dt^2} = -\frac{1}{2} e r \varphi''(z)$$  \hspace{1cm} (4.7)

(remembering that the charge on the electron is $-e$).

The radial force has a focusing action when $\varphi''(z) > 0$ and a defocusing action when $\varphi''(z) < 0$. If the curve which represents $\varphi(z)$ is plotted with the cathode potential as the origin of coordinates ($\varphi = 0$), the lens will be locally convergent at a point with abscissa $z$ if the curve is concave upwards at this point, and divergent if the curve is concave downwards (Fig. 22).

For the motion in the $z$ direction, we obtain a first order differential equation directly from the conservation of energy:

$$\frac{1}{2} m \left( \frac{dz}{dt} \right)^2 = e \varphi.$$  \hspace{1cm} (4.8)

With the aid of equations (4.7) and (4.8), we can study such electron trajectories as remain within a meridian plane; this is the case for all rays originating from an axial object point, and for a particular class of rays originating in an off-axial object point. For this case, we find, eliminating the time variable between the equations,

$$\sqrt{\varphi} \cdot \frac{d}{dz} \left( \sqrt{\varphi} \cdot \frac{dr}{dz} \right) = -\frac{1}{4} r \varphi'',$$

or

$$\frac{d^2 r}{dz^2} + \frac{r'}{2\varphi} \frac{dr}{dz} + \frac{r''}{4\varphi} r = 0.$$  \hspace{1cm} (4.9)

Skew trajectories are, in fact, no problem as it is simple to transform their differential equations into an equation of the form (4.9) or (4.10). If $Ox$ and $Oy$ are mutually perpendicular axes in a plane perpendicular to $Oz$, we can describe quantities which depend linearly upon $x$ and $y$ in a condensed fashion by introducing the complex variable $u = x + jy$:

$$\frac{du}{dt} = \frac{dx}{dt} + j \frac{dy}{dt},$$

$$\frac{d^2 u}{dt^2} = \frac{d^2 x}{dt^2} + j \frac{d^2 y}{dt^2},$$  \hspace{1cm} (4.11)

$$E_u = E_x + j E_y = -\frac{1}{2} u \frac{\partial E_z}{\partial z}.$$
The equation which relates \( u \) and \( z \) in space relates \( r \) and \( z \) in a meridian plane. Writing \( u \) for \( r \) and interpreting the results with the aid of equations (4.11), oblique trajectories are found to have the same optical properties as the meridian trajectories to which we had previously restricted ourselves.

The similarity properties are immediately obvious if we notice that the equations are homogeneous and of zero degree with respect to \( \varphi \), and that neither \( e \) nor \( m \) appears (we shall consider the effect of relativity upon the homogeneity later).

We now return to the Gaussian equation in the form (4.9), and integrate between object space (characterized by the index \( a \)) and image space (index \( b \)). We find

\[
\left( \sqrt{\varphi} \frac{dr}{dz} \right)_b - \left( \sqrt{\varphi} \frac{dr}{dz} \right)_a = -\frac{1}{4} \int_a^b r \frac{\varphi''}{\sqrt{\varphi}} \, dz.
\]

For an incident trajectory parallel to the axis, \( r'_a = 0 \), and hence

\[
r'_b = -\frac{1}{4\sqrt{\varphi}_b} \int_a^b r \frac{\varphi''}{\sqrt{\varphi}} \, dz.
\]

The image focal length, \( f_b \), is given by

\[
\frac{1}{f_b} = -\frac{r'_b}{r_a},
\]

and hence

\[
\frac{r_a}{f_b} = \frac{1}{4\sqrt{\varphi}_b} \int_a^b r \frac{\varphi''}{\sqrt{\varphi}} \, dz. \tag{4.12a}
\]

In a weak lens, \( r \) varies only slightly through the lens, and we can write approximately

\[
r = r_a.
\]

For a thin lens, therefore, we obtain the following expression for the image focal length, \( f_b \):

\[
\frac{1}{f_b} = \frac{1}{4\sqrt{\varphi}_b} \int_a^b \frac{\varphi''}{\sqrt{\varphi}} \, dz. \tag{4.12b}
\]

If \( \varphi' \) is zero on both sides of the lens, the second derivative can be eliminated on integrating by parts, and we obtain

\[
\frac{1}{f_b} = \frac{1}{8\sqrt{\varphi}_b} \int_a^b \frac{\varphi'^2}{q^{3/2}} \, dz. \tag{4.12c}
\]
4.2.3 The Reduced Equation

The Gaussian equation is difficult to use due to the presence of $\varphi''$, as for most lenses the only information available about the potential distribution has been obtained from an electrolytic tank or a resistance network and a second derivative determined graphically is far too inaccurate. Instead, therefore, it is often considered preferable to transform the Gaussian equation (Picht, 1939) by writing

$$R = r\varphi^{3/4}.$$  \hspace{1cm} (4.13)

The curve $R(z)$ is called a “reduced” ray. Equation (4.10) then becomes

$$\frac{d^2R}{dz^2} + \frac{3}{16} \left( \frac{\varphi'}{\varphi} \right)^2 R = 0,$$  \hspace{1cm} (4.14)

which, apart from the absence of an $R'$ term, displays the interesting feature of possessing a single expression characteristic of the lens. All the Gaussian optical properties of the lens are determined by the characteristic function

$$T(z) = \frac{\varphi'}{\varphi}.$$ 

For a weak lens, we can treat $R$ in the same way as earlier we treated $r$. A ray incident parallel to the axis corresponds to $R_a = \text{constant}$ and $R'_a = 0$. The “reduced focal length”, $F$, is defined in terms of the quotient $R'_b/R_a$ thus:

$$\frac{1}{F} = -\frac{R'_b}{R_a}.$$ 

If we assume that $R \simeq \text{constant} \simeq R_a$ throughout the lens, we find

$$\frac{1}{F} = \frac{3}{16} \int_a^b \left( \frac{\varphi'}{\varphi} \right)^2 dz.$$  \hspace{1cm} (4.15a)

If the potentials are constant on both sides of the lens, we have

$$R'_b = r'_b \varphi'_b^{3/4},$$

$$f_b = \frac{r'_b}{r_a} = \left( \frac{\varphi'_b}{\varphi_a} \right)^{1/4} F_b,$$

and hence

$$\frac{1}{f_b} = \frac{3}{16} \left( \frac{\varphi_a}{\varphi_b} \right)^{1/4} \int_a^b \left( \frac{\varphi'}{\varphi} \right)^2 dz,$$  \hspace{1cm} (4.15b)

which reduces, if $\varphi_a = \varphi_b$, to the following expression:

$$\frac{1}{f_b} = \frac{3}{16} \int_a^b \left( \frac{\varphi'}{\varphi} \right)^2 dz.$$  \hspace{1cm} (4.15c)

Formula (4.15c) differs from formula (4.12c) even though a priori equivalent approximations have been made; in practice, (4.12c) is found very often
to lead to incorrect results, whereas (4.15b) yields values of $f_k$ which are very close to those which are obtained by integrating the equations of motion numerically. This somewhat subtle point has been discussed by Sturrock at the London Conference on Electron Optics (1954) and more recently by Felici (1959), who gives the limits within which these formulae are in practice valid.

![Diagram](image)

**Fig. 22.** The relation between the form of $\psi(z)$ and the force $F$ which acts on an electron. (a) "Einzel" lens. (b) Accelerating lens.

The approximation which leads to formula (4.12c) is valid only when $\psi''(z)$ has the same sign throughout the region of integration, and this is never the case with ordinary lenses—only lenses with a central grid satisfy this requirement (§ 8.4). If, however, we break down the curve $\psi''(z)$ into regions where $\psi'' > 0$ and regions where $\psi'' < 0$, we can find a thin lens equivalent to each region and hence calculate the focal length. The thin lens will lie at the centre of gravity of the curve (that is, the point such that the areas on either side are the same). By applying the rules obeyed by combinations of thin lenses, we can determine the elements of the real lens. This method has been used mostly for two-cylinder lenses which reduce to a convergent lens followed by a divergent lens when the complete lens is accelerating (Spangenberg and Field, 1942).

The same method of breaking down into thin lenses can be applied to the function $T^2(z) = \left( \frac{\psi'}{\psi} \right)^2$. We first obtain the centre of gravity, $G$, of this function and then the centres of gravity of the two regions of the curve which lie on either side of $G$; at the latter we place two thin lenses $L_1$ and $L_2$, with the focal lengths given by (4.15b). The real lens is equivalent to the combination of $L_1$ and $L_2$, to within a few per cent, even for very strong lenses (Felici, 1959). The various integrations can be performed graphically and the method is particularly rapid in the symmetrical einzel lens case.

### 4.2.4 The Immersion Objective

A new difficulty appears when the object itself is a source of thermal electrons, or of electrons produced either photo-electrically or by secondary emission, for these particles are emitted in all directions and the conditions under which the Gaussian approximation is valid seem no longer to be
satisfied. The problem is only apparent, however, as the initial velocities are very low (corresponding to energies of the order of a few tenths of an electron volt) and the field is initially normal to the surface (which is equivalent to being parallel to the axis), very intense and almost uniform. In such conditions, the initial trajectory is parabolic in shape, and the velocity changes direction very rapidly along the curve until it becomes nearly parallel to the axis in a plane $z = z_1$, a few tenths of a millimetre from the cathode.

At this plane, rays which have in fact originated in a point on the cathode such as $A$ (Fig. 23) appear to be coming in a straight line from a small disc around $A_1$, situated at $z = -z_1$; the radius of this disc is $q = \frac{2\varphi_c}{E_c}$, in which $\varphi_c$ is the voltage equivalent to the most probable initial electron velocity ($\varphi_c = T/11600$ where $T$ is the absolute cathode temperature) and $E_c$ is the field on the cathode (Langmuir, 1937).

The radius of this disc is usually of negligible extent, lying as it does between $2\, \mu m$ and $0.02\, \mu m$. The angular aperture of the beam corresponding to each point on the cathode is also small, of the order of some hundredths of a radian: $2\alpha = 2\sqrt{\frac{\varphi_c}{E_c z_1}}$. Beyond $z = z_1$, the conditions are totally Gaussian, and the lens needs no exceptional treatment. Experimental work has confirmed this theory, and electron or ion emission microscopes working at $30\, kV$ give finer images than the best optical microscope (Mecklenburg, 1939).
1942; Kinder, 1944; Septier and Gauzit, 1950; Couchet, Gauzit and Septier, 1951, 1952, 1954; Septier, 1953; Möllenstedt and Hubig, 1954; Fert, 1956; see also Chapter 13).

4.2.5 Electron Mirrors

The usual reasoning seems to be inapplicable here as the region in which the electrons are reflected is a region in which the momentum is perpendicular to the axis, so that although $dr/dz$ is initially finite, it becomes infinite at the point of reflexion. In any event, it certainly does not remain infinitesimally small while the electrons are within the lens. Since it was as a consequence of this hypothesis that we were able to establish equation (4.9), it is to be feared that neither (4.9) nor the reduced equation which has been derived from it is capable of describing the behaviour of electrons in a mirror. However, only the reduced equation proves to be invalidated, while, as a careful examination of the higher order terms shows, equation (4.9) remains sound.

There are two alternative ways of studying mirrors. In the first, the system is decomposed into two regions, a narrow zone around the reflexion equipotential, where the field is regarded as uniform and the trajectories as parabola, and a large zone consisting of the remainder of the system where the Gaussian conditions are satisfied; the usual equations can then be used with impunity, and by joining up the Gaussian and parabolic trajectories continuously, we can enumerate the optical properties.

The alternative way of studying electron mirrors is to choose another parameter to describe the trajectory. Up till now, we have used as abscissa the distance $z$, which increases monotonically with time within the lens. In mirrors, on the contrary, $z$ is no longer a monotonic function of time, but it would be advantageous to choose one that is, and to express $z$ and $r$ in terms of this parameter. The time itself could be selected for this parameter, or, as Bernard has suggested (1952a), we could make the substitution $z = s^2$, with the origin of $z$ at the point of reflexion so that the incident ray is characterized by $-\infty < s < +0$. The quantity $dr/ds$ remains small, and never goes to infinity. An equation giving $r$ as a function of $s$ is easily derived, with solutions which are parametric representations of the trajectories.

Again, it is possible to obtain a reduced equation, analogous to the Pichg equation, from which the optical properties of the mirror prove to be wholly defined by the single function:

$$\mathcal{L} = \left( \frac{1}{\varphi} \frac{d\varphi}{ds} \right)^2 - \frac{4}{s^2},$$

where $\varphi(s)$ is the axial potential, with origin at the point of reflexion.
Electron mirrors are rarely used in practice, interesting though some of their properties are as we shall show later, since it is difficult to insert a receiver (fluorescent screen or photographic plate) into the path of the reflected beam which is in practice mingled with the incident beam. Mahl (1943), using electron mirrors in a microscope, obtained images comparable in quality with the images given by an ordinary lens.

The analysis when the point of reflexion is close to the centre of a symmetrical lens is more difficult, as the field falls to zero with the potential. This is a very rare situation in practice, however, as the images which result are unstable.

4.3 OPTICAL FORMALISM

4.3.1 Image Formation

For every thick lens, the ability to form an image is a general consequence of the linearity of equation (4.10).

If we know the trajectories \( r_1 = X(z) \), \( r_2 = Y(z) \) of two independent rays (rays, that is, between the equations of which no linear relation exists), we can deduce immediately from the theory of linear differential equations that the equation of every other ray is of the form

\[
r = \lambda X(z) + \mu Y(z)
\]

where \( \lambda \) and \( \mu \) are arbitrary constants.

![Diagram](image)

Fig. 24. The two characteristic rays of Gaussian optics (\( G \) is the magnification).

\textit{A priori}, the trajectories \( r_1, r_2 \) are arbitrary, but they will reappear frequently, and it is of interest to select the two particular functions shown in Fig. 24: \( X(z) \) intersects the axis at \( z = z_0 \) (where the object plane intersects the \( z \)-axis) with slope unity—i.e. \( \left( \frac{dX}{dz} \right)_{z = z_0} = 1 \); \( Y(z) \) passes through the object plane at unit distance from the axis, which it subsequently intersects in the plane of the aperture diaphragm, the function of which is to restrict the lateral extent of the beam.
This choice of \( Y(z) \) is standard in electron optical calculations, but it is not, it must be noticed, the second fundamental trajectory habitually used in elementary optics for graphical constructions. There, a trajectory which leaves the object plane parallel to the axis is chosen.

We can make a number of straightforward deductions from equation (4.16). Every ray leaving a given point object \( A \) converges into a real point image \( A' \), after passing through the lens. If the rays do not converge, then we can produce back the rays which in fact emerge from the lens as straight lines (these lines are broken in the diagrams, to conform with the glass optical convention); mathematically, \( X(z) \) and \( Y(z) \) represent both the actual rays and their produced sections, and the "virtual" rays too, therefore, obey a linear relation similar to equation (4.16) (no problem arises from the double role which \( X \) and \( Y \) play), and converge into a virtual image point. Object point and image point are known as "conjugate points". The points which are conjugate to the points on the axis at plus or minus infinity are called the two "foci" of the lens.

From equation (4.16), we can also deduce information about the position and magnification of the image. The image of some object lying in a plane perpendicular to the axis is situated in the plane, perpendicular to the axis, containing the conjugate point (the frontal plane); and the image is similar to the object. For every pair of conjugate planes, the linear magnification of the image is a constant; the angular magnification, which is defined as the ratio of the angular aperture of an elementary pencil at the image to that of the corresponding pencil at the object, is also constant between conjugate planes.

We shall now establish an interesting relation between \( X(z) \) and \( Y(z) \), which will prove to be useful when we calculate the spherical aberration in § 7.3. \( X(z) \) and \( Y(z) \) are both solutions of equation (4.9), thus:

\[
\frac{d}{dz} \left( \sqrt{\varphi} \frac{dX}{dz} \right) + \frac{1}{4} \frac{\varphi''}{\varphi} X = 0,
\]

\[
\frac{d}{dz} \left( \sqrt{\varphi} \frac{dY}{dz} \right) + \frac{1}{4} \frac{\varphi''}{\varphi} Y = 0.
\]

We can eliminate the term which contains \( \varphi'' \) by multiplying the equations by \( Y(z) \) and \( X(z) \) respectively and subtracting:

\[
Y \frac{d}{dz} \left( \sqrt{\varphi} \frac{dX}{dz} \right) - X \frac{d}{dz} \left( \sqrt{\varphi} \frac{dY}{dz} \right) = 0,
\]

or

\[
\frac{d}{dz} \left[ \sqrt{\varphi} \left( Y \frac{dX}{dz} - X \frac{dY}{dz} \right) \right] = 0,
\]

or, integrating,

\[
\sqrt{\varphi} \left( Y \frac{dX}{dz} - X \frac{dY}{dz} \right) = \sqrt{\varphi_0} \left[ Y_0 \left( \frac{dX}{dz} \right)_0 - X_0 \left( \frac{dY}{dz} \right)_0 \right],
\]

in which the index zero refers to the object point in question.
For the two particular trajectories which we envisage here, we have
\[ X_0 \left( \frac{dY}{dz} \right)_0 = 0 \]
and
\[ Y_0 \left( \frac{dX}{dz} \right)_0 = 1, \]
so that finally,
\[ Y \frac{dX}{dz} - X \frac{dY}{dz} = \sqrt{\frac{q_0}{q}}. \] (4.16‴)

4.3.2 The Standard Cardinal Points

In classical electron optics, we postulate that the object and the image shall lie in regions where the field is zero and the trajectories are straight lines. Physically, this is a most useful approximation as the field has effectively vanished beyond a few lens-diameters from the lens, although it still exists mathematically. In general, in electron optics, lenses have a finite and well-defined thickness, which is the region throughout which \( E \)—and later, \( B^2 \)—has an appreciable effect upon the electrons. Mathematically speaking, electron optics treats not the trajectories themselves, but only their asymptotes, the straight lines to which the rays tend. The classical method leaves us with no information about the curved parts of the trajectories which lie within the lens; nor may we consider objects or images lying within the lens, for in this region we can only deal with virtual objects and images formed by producing real rays in straight (dotted) lines. Provided, therefore, we are willing to sacrifice real immersed objects and images, the standard cardinal points provide a most fruitful method of characterizing the optical properties of a lens system. Nevertheless, immersed virtual images and objects are intrinsically interesting—the objective of an electron microscope, for example, produces an intermediate image which is usually located within the projector lens, and plays, for this lens, the role of a virtual object.

Let us now consider a ray \( y(z) \) which is parallel to the axis in object space; in image space, it intersects the axis at the image focus. Another ray, \( U(z) \), also parallel to the axis in object space, will then be defined by \( r = U(z) = \mu y(z) \). Whatever the value of \( \mu \) or \( U \), therefore, \( U(z_0)/U'(z_i) = \text{constant} \). In optical notation, this is equivalent to \( h_0/s_l = f_i \), which defines the image side focal length \( f_i \) (Fig. 25). These equations express the fact that if the beginning and end of each ray \( U \) is produced, the points of intersection lie in a fixed plane, known as the image (side) principal plane; an object (side) principal plane is defined in an exactly similar way by considering the family of rays which are concurrent at the object focus and parallel to the axis in image space \([V(z)]\).

Once the two principal planes have been located and the two focal lengths are known, the optical properties of the system are completely
determined; with the aid of simple geometrical constructions and the formulae to be found in every elementary book on optics, we can obtain the emergent ray corresponding to any given incident ray, and the position and magnification of the image of a given object. The relevant formulae are the following:

(i) With the origins at the principal planes,

\[
\frac{F_0}{p} = f_0, \quad \frac{F_i}{p'} = f_i, \quad \frac{F_0 A}{p_1} = p_1, \quad \frac{F_i A'}{p'} = p',
\]

\[
f_0 + \frac{f_i}{p'} = 1, \quad \frac{h_i}{h_0} = -\frac{f_0}{f_i} \frac{p'}{p}, \quad \frac{f_i}{f_0} = -\sqrt{\frac{\varphi_i}{\varphi_0}}.
\] (4.17)

(ii) With the origins at the foci,

\[
\frac{F_0}{p} = f_0, \quad \frac{F_i}{p'} = f_i, \quad \frac{F_0 A}{\pi} = \pi, \quad \frac{F_i A'}{\pi'} = \pi', \quad \pi \pi' = f_i f_0,
\]

\[
\frac{h_i}{h_0} = \frac{f_0}{\pi} = \frac{\pi'}{f_i}, \quad \frac{f_i}{f_0} = -\sqrt{\frac{\varphi_i}{\varphi_0}}.
\] (4.18)

A rather more subtle consequence of the differential effect which has already been described in the elementary explanation of lens convergence is that the principal planes are always crossed, as in Fig. 25. The reader will find a proof of this property, which follows straightforwardly from the structure of the reduced equation, in the book by Zworykin et al.

4.3.3 The Lagrange–Helmholtz Relation

This is the name given to the formula

\[
\frac{f_i}{f_0} = -\sqrt{\frac{\varphi_i}{\varphi_0}},
\] (4.19)
which relates the two focal distances. It is obtained by combining the differential equations (4.10) for the rays $U(z)$ and $V(z)$ symmetrically, thus eliminating the term in $r$, and integrating, which gives

$$U \frac{dV}{dz} \sqrt{\nu} - V \frac{dU}{dz} \sqrt{\nu} = \text{constant.}$$

(4.20)

Writing out this expression in the object and image planes in optical notation (Fig. 25), we find:

$$\sqrt{\nu_o} \cdot h_o \alpha_o = \sqrt{\nu_i} \cdot h_i \alpha_i.$$  

(4.21)

![Figure 26. Image cardinal elements (F and P are the normal cardinal elements, G and Q their immersion counterparts).](image)

Equation (4.21) summarizes the fact that the product of angular magnification ($\alpha_i/\alpha_o$) and linear magnification ($h_i/h_o$) is equal to the constant $\sqrt{\nu_o/\nu_i}$. Expressing the magnifications in terms of the cardinal elements, we obtain equation (4.19); two values of $\varphi$ suffice to characterize object space and image space.

### 4.3.4 Objects and Images Lying within the Lens Field (“Immersed” Elements)

In glass optics, the curved parts of the rays are of no special interest, since they degenerate into sudden bends at the boundaries between each pair of media. The standard cardinal points are wholly adequate. In electron optics, on the other hand, it often happens that a lens is thick and sufficiently convergent to produce an image which possesses all the properties described in section 4.3.1 inside the lens. For example, a symmetrical lens may be perfectly capable of refracting a ray which is incident parallel to the axis so sharply that it intersects the axis inside the lens at some point $G$, which can be thought of as a real immersed focus (Fig. 26). This class of images and objects is of practical interest only for magnetic lenses; in electrostatic lenses, a specimen (the object), a fluorescent screen, or a photographic plate (as receiver) would perturb the electric field to such an extent that an image such as $A_1 B_1$ (Fig. 27) could not be observed.

So long as we consider only the rectilinear asymptotes to the rays, the analysis remains simple, concerned as it is with linearly related straight
lines which cut the axis in points which are defined mathematically by an elementary homographic relation. Generally speaking, however, no method of this kind can be devised which will allow us to consider effects produced by the curved part of the rays. No simple relation exists between the real immersed images (or objects) of the preceding paragraph and their corresponding objects (or images). If we wish to consider image formation actually inside the lens, we must resort to direct calculation of the two rays, starting from the differential equation. The Lagrange–Helmholtz relation (4.21) is still valid, but $q_0$ and $q_i$ now represent $\varphi$ at the particular pair of conjugate points in question.

![Fig. 27. The “classical” image $A'B'$ and the immersion image $A_1B_1$ of the virtual object $AB$.](image)

A class of fields exists, however, known as “Newtonian”, for which relations of the classical form remain applicable, even for immersion objects and images (Glaser, 1941b; Glaser and Lammel, 1941; Hutter, 1945; Glaser, 1950b). The most interesting feature of this theory is that one of the members of the class provides rather a good approximation to the field within an ordinary magnetic lens. Whether the points lie within or without the lens, the relations remain valid, and the Newtonian expressions give the magnification and position of the image.

The relations can be represented geometrically by the usual constructions, provided that we define the foci as the two real points $G_i$ and $G_0$ in which rays, initially parallel to the axis, effectively cross the axis for the first time, and that we define $Q_i$ and $Q_0$, the principal planes, by $\overrightarrow{G_0Q_0} = g_0$ and $\overrightarrow{G_iQ_i} = g_i$; the quantities $g_0$ and $g_i$ are the focal lengths defined by

$$g_0 = \frac{r(+\infty)}{r'(G_0)}, \quad g_i = \frac{r(-\infty)}{r'(G_i)}$$

(see Fig. 27). It is to be noticed, however, that these straight lines are not at all an indication of the path of the actual curved rays, and are not related in any direct way to the classical points $F_i, F_0, P_i, \text{ and } P_0$. The Lagrange–Helmholtz relation (4.21) may no longer be put into the form (4.19).
4.3.5 Photometry; Brightness

Electron photometry can be developed from the Lagrange–Helmholtz relation in the form (4.21)—the angular magnification constant takes the same value for oblique rays as it does for meridian rays.

A more suggestive form of the result is obtained by considering the area which the whole beam covers in a particular forward plane; this surface can be regarded as a diaphragm (of radius R) in a plane. If we introduce the image D' (of radius R') of D, which is defined by that part of the system which precedes D, the rays between D' and the object are straight lines; if d is the distance between the object and D', the value of α at the object is α₀ = R'/d (Fig. 28). We can consider the first term, \( \sqrt{\varphi_0} h_0 \frac{R'}{d} \), of the Lagrange–Helmholtz relation as the product of \( R' \sqrt{\varphi_0} \) and the aperture \( \alpha' = h_0/d \) of an elementary beam in the D' plane:

\[ h_0 \alpha_0 \sqrt{\varphi_0} = R' \alpha' \sqrt{\varphi_0}. \]

Applying the Lagrange–Helmholtz relation to the conjugate planes D and D', and recalling that D' lies in object space, we obtain

\[ R' \alpha' \sqrt{\varphi_0} = R \alpha_D \sqrt{\varphi_D}, \]

and thus

\[ h_0 \alpha_0 \sqrt{\varphi_0} = R \alpha_D \sqrt{\varphi_D}. \] (4.22)

The product \( R \alpha_D \sqrt{\varphi_D} \) is constant irrespective of the choice of the plane D (R is the radius of the cross-section of the beam, and \( \alpha_D \) is the angular aperture of an elementary beam in the plane D). The most interesting application of formula (4.22) is to electron sources (definitions are given in § 8.3). If we assume that the cathode is effectively plane, and of radius \( r_c \), and take into account the features peculiar to immersion objectives already
discussed \((h_0 = r_c; \alpha_0 = \sqrt{\varphi_c/E_c z_1}; \varphi_0 = E_c z_1)\), we obtain an expression for the minimum possible radius \(R_s\) of the spot in a cathode ray tube working at final voltage \(\varphi_s\) and final angular aperture \(\alpha_s\), which is of the form

\[
r_c \sqrt{\frac{T}{11,600}} = R_s \alpha_s \sqrt{\varphi_s}
\]

where we have substituted \(\varphi_c = \frac{kT}{e} = \frac{T}{11,600}\).

In terms of specific emission, \(i_e\), and current density at the screen, \(i_s\) (defined by \(I_s = \pi r_c^2 i_e = \pi R_0^2 i_s\), \(I_s\) being the actual total current), we find

\[
\frac{i_s}{i_e} = \left(\frac{11,600 \varphi_s}{T}\right) \alpha_s^2,
\]

the celebrated formula of Langmuir (1937), perfected by Pierce (1939–45). Since \(\alpha_s\) and \(r_s\) are related, however, (4.24) can be simplified. Let us consider, first of all, the image focal plane of an immersion objective, where the beam crosses the axis for the first time (at the "crossover"). Applying (4.23), we obtain

\[
\alpha_f = \frac{r_c}{f}, \quad r_f = f \sqrt{\frac{T}{11,600 \varphi_f}}
\]

where \(f\) is the image focal distance, \(r_f\) the crossover radius, \(\alpha_s\) the angular aperture of the beam at the focus, and \(\varphi_f\) the potential at this point.

If the second immersion lens divides the distance between the image focal plane and the screen into two parts, the ratio of whose lengths is \(m\), then (see equation 4.17)

\[
R_s = m r_f \sqrt{\frac{\varphi_f}{\varphi_s}} = m f \sqrt{\frac{T}{11,600 \varphi_s}}.
\]

This expression is valid so long as \(r_c\) and \(\alpha\) are small enough for spherical aberration to be negligible in the objective and in the second lens (otherwise, the Lagrange–Helmholtz relation loses its simple first order form). We can satisfy this requirement, at least partially, by choosing a suitable cathode design, or by stopping down the beam in the second lens.

Following Law (1937), a more detailed study would take into consideration the Maxwellian distribution of the emission velocities, and give us the current distribution in the spot. The result is that at a distance \(r\) from the centre, the density is given by

\[
i(r) = \left(\frac{I_s}{\pi m^2 f^2}\right) \frac{11,600 \varphi_s}{T} \exp \left\{\left(\frac{11,600 \varphi_s}{T}\right) \left(1 - \frac{r^2}{m f^2}\right)\right\}
\]

or

\[
i(r) = i_s e^{-r/\langle R_0 \rangle^2}.
\]
The corresponding curve is bell-shaped, with a "natural width" of $2R_s$. In the crossover plane, we have a similar expression: $i(r) = i_{f}e^{-\alpha/r_{f}}$. Experiment confirms this result perfectly, save for the absolute value of $f$ which depends upon the space charge correction which is applied (Dosse, 1940; Jacob, 1939). The saturation current density $i_s$ has been obtained in a number of guns with tungsten hairpin filaments; so strong is the influence of space charge on oxide-coated cathodes, however, that the emission is controlled by the quality of the vacuum within the sealed-off tube (Klemperer and Mayo, 1948). Finally, complex oscillatory phenomena may appear, which are only just beginning to be understood, which break up the beam into separate pencils (Klemperer, 1947; Ellis, 1947). Another example of the application of this geometrical photometry is the reduction of the size of the spot by "post-acceleration" (see equation 4.27); Pierce (1942) has furnished a complete treatment of this effect.

Just as in ordinary optics, we can define the brightness of an electron source as the electron intensity emitted per unit area of the surface into unit solid angle. In many electron optical instruments, the crossover can be regarded as the real source of the electrons. Its theoretical brightness in an infinitely small central zone of radius $r_s$ is defined by

$$B_{th} = \frac{I}{\pi r_s^2 \pi \alpha^2} = \frac{i_s}{\pi \alpha^2},$$

(4.27')

or using (4.24),

$$B_{th} = \frac{i_c}{\pi} \frac{q_s}{q_e},$$

or finally

$$B_{th} = i_c \frac{e q_s}{\pi k T} = i_c \left(\frac{11,600}{\pi T}\right) q_s.$$

(4.27'')

As an example, suppose we have a tungsten filament, at a temperature of 2700 K, and $i_c = 2 \text{ A cm}^{-2}$; if $q_s = 30 \text{ kV}$, we find

$$B_{th} \sim 8 \times 10^4 \text{ A cm}^{-2} \text{ sterad}^{-1}.$$

Similarly, we can define a mean brightness $B_m$ for the whole area of the crossover; experiment, however, shows that $B_m$ is considerably lower than $B_{th}$:

$$B_m \sim \frac{1}{10} B_{th}.$$

If we form successive images of the crossover with the aid of an aberration-free optical system in regions in which the potential $q_s$ is the same, the theoretical brightness of the successive images remains constant and equal to $B_{th}$; this is, in fact, a consequence of the Lagrange–Helmholtz relation.

To increase $B_{th}$ for a given value $q_s$, we can only increase $i_c$, the specific emission of the cathode, and reduce $T$. If an extremely bright source is
required, we use only the central region of the crossover, rejecting the remainder with the aid of a diaphragm very small in diameter \( d \leq \frac{1}{10} \text{ mm} \); the measured brightness is of the order of 80 or 90 per cent of \( B_{th} \). Brightnesses of the order of \( 2 \times 10^8 \text{ A cm}^{-2} \text{ sterad}^{-1} \) have recently been attained, with the aid of heated point cathodes (T.F. emission, see §17.1.2).

4.4 RELATIVISTIC LENSES

At high speeds, relativistic mechanics is necessary to describe the behaviour of the electrons. Although the equations are more complex, the general line of reasoning and the results are the same. Starting rather from the Einstein equations than from the Newtonian ones, the existence of first order focusing can be demonstrated, whatever the electron velocity. In effect, the differential equations remain linear.

If we write

\[
\varepsilon(z) = \left( \frac{e}{2m_0 c^2} \right) \varphi(z)
\]

with

\[
\frac{e}{2m_0 c^2} = 0.977 \times 10^{-3} \text{ (kV)}^{-1},
\]

the relativistic Gaussian equation has the form

\[
\frac{1 + \varepsilon}{1 + 2\varepsilon} \varphi r'' + \frac{1}{2} \varphi' r' + \frac{1}{4} \varphi'' r = 0. \tag{4.28}
\]

The loss of homogeneity appears in the coefficient of the first term. If we substitute

\[
R = r \varphi^{3/4}(1 + \varepsilon)^{1/4}, \tag{4.29}
\]

we obtain a reduced equation which resembles (4.14) save that the coefficient \( \frac{3}{16} \left( \frac{\varphi'}{\varphi} \right)^2 \) which appears is now multiplied by the factor \( K(\varepsilon) \), defined by

\[
K(\varepsilon) = \left( \frac{1 + \frac{4\varepsilon}{3} + \frac{4\varepsilon^2}{3}}{(1 + \varepsilon)^2} \right). \tag{4.30}
\]

As a first approximation, we can neglect the variation of \( z \) in \( \varepsilon(z) \), and replace the latter everywhere by a mean value \( \bar{\varepsilon} \). Equation (4.30) shows that if this mean value rises, the correction factor \( K \) decreases from the value unity at low voltages to a minimum of 0.88 at 600 kV, then begins to rise, passing unity again at 2 MV and tending towards 4/3 at still higher voltages. We ought, therefore, to be able to produce an initial increase of focal length followed by a slow decrease towards an asymptotic value by multiplying
gun and lens potentials by the same (increasing) factor. Exact calculation
confirms this qualitative prediction, as shown by Fig. 29 which represents
the behaviour of a three-electrode unipotential lens (Laplume, 1947; Cotte, 1939).

The danger of breakdown prevents the use of very high potential differ-
ences. The systematic investigations of Arnal (1953) allow the optimum
choice of potential to be made in a given electrostatic system. It seems

![Graph](image)

**Fig. 29.** The focal length and the position of the focus as a function of the maxi-
mum potential in the system; the other parameters are constant.

that adjacent electrodes will not support much more than 100 kV. In
the megavolt region, therefore, grid lenses, with their lower potentials,
become of interest (Gabor, 1947; Cartan, 1937; Knoll and Weichart,
1938; Bernard, 1951a, 1952a, 1953a).

### 4.5 PRACTICAL DETERMINATION OF ELECTRON
TRAJECTORIES

#### 4.5.1 Integration of the Ray Equation

The expression for the axial potential is rarely sufficiently simple for the
Gaussian equation which results from it to be directly integrable. Worse,
in most of the experimental arrangements of interest, it is not even pos-
sible to obtain an expression for the potential; only the experimental
curves provided by an analogue method are available. On the other hand,
it is often possible to find an expression which is both a very good approxi-
mation to the (calculated or measured) axial potential and sufficiently
simple for the Gaussian equation to be soluble in terms of tabulated func-
tions. This solves the problem, as we then have expressions for the tra-
jectories.
In the chapter devoted to lens models, more details of these representations are to be found, and in particular, details of the procedures by which model and experiment may be matched. Here, we shall simply mention the works of: Rüdenberg (1948) and Regenstreif (1951a), whose representation of the axial potential by a chain of parabolic segments allows the Gaussian equation (4.10) to be solved; of Hutter (1945) and Bernard and Grivet (1952a), who have expressed the trajectories in a two-electrode lens in terms of sinusoidal functions; and of Grivet (1952a) and Lenz (1951) who have obtained the trajectories in the same lens in terms of Legendre polynomials.

The latter methods effect an integration of the reduced equation (4.14).

4.5.2 Methods of Step-by-step Integration

In this section, the necessary practical information, with the aid of which the electron trajectories in a known potential distribution \( q(z) \) can be calculated with reasonable precision, is set out. Only the simplest methods are described; either the basic function \( q(z) \), or the reduced "characteristic" \( T(z) \) can be used as a point of departure. \( T(z) \), now defined by

\[
T(z) = \frac{3}{16} \left( \frac{q'}{q} \right)^2,
\]

is the function which appears in the reduced equation

\[
R'' + T R = 0; \quad R = r q^{1/4}.
\]

(1) The Störmer–Adams method

This is a method of solving equations of the form

\[
\frac{d^2 y}{dt^2} = f(x, y, t).
\]

The first detailed discussions of the method were published by Störmer (1920) and Kryloff (1927); four points at the origin of the trajectory must be known, and as the trajectories of electron rays are rectilinear outside the lens, this method is most suitable for their calculation. It is less convenient for strongly convergent immersion objectives, where the beginning of the trajectory, close to the cathode, is a parabola.

Regenstreif (1947) has shown how the method may be adapted for electron mirrors. Time \( t \) is chosen as the parameter, and is measured in units of magnitude \( \sqrt{m/2e} \). The Gaussian equations of motion are then of the form:

\[
\frac{dz}{dt} = q^{1/2},
\]

\[
\frac{d^2 r}{dt^2} = \frac{1}{2} \frac{\partial \Phi}{\partial r} = -\frac{r}{4} q''.
\]
Labelling the abscissa at the point of reflection $z_i$, the value of $t$ at this point $t_i$, and the value of $z$ at $t = 0$, $z_0$, we have

$$t(z) = \int_{z_0}^{z} \frac{dz}{\varphi_i^{1/2}} \text{ for } t < t_i$$

and

$$t_i - t = \pm \int_{z}^{z_i} \frac{dz}{\varphi_i'(z - z_i)^{1/2}} = \pm 2 \left[ \frac{z - z_i}{\varphi_i'} \right]^{1/2}.$$  

In the neighbourhood of $t = t_i$, $\varphi$ is replaced by an approximate expression, and

$$t = 2t_i - \int_{z_0}^{z} \frac{dz}{\varphi_i'^{1/2}} \text{ for } t > t_i.$$  

Having once obtained $t(z)$ and $\varphi(z)$ we can calculate $z(t)$ and $\varphi''(t)$; $r(t)$ is obtained by integrating

$$\frac{d^2r}{dt^2} = -\frac{1}{4} \frac{r \varphi''(t)}{r}$$

by Störmer’s method. $z(t)$ and $r(t)$ wholly define the trajectory.

(2) The Runge–Kutta method (1901) (see Scarborough, 1930)

This method requires only a knowledge of a single point on the trajectory and of the slope at this point. We use the reduced equation:

$$R'' = R T(z)$$

and an integration interval of length $\Delta z = h$. We simply quote the formulae with which the values of $R$ and $R'$ (the trajectory and its slope, in normalized coordinates) can be calculated at a point with abscissa $(z_i + h)$ when the values of these functions are known at the point with abscissa $z_i$, and also the values of the functions $T$, $T_n$ and $T_{n+1}$, at $z_i$ and $z_{i+h}$ respectively (Lapeyre and Laudet, 1960). If an accuracy of one part in $10^3$ is adequate, we use:

$$R_{n+1} = R_n \left[ 1 - \frac{h^3}{6} \left( 2T_n + T_{n+1} \right) \right] + R'_n h \left[ 1 - \frac{h^2}{6} T_{n+1} \right],$$

$$R'_{n+1} = -R_n \frac{h}{2} \left( T_n + T_{n+1} \right) + R'_n \left[ 1 - \frac{h^2}{2} T_n \right].$$

We can obtain an accuracy of better than one part in $10^6$, however, if we use the following expressions which require the value of the function $T$
at the centre of the integration intervals to be known:

\[ R_{n+1} = R_n \left[ 1 - \frac{\hbar^2}{6} \left( T_n + 2T_{n+1/2} - \frac{\hbar^2}{4} T_{n+1/2} \right) \right] \]

\[ + \quad R_n \hbar \left[ 1 - \frac{\hbar^2}{6} T_{n+1/2} \right], \]

\[ R'_{n+1} = - \frac{R_n \hbar}{6} \left[ T_n + 4T_{n+1/2} - \frac{\hbar^2}{2} T_{n+1/2} \right] \]

\[ + \quad R_n \left[ 1 - \frac{\hbar^2}{3} T_{n+1/2} \right] - \frac{\hbar}{6} T_{n+1} R_{n+1}. \]

The accuracy of the method is discussed in an article by Lapeyre (1961). This method is particularly well adapted for calculations on magnetic lenses, which are described in Chapter 5, as the form of the equation of motion for a trajectory \( r(z) \) resembles that of the Picht equation.

(3) **Liebmann's method (1949)**

The starting point here is the more complete equation (see §7.3.1):

\[
\frac{r''}{U} - \frac{1 + r'^2}{2U} \left( \frac{\partial U}{\partial r} - r' \frac{\partial U}{\partial z} \right) = 0
\]

with

\[ U = \Phi(r, z), \]

in which \( r \) is the radial coordinate of a paraxial trajectory. Using the Taylor series expansion, and integrating step-by-step, the method allows us both to pass from a point \( P_n(z_n, r_n, r'_n) \) to a point \( P_{n+1}(z_{n+1}, r_{n+1}, r'_{n+1}) \) and to calculate the error \( \Delta r_n \) due the spherical aberration. The curves which properly represent \( U \) and its derivative are replaced by step functions, each "step" of constant value coinciding with an interval \( \Delta z \). This interval must be chosen to be, at most, one-tenth of the radius of the smallest of the apertures, to attain a very high precision. Equations of the form

\[ r_{n+1} = Q_1 r_n + Q_2 r'_n, \]

\[ r'_{n+1} = Q_3 r'_n + Q_4 r_n, \]

\[ \Delta r_{n+1} = Q_1 r'_n + Q_2 r'_n, \]

\[ \Delta r'_{n+1} = Q_3 r'_n + Q_4 r_n + Q_5 \]

appear, with coefficients—relatively simple for electrostatic lenses—which are functions of

\[ \frac{\varphi'}{\varphi}, \quad \frac{\varphi''}{\varphi}, \quad \text{and} \quad \frac{\varphi'''}{\varphi}. \]
(4) Gans’ method (1937)

Only $\varphi(z)$ need be known for this simple method, which is suitable for the electrostatic case. $\varphi(z)$ is represented as nearly as possible by a number of rectilinear segments. In each region $z_j$, the potential varies linearly and the equation of the Gaussian trajectories is simple to integrate; the length of each region may be varied, according to the shape of the curve $\varphi(z)$ in the vicinity, to increase the accuracy. The trajectories will be composed of a series of parabolic segments. Across each inter-regional boundary, $\varphi$ is continuous but $\varphi'$, and hence $r'$, is not. Integration of the trajectory equation over a limitingly small region around this discontinuity is equivalent to calculating the refraction at this point (see §8.4.1). Finally, with the notation of Fig. 30a, we obtain the formulae

$$
 r_j = r_{ij} + \frac{2C_j}{\varphi_j'} (\varphi_j - \varphi_{ij}) \quad \text{if } \varphi_j' \neq 0,
$$

or

$$
 r_j = r_{ij} + r_{ij+}' \times (z_j - z_{ij}) \quad \text{if } \varphi_j' = 0,
$$

in which

$$
 C_j = (r_j' \varphi_j)^{1/2} = (r_{ij+}' \varphi_{ij})^{1/2}
$$

and also

$$
 r_{ij+}' = r_{ij-}' + \frac{\varphi_{ij-}' - \varphi_{ij+}'}{4\varphi_{ij}} r_{ij}.
$$

We should also mention the methods due to Bertein (1952 b), Burfoot (1952), and Laudet (1953).
4.5.3 Graphical Trajectory-tracing

A great deal of information about the properties of electrostatic lenses can be obtained by tracing out the trajectories on a large scale chart of the potential. Tracing in this way gives less precise trajectories than do the methods discussed in the last section; it is most useful, therefore, for rays far from the axis, for which numerical methods of calculation are extremely laborious (see for example Duchesne, 1953).

(1) The radius of curvature procedure

We consider two equipotentials, \((\Phi_1)\) and \((\Phi_2)\), sufficiently close that the potential in the region between them may be considered as effectively equal to the mean value \(\frac{\Phi_1 + \Phi_2}{2}\). Within this region, the trajectory originating at some point \(A\) can be regarded as a circular arc, the radius, \(R\), of which

![Diagram showing a trajectory tracing by the radius of curvature method.](image)

Fig. 30b. Tracing a trajectory by the radius of curvature method.

is obtained from the equation which expresses the fact that the normal forces balance:

\[
\frac{m v^2}{R} = eE_n,
\]

where \(E_n\) is the field component normal to the direction of motion. To a first approximation, we write

\[
E_n = \frac{\Phi_1 - \Phi_2}{AB},
\]

where the direction of \(AB\) is normal to that of the initial electron velocity (Fig. 30b).
(2) The parabola method

The space between the two equipotentials is now regarded as a region in which the potential varies linearly along the normals to these surfaces; the trajectories are therefore parabolae. If we consider a particle which arrives at a point $I_1$ on the equipotential $(\Phi_1)$ at an angle of incidence $i_1$ with velocity $V_1$ (Fig. 30c), we have to determine the point $I_2$ at which it reaches $(\Phi_2)$, and its direction of motion and velocity $V_2$ at this point. Since $E$ is normal both to $(\Phi_1)$ and to $(\Phi_2)$, the tangential component of $V$ is constant, and we have

$$V_1 \sin i_1 = V_2 \sin i_2,$$

But

$$\frac{1}{2} m V_1^2 = e \Phi_1,$$

and

$$\frac{1}{2} m V_2^2 = e \Phi_2,$$

so that

$$\sqrt{\Phi_1} \sin i_1 = \sqrt{\Phi_2} \sin i_2.$$

Between the two equipotentials $(\Phi_1)$ and $(\Phi_2)$, the particles are uniformly accelerated, since the field is constant. The velocities, therefore, are related in the following way:

$$V = V_1 + \gamma t,$$

$$V_2 = V_1 + \gamma \tau.$$

The chord $I_1 I_2$ has the direction of the mean velocity $V_m$ between the two equipotentials. $V_m$ is given by

$$V_m = V_1 + \frac{1}{2}(V_2 - V_1).$$

It is now possible to devise a construction to give the directions of $I_1 I_2$ and $V_2$ (Fig. 30c). We draw two concentric circles, radii $\sqrt{\Phi_1}$ and $\sqrt{\Phi_2}$, with centre $O$; we draw $\overrightarrow{OV_1}$ at an angle $i_1$ to $\overrightarrow{OX}$ (the normal at $O$ to the
equipotential ($\Phi_1$) and then $V_1V_2$ parallel to $OX$; $OM$ provides us with the direction of $I_1I_2$ while $OV_2$ gives that of $V_2$. It is convenient to draw $OX$ and a family of circles with radii $\sqrt{\Phi_1}$ on transparent paper, which can then be moved along the trajectory which is being studied.

4.5.4 Mechanical Integration; the Use of Electrical Analogues

Electrical analogues have been successfully used to solve the Laplace equation for the potential; it is, however, also possible to find analogues by means of which the trajectory equation can be integrated. Consider,

![Diagram]

**Fig. 31.** The underlying principle of an analogue device to integrate the equation $y'' + f(x) y = 0$.

for example, a transmission line of the form shown in Fig. 31 (Grivet and Rocard, 1949); the equations which describe the propagation of current and potential along the line are

$$\frac{\partial i}{\partial z} = j \frac{C}{h} \omega V, \quad \frac{\partial V}{\partial z} = j \frac{L}{h} \omega i,$$

in which $h$ is the length of each section of line.

Eliminating $i$, we obtain

$$\frac{\partial^2 V}{\partial z^2} + \frac{L C \omega^2}{h^2} V = 0$$

which is formally identical to the Picht equation provided

$$\frac{3}{16} \left( \frac{q'}{q} \right)^2 = \frac{L C \omega^2}{h^2} \quad \text{and} \quad V = R.$$

An integrating machine of this kind is in use at the Laboratoire de Radioélectricité (Faculty of Sciences, University of Paris), in which the characteristic function is simulated as nearly as possible with variable condensers. A low frequency generator supplies the potential, and the po-
potential at each point is measured with a voltmeter. The trajectories can be constructed very rapidly (Hampikian, 1953).

The "trajectory-tracers" which have been current now for a decade (see, for example, Gabor and Langmuir) are not, however, particularly well adapted for electron optics, as the analogy upon which these devices rest depends upon angular relationships; as the angles in the Gaussian approximation are so slight, any attempt to obtain the point of intersection of trajectory and axis can only be imprecise.

Machines could be designed, however, which were specially adapted for the integration of this equation; we single out the work of Schiekel (1952) on this problem, using the mechanical integrator of Walther and Dreyer (1949), that of Barber and Sanders (1959), using the electrolytic tank described in Sanders and Yates (1953, 1956), and that of Verster (1961).

4.6 EXPERIMENTAL DETERMINATION OF THE
STANDARD CARDINAL ELEMENTS

The positions of the asymptotic principal planes and foci (Fig. 25) can be rapidly calculated from a knowledge of particular trajectories in object and image space.

4.6.1 The Parallel Beam Method

In this, the simplest method, a parallel beam of electrons is incident upon the lens; the radius, \( r_0 \), of the beam is small in comparison with the radii of the apertures in the lens. After passing through the lens the beam

\[ \tan \beta \sim \beta = \frac{R_2 - R_1}{D} \]

\[ f_1 = \frac{r_0}{\beta} = \frac{r_0 D}{R_2 - R_1} \]

produces a spot of radius \( R_1 \) on a fluorescent screen at \( E_1 \). The screen is then moved to a new position \( E_2 \) a distance \( D \) away from \( E_1 \); the radius of the spot is here \( R_2 \) (see Fig. 32). We can now deduce the semi-aperture of the beam and the image focal length:
The position of the image principal plane $H_i$ is marked with respect to an arbitrary plane $P_0$ (which will be the symmetry plane when the lens is symmetrical). We then have

$$h_1 = H_i P_0 = H_i F_i - P_0 F_i,$$

or

$$h_1 = f - (P_0 E_2 - F_i E_2).$$

We know that the distance $g = P_0 E_2$ and $F_i E_2 = \frac{R_2}{\tan \beta}$; finally, therefore

$$h_1 = f - \left( g - \frac{R_2}{\tan \beta} \right).$$

If the lens is either electrically or geometrically unsymmetrical, a fresh set of measurements is made after turning the lens round without disturbing the plane $P_0$. In this way, the "object" elements are also determined (Septier, 1960).

The accuracy of the method depends upon the accuracy with which $r_0$, $R_1$ and $R_2$ are measured; it can only be used for large lenses, or for large-scale models of normal lenses.

We might equally well consider locating the position of the focus directly with a mobile screen, but this is a far less accurate method than the one described above.

### 4.6.2 Methods Using the Shadows of Grids

For microscope lenses, we can use the methods described by Spangenberg and Field (1942), Heise and Rang (1949), Everitt and Hanssen (1956) and Septier (1960). All of these methods are based upon the observation of the shadows of two grids $G_1$ and $G_2$ which are placed respectively in front of and behind the lens in question; the observations are made on a fixed fluorescent screen (see Fig. 33) which remains at a constant distance from a point electron source which is placed on the axis in front of the lens.

This method is particularly easy to apply to einzel lenses ($\Phi_{\text{object}} = \Phi_{\text{image}}$) and to symmetrical lenses (such as those illustrated in Figs. 20b, 26 and in § 8.2), as the object and image elements are symmetrically placed with respect to the geometrical centre of the lens.

The grids are composed of parallel wires; they are crossed, so that we shall easily be able to identify their respective shadows on the screen. We consider Figs. 33a and 33b in which the experimental arrangement is represented in the planes $xOz$ and $yOz$. The image of the point source $P$ is $Q$; an incident ray inclined at an angle $\alpha_0$ produces an emergent ray with inclination $\alpha_1$, which falls on $G_2$ at a distance $x_i$ from the axis and on the screen at a distance $X_i$. An identical ray which passes through the grid $G_1$ at a distance $y_0$ from the axis meets the screen at a distance $Y_0$. 

We can thus construct

\[ OG_1 = g_0, \quad OG_2 = g_l, \]
\[ OP = p, \quad OE = s, \]

where \( O \), the centre of the lens, lies in the plane \( P_0 \).

\[ \]

Fig. 33. The method using the shadow of a grid. (a) Ray-path in the \( xOz \) plane; (b) Path of an identical ray in the \( yOz \) plane.

If \( f \) is the focal length and \( \Delta \) is the distance between the two principal planes, we can write:

\[
\frac{\tan \alpha_i}{\tan \alpha_0} = \frac{\lambda_0}{f} = \frac{f}{\lambda_l} \quad \text{with} \quad \begin{cases} 
\tan \alpha_i = \frac{X_l}{s - g_l}, \\
\tan \alpha_0 = \frac{Y_0}{p - g_0}, 
\end{cases}
\]

and (see Fig. 33)

\[ q = f - \frac{\Delta}{2} + \lambda_l, \]
\[ p = f - \frac{\Delta}{2} + \lambda_0. \]

At the screen, we measure \( Y_0 \) and \( X_l \); since we know the spacing between the wires of the grids, we can immediately calculate the quantities \( V_l = X_l/x_l \) and \( V_0 = Y_0/y_0 \).

The distance \( q \) is given by

\[ q = \frac{s - V_l g_l}{1 - V_l}, \]
and combining all these relations, we arrive finally at the following expressions:

\[ f = \frac{V_0(p - q)(s - q)(p - g_0)}{V_0^3(p - g_0^3) - (s - q)^2}, \]

\[ \frac{\Delta}{2} = p - f \left( 1 + V_0 \frac{p - g_0}{s - q} \right). \]

For lenses which are either geometrically or electrically asymmetrical, the plane \( P_0 \) is selected arbitrarily. Two series of measurements are necessary: either the lens is reversed without disturbing the position of the plane \( P_0 \) with respect to \( P \) and \( E \) (Everitt and Hanssen) or it is displaced longitudinally a known distance without altering the distance between \( P \) and \( E \) (Spangenberg and Field, Septier). We can then determine:

(i) the focal lengths \( f_i \) and \( f_o \) (if the lens is an einzel lens, \( f_i \) and \( f_o \) are always equal), and

(ii) the distances \( h_i \) and \( h_o \) between the principal planes and \( P_0 \).
CHAPTER 5

THE OPTICAL PROPERTIES OF MAGNETIC LENSES

5.1 THE FORM OF THE TRAJECTORIES IN AN AXIALY SYMMETRIC FIELD

5.1.1 The Magnetic Force

The force which a magnetic field $\mathbf{B}$ exerts on an electron of charge $-e$ moving with velocity $\mathbf{v}$ is given by

$$\mathbf{F} = -e(\mathbf{v} \times \mathbf{B}).$$

The relative orientation of the various vectors is shown in Fig. 34; it illustrates the sign convention which is described by such well-known mnemonics as Maxwell’s right-hand rule, or the motion of a right-handed screw.

![Diagram showing the deflexion of a charged particle by a magnetic field.](image)

**Fig. 34.** The deflexion of a charged particle by a magnetic field.

In a uniform field, the particle follows a circular trajectory of radius $r = \frac{mv}{eB}$; if this field acts only over a region of length $l$, the resulting deviation, $\alpha$, is given by

$$\sin \alpha = \frac{eB l}{mv} = \alpha.$$
We shall find subsequently that it is often convenient to characterize the electron velocity by the product $B \varphi$ to which it corresponds rather than by the potential $\Phi$ by which it was produced. These quantities are related by

$$v = \frac{e}{m} B \varphi = \sqrt{\frac{2e\Phi}{m}}, \quad (5.1)$$

or for a non-relativistic electron in particular:

$$v \text{ (m sec}^{-1}) = 1.77 \times 10^{11} B \rho \text{ (webers m}^{-1}) = 5.9 \times 10^5 \sqrt{\Phi} \text{ (volts)}.$$  

For relativistic electrons, we should have

$$p = mv = \sqrt{2em\Phi(1 + e\phi)},$$

$$(B\rho) = \frac{p}{e} = \sqrt{\frac{2m_0\Phi}{e}(1 + e\phi)}.$$  

5.1.2 Trajectories Inside a Lens

In every round lens, the magnetic field $B$ has a symmetry axis $Oz$, along the direction of the overall motion of the electron; two components, therefore,

![](image.png)

**Fig. 35.** Cross-section of a screened magnetic lens and the field distribution inside it.

suffice to define it: one radial $B_r(r, z)$, the other axial $B_z(r, z)$. The tangential component $B_\theta$ is zero and so the force on the electron has components

$$f_r = -e v_\theta B_z, \quad f_\theta = -e(v_z B_r - B_z v_r), \quad (5.2)$$

$$f_z = +e v_\theta B_r.$$

Most magnetic lenses consist simply of short coils, either unshielded or, more often, provided with a magnetic circuit of iron which concentrates the field into a small region. Figure 35 represents a cross-section of a common design. The distribution of lines of force shows that the electrons will be affected simultaneously by the radial component $B_r$ and equally strongly by the axial component $B_z$. An electron emitted at a point $A$ follows a rectilinear path until it reaches the plane $H$ in which the lens action effectively begins (Fig. 36); at this point, the field is predominantly radial (it is the region $\alpha$, $\beta$ on the frontal projection of the trajectory, where the curvature becomes large), and the trajectory is forced to rotate sideways out of the plane of the paper in Fig. 36. Therefore a transverse component of velocity $v_r$ appears, perpendicular to the axis, which combines with the longitudinal field component to produce a radial force which draws the electron towards the axis. The particle then enters a region in which the field is effectively longitudinal and uniform—this is the region $\beta$, $\gamma$ of the frontal projection which has become approximately circular. Finally, the electron passes through the field region in which $B_r$ is once again dominant—conditions similar to those at the beginning of the lens obtain. The curvature increases again, but $B_r$ has altered in direction and the resultant force draws the trajectory towards the axis. The relative intensities of these effects are such that the emergent ray always intersects the axis a second time. The radial component of the field as the electron leaves the lens “repels” the trajectory in a meridional plane, which is itself inclined at an angle $\theta_{ol}$ to the plane of incidence, and the image of an extended object will therefore appear to have been rotated relative to the object. This qualitative analysis shows that the trajectory is complex, and it seems a difficult task to give a simple description of the motion of an electron. The problem is, however, considerably simplified by taking into account the existence of a new first integral, while the conservation of energy relation remains simple.

We notice first that the integral of the conservation of energy equation remains valid, since the force due to the magnetic field is perpendicular

![Fig. 36. The path of a ray through a magnetic lens. The field is effectively restricted to the region between the two planes $H$ and $H'$. The projection of the trajectory back onto the object plane is also shown.](image-url)
to the velocity, and there is therefore no tangential acceleration; the velocity and energy of the electron are constant. Equation (4.3) remains valid even if the lens uses a combination of electrostatic and magnetic fields. Further, there is another first order integral which Störmer discovered at the beginning of the century, but which was subsequently forgotten, until Busch rediscovered it in 1926. This Störmer integral is quite general, but we shall first consider only the Gaussian approximation, where the calculations are simpler; the full proof is to be found in § 6.2.2.

5.2 THE GAUSSIAN APPROXIMATION

5.2.1 The Equations of Motion in Fixed Axes

Just as the axial potential \( \varphi(z) \) is sufficient to describe the electric field near the axis of an electrostatic system, so the axial field \( B_z(z) \) is adequate to describe the radial field near the axis of a magnetic lens. The Maxwell equation \( \text{div} \, B = 0 \) immediately implies

\[
B_r(z, r) = -\frac{1}{2} \frac{\text{d} B_z(z)}{\text{d} z}.
\]

(5.3)

If we write out the general equation of motion for an electron in a magnetic field

\[
\frac{\text{d}(m v)}{\text{d} t} = m \frac{\text{d} v}{\text{d} t} = -e (v \times B)
\]

(5.4)

in cylindrical polar coordinates, we obtain

\[
m \left[ \frac{\text{d}^2 r}{\text{d} t^2} - r \left( \frac{\text{d} \theta}{\text{d} t} \right)^2 \right] = -e \frac{\text{d} \theta}{\text{d} t} B,
\]

\[
m \left[ r \frac{\text{d}^2 \theta}{\text{d} t^2} + 2 \frac{\text{d} r}{\text{d} t} \frac{\text{d} \theta}{\text{d} t} \right] = e \left( B \frac{\text{d} r}{\text{d} t} + \frac{1}{2} r \frac{\text{d} B}{\text{d} z} \frac{\text{d} z}{\text{d} t} \right),
\]

(5.5)

\[
m z'' = 0.
\]

The third equation expresses the fact that the velocity is constant in the Gaussian approximation—this is a direct consequence of conservation of energy. The second equation can be written

\[
\frac{\text{d}}{\text{d} t} \left( r^2 \frac{\text{d} \theta}{\text{d} t} \right) = \frac{e}{m} \left( r B \frac{\text{d} r}{\text{d} t} + \frac{1}{2} r^2 \frac{\text{d} B}{\text{d} z} \frac{\text{d} z}{\text{d} t} \right),
\]

which can be integrated directly, to give

\[
\frac{\text{d} \theta}{\text{d} t} = \frac{e}{2m} B + C \frac{1}{r^2},
\]

(5.6)

\( C \) being zero when the ray intersects the axis. This can occur in any of three ways; first, the ray, of arbitrary direction, may originate in an axial point.
object \((r_0 = 0)\); alternatively, the object may have any position outside the lens, or may be a virtual immersed object, provided the rays lie in meridian planes \(r_0 \neq 0\) but \(\left(\frac{d\theta}{dr}\right)_0 = 0\); finally, for a real immersed object, rays which crossed the axis before they reached the lens will intersect it again after leaving the lens and entering free space once again. Such a case occurs in the magnetic microscope objective, where the rays are emitted by an axial point source in the electron gun.

In any of the cases for which \(C = 0\), Störmér's equation reduces to the simple form:

\[
\frac{d\theta}{dt} = \frac{e}{2m} B \quad \left( \frac{e}{2m} = 8.797 \times 10^{10} \text{ C kg}^{-1} \right).
\] (5.7)

Any ray of this kind always leaves the lens as a straight line which intersects the axis (which lies, that is, in a meridian plane), since the condition \(B = 0\) implies \(\frac{d\theta}{dt} = 0\).

### 5.2.2 The Equations of Motion in the Rotating Coordinate System of Larmor

Larmor's theorem suggests that we should simplify our description of the motion still further by using a reference system which rotates about the axis with the Larmor angular velocity:

\[
\omega = \frac{e}{2m} B.
\]

For the family of rays for which \(C = 0\), the simplification is obvious, as the various trajectories appear to be planar and meridional in the rotating system, for the new angular velocity of the particles is \(\frac{d\theta'}{dt} = \frac{d\theta}{dt} - \omega = 0\).

For the rays for which \(C \neq 0\), however, the transformation is still more interesting, as we shall now show.

To apply Newton's laws of motion in a rotating system, we must take into account two artificial forces in addition to the force \(f\). These are the Coriolis force, \(-2m\omega \times \mathbf{v}'\), and the drag (which corresponds to the "drag" acceleration) which has two components: the radial centrifugal force \(m\omega^2 r\) and the tangential force \(-m r \left( \frac{d\omega}{dt} \right)\). By a proper choice of \(\omega\), we can cancel all the force components save one. This can be seen from the following table in which the velocity and its transverse component are written \(\mathbf{v}, \mathbf{v}_\theta\), in the real motion, and \(\mathbf{v}', \mathbf{v}'_\theta\) in the relative motion; \(\omega\) is a vector of length \(\frac{e}{2m} B\) along \(Oz\) (Fig. 37); the longitudinal force component is negligible in the Gaussian approximation, as it is proportional
to \( r' \) and is thus second order. The relation \( v_0' = v_0 - \omega r \) considerably simplifies the calculation of the resultant radial force:

\[
f_r' = -e (v_0' + \omega r) B + e v_0' B + \frac{e^2}{4m} r B^2 = -e \omega r B + \frac{e^2}{4m} r B
\]

\[
= -\frac{e^2}{4m} r B.
\]

![Fig. 37. Larmor's rotating coordinate system.](image)

The equations of motion transform, in the frontal projection, into

\[
\frac{d^2 r}{dt^2} + \frac{e^2}{4m^2} B^2 r = 0,
\]

and along the axis \( Oz \), into

\[
\frac{dz}{dt} = \sqrt{2e \varphi_0/m},
\]

where \( \varphi_0 \) is the constant electric potential which obtains throughout the whole lens. Provided \( \varphi'' \) is everywhere replaced by \( (e/2m) B^2 \), equation (5.8)
is identical to (4.7) of § 4.2.2, which is its counterpart for electrostatic lenses.

(i) \( C = 0 \). For this family of rays, the trajectory is planar; the equation of motion in the rotating meridian plane is obtained by eliminating the time, thus:

\[
\frac{d^2r}{dz^2} + \frac{e}{8m\varphi_0}B^2r = 0. \tag{5.9}
\]

It is useful to rewrite the quantity \( (e/8m\varphi_0) \) in terms of the radius \( \varrho \) of the circle which the electrons would follow if they were introduced normally into a uniform magnetic field of intensity \( B \) with a velocity corresponding to an accelerating potential \( \varphi_0 \). We should have

\[
\left( \frac{1}{\varrho} \right)^2 = \frac{eB^2}{2m\varphi_0},
\]

whence

\[
\varrho = 3.371 \sqrt{\varphi_0/B} \text{ (cm; V; gauss)},
\]

or

\[
\varrho = 3.371 \times 10^6 \sqrt{\varphi_0/B} \text{ (m; V; tesla)}.
\]

and equation (5.9) becomes:

\[
\frac{d^2r}{dz^2} + \frac{r}{4\varrho^2} = 0. \tag{5.10}
\]

(ii) \( C \neq 0 \). Even in the rotating system, these rays remain skew. However, by using a complex coordinate to describe an off-axial position as in the parallel electrostatic situation, it is not difficult to show that these oblique rays also coincide, at the same point as the meridian rays, to form a point image. This involves writing \( u = r e^{i\theta_1} \), in which \( \theta_1 \) is the angle between the Larmor rotation and the actual rotation, defined by

\[
\theta_1 = \int_{z_0}^{z_1} \frac{d\theta}{dt} \frac{dt}{dz} \, dz - \int_{z_0}^{z_1} \frac{e}{2m}B(z) \, dz.
\]

After a certain amount of calculation, which is set out in detail in the book by de Broglie, \( u \) is found to satisfy equation (5.9) for all values of \( C \).

(iii) The image rotation. The rotation of the meridian plane between object and image is given by

\[
\theta_{01} = \theta_1 - \theta_0 = \sqrt{\frac{e}{8m\varphi_0}} \int_{z_0}^{z_1} B(z) \, dz = \int_{z_0}^{z_1} \frac{1}{2\varrho} \, dr. \tag{5.11}
\]

(iv) The stages in the geometrical argument which the table summarizes can be interpreted very straightforwardly by using a complex notation for the forces, as in § 4.2.2,

\[
f_u = j e \left( B u' + \frac{1}{2} u \frac{dB}{dz} \frac{dz}{dt} \right).
\]
Writing

\[ W = u e^{i \Omega}, \quad \Omega = \int_{i_n}^{i} \omega \, dt, \]

we obtain the trajectory directly.

If the system combines electrostatic and magnetic fields, we obtain an equation of motion

\[ \frac{d}{dz} \left( \frac{1}{\sqrt{\varphi}} \frac{d\varphi}{dz} \right) + \frac{1}{4\sqrt{\varphi}} \left( \varphi'' + \frac{e}{2m} B^2 \right) r = 0. \quad (5.12) \]

Making the simplifying assumptions listed in § 4.2.2, we find that the focal length of a thin weak electromagnetic lens is given approximately by

\[ \frac{1}{f} = \frac{1}{4\sqrt{\varphi_0}} \int_{-\infty}^{\infty} \left( \frac{\varphi''}{\sqrt{\varphi}} + \frac{e}{2m} \frac{B^2}{\sqrt{\varphi}} \right) \, dz, \quad (5.13) \]

in which \( \varphi'' \), \( \varphi \) and \( B \) are functions of \( z \). For a purely magnetic lens, \( \varphi(z) = \varphi_0 \), so that

\[ \frac{1}{f'} = \frac{e}{8m \varphi_0} \int_{-\infty}^{\infty} B^2(z) \, dz. \quad (5.13') \]

### 5.3 PROPERTIES OF MAGNETIC LENSES

#### 5.3.1 The Optical Formalism

All the optical formalism of electrostatic lenses arises from the fact that the equations of motion are linear. As equation (5.12) is linear, therefore, all the reasoning presented in § 4.3 can be carried over directly into the magnetic or electromagnetic lens situation. The cardinal points, for example, are defined perfectly straightforwardly by a pair of rays in a rotating meridian plane.

![Fig. 38. An immersion object on the axis of a magnetic lens (the image rotation is not shown).](image-url)
The only important difference between the electrostatic and electromagnetic situations lies in the Larmor rotation of the image which appears in the latter, and which can be evaluated with the aid of formula (5.11). It is interesting to notice that immersion elements are now of considerable importance, as a magnetic field will not be perturbed by the intrusion of an object, or a screen with which to inspect the image. In Figure 38, two of the trajectories which form the real image, outside the lens, of an object immersed within it, are shown.

5.3.2 Physical Properties

Unlike the optical formalism, however, the physical properties are very definitely different, as we might have guessed from the elementary deviation formula quoted earlier,

$$\alpha = \frac{l}{\rho} = \sqrt{\frac{e}{2m \varphi_0}} Bl.$$  

The value of \(e/m\) now effects the deviation, and the nature of the particle affects the trajectory it follows—the focal length of a magnetic lens is much longer for ions, with their greater weight, than for electrons, and a magnetic lens is of little use for fast heavy ions.

No longer are the coefficients homogeneous with respect to the electric potential \(\varphi(z)\). The positions of the cardinal elements now vary, therefore, with the applied potential and with the lens currents.

In particular, accidental fluctuations of the potential will mar the image, and very careful stabilization of the accelerating high tension which is applied is necessary—similar precautions must be taken with the lens currents.

Yet another difference appears if we consider a beam of electrons moving away from an observer; the electrons will appear to rotate in a clockwise sense if the magnetic field points in the same direction as the motion of the electrons, but in an anticlockwise sense if it points in the opposite direction. The trajectory is dependent upon the direction of motion of the electrons relative to the field. Even though two trajectories pass through the same point with the same tangent, they will be different elsewhere if their velocities are in opposite directions (Fig. 39). The optical reversibility principle is no longer valid, and a close analogy with the optics of crystalline media is seen to be impossible.

In practice, however, it is the image itself in which we are interested, rather than the rays which produce it, so that this feature of magnetic lenses is not particularly disturbing.

The sense of the rotation is reversed by a change of sign of the magnetic field, while the convergence, depending as it does upon \(B^2\), is unaffected. The image rotation, therefore, could be suppressed without interfering
with the focusing action. All that would be required would be to couple together a pair of lenses in which the currents flow in opposite directions. This is in fact rarely done, however, as, for a given focal length, the power expended in the magnetizing coils is greater, and some of the aberrations of the system are accentuated (Becker and Wallraff, 1940). It is much more satisfactory to break the system down into two separate lenses, the objective and projector in the microscope, for example.

5.4 RELATIVISTIC LENSES

The effect of relativity is to replace \( \varphi_0 \) or \( \varphi^2 \) by \( \varphi_0 (1 + \epsilon_0) \) or \( \varphi^2 (1 + \epsilon_0) \) respectively, in the equations of motion (5.9) and (5.10), where we have written \( \epsilon_0 \) for \( \frac{e}{2m_0 c^2} \varphi_0 \). The focal length increases steadily with \( \varphi_0 \) (Fig. 40). At velocities corresponding to energies in the neighbourhood of one MeV, the magnetic lens becomes impracticable, as saturation in the iron prevents \( B \) from exceeding some 26,000 gauss.
5.5 METHODS OF DETERMINING THE FUNDAMENTAL RAYS

Equation (5.12), which defines the rays in a rotating meridian plane, is formally identical with the Picht equation which gives the reduced rays of an electrostatic system. All the methods listed in § 4.5, therefore, can be invoked to help provide a solution of it. The equation can be integrated by using either a suitable mathematical model of the field—two such models, due to Glaser (1942) and to Grivet (1950b, 1952a, b) and Lenz (1951) are of particular importance—or one of the step-by-step methods already discussed (Störmer—Adams, Runge—Kutta, Liebmann, for example; see 4.5.2); or, again, by using an electrical analogue—the values of the condensers along the transmission line now being given by:

$$L C \omega^2 = \frac{e}{8 m \varphi_o} B^2.$$

There are, however, other methods designed specifically for magnetic lenses; a very curious analogue method has been suggested by Vineyard (1952) in which the magnetic force is modelled by a Coriolis force. Unfortunately, the practical construction is very complicated, especially in a three-dimensional system, but we have already seen how useful this analogy is, permitting us as it does to integrate the equations of motion quite simply.

Another method, suggested by Loeb (1947), is known as the "hodoscope". This method rests on the analogy which exists between the trajectory of a particle and the equilibrium form of a conducting wire which is traversed by a current \( i \) and placed in the same magnetic field. (See Fig. 41.) From the equations

$$\frac{d(mv)}{dt} = \frac{dp}{dt} = -e v \times B$$

![Fig. 41. The "Hodoscope" or "Floating Wire". The incident trajectory is fixed by the point \( B \) and the centre of the diaphragm \( D_2 \), for a given current \( i \) and induction \( B \); the emergent trajectory is likewise determined by \( D_1 \) and \( A \). The parts such as \( D_1 \) and \( D_2 \) which support the wire are mobile in every direction. The positions can be made precise microscopically. The tension is provided by the weight \( T \); the pulley \( P \) must be frictionless.](image-url)
we derive \[ dp + e(ds \times B) = 0 \]
in which \( p \) denotes the momentum of the particle and \( ds \) an elementary vector, tangent to the trajectory.

If \( T \) is the tension of the wire and \( i \) is the intensity of the current flowing through it,
\[ dT + i(ds \times B) = 0, \]
provided we neglect the other forces which may act upon the wire (such as its weight and rigidity); \( dT \) represents an elementary increase in the tension. The vectors \( p \) and \( T \) are parallel to \( ds \), so that we can write:
\[ p = p \frac{ds}{ds}, \quad T = T \frac{ds}{ds}, \]
and hence
\[ dT = T d\left( \frac{ds}{ds} \right) + \frac{ds}{ds} dT. \]

Since \( (ds \times B) \) is perpendicular to \( s \), \( dT \) is perpendicular to \( s \) also; it follows therefore, that \( dT = 0 \), since \( \frac{ds}{ds} \) is parallel to \( s \)—only the first term which represents a vector which is normal to \( s \) remains. Similarly, we show that \( dp = 0 \). In a magnetic field, both the momentum \( p \) of a particle and the tension \( T \) of a wire remain constant.

We thus obtain the following equations:
\[ \frac{ds}{ds} \times B = -\frac{q}{e} \frac{d}{ds} \left( \frac{ds}{ds} \right), \]
\[ \frac{ds}{ds} \times B = -\frac{T}{i} \frac{d}{ds} \left( \frac{ds}{ds} \right). \]

If \( p/e = T/\rho \) and the initial conditions are the same, these equations have identical solutions throughout the region.

The ratio \( \frac{p}{e} = B \rho \) is the "rigidity" of the particle. It can easily be expressed as a function of the accelerating voltage of the particles.
\[ (B \rho)^2 = \frac{2m_0 q_0}{e} \left( 1 + \frac{e q_0}{2m_0 c^2} \right) = \frac{2m_0}{e} \Phi_0^* \]
in which \( \Phi_0^* = q_0 \left( 1 + \frac{e q_0}{2m_0 c^2} \right) \) as in all magnetic systems. For non-relativistic particles, \( (B \rho)^2 = \frac{2m_0 q_0}{e} \), where \( e \) denotes the absolute value of the charge. The condition to be satisfied, therefore, is
\[ T = (B \rho) \rho. \]

If, for example, \( B \rho = 0.3 \) M.K.S. units (for 5 MeV protons) and \( \rho = 1 \) A, we find that \( T = 0.3 \) newton, or about 30 gm weight. Since the tension
is constant over the whole length of the wire, it is sufficient to apply this tension at a single point only, by attaching a stretched spring or a weight $T$ to the end of the "trajectory", for example.

There are several possible sources of error, which arise from the action of perturbing forces on the wire:

(i) The force exerted by the image of the wire which is produced in the magnetic material, and hence in the four poles.

(ii) The force due to the rigidity of the wire.

In a well-annealed wire these first two causes are negligible provided $T$ is sufficiently large.

(iii) The weight of the wire. This cannot be neglected, but its effect can be eliminated if we are looking only for plane trajectories, by arranging that the plane of the trajectories is horizontal. The friction of the system by means of which the force provided by a weight is converted into a horizontal tension should be eliminated as completely as possible.

Finally, for the given initial conditions, the wire may be in stable or unstable equilibrium in the magnetic field.

Detailed studies of these questions are to be found in the articles by Carlile (1957), by Citron et al. (1959) who have used this method at CERN to study the trajectories of protons and mesons within the magnet of the 600 MeV synchrocyclotron, and by Pinel (1959, 1960). Although this device was first suggested as a means of studying highly convergent, axially symmetric magnetic lenses, it has been used mainly for the determination of the cardinal elements and aberrations in magnetic prisms (see Chapter 22) and in the strong focusing lenses which are described in Chapter 10 (Carlile, 1957).

5.6 Calculation of the Trajectories and Optical Elements

5.6.1 Integration of the Equation of Motion

The form of the equation of motion:

$$\frac{d^2r}{dz^2} + \frac{e}{8m\Phi^*_0} B^2(z) r = 0,$$

in which $B(z)$ represents the induction on the axis and $\Phi^*_0$ the accelerating potential, relativistically corrected if necessary, is identical to that of the reduced (Picht) equation which we have examined in connexion with electrostatic lenses, namely:

$$\frac{d^2R}{dz^2} + T^2(z) R = 0.$$
The methods of integration listed in the preceding chapter are all, in consequence, applicable to equation (5.14); in particular, the Runge–Kutta method has been used by Lapeyre and Laudet (1960) in the present case of magnetic lenses. If we refer back to the equations of § 4.5.2 we have simply to effect the following changes:

\[ T(z) = \frac{e}{8m\Phi_0^s} B^2(z), \]
\[ R = r, \]
\[ R' = r'. \]

In addition, we should mention a method which has been described by Gautier (1953), which is quicker to use than the Runge–Kutta method, and which provides very fair results. We write

\[ k^2 = \frac{e B_0^2}{8m\Phi_0^s}, \quad x = \frac{z}{a}, \quad y = \frac{r}{a} \]

\((a \) is some arbitrary dimension of the lens), and

\[ B(x) = B_m b(x), \]

where \( B_m \) is the maximum induction on the axis. This implies

\[ T(x) = k^2 b^2(x). \]

The values of \( y_{n+1}, y_n \ldots \) at points with abscissae \( x_{n+1}, x_n \ldots \) are obtained by a Taylor series expansion:

\[ y_{n+1} = y_n + A x \cdot y'_n + \frac{(A x)^2}{2} y''_n + \frac{(A x)^3}{6} y'''_n + \cdots, \]
\[ y_{n-1} = y_n - A x \cdot y'_n + \frac{(A x)^2}{2} y''_n - \frac{(A x)^3}{6} y'''_n + \cdots, \]

and from these expressions we obtain

\[ y'_n = \frac{y_{n+1} - y_{n-1}}{2A x} + \alpha', \]
\[ y''_n = \frac{y_{n+1} + y_{n-1} - 2y_n}{(A x)^2} + \alpha'', \]

in which \( \alpha' \) and \( \alpha'' \) are second order in \( \Delta x \). On replacing the differentials by finite differences, the trajectory equation \( y'' + T(x)y = 0 \) becomes

\[ y_{n+1} = [2 - T_n(A x)^2] y_n - y_{n-1} + \varepsilon, \]

in which \( \varepsilon \) is of the fourth order in \( \Delta x \) and can thus be neglected. We thus have a recurrence formula from which we can calculate \( y_{n+1} \), once \( y_n \) and \( y_{n-1} \) are known. In practice, we evaluate the function \([2 - T_n(A x)^2]\) throughout the useful region, after selecting the integration strip-width, \( \Delta x \). At
the beginning, the value of $y$ must be known at $x_0$ and $x_1 = x_0 + \Delta x$. If $y = y_0$ and $y' = y'_0$ at $x_0$, we can write

$$y_1 = y_0 + y'_0 \Delta x - y_0 T_0 (\Delta x)^2$$

and proceed with the calculation in this way. If we select trajectories incident parallel to the axis, we can easily establish the positions of the object and image cardinal points.

### 5.6.2 Direct Evaluation of the Optical Quantities

The convergence of any magnetic lens can be calculated quickly and simply (Durand and Fert, 1957), once the distribution of the induction on the axis, $B(z)$, has been measured. The real lens is replaced by an ideal lens of length $L$, such that $B(z) = 0$ outside and $B(z) = B_0 = \text{constant}$ within.

$L$ is the "equivalent length" which is here defined by

$$L = \frac{1}{B_0} \int_{-\infty}^{\infty} B(z) \, dz,$$

or if $NI$ is the number of ampère-turns of the lens

$$L = \frac{\mu_0 NI}{B_0}.$$

The equation of motion in a rotating coordinate system can be integrated straightforwardly. Within the fictitious lens, of length $L$, we have

$$\frac{d^2 r}{dz^2} + \frac{e B_0^2}{8 m \Phi_0^*} r = 0,$$

and writing

$$\frac{e B_0^2}{8 m \Phi_0^*} = k^2,$$

this becomes

$$\frac{d^2 r}{dz^2} + k^2 r = 0.$$

If the lens lies between the planes $z = 0$ and $z = L$, the values of $r$ and $r'$ at the exit will be given by

$$r_s = r_0 \cos kL + \frac{r'_0}{k} \sin kL,$$

$$r'_s = -r_0 k \sin kL + r'_0 \cos kL.$$

$r_0$ and $r'_0$ are the initial conditions at $z = 0$. If $r'_0 = 0$ we obtain

$$r_s = r_0 \cos kL,$$

$$r'_s = -r_0 k \sin kL.$$
When \( kL < \frac{\pi}{2} \), the real focus lies outside the lens. If we calculate the expressions for the focal length \( f_i \) and the abscissa of the image focal point \( z_{f'} \) with respect to the fictitious exit plane, we obtain

\[
\begin{align*}
    f_i &= -\frac{r_0'}{r_s'} = \frac{1}{k \sin kL}, \\
    z_{f'} &= -\frac{r_s}{r_s'} = \frac{\cos kL}{k \sin kL}.
\end{align*}
\]

We can determine the position, \( z_{H'} \), of the principal plane immediately: the object focus and principal plane are symmetrically placed with respect to their image counterparts about the point \( z = \frac{L}{2} \), the centre of gravity of the curve which represents \( B(z) \).

If \( kL > \frac{\pi}{2} \), the preceding formulae give the normal cardinal elements.

The immersion elements can be determined with the aid of the expression for the trajectory within the lens:

\[
r = r_0 \cos kz + \frac{r_0'}{k} \sin kz,
\]

by writing \( r(z_1) = 0 \) and calculating the slope \( r'(z_1) \) at this point. In this way, we find

\[
f_{\text{min}} = -\frac{r_0}{r'(z_1)}.
\]

For a weak lens, we should write

\[
\sin kL \simeq kL, \quad \cos kL \simeq 1 - \frac{k^2 L^2}{2},
\]

so that

\[
\frac{1}{f} \sim k^2 L,
\]

and

\[
1 - \frac{k^2 L^2}{2} < \frac{z_{f'}}{k^2 L},
\]

and hence

\[
z_{H'} = z_{f'} - f = -\frac{L}{2}.
\]

The principal planes coincide at the centre of the model lens. The rotation of the coordinate system to which the trajectory is referred, and hence the rotation of the image with respect to the object, is given very simply by

\[
\theta = kL
\]

(see equation (5.11) of § 5.2.2).
CHAPTER 6

SIMILARITIES AND DIFFERENCES
BETWEEN GLASS OPTICS AND
ELECTRON OPTICS

6.1 ELECTROSTATIC LENSES

6.1.1 The Concept of Refractive Index in Electrostatic Electron Optics

Equation (4.21) and the Lagrange–Helmholtz relation for glass lenses become identical if we write \( \sqrt{\varphi_0} = n_0 \) and \( \sqrt{\varphi_1} = n_1 \). This is a special case of a very general formal analogy between the distribution of refractive index \( n \) in a glass lens and the distribution of the function \( \sqrt{\varphi} \) in an electrostatic lens. If the two distributions are identical, the trajectory of an electron and that of a light ray, originating at the same point in their respective sources, are identical also.

The zero of potential is chosen to be the cathode potential; \( n' = K'n \) is equally suitable as refractive index, and a common choice is \( K^2 = \frac{e}{mc^2} \), so that \( n' = \frac{v}{c} \), where \( c \) is the velocity of light in free space. The refractive index, therefore, has become a dimensionless number, just as in glass optics. It is to be noted, however, that the ratio of the refractive index at the object to that at the image, \( \frac{n_0}{n_1} \), can reach values far higher than those which can be achieved in glass optics.

6.1.2 Snell’s Law and Electron Optics

In glass optics, the rays are bent straight lines, the sudden changes in gradient occurring when the ray crosses from one medium into another. The deviation is given by the elementary Snell law

\[ n_1 \sin i_1 = n_2 \sin i_2, \]

and it is natural to enquire whether there is a similar law for electron optics. In this case, the boundary between two optical media becomes the boundary
between two media of different potential \( q_1 \) (\( n_1 = \sqrt{q_1} \)) and \( q_2 \) (\( n_2 = \sqrt{q_2} \)). Such a boundary could be formed with the aid of two metal membranes distance \( l \) apart; between the membranes, the field is \( E = \frac{q_2 - q_1}{l} \). If an electron arrives at the first membrane with velocity \( \sqrt{\frac{2e}{m} q_1} \) and at an angle of incidence \( i_1 \), it can proceed in either of two ways.

If \( i_2 \leq \sqrt{q_2/q_1} \) the electron leaves the membrane at an angle of refraction \( i_2 \), such that \( \sqrt{q_1} \cdot \sin i_1 = \sqrt{q_2} \cdot \sin i_2 \), and thus obeys Snell's law exactly. Whatever the angle of incidence, the condition is satisfied

![Diagram](https://via.placeholder.com/150)

**Fig. 42.** A transmitted ray and a totally reflected ray at the boundaries between regions of different refractive index in electron optics.

provided \( \sqrt{q_2} > \sqrt{q_1} \) (provided that the less refringent or "optically rarer" medium precedes the optically denser one); if the denser medium precedes the rarer, the condition is satisfied only for angles below a limiting value.

If, on the other hand, \( \sin i_1 > \sqrt{q_2/q_1} \), which happens for large angles in the case in which \( \sqrt{q_2} < \sqrt{q_1} \), total reflexion occurs; the rays turn back upon themselves, and leave the membrane on the same side as they were incident, at an angle of reflexion equal to their angle of incidence. Figure 42 illustrates the various possibilities.

It might seem reasonable, therefore, to try to build lenses from suitably curved membranes separating regions of different refractive index. In the early days of electron optics, Knoll and Ruska (1932) suggested that such a lens might be feasible, but in practice, it is very difficult to construct, as no membrane is available with which it might be built which is adequately transparent to electrons. One is forced to resort to grids, which produce marked optical defects as the holes perturb the potential and prevent a clear image from being formed.

Save for a suggestion by Ruska (1933) that they might be used to focus very wide beams, and as aids in teaching electron optics, therefore, lenses of this type have virtually no application.
6.1.3 The Refractive Index Applied to the Study of Lenses

If we examine a chart of the equipotential surfaces in an electrostatic lens (Fig. 20), a striking analogy is apparent between these surfaces (which become almost indistinguishable from spheres, near the axis) and the intermediate boundaries of a centred glass optical system. It is natural, therefore, to try to transfer the formulae for the cardinal elements in glass optics into electron optics.

As an example of how this can be done, we shall calculate the focal length of a thin lens. In glass optics, the power of a succession of regions of different refractive index, \( n_0, n_1, n_2, \ldots, n_i, \ldots \) separated by surfaces of radius of curvature \( R_1, R_2, R_3, \ldots, R_i, \ldots \) is given by

\[
n_i C_i = \frac{n_i - n_0}{R_1} + \frac{n_2 - n_1}{R_2} + \cdots + \frac{n_i - n_{i-1}}{R_i} = \sum \frac{\Delta n}{R}.
\]  

(6.1)

To transfer this into electron optics, we have to integrate over an infinite number of regions instead of summing over a finite number, to replace the refractive index by \( \sqrt{\varphi} \), and to replace the radii of curvature by those of the equipotentials, \( R = 2\varphi''/\varphi'' \), a standard result (\( \varphi \) is the axial potential). The power, therefore, is given by

\[
C_i = \frac{1}{n_i} \int_{-\infty}^{\infty} \frac{dn}{R} = \frac{1}{4 \sqrt{\varphi_i}} \int_{-\infty}^{\infty} \frac{\varphi'''}{\sqrt{\varphi}} \, dz.
\]  

(6.2)

Other formulae are susceptible to this transition; the well-known Cotes formula, as Bernard (1951) has shown, can be translated in this way and thus be used for the step-by-step tracing of electron rays. The trajectory is represented not by a curve but by a bent straight line, a change of slope occurring at each of the equipotentials on the diagram. The magnitude of the change can easily be determined by a geometrical construction which stems from the analogous procedure (due to Huyghens) of glass optics. For further details, the reader is referred to the articles by Gans (1937), by Knoll (1935) and more recently, by Fert (1952).

A number of points are worthy of special emphasis. First, the converging power of an elementary intermediate boundary depends both upon the change in refractive index \( n \) across it and upon the radius of curvature \( R \); in electron optics, \( R \) and \( n \) are no longer independent, but are related by the expression \( R = \frac{2\varphi'}{\varphi''}, \) and it is due to this lack of freedom that the correspondence between glass and electron optics is necessarily incomplete (so that, for example, divergent lenses are in general impossible to obtain in electron optics and spherical aberration cannot be eliminated).

In addition, there are no reflected electrons comparable to the finite proportion of light which is reflected at a glass optical boundary, as it is
not possible to obtain an electric field strong enough to have a perceptible effect over a distance as short as the wavelength of an electron (some thousandths of a millimicron).

Further, the refractive index \( n \) in electron optics is proportional to the particle velocity \( v \). In glass optics, on the other hand, \( n = c/v' \) where \( v' \) is the phase velocity of the light wave. These two statements no longer seem contradictory, however, if we recall that the particle velocity and the phase velocity of the associated wave are related by \( v \cdot w = c^2 \). This relation is of extremely general validity, as it need no longer be based upon an analysis of the motion of wave-groups, such as that due to Rayleigh, Sommerfeld and L. Brillouin towards the beginning of the century, but can be derived directly from the principle of relativity (Ditchburn, 1948; Synge, 1952).

It is interesting to recall that even in the seventeenth century, a time when the distinction between \( v \) and \( w \) was still unknown, Fermat was aware of the disparity between the relations which related the index to the velocity in mechanics and in optics, and considered this sufficient reason to delay publication of the principle which bears his name until 1661, although he had discovered it in 1657. For more information on this point, and for a description of the discovery of the refractive index by Descartes, the reader is referred to the interesting historical study which comprises the first ten pages of *Geometrische Optik* by C. Carathéodory (1937).

### 6.1.4 Fermat's Principle and the Principle of Least Action

We can express the analogy between electron and glass optics most concisely and rigorously by a comparison of Fermat's principle and the principle of least action. We shall content ourselves here with a summary outline of the way in which this relation may be demonstrated; a more rigorous discussion is to be found in the book by de Broglie and in two specialized articles by Ehrenberg and Siday (1949) and by Glaser (1950).

It is well known that the path of a light ray is dictated by Fermat's principle, expressing the fact that the total optical path will have a stationary value

\[
\frac{\partial}{\partial \lambda} \int_{\lambda_1}^{\lambda_2} n \, ds = 0.
\]

The motion of an electron, on the other hand, is governed by the principle of least action,

\[
\frac{\delta}{\delta \lambda} \int_{t_1}^{t_2} (T + U) \, dt = 0,
\]

where \( U \) is the potential which defines the force applied to the particle and \( T \) the kinetic energy, \( T = \frac{1}{2} m v^2 \). As the conservation of energy relation,
\[ T - U = \text{constant}, \text{ is satisfied, we deduce immediately that} \]

\[ \int_{t_u}^{t_i} (T + U) \, dt = 2 \int_{t_v}^{t_i} T \, dt + \text{constant} = \int_{A}^{B} m \, v \, ds + \text{constant}. \]

Rewriting the velocity in terms of the corresponding electric potential, we obtain a modified version of the principle of least action:

\[ \delta \int_{A}^{B} v \, ds = \delta \int_{A}^{B} K \sqrt{\Phi} \, ds = 0, \tag{6.5} \]

which becomes exactly parallel to Fermat’s principle if we write

\[ n = K \sqrt{\Phi} (r, z), \tag{6.6} \]

\( K \) being an arbitrary constant.

So far, we have restricted the argument to non-relativistic mechanics; if, however, we substitute the relativistic value of the function \( T \), the refractive index becomes

\[ n = K \sqrt[4]{\Phi (1 + e \Phi)}, \quad e = \frac{e \Phi}{2 m_0 c^2}. \tag{6.7} \]

Although it is possible to apply Fermat’s principle directly, in certain numerical methods for example, it is more usual to work from the Euler equations which are derived from it, and which in fact prove to be the familiar differential equations for the electron trajectories. If we replace \( \Phi \) in these equations by \( n^2 \), we regain the Bouguer formula with which it is possible to calculate the path of a light ray in an inhomogeneous medium (and to explain such phenomena as mirages). Here, the formula applicable to media with a symmetry axis is given. All these points are very fully dealt with in de Broglie’s book on particle optics.

6.1.5 The Limits of Electron Optics

The preceding results represent only one particular aspect of a very general analogy, due to L. de Broglie, between wave-propagation and particle motion. We shall do no more than underline the fact that this analogy is maintained even when geometrical optics becomes invalid. In glass optics, the laws of geometrical optics (and Fermat’s principle in particular) cease to be valid when the refractive index varies appreciably over a distance of the order of magnitude of the wavelength, and we have to resort to wave theories of light to resolve the problem.

Similarly, if the electric potential varies appreciably over a distance of the order of magnitude of the wavelength of the wave associated with the electrons, \( \lambda = h/m \, v \), the laws of electron optics which we have enun-
ciated are no longer valid, and the phenomena which appear can be described only in terms of the wave nature of the electron. This is precisely the case when an electron passes through a solid body—the object in an electron microscope, or the specimen in a diffraction camera, for example—and the attractive phenomenon of electron diffraction is observed. Diffraction patterns can be observed, too, in the geometrical shadow cast by a sharp edge illuminated by electrons, but in this case they are extremely tiny and can be distinguished only with the aid of a powerful microscope (Boersch, 1940). A similar effect is to be found in the very fine beams which are used in electron microscopes, and in fact, it is this which provides a limit to the resolution which can be obtained. All these questions are of cardinal importance in the electron microscope, to such an extent that when such an instrument is being designed using formulae such as those established in the present work, allowance must be made for wave mechanical effects.

6.2 MAGNETIC LENSES

6.2.1 The Refractive Index of the Fictitious Motion; the Gaussian Approximation

If we attempt to use the notion of refractive index to describe the focusing properties of a magnetic lens, a surprising feature of this focusing is revealed—the refractive index which characterizes the presence of a magnetic field is constant on the axis, and is of interest, therefore, only off the axis.

If we examine the ray equations of Chapters 4 and 5, we see that the transition from electrostatic focusing (equation 4.9) to mixed electromagnetic focusing (equation 5.12) is effected by replacing $\varphi''$ by $\varphi'' + \frac{e}{2m} B^2$.

This suggests that we should make the same substitution in the Gaussian electrostatic refractive index:

$$n^2 = \varphi(z) - \left(\frac{r^2}{4}\right) \varphi''(z), \tag{6.8}$$

and regard

$$n^2 = \varphi(z) - \left(\frac{r^2}{4}\right) \left[ \varphi'' + \left(\frac{e}{2m}\right) B^2 \right] \tag{6.8'}$$

as the general expression for the Gaussian refractive index.

It is easy to verify that the correct ray-path is obtained if we trace the refraction through a medium of fictitious refractive index $n$ as given by (6.8'). In fact, applying the electrostatic equations to a medium with fictitious potential $\Phi_1(r, z) = n^2(r, z)$ gives the correct equations of motion in coordinates rotating with the Larmor precession velocity.
6.2.2 The Refractive Index of the Fictitious Motion: the General Case (Störmer, 1933)

Assuming that the observer too is rotating at the Larmor precession velocity, we attempt to establish a distribution of refractive index which will describe the motion in the rotating meridian plane, according to the rules of geometrical optics. To do this, we substitute for $B$ in the general equations of motion in terms of the vector potential $A$, where $B = \text{curl} \, A$, so that

$$m \frac{dv}{dt} = e(v \times \text{curl} \, A)$$  

(6.9)

(in which $e$ is the algebraic magnitude of the charge). $Oz$ is the symmetry axis of the magnetic field, so that $A$ is reduced to single component $A_\theta$. We find, then:

$$\frac{d^2 r}{d t^2} - r \left( \frac{d \theta}{d t} \right)^2 = - \frac{e}{m} \frac{1}{r} \frac{\partial (r A)}{\partial r} r \theta',$$

$$\frac{1}{r} \frac{d}{dt} \left( r^2 \theta' \right) = \frac{e}{m} \left[ \frac{dr}{dt} \frac{1}{r} \frac{\partial (r A)}{\partial r} + \frac{dz}{dt} \frac{\partial A}{\partial z} \right],$$

(6.10)

$$\frac{d^2 z}{d t^2} = \frac{e}{m} \frac{\partial A}{\partial z} r \theta'.$$

As Störmer pointed out, the second equation integrates directly to

$$\theta' = \frac{e}{m} \left( \frac{A}{r} + \frac{C}{r^2} \right),$$

(6.11)

in which $C$ is a constant of integration. Substituting this value of $\theta'$ into the other two equations, we find

$$\frac{d^2 r}{d t^2} = \frac{e}{m} \frac{\partial}{\partial r} \left[ \frac{e}{2m} \left( A + \frac{C}{r^2} \right)^2 \right],$$

$$\frac{d^2 z}{d t^2} = \frac{e}{m} \frac{\partial}{\partial z} \left[ \frac{e}{2m} \left( A + \frac{C}{r^2} \right)^2 \right].$$

(6.12)

Comparing equations (6.12) with

$$\frac{d^2 r}{d t^2} = - \frac{e}{m} \frac{\partial \Phi}{\partial r},$$

$$\frac{d^2 z}{d t^2} = - \frac{e}{m} \frac{\partial \Phi}{\partial z},$$

(6.13)

which would describe the motion of the same particle in an electric field of potential $\Phi$, the analogy is obvious.

The motion of the electrons with respect to a rotating observer is analogous to that of a light ray in a medium of refractive index (variable but axially symmetrical) $n$ given by

$$n^2 = \Psi(z, r) = \Psi_0 - \frac{e}{2m} \left( A_\theta + \frac{C}{r} \right)^2.$$  

(6.14)
The Gaussian approximation to this formula gives the expression found in the preceding paragraph; near the axis, \( A_0 = \frac{Br}{2} \), which gives the refractive index of \((6.8')\) apart from a constant factor, in the case \( C = 0 \).

The trajectory as seen by an observer in a rotating meridian plane, therefore, is precisely the same as the real trajectory in the corresponding hypothetical electrostatic case. If the constant \( C \) of equation \((6.14)\) is not zero, however, the velocity of rotation—even to the Gaussian approximation—differs from the value given by equation \((5.11)\).

The refraction of the ray can be traced through the combined electric and magnetic fields, step-by-step, by using a refractive index of the form

\[
n^2 = \Phi(r, z) - \left( \frac{e}{2m} \right) \left[ A_0 + \frac{C}{r} \right]^2.
\]

(6.15)

In Figure 43a, the general form of the surfaces \( n = \) constant is shown, for a purely magnetic lens. The index defined by \((6.15)\) has been used by Sandor (1941) to trace the trajectories through a magnetic lens; he also described a device to measure \( n^2 \) directly, but his work has not been followed up. The value of Störrmer's theory lies in the light which it sheds upon the relation between electrostatic and magnetic systems; this will prove to be valuable in aberration theory, where the aberrations of magnetic and of electric systems can be related quantitatively by using \( \Psi' \).

6.2.3 The Refractive Index for a Stationary Observer

It is equally possible to use a refractive index to determine directly the motion relative to fixed axes. Schwarzschild (1905) discovered this representation in the course of his extension of the variational principle to motion in electric and magnetic fields. We shall consider a mixed system, in which
the electrons experience the influence both of an electric field derived from a potential $\Phi(x, y, z)$ and of a magnetic field derived from a vector potential $A$ [such that $B(x, y, z) = \text{curl } A(x, y, z)$].

As the electrons experience a force whose potential is not simply scalar, the principle of least action cannot be applied as straightforwardly as in the electrostatic case; Schwarzschild showed, however, that the principle is still applicable provided we take into account the fact that the kinetic energy $T$ is no longer simply $m v^2/2$, but contains a further term due to the magnetic field.

The presence of this additional term can be made plausible by considering the kinetic energy in a system of coupled circuits. The electromagnetic energy stored in the field which is produced by two circuits of self-inductance $L_1$ and $L_2$ and mutual inductance $M$ is

$$W = \frac{1}{2} L_1 i_1^2 + \frac{1}{2} L_2 i_2^2 + M i_1 i_2.$$

Let us now consider a very special case of such a pair of circuits, in which one circuit is the lens itself, and the other the beam of electrons which proceeds virtually along the axis. In this case, $\frac{1}{2} L_1 i_1^2$ is the electromagnetic energy stored in the self-inductance which constitutes the coil-windings; it is a constant, independent of the second circuit, which is formed by the electron beam. $\frac{1}{2} L_2 i_2^2$ represents the energy of an electron due to its motion, namely $\frac{1}{2} m v^2$. The term $M i_1 i_2$, however, represents energy which is present when an electron moves in a magnetic field. It is precisely the additional term for which we are searching—all we have to do is to transform it into a more suggestive form, in which it can be more easily manipulated.

For this purpose, we introduce the flux $\mathcal{F}_{12}$ which threads the second circuit (the electron trajectory) as a consequence of the presence of the first (the coil). In terms of vector potential $A$:

$$\mathcal{F}_{12} = \oint_{C_2} A \cdot dl_2$$

(where $dl_2$ represents a line element of the second circuit), which allows us to define an energy of coupling for each element of the second circuit

$$dW = A \cdot dl_2 i_2,$$

but $i_2 dl_2 = e v$, so that the energy stored in the electron by virtue of its motion through a magnetic field is of the form

$$W = A \cdot v e.$$

Returning to the principle of least action in the form (6.4), we find

$$\delta \int_{t_0}^{t_1} \left( \frac{1}{2} m v^2 + e v \cdot A + U \right) dt = 0,$$
which can be simplified, using $\frac{1}{2}mv^2 = e\Phi$ and $U = e\Phi$, into

$$\delta \int_{t_0}^{t_1} (2\Phi + v \cdot A) \, dt = 0,$$

or

$$\delta \int_{s_0}^{s_1} \left[ \sqrt{\Phi} + \sqrt{\frac{e}{2m}} \frac{v \cdot A}{|v|} \right] \, ds = 0.$$

In order to render this integral identical to the integral which forms the basis of Fermat’s principle, for refractive index we must use

$$n = K' \left[ \sqrt{\Phi} + \sqrt{\frac{e}{2m}} A \cdot \cos \alpha \right],$$

in which $\alpha$ is the angle between the vector potential and the direction of the trajectory. $K'$ is a constant, whose choice is arbitrary but which can with advantage be chosen to be $\sqrt{\frac{2e}{mc^2}}$, in order that the refractive index be dimensionless. We then find

$$n = \frac{1}{mc} \left\{ mv + e(A \cdot ds) \right\}$$

(ds being an element of length along the trajectory), in which $c$ denotes the velocity of light in vacuo; $e$ is the algebraic value of the charge on the particle (for electrons, $e = -|e|$) but we adopt the convenient convention that $e$ is to be replaced by $|e|$ whenever it occurs under a square root, as in $\sqrt{(e/2m)}$.

Seen from this point of view, the magnetic field behaves like an anisotropic optical medium of a rather curious kind. The refractive index has different values for electrons moving in opposite directions, and the optical reversibility principle is no longer valid. $n$, on the other hand, can be adjusted at will by adding to $A$ any function of the form $\text{grad} f(x, y, z)$ without the trajectories being affected at all, since any function of the form $\text{grad} f$ vanishes identically if subjected to the curl operator. The refractive index has the sole simple property that it is constant for all rays lying in a plane perpendicular to $A$. As $A$ is perpendicular to the meridian planes in a rotationally symmetrical system, the refractive index remains constant for rays moving in a meridian plane. It is upon this result that the success of the theory of magnetic lenses using a fictitious refractive index is based.

Because of these complications, the general expression for refractive index is rarely used. It is, however, possible to obtain a more general, relativistically exact expression of this form, and it is in this extended range of its validity that the interest of so general a refractive index resides. Every problem of electron motion can be attacked in the same way by using a standard procedure, namely, by obtaining the Lagrangian equations.
6.3 THE REFRACTIVE INDEX IN INTERFERENCE PHENOMENA AND THE "WAVE" MECHANICS OF THE ELECTRON

6.3.1 Relative and "Absolute" Definitions in Light Optics; Special Features of Electron Optics

In this chapter, we have hitherto been concerned with geometrical optics: the problem has always involved tracing a ray through a refracting medium, characterized by its refractive index \( n(x, y, z) \). This never entailed more than a knowledge of the relative magnitude of the refractive index; in other words, if a "ray" or "trajectory" has been determined through a medium of index \( n(x, y, z) \), we should obtain the same solution if we were to consider the distribution \( n'(x, y, z) = kn(x, y, z) \), where \( k \) is a constant. This is clearly a consequence of Snell's law,

\[
n_1 \sin i_1 = n_2 \sin i_2,
\]

in which only the ratio, \( n_2/n_1 \), of the refractive indices of two successive elementary layers of the medium is involved. Since this law is valid in both light and electron optics, there seems at first sight to be a complete analogy between the laws of refraction in these two domains.

In the case of light, however, the possibility of transferring from \( n \) to \( kn \) (\( k \) being a constant) is never mentioned; a basic reason for this is that there is one medium with a privileged refractive index: the vacuum. The problem of changing from \( n \) to \( kn \) does not need to be considered because it seems perfectly natural to attribute to the vacuum—which is the simplest, best defined and most easily reproducible medium—the refractive index

\[
n_{\text{vacuum}} = 1.
\]

In electron optics, the vacuum is no longer such a simple medium. For, let us consider the common situation, which is also the simplest, in which the object scatters electrons from an electron gun and the focusing is of the magnetic type. The electrons emerge from the gun with velocity \( v_0 \), which is usually measured in terms of the equivalent potential, \( \Phi_0 \); \( v_0 \) plays the role of initial velocity for the electron scattered by the object. The velocity \( v_0 \) is involved in calculations of the trajectory (in the coefficient of the second term of equation (5.9), for example) and hence, in the general expression for the refractive index which in this case takes the form

\[
n = \left( \sqrt{\frac{e}{2m}} \Phi_0 + \sqrt{\frac{|e|}{2m}} A \cos \alpha \right) K'.
\]  

(6.16)

In the "vacuum", there is thus no unique reference medium; instead, we have a family of media, each element of which is defined by a value of
\( \Phi_0 \). This leads to another convention for defining a dimensionless refractive index: \( K' \) is chosen to be \( 1/\sqrt{\Phi_0} \); in the commonest cases this leads to the following expressions for \( n \):

**electrostatic focusing**

\[
n = \sqrt{\Phi/\Phi_0}, \tag{6.17}
\]

**magnetic focusing**

\[
n = 1 + \sqrt{\frac{e}{2m\Phi_0}} A_v. \tag{6.18}
\]

\( A_v \) denotes the component of vector potential in the direction of the particle velocity \( v \).

### 6.3.2 The Refractive Index in Interference Phenomena; Electron Optics

For light, however, the refractive index is also involved in interference phenomena. Let us consider, for example, the classic experiment of Young’s slits, in the improved form due to Lord Rayleigh (Fig. 43b). Thus if we select \( a \) as the reference tube and maintain the vacuum within it, we can measure the refractive index of a gas introduced into tube \( b \), identical with tube \( a \), by measuring the shift of the interference fringes caused by the presence of gas. This shift is a consequence of the phase difference between the two waves \( a \) and \( b \), resulting from the gas, and this phase difference is given by

\[
\delta \phi = 2\pi l \left( \frac{1}{\lambda_b} - \frac{1}{\lambda_a} \right) = 2\pi \frac{(n_b - 1)l}{\lambda_0}. \tag{6.19}
\]

The length of tube \( b \) is denoted by \( l \), \( \lambda_0 \) is the wavelength \textit{in vacuo} of the monochromatic light employed, \( \lambda_b \) is the wavelength in the gas \( b \), \( n_b \) is the refractive index of gas \( b \) and the refractive index of the vacuum is unity.

![Diagram](image)

**Fig. 43b.** Interference between two coherent light beams, by means of which the refractive index of a gas can be measured. \( a \): evacuated tube; \( b \): gas-filled tube.
In this experiment, and in equation (6.19), therefore, we perceive the fundamental reason why the ordinary definition of the refractive index (with \( n = 1 \) for a vacuum) is regarded as an "absolute" definition.

We now know how to translate this experiment into electron optics, and several experimental workers have followed Chambers (1960) in attempting this. Möllenstedt and Bayh (1962) in particular have succeeded in producing fringes comparable in quality with those regularly obtained with light. This entailed overcoming the experimental problems caused by the extreme shortness of the electron wavelength (0.05 Å in a typical situation). Their experiment is described in §19.2.5 and Fig. 234g shows the fringe shift obtained in the electron analogue of Rayleigh’s experiment; the thin graphite flake replaces the medium in tube \( b \) in the light optical situation.

This experiment can be analysed by transforming formula (6.19) into electron optical terms. Thus, if the de Broglie wavelength at a point \( M \) is \( \lambda \), we can write \( \lambda = \lambda_0/n \), if we introduce a refractive index \( n \) for the electrons, just as we do for light in the second part of equation (6.19). The wavelength is given by

\[
\frac{1}{\lambda} = \frac{mv}{h} = \frac{\sqrt{2e|m|\Phi}}{h} \tag{6.20}
\]

(where \( h \) is Planck’s constant) in the simple case when the effects are wholly electrostatic. This is indeed the case in the experiment we are considering, since the graphite flake has the effect of modifying the electrical potential. The so-called "internal" potential is slightly different from that of the same point in space before the flake was inserted; this effect is adequately explained in the theory of solids. We obtain a general formula for the electrostatic case by supposing that the refractive index varies continuously (at the edge of the flake, we then assume that there are two "transition" zones in which the variation is rapid) and that the phase change \( d\phi \) produced by an element of length \( ds \) is

\[
d\phi = \frac{2\pi ds}{\lambda} = \frac{2\pi n ds}{\lambda_0}, \tag{6.21}
\]

which leads to

\[
n = \frac{\lambda_0}{\lambda} \tag{6.22}
\]

and, using (6.20), to

\[
n = (\Phi/\Phi_0)^{1/2}. \tag{6.23}
\]

There are other experiments in which magnetic effects occur along the path \( b \), and it is then interesting to consider the theory of the general case of mixed electric and magnetic fields. In this situation, the books on wave mechanics do not give a general expression for \( \lambda \) but instead, they
derive a relation giving the phase change of the wave \((\Delta \phi)^B_A\) when we pass from a point \(A\) to a point \(B\). This branch of wave mechanics is explained in a particularly careful fashion by Professor Feynman and his collaborators (1964) in \S\ 15.5 of the second volume of their \textit{Lectures on Physics}: a simple discussion is also to be found in the opening chapters on quantum mechanics of Blokhintsev (1964), which is a good introduction to the fundamental article by Aharanov and Bohm (1959). The phase difference is given by
\[
(\Delta \phi)^B_A = \frac{2\pi}{\lambda_0} \int_A^B \left[ \left( \frac{\Phi}{\Phi_0} \right)^{1/2} ds - \left( \frac{|e|}{2m\Phi_0} \right)^{1/2} A \cdot ds \right],
\]
(6.24)
in which \(A\) is the vector potential of the magnetic field and \(\lambda_0\) is the de Broglie wavelength which in the purely electrostatic case corresponds to the potential beyond the gun and is given by formula (6.20) with \(\Phi = \Phi_0\).

A straightforward calculation shows that the wave mechanical expression (6.24) leads to the same result as we should obtain in light optics if the wavelength in the general mixed case is given by
\[
\frac{1}{\lambda} = \frac{mv + eA_v}{h}.
\]
(6.25)

For the refractive index, we then have
\[
n = \frac{\Phi^{1/2} + eA_v}{\Phi_0^{1/2}}.
\]
(6.26)

\(A_v\) denotes the component of vector potential \(A\) in the direction of the velocity \(v\) of the particle, the charge of which is equal to \(e\) in algebraic value (for the electron, \(e = -|e|\)).

In the geometrical optical approximation, for which the refractive index does not perceptibly vary over distances of the order of a wavelength, \(\lambda_0\), we see that wave mechanics leads to the same expression for the refractive index as classical mechanics. The former is more exact, however, and the value of \(K'\) in equation (6.16) is determined: \(K' = 1/\sqrt{\Phi_0}\). In addition, these wave mechanical arguments have the virtue of bringing out the character of the vector potential \(A\) as the \textit{local physical quantity} representing magnetic effects in formulae (6.18) and (6.26).

This point has been clearly and exactly discussed by Feynman \textit{et al.} (\textit{loc. cit.}) and in the following section, we attempt no more than to describe the interference phenomena upon which this analysis reposes.

\textbf{6.3.3 The “Reality” of the Vector Potential}

In the foregoing formulae, the magnetic character is defined by the single condition \(A \neq 0\). In other words, if in some region within our
Experimental space a vector potential $A$ is present, even though the magnetic field vanishes there ($B = 0$), we are confronted with a "magnetic" case in which $A_\theta$ must appear in the refractive index. In other words again, an electron interference experiment may reveal the existence of $A$ locally, irrespective of the presence of $B = \text{curl } A$, and hence even when $\text{curl } A = 0$. It is, moreover, very easy to create two regions with a solenoid, in one of which $B = 0$, $A \neq 0$; this region is the space outside an idealized solenoid of infinite length, or in practice, very long (there is no need to adopt a toric winding). Inside the solenoid, both $A$ and $B$ exist. The experiment is thus in the form illustrated in Fig. 43c, and the sole complication in comparison with Rayleigh's experiment is that $A_\theta$ is not constant in the region in which the electrons interact with $A$: in the space outside the

![Diagram](image)

**Fig. 43c.** Interference between two electron beams, displaying the effect of the vector potential $A_\theta$ of the solenoid $S_\theta$. The two beams (1) and (2) are emitted from $A$, and deflected by an electric field between $PP'$ and $S_\theta$. (This experiment is identical with that using the Fresnel biprism.)
solenoid, the refractive index of the medium is variable. The dimensions of the solenoid must also be minute and \( A \) will affect both waves simultaneously but in opposite senses; the term \( A_v \) will be of opposite sign at two corresponding points of the two regions, and the fringe shift for a given current is thus doubled.

It is always extremely difficult to obtain an appreciable spatial separation between the two coherent beams that are to interfere. It was for this reason that the first experimentalists, Chambers (1960) and Fowler et al. (1961), employed not a solenoid but a very slender needle of magnetized iron (a whisker) and were only able to confirm the theory rather crudely because of uncertainty about the magnetization of the iron.

In 1962, Möllnenstedt and Bayh succeeded in making a genuine solenoid of very small diameter \((2a = 10 \mu m)\), and by means of their biprism they were able to separate the beams sufficiently to insert this microscopic coil. With this, they obtained good quantitative confirmation of the theory. If \( \Psi \) is the magnetic flux inside the solenoid \((\Psi = B\pi a^2)\), then at a point \( M \) distance \( r \) from the axis of the coil the vector potential has a component \( A_\theta \) given by

\[
A_\theta = \frac{\Psi}{2\pi r}.
\]

There is no need to know this expression, however, since the fundamental property of the vector potential gives us the quantity of interest directly:

\[
\oint A \cdot ds - \oint A \cdot ds = \Psi. \tag{6.27}
\]

The wave mechanical formula (6.24) then shows directly that the phase difference between the two beams, \( \delta \phi \), is given by

\[
\delta \phi = \frac{2\pi e}{h} \Psi, \tag{6.28}
\]

as Aharonov and Bohm showed for the first time in 1959 and as Professor Feynman explains in detail in chapter 15 of volume II of his Lectures. For \( \delta \phi = 2\pi \), the interference pattern is shifted by one fringe-width, which in theory corresponds to flux \( \Psi_{th} \):

\[
\Psi_{th} = \frac{h}{e} = 4.13 \times 10^{-15} \text{ tesla} = 4.13 \times 10^{-7} \text{ maxwell}. \tag{6.29}
\]

Experimentally, Möllnenstedt and Bayh found

\[
\Psi_{exp} = 4.07 \times 10^{-7} \text{ maxwell}.
\]

The sources of error have been analysed in detail by Werner and Brill (1960), who showed in particular that the presence of time-dependent parasitic fields would not modify the theory in any way.

These interference effects were predicted as early as 1949 by Ehrenberg and Siday, who employed purely optical arguments.
CHAPTER 7

LENS DEFECTS

7.1 THE LIMITS OF THE GAUSSIAN APPROXIMATION

In practice, the Gaussian theory is of considerable importance, as there are a great many electron optical instruments which obey the laws to which it leads, and which form images which seem both well-defined and faithful. Nevertheless, a meticulous examination of the image always reveals defects, so that instead of a point image, we find a small spot which need not even be located at the Gaussian image point if the object is not situated on the axis. Whether, however, the origin of these defects is to be sought in the terms which we have neglected in establishing the Gaussian theory, or whether they are due to other causes which we have not as yet mentioned can only be decided by a careful study of each individual instrument, and is not known in every case.

We shall not expatiate here upon the effects caused by the mutual repulsion of the electrons, which are known as “space charge effects”, as the effect of space charge—which has no glass optical parallel—on image formation is, as a general rule, negligible in the instruments which concern us here; the reason for this is simply that, since we can obtain very brilliant images and intense photographic effects with very low current densities, the electrons will be widely separated in space, and their interaction negligible. Space charge is of prime importance, however, in certain electron guns and in electron tubes, but this represents only a very specialized category of phenomena, and we shall simply refer the reader to §8.5.3 and to the relevant books by Pierce (1949) and Klemperer (1953).

We shall, however, study in some detail the third order geometrical aberrations (which arise from terms in \( r^3 \), which we ignored in the Gaussian approximation), chromatic aberration, and the mechanical aberrations which arise from imperfections in the assembly and machining of the various parts of the instrument. It is the magnitude of these defects which decides the width of the tubular region around the axis within which the Gaussian approximation is valid; this, in practice, means the maximum diameters of the metal apertures which define this region.

* A priori * it would seem that the quality of an image can be improved to any desired extent simply by reducing the diameter of these apertures.
This cannot be continued indefinitely, however, owing to the corresponding reduction of luminous intensity (in a cathode ray tube, for example, we have to choose between spot size and luminous intensity), or more rarely, to diffraction; the latter is particularly marked in electron microscopy.

The angular aperture $\alpha$ of the elementary pencils which are emitted by a point on a microscope object is of the order of $5 \times 10^{-4}$ radians. The image of a point is thus a small spot, of which the diameter, divided by the magnification, is

$$d = \frac{24}{\alpha \sqrt{\Phi}}$$

(in which $d$ is measured in millimicrons, $\alpha$ in radians and $\Phi$ in kV).

The diffraction is a consequence of the wave-like qualities of the electron, which ceases to behave like a classical particle when its trajectory is too narrowly restricted by small diaphragms. This is an example of the existence in instrumental optics of the electron waves of L. de Broglie, which behave like light waves save in that their wavelengths are usually some 100,000 times smaller than those of light in the visible region.

### 7.2 Third Order Aberrations

#### 7.2.1 Definition of the Aberration Coefficients

We define $x_0$, $y_0$ to be the coordinates of some point $A$ in the object plane $z = z_0$ and $x_i$, $y_i$ to be the coordinates of $A'$, the Gaussian image of $A$, in the plane $z = z_i$ (Fig. 44). The point $N$, at which a ray from $A$

![Diagram of electron optics](image)

Fig. 44. The parameters which characterize a trajectory.

passes through the diaphragm which defines the extent of the beam, has coordinates $x_d$, $y_d$; if the actual aperture is situated within the lens, however, it is replaced by a hypothetical equivalent aperture, in the region in which
the field is zero. We may for example, select the image of the real stop
formed by the part of the lens which precedes it, which gives us the “entry
pupil”, the plane $D'$ in Fig. 44. This ray through $A$ and $N$ intersects
the image plane $z = z_i$ not at the Gaussian image point $A'(x_i, y_i)$ but at
the point $A''(x'_i, y'_i)$. The aberration is characterized by the vector $A'A''$, thus:

\[
\begin{align*}
\xi &= x'_i - x_i = f(x_0, y_0, x_d, y_d), \\
\eta &= y'_i - y_i = g(x_0, y_0, x_d, y_d),
\end{align*}
\]  

(7.1a)
in which $x_i = G x_0$ and $y_i = G y_0$ ($G$ is the linear magnification).

We now express $f$ and $g$ as series expansions up to terms of the third
order. All the second order and other terms of even order drop out, as all
the quantities $x_0, y_0, x_d$ and $y_d$ must change sign simultaneously in a rotation of 180°.

The general expression for the third order terms contains twenty
coefficients for $x_i$ and twenty for $y_i$; for example:

\[
\begin{align*}
x'_i - G x_0 &= a_1 x_0^3 + a_2 x_0^2 y_d + a_3 y_0^2 x_d + a_4 y_d^3
+ b_1 x_0 x_0^2 + b_2 x_0 x_d y_d + b_3 x_0 y_0^2 + b_4 y_0 x_d^2
+ b_5 y_0 x_d y_d + b_6 y_0 y_d^2 \\
+ c_1 x_0^3 x_d + c_2 x_0^2 y_d y_d + c_3 y_0 x_0 x_d + c_4 x_0 y_0 x_d
+ c_5 y_0 x_d y_d + c_6 y_0 y_d^2 \\
+ d_1 x_0^3 y_d + d_2 x_0^2 y_d y_d + d_3 x_0 y_0^2 + d_4 y_0 y_d^2.
\end{align*}
\]  

(7.1b)

We shall show that the presence of axial symmetry implies a large reduc-
tion in the number of terms; in the general case, only twelve aberration
terms remain.

If we write

\[
\begin{align*}
u &= x_0 + j y_0, \\
v &= x_d + j y_d, \\
w &= \xi + j \eta,
\end{align*}
\]  

(7.2)

then $w$ is a function of $u, v$, and their complex conjugates

\[
\begin{align*}
\bar{u} &= x_0 - j y_0, \\
\bar{v} &= x_d - j y_d
\end{align*}
\]  

(7.3)

($\bar{u}$ and $\bar{v}$ are necessary, since the relation between $w$ and $x_0, y_0, x_d$ and $y_d$
is too general for $u$ and $v$ alone to be sufficient). An arbitrary rotation of
the coordinates about the axis transforms $u$ and $v$ into $u' = u e^{i\theta}$ and $v' = v e^{i\theta}$
respectively (with complex conjugates $\bar{u}' = \bar{u} e^{-i\theta}$, $\bar{v}' = \bar{v} e^{-i\theta}$) and $w$ into
$w' = w e^{i\theta}$. The relation between $w'$ and $u'$, $v'$ must be identical to the relation
between $w$ and $u$, $v$. This condition is satisfied if we write the third order
terms in the form

\[
w = (a u + b v) \cdot P_2(u, \bar{u}, v, \bar{v}),
\]
in which $P_2$ is a polynomial of the second degree which is invariant to the
transformation above, and which is therefore of the form
\[ P_2 = cu \bar{u} + du \bar{v} + ev \bar{u} + fv \bar{v}. \]

$w$, therefore, is of the form
\[
w = (A + jA')v^2 \bar{v} + (B + jB')u^2 \bar{v} + (C + jC')uv \bar{u}v
+ (D + jD')\bar{u}v^2 + (E + jE')uv \bar{v} + (F + jF')u^2 \bar{u}.
\]

This function contains twelve real coefficients, but we can simplify it
still further by taking an important physical property of the system into
account; according to the theorem of Malus and Dupin, every ray is a
normal to each member of a family of surfaces (a statement and proof
of this theorem is to be found in books on geometrical optics, starting
from the laws of refraction). This physical property is contained in the
statement
\[
\text{Im}\left(\frac{\delta w}{\delta v}\right) = 0
\]
(cf. *Optique Electronique et Corpusculaire*, by Louis de Broglie).

Applying this condition, we find
\[
A' = C' = 0,
\]
\[
2D = E,
\]
\[
2D' = -E'.
\]

In the general case, therefore, we have only eight aberration coefficients,
and $w$ becomes:
\[
w = Av^2 \bar{v} + (B + jB')u^2 \bar{v} + Cu \bar{u}v + (D + jD')\bar{u}v^2
+ 2(D - jD')uv \bar{v} + (F + jF')u^2 \bar{u}.
\]  \hspace{1cm} (7.4)

This formula is in the form suitable for magnetic lenses; in electrostatic
lenses, however, there is no force which would make an electron leave
its meridian plane, which behaves therefore like a symmetry plane. $w$, in
consequence, must be unaffected when $u, \bar{u}, v$ and $\bar{v}$ are replaced by $u, v$
and $\bar{v}$ respectively, which implies
\[
B' = D' = F' = 0.
\]  \hspace{1cm} (7.5)

The aberration coefficients are now, as in glass optics, only five in number.

**7.2.2 The Form of the Aberration Figures**

With the aid of the formulae of the preceding section, we can determine
both the shape and the position of the small spot which plays the role of
image point. If all the terms are of comparable importance, the result
is extremely complicated (Bricka and Bruck, 1948). It is, however, possible
to choose the experimental conditions in such a way that all the terms
except one are negligible. We shall, in fact, select these conditions in such
a way that the spots have shapes already familiar in glass optics, made
slightly more complicated by the effect of a magnetic field.

Aberrations produced by the aperture

The lens is assumed to be operating at maximum aperture, with the whole
surface of the diaphragm illuminated, but the object is to remain in the
neighbourhood of the axis. Polar coordinates in the aperture plane are
denoted by $r, \varphi (v = r e^{i\varphi})$ and those in the object plane by $\rho, \theta (u = \rho e^{i\theta})$;
we shall always select $\theta = 0$ to simplify the analysis.

![Diagram](image)

**Fig. 45.** Coma. (a) The circle described by those rays which pass through the
diaphragm, at the same distance from the axis. (b) The appearance of the aberration
figure.

(1) *Spherical aberration.* If we consider a point object on the lens axis,
$u = 0$, equation (7.4) is reduced to

$$w = A r^3 e^{i\varphi}.$$  

(7.6)

If we consider the rays which pass through the entry pupil along a circle
of radius $r$, the angle $\varphi$ varies between 0 and $2\pi$; the points $A''$ in which
these rays intersect the Gaussian image plane lie on a circle whose centre is the
Gaussian image point and whose radius is $A r^3$. If the whole diaphragm is
irradiated, the family of circles will coalesce into a small circular spot, of
radius $A R^3$, where $R$ is the radius of the diaphragm, and $A$ is the *spherical
aberration coefficient*.

(2) *Coma.* Let us suppose that the spherical aberration is negligible for
a pair of conjugate planes $z_0$ and $z_1$; we consider an object sufficiently
close to the axis for $u^2$ and $u^3$ to be negligible. The aberration, therefore,
is given by

$$w = R_D r^2 e^{i(2\varphi + \varphi_D)} + 2e^{-i\varphi_D},$$  

(7.7)
where the substitution

\[ D + jD' = R_D e^{j\varphi_D} \]

has been made.

Once again, we consider those rays which pass through the entry diaphragm around a circle of radius \( r \), over which \( \varphi \) varies from 0 to \( 2\pi \); the points \( A'' \) lie on a circle, with radius \( R_D r^2 \theta \) from the Gaussian image point at an azimuthal angle \( \varphi_D \) from the radius vector \( OA' \). It is to be noticed that as \( \varphi \) varies from 0 to \( 2\pi \), this circle is described twice (because of the presence of \( 2 \varphi \) in the exponent). This is illustrated in Fig. 45a.

If we consider the rays which pass through the remainder of the diaphragm, we obtain a family of superimposed circles, giving an aberration figure of the form shown in Fig. 45b.

\( D \) is known as the \textit{coma coefficient}, while \( D' \), which is responsible for the rotation and is peculiar to magnetic lenses, is the \textit{coefficient of anisotropic coma}.

In general, the aberration figure which is observed will not be of this simple form, as the spherical aberration, which cannot be eliminated from electron lenses, will be superimposed onto the coma; preponderant as the spherical aberration often is, the coma is quite swamped. In glass optics, on the contrary, objectives rigorously corrected for spherical aberration can be obtained, and it is then possible to observe pure coma.

\textit{Field aberrations}

We suppose now that the object is far from the axis, so that all the powers of \( u \) have to be taken into consideration, but that the system is stopped, on the other hand, in such a way that powers of \( v \) higher than unity can be neglected. The general formula (7.4) becomes

\[ w = R_B \varrho^2 r e^{j(\varphi_B - \varphi)} + C \varrho^2 r e^{j\varphi} + R_F \varrho^2 e^{j\varphi_F}, \]  

(7.8)

in which we have rendered the notation more consistent by writing

\[ B + jB' = R_B e^{j\varphi_B}; \quad F + jF' = R_F e^{j\varphi_F}. \]

(1) \textit{Distortion:} The last term is independent of the point in which the ray passes through the diaphragm—it produces, therefore, not a spot but an overall displacement of the image, which is subjected to a translation. \( F \), from now on, will be known as the \textit{distortion coefficient}, while \( F' \) which produces a rotational effect as well, will be called the \textit{coefficient of anisotropic distortion}. (See Fig. 46.)

If the object is a square mesh, it is easy to verify that the image is deformed into a "cushion" or a "barrel" according to the sign of \( F \). In a magnetic lens, the presence of \( F' \) produces, in addition, strain across the image. These results are illustrated in Fig. 47.
(2) *Astigmatism and field curvature*: If we consider the aberration from the point of view of the new image which has undergone translation as a result of distortion, the remaining term of equation (7.4) is simply

\[ w = \theta^2 r \left[ R_B e^{i(\psi_B - \phi)} + C e^{i\phi} \right]. \quad (7.9) \]

Once again we consider the family of rays which pass through the entry pupil along a circle of radius \( r \). It is easy to show that the points \( A'' \) define an ellipse, the characteristics and position of which are set out in Fig. 48. When the whole diaphragm is illuminated, the spot is found to be elliptical, with major axis \( 2\theta^2 R(R_B + C) \) and minor axis \( 2\theta^2 R(R_B - C) \), and with the major axis inclined to the radius vector \( OA' \) at an angle \( \psi_B/2 \).

If we calculate the shape of the pencil in image space, we find that it collapses into two focal lines—the lens is said to be essentially astigmatic. The standard procedures of geometrical optics show that the distance between the line foci is proportional to half the sum of the axes of the ellipse; the astigmatic difference depends, therefore, only upon \( R_B e^{i\psi_B} \). \( B \), then, is the *coefficient of astigmatism*, while \( B' \), peculiar to magnetic lenses, which produces a rotation of the directions of the line foci, is the
coefficient of anisotropic astigmatism. The line foci associated with a magnetic lens are no longer "sagittal" and "tangential", but are inclined to the reference direction at an arbitrary angle (Fig. 49).

Except in the special case of a rectilinear object, the best image which is obtainable when astigmatism is present is the circle of least confusion

![Diagram](image)

**Fig. 48.** The ellipse which is traced out by rays which pass through the diaphragm at the same distance from the axis.

![Diagram](image)

**Fig. 49.** The structure of an astigmatic pencil.

of the beam. It is of some interest to calculate the distance between this circle and the Gaussian image plane; this distance depends only upon the difference between the axes of the ellipse, upon $C$ that is, as can be verified without difficulty.

$C$ is the field curvature coefficient. Its presence in the formula implies that the best image is to be obtained not on a plane screen passing through the Gaussian image plane, but on a curved screen, like the ones used in cathode ray tubes.
7.2.3 A Comparison of Magnetic and Electrostatic Lenses

Both types of lens display the same basic defects: spherical aberration, coma, distortion, astigmatism and field curvature.

In addition to these, three further aberrations, which are said to be "anisotropic", appear in magnetic lenses as a result of the difference between the exact rotation of the meridian plane and the rotation of the Gaussian image, which are equal only in the first order. The signs of these aberrations depend upon that of \( B(z) \), and on reversing the direction of the current, they take up a configuration which is symmetrical to the previous one with respect to the meridian plane. The coefficients \( B', D' \) and \( F' \) are peculiar to magnetic lenses.

With the aid of the Störmer potential, \( \Psi \), we can compare the coefficients \( A, B, C, D \) and \( F \) which are common to both kinds of lens, and pass from the coefficients calculated for the electric lens to the corresponding quantities for a magnetic lens. We make a distinction between the two cases in which the integration constant appears and does not appear.

The constant of integration \( C \) does not affect three of the aberrations: spherical aberration, distortion and coma. For spherical aberration, \( C = 0 \) for all the rays when the point object is situated on the axis. Distortion only deforms the image, and does not blur it; rays for which \( C = 0 \) therefore converge onto the same point as those for which \( C \neq 0 \), and hence \( C \) cannot appear in the expression for this aberration.

For coma, finally, we know that two rays whose directions are initially symmetrical with respect to the meridian plane intersect again in the image plane—since these rays correspond to different values of \( C \), however, the latter cannot appear in the general expression.

In these three cases, the transfer from the expressions for the coefficients \( A, D \) and \( F \) which are valid for an electrostatic lens over to those which are valid for a mixed lens, which produces a combined electric and magnetic field, is effected by replacing

\[
\Phi(r, z) \quad \text{by} \quad \Phi(r, z) - \frac{e}{2m} A_0^2
\]

(see § 6.2.2).

Since

\[
A_0 = \frac{1}{2} r B - \frac{1}{16} r^3 B'',
\]

we have

\[
A_0^2 = \frac{r^2}{4} B^2 - \frac{r^4}{16} B' B''.
\] (7.10)

Hence

\[
\Phi(r, z) = \varphi(z) - \frac{r^2}{4} \left( \varphi'' + \frac{e}{2m} B^2 \right) + \frac{r^4}{64} \left( \varphi''(z) + B B'' \frac{2e}{m} \right),
\] (7.11)
so that in practice the following substitutions have to be made:

\[ \varphi(z) \rightarrow \varphi(z), \]

\[ \varphi''(z) \rightarrow \varphi''(z) + \frac{2e}{m} B^2, \]

\[ \varphi^{(iv)}(z) \rightarrow \varphi^{(iv)}(z) + \frac{2e}{m} B B''. \]

In the other situation, in which \( C \) has an effect, it is no longer possible to pass easily from the expressions for the aberrations in an electrostatic lens to those in a magnetic lens. This is the case for field curvature and astigmatism, where the corresponding coefficients contain supplementary terms in the magnetic case due to the presence of the constant \( C \).

### 7.3 Calculation of the Aberration Coefficients

#### 7.3.1 The Trajectory Method

The aberration coefficients can be fully evaluated by calculating the electron trajectories when the terms in \( r^3 \) which we neglected in the Gaussian approximation are taken into account. First, however, we must determine the terms up to and including the third order in the series expansion of the electric or magnetic field.

We shall show in detail how the spherical aberration is calculated for the simpler case of an electrostatic field. This means that we still have cylindrical symmetry about the axis, and it is profitless to concern ourselves with rays which do not lie in a meridian plane. We can, therefore, perform the whole calculation, as in the Gaussian approximation, simply by determining a meridian ray.

For the electrostatic potential, we find

\[ \Phi(r, z) = \varphi(z) - \frac{r^2}{4} \varphi''(z) + \frac{r^4}{64} \varphi^{(iv)}(z). \]  

(7.12)

The equations of motion, therefore, must be written in the following form, in which terms in \( r^3 \) and \( r'^2 \), hitherto neglected, have been retained:

\[ m \frac{d^2r}{dt^2} = e \left[ \frac{r}{2} \varphi'' - \frac{r^2}{16} \varphi^{(iv)} \right]; \quad m \frac{dz}{dt} = -e \left[ \varphi' - \frac{r^2}{4} \varphi''' \right]. \]  

(7.13)

Replacing the second of these equations by the conservation of energy relation:

\[ \frac{1}{2} m (z_t' + r_t'^2) = e \left[ \varphi(z) - \frac{r^2}{2} \varphi'' \right], \]  

(7.14)
we obtain the third order approximation to $\frac{dz}{dt}$:

$$\frac{d}{dt} = \sqrt{\frac{2e\varphi}{m}} \left(1 - \frac{r^2}{4 \varphi'} \right) \left(1 - \frac{r'^2}{2} \right) \frac{d}{dz}.$$

Finally, we obtain the trajectory equation for an electrostatic lens:

$$\frac{d^2r}{dz^2} + \frac{1}{2} \frac{q'}{\varphi} \frac{dr}{dz} + \frac{1}{4} \frac{q'''}{\varphi} r = \left[ \frac{1}{32} \frac{q^{(iv)}}{\varphi} - \frac{1}{16} \left( \frac{q'''}{\varphi} \right)^2 \right] r^3$$

$$+ \frac{1}{8} \left[ \frac{q''''}{\varphi} - \frac{q'''}{\varphi^2} \right] r^2 r' - \frac{1}{4} \frac{q'''}{\varphi} r r'^2 - \frac{1}{2} \frac{q'}{\varphi} r'^3 \ldots \tag{7.15}$$

The reader can establish without difficulty the trajectory equation for an axially symmetric system containing both an electric field and a magnetic field. The full expression is to be found in the book by de Broglie (1950).

Since the aberration terms are small, we shall integrate this equation by a perturbation method. We write

$$r = r_0 + \varepsilon$$

in which $r_0$ represents the Gaussian trajectory, originating in a point object, in some chosen direction. As was shown in § 4.3.1, it is possible to find the equation of this trajectory if we know two particular trajectories $X(z)$ and $Y(z)$. The perturbation $\varepsilon$ is sufficiently small for powers greater than unity to be negligible.

The equation for the perturbation, therefore, is

$$\frac{d^2\varepsilon}{dz^2} + \frac{1}{2} \frac{q'}{\varphi} \frac{d\varepsilon}{dz} + \frac{1}{4} \frac{q'''}{\varphi} \varepsilon = \left[ \frac{1}{32} \frac{q^{(iv)}}{\varphi} - \frac{1}{16} \left( \frac{q'''}{\varphi} \right)^2 \right] r_0^3$$

$$+ \frac{1}{8} \left[ \frac{q''''}{\varphi} - \frac{q'''}{\varphi^2} \right] r_0^2 r'_0 - \frac{1}{4} \frac{q'''}{\varphi} r_0 r_0'^2 - \frac{1}{2} \frac{q'}{\varphi} r_0'^3, \tag{7.16}$$

a linear second order differential equation with a term independent of $\varepsilon$ on the right-hand side. Provided two solutions of the corresponding homogeneous equation are known, this equation can be solved by the method of “variation of parameters”. These two solutions are, however, already known, since the homogeneous equation is none other than the equation for the Gaussian trajectories—Fig. 24 illustrates their behaviour.

The full calculations are lengthy, but not particularly difficult—we shall only give them in outline.

We consider a Gaussian ray which leaves an axial point object which has abscissa $z_0$, at an inclination $\alpha$, so that

$$r_0 = \alpha X(z), \tag{7.17}$$

in which $X(z)$ is the particular solution shown in Fig. 24. Both the perturbation and its derivative must be zero at $z_0$, since the third order trajectory certainly coincides with the Gaussian trajectory at the object. $\varepsilon(z_0) = \varepsilon'(z_0) = 0$, therefore. Elsewhere, we write

$$\varepsilon = \alpha(z) X(z) + \beta(z) Y(z) \tag{7.18}$$
and apply the method of variations of parameters. We have

\[ \alpha' (z) X(z) + \beta' (z) Y(z) = 0, \]

\[ \alpha' (z) X'(z) + \beta' (z) Y'(z) = \alpha^3 P(X, X', \varphi), \]

in which \( P \) represents the inhomogeneous term of the differential equation for \( \varepsilon \) into which the value of \( r_0 \) has been substituted.

The spherical aberration represents the value of \( \varepsilon \) in the Gaussian image plane, \( z = z_i \). But \( X(z_i) = 0 \), and the radius of the aberration disc is given, therefore, by the expression:

\[ \varepsilon(z_i) = |w| = \beta(z_i) Y(z_i). \]

Solving the pair of simultaneous equations for \( \alpha' \) and \( \beta' \) we obtain

\[ \beta'(z) = \frac{\alpha^3 X P(X, X', \varphi)}{X'Y - XY'}. \]

But according to (4.16''')

\[ X'Y - XY' = \sqrt{\varphi_0 \varphi}, \]

so that

\[ \beta'(z) = \frac{\alpha^3}{\sqrt{\varphi_0 \varphi}} X \sqrt{\varphi} \cdot P(X, X', \varphi), \]

and integrating, this gives

\[ \beta(z_i) - \beta(z_0) = \frac{\alpha^3}{\sqrt{\varphi_0 \varphi}} \int_{z_0}^{z_i} X \sqrt{\varphi} P(X, X', \varphi) \, dz. \]

Since \( \varepsilon(z_0) = 0 \), however, \( \beta(z_0) \) is necessarily zero, so that finally we obtain

\[ |w| = \frac{\alpha^3}{\sqrt{\varphi_0 \varphi}} \int_{z_0}^{z_i} X \sqrt{\varphi} \cdot P(X, X', \varphi) \, dz. \]  \hspace{1cm} (7.19)

In order to express this result in the notation which we use in the general study, we introduce an entry pupil, radius \( R \), which is placed a distance \( \delta \) from the object. \( R \) and \( \alpha \) are related by the equation \( R = \alpha \delta \) (we have in fact two points in object space, between which the rays are straight lines). The transverse linear magnification, \( G \), is given by \( Y(z_i) \), as Fig. 24 shows. We know that

\[ |w| = A R^3, \]

and hence

\[ A = \frac{G}{\delta^3 \sqrt{\varphi_0}} \int_{z_0}^{z_i} X^4 \sqrt{\varphi} \left\{ \frac{1}{32} \varphi^{(iv)} - \frac{1}{16} \left( \frac{\varphi'''}{\varphi} \right)^2 + \frac{1}{8} \left( \frac{\varphi''}{\varphi} - \frac{\varphi'}{\varphi^2} \right) \frac{X'}{X} \right. \]

\[ - \frac{1}{4} \frac{\varphi''}{\varphi} \left( \frac{X'}{X} \right)^2 - \frac{1}{2} \frac{\varphi'}{\varphi} \left( \frac{X'}{X} \right)^3 \} \, dz. \]
Integration by parts removes $\varphi'''$ and $\varphi' \varphi''$; for example,
\[
\int_{z_0}^{z_i} \frac{\varphi'''}{\sqrt[3]{\varphi}} X^3 \, \mathrm{d}z = \left[ \frac{\varphi'''}{\sqrt[3]{\varphi}} X' X^3 \right]_{z_0}^{z_i} + \int_{z_0}^{z_i} X^4 \sqrt[3]{\varphi} \left[ \frac{1}{4} \left( \frac{\varphi'''}{\varphi} \right)^2 + \frac{\varphi' \varphi''}{\varphi^2} X' X^3 - 3 \frac{\varphi'}{\varphi} \left( \frac{X'}{X} \right)^2 \right] \, \mathrm{d}z,
\]
and finally
\[
A = \frac{G}{\delta^3 \sqrt[3]{\varphi}} \int_{z_0}^{z_i} X^4 \sqrt[3]{\varphi} \left[ \frac{1}{32} \frac{\varphi^{(iv)}}{\varphi} - \frac{1}{32} \left( \frac{\varphi'''}{\varphi} \right)^2 - \frac{1}{4} \frac{\varphi''}{\varphi} \left( \frac{X'}{X} \right)^2 - \frac{1}{2} \left( \frac{X'}{X} \right)^4 \right] \, \mathrm{d}z. \tag{7.20}
\]

A different aberration constant is frequently used, which is defined by the relation
\[
|w| = G C_3 X^3,
\]
$C_3$ and $A$ being connected by the relation
\[
C_3 = \frac{A R^3}{G \alpha^3} = \frac{A \delta^3}{G}.
\]

$C_3$ can therefore be calculated directly from (7.20) by numerical integration; it is, however, better to use the formula below, which has been taken from the list given in § 7.3.2 as it is equivalent to (7.20) but no longer contains terms in $\varphi^{(iv)}$.
\[
C_3 = \frac{1}{16} \frac{1}{\sqrt[3]{\varphi_0}} \int_{z_0}^{z_i} X^4 \sqrt[3]{\varphi} \cdot \left[ \frac{5}{4} \left( \frac{\varphi''}{\varphi} \right)^2 + \frac{5}{24} \left( \frac{\varphi'}{\varphi} \right)^4 \right.
\]
\[
+ \frac{14}{3} \left( \frac{\varphi'}{\varphi} \right)^3 X' X - \frac{3}{2} \left( \frac{\varphi'}{\varphi} \right)^2 \left( \frac{X'}{X} \right)^2 \right] \, \mathrm{d}z. \tag{7.20'}
\]

We could equally well use yet another equivalent formula in which only the function $T(z) = \frac{\varphi'}{\varphi}$ and its derivative $T'(z)$ appear (see Glaser, 1956, p. 229).
\[
C_3 = \frac{1}{64} \frac{1}{\sqrt[3]{\varphi_0}} \int_{z_0}^{z_i} X^4 \sqrt[3]{\varphi} \cdot \left[ \left( 3 T^4 - \frac{9}{2} T^2 T' + 5 T'^2 \right) + 4 T T' X' \right] X \, \mathrm{d}z. \tag{7.20'}
\]

Using Störmer's analogy we can transfer this to the magnetic lens situation, to give:
\[
A = \frac{G}{\delta^3 \sqrt[3]{\varphi}} \int_{z_0}^{z_i} X^4 \sqrt[3]{\varphi} \left\{ \frac{1}{32} \frac{\varphi^{(iv)}}{\varphi} + \frac{1}{16} \frac{e}{m} \frac{BB''}{\varphi} - \frac{1}{32} \left( \frac{\varphi'''}{\varphi} + \frac{e}{2m} B^2 \right) \right\} \, \mathrm{d}z.
\tag{7.21}
\]
For a purely magnetic lens, in which \( \varphi = \varphi_0 \), the expression for \( C_3 \) will become

\[
C_3 = \frac{e}{m} \int_{z_o}^{z_i} X^4 \left[ \frac{1}{16} \frac{B B''}{\varphi_0^*} - \frac{1}{32} \left( \frac{e}{2m} \frac{B}{\varphi_0^*} \right)^2 - \frac{1}{8} \frac{B^2}{\varphi_0^*} \frac{X''^2}{X^2} - \frac{1}{2} \frac{X'^4}{X^4} \right] \, dz,
\]

(7.21')

and this can be transformed into

\[
C_3 = \frac{e}{96 m \varphi_0^*} \int_{z_o}^{z_i} X^4 \left( \frac{2e}{m \varphi_0^*} B^4 + 5 B'^2 - B B'' \right) \, dz,
\]

(7.21'')

and hence into

\[
C_3 = \frac{e}{128 m \varphi_0^*} \int_{z_o}^{z_i} X^4 \left( \frac{3e}{m \varphi_0^*} B^4 + 8 B'^2 - 8 B^2 \frac{X''^2}{X^2} \right) \, dz.
\]

In all these formulae, the potential \( \varphi_0^* \) represents the relativistically corrected value of the acceleration potential. It is (7.21'') which is used above all for the numerical calculation of \( C_3 \).

The complete calculation of the coefficients has been carried out by Glaser (1933) and by Scherzer (1936). The latter had the idea of eliminating all the terms containing third or fourth derivatives by a series of partial integrations. The integrated terms disappear, a consequence of the Lagrange–Helmholtz relation, and the formulae listed in the next section result.

Recknagel (1940) attempted to obtain the aberration coefficients of an electron lens, starting from the reduced equations (§ 4.2.2), which give a simpler Gaussian equation. Later, however, Sturrock (1951) established these formulae by a more direct method, the principle of which we shall explain later.

### 7.3.2 List of the Aberration Coefficients

We give below the formulae obtained by Scherzer, in a notation which is in conformity with that which we have used earlier; \( X(z) \) and \( Y(z) \) represent the two Gaussian trajectories defined in § 4.3.1, \( \varphi(z) \), \( B(z) \) and \( C \) represent the axial potential, the axial magnetic field, and the integration constant (not to be confused with the field curvature coefficient), and \( G \) the magnification; \( \delta \) is the distance between the object plane and the entry diaphragm.
We obtain
\[ A = \frac{G}{16 \delta^2 \sqrt{\varphi_0}} \int_{z_0}^{z_i} X^4 \sqrt{\varphi} \left[ f + 4g \frac{X'}{X} + 2h \left( \frac{X'}{X} \right)^2 \right] dz, \]
\[ B = \frac{G}{16 \delta \sqrt{\varphi_0}} \int_{z_0}^{z_i} X^2 Y^2 \sqrt{\varphi} \cdot \left[ f + 2g \left( \frac{X'}{X} + \frac{Y'}{Y} \right) + \left( 3h + \frac{q^2}{q^2} \right) \left( \frac{Y'}{Y} - \frac{X'}{X} \right)^2 + 2h \frac{Y' X'}{Y X} \right] dz + \frac{G}{4 \delta} \left[ X' Y' \right]_{z_0}^{z_i}, \]
\[ B' = \frac{G}{8 \delta} \sqrt{\frac{2e}{m}} \int_{z_0}^{z_i} X Y B \sqrt{\varphi} \cdot \left[ k - \frac{q^2}{4} \left( \frac{Y'}{Y} + \frac{X'}{X} \right) + \varphi \frac{X' Y'}{X Y} \right] dz \]
\[ + \frac{G}{16 \delta} \sqrt{\frac{2e}{m}} \sqrt{\varphi_0} \left[ B' \right]_{z_0}^{z_i}, \]
\[ C = \frac{G}{8 \delta \sqrt{\varphi_0}} \int_{z_0}^{z_i} Y^2 X \sqrt{\varphi} \cdot \left[ f + 2g \left( \frac{X'}{X} + \frac{Y'}{Y} \right) - \left( h + \frac{q^2}{2q^2} \right) \left( \frac{X'}{X} - \frac{Y'}{Y} \right)^2 + 2h \frac{X' Y'}{X Y} \right] dz + \frac{G}{8 \delta} \left[ \sqrt{\varphi_0} \frac{q'}{q^{3/2}} 4 X' Y' \right]_{z_0}^{z_i}, \]
\[ D = \frac{G}{16 \delta^2 \sqrt{\varphi_0}} \int_{z_0}^{z_i} Y X^3 \sqrt{\varphi} \cdot \left[ f + g \left( 3 \frac{X'}{X} + \frac{Y'}{Y} \right) + h \left( \frac{X' Y'}{X Y} + \left( \frac{X'}{X} \right)^2 \right) \right] dz + \frac{G}{8 \delta^2} \left[ X' Y \right]_{z_0}^{z_i}, \]
\[ D' = \frac{G}{16 \delta^2} \sqrt{\frac{2e}{m}} \int_{z_0}^{z_i} \frac{B X^2}{q^2} \sqrt{\varphi} \cdot \left[ h - \frac{q'}{2} \frac{X'}{X} + \varphi \left( \frac{X'}{X} \right)^2 \right] dz, \]
\[ F = \frac{G}{16 \sqrt{\varphi_0}} \int_{z_0}^{z_i} Y^3 X \sqrt{\varphi} \cdot \left[ f + g \left( 3 \frac{Y'}{Y} + \frac{X'}{X} \right) + h \left( \left( \frac{Y'}{Y} \right)^2 + \frac{Y' X'}{X Y} \right) \right] dz \]
\[ + \frac{G}{4} \left[ Y^2 \left( \frac{q''}{8q^2} + \frac{5}{32} \left( \frac{q'}{q} \right)^2 \right) + \frac{q'}{2q} \frac{Y'}{Y} + \frac{3}{2} \left( \frac{Y'}{Y} \right)^2 + \frac{e B^2}{8m \varphi} \right]_{z_0}^{z_i}, \]
\[ F' = \frac{G}{16} \sqrt{\frac{2e}{m}} \int_{z_0}^{z_i} \frac{B Y^2}{q^2} \sqrt{\varphi} \cdot \left[ k - \frac{q'}{2} \frac{Y'}{Y} + \varphi \left( \frac{Y'}{Y} \right)^2 \right] dz \]
\[ + \frac{G}{32} \sqrt{\frac{2e}{m}} \left[ \frac{B Y^2}{q} \left( \frac{3 q'}{2 q} + 2 \frac{Y'}{Y} - B' \right) \right]_{z_0}^{z_i}, \]
in which

\[ f = \frac{5}{4} \left( \frac{q''}{q} \right)^2 + \frac{5}{24} \left( \frac{q'}{q} \right)^4 + \frac{e}{m} \frac{B^2}{q} + \frac{35}{16} \frac{e}{m} \left( \frac{q'}{q} \right)^2 \frac{B^2}{q}, \]

\[ g = \frac{7}{6} \left( \frac{q'}{q} \right)^3 - \frac{e}{2m} \frac{q'}{q} \frac{B^2}{q}, \]

\[ h = -\frac{3}{4} \left( \frac{q'}{q} \right)^2 - \frac{e}{2m} \frac{B^2}{q}, \]

\[ k = \frac{3}{8} \frac{e}{m} B^2 + \frac{9}{8} \frac{q'}{q}^2 - \frac{q'}{q} B'. \]

7.3.3 The Eikonal Method

The essence of the method

An electrical system is wholly defined if its potential function is known; likewise, the relevant thermodynamic potential is sufficient for all the principal characteristics of a chemical or physical system to be calculable. Hamilton had the idea of trying to find a function from which all the properties of an optical system could be simply derived. Such functions (for many equivalent definitions have been suggested since that due to Hamilton) bear the generic title of "eikonals". This word is surrounded by an air of mystery, which conjures up, quite wrongly, pictures of bizarre calculations. It is important, therefore, to convince oneself that the eikonal is, in principle, no more complicated a concept than the potential in electricity.

Practical calculations, however, will be more difficult, since there is, in optics, no simplifying principle analogous to the conservation of energy in electrostatics, for example. In consequence, the difference between the values of the eikonal function at two points will depend upon the actual trajectory along which the particle travels from one point to the other, unlike electric potential, which is independent of path. It seems, therefore, that the eikonal function can be calculated only if we already know the light rays or corpuscular trajectories. Since the whole point of the function is precisely to calculate these rays, its introduction seems rather to beg the question than to answer it. It is doubtless due to this aspect of the eikonal method that it has had little success as a basis for teaching geometrical optics.

In fact, we shall show that it is possible to surmount this difficulty by taking into account the fact that the optical trajectory between two points \( A \) and \( B \) has a "stationary" value—if a ray is displaced slightly from its real position, only a second order change in the length of the trajectory ensues. This property is sufficient for us to be able to extract a great deal
of information from the eikonal, and in particular, to calculate the aberration coefficients. The results can be obtained much more easily and quickly than by using the trajectory method. It is, however, to be emphasized that the eikonal method does not involve any numerical calculation, so that the final formulae are no less complicated irrespective of the method with which they have been established.

The Hamilton eikonal and the Schwarzschild eikonal

We consider two points, \( M_0 \), situated in object space, with coordinates \( x_0, y_0, z_0 \), and \( M_i \), lying in image space, with coordinates \( x_i, y_i, z_i \). In general, only one light ray, which we shall assume to be known, links these points. Along this ray, we calculate the "optical distance"

\[
\int_{M_0}^{M_i} n \cdot ds = V,
\]

in which \( ds \) is the element of arc-length along the ray. It is possible to express this quantity in terms of the coordinates of \( M_0 \) and \( M_i \); we shall denote it, therefore, by

\[
V(x_0, y_0, z_0; x_i, y_i, z_i)
\]

and refer to it as the "Hamilton eikonal".

We consider now two neighbouring points \( M'_0 \) and \( M'_i \), lying on the same ray, with coordinates \( x_0 + dx_0, y_0 + dy_0, z_0 + dz_0 \), and \( x_i + dx_i, y_i + dy_i, z_i + dz_i \). The difference between the optical paths \( M_0 M_i \) and \( M'_0 M'_i \) is given by

\[
dV = \frac{\partial V}{\partial x_0} dx_0 + \frac{\partial V}{\partial y_0} dy_0 + \frac{\partial V}{\partial z_0} dz_0 + \frac{\partial V}{\partial x_i} dx_i + \frac{\partial V}{\partial y_i} dy_i + \frac{\partial V}{\partial z_i} dz_i,
\]

or alternatively, if we introduce the direction cosines of the ray — in object space \( \alpha_0, \beta_0, \gamma_0 \), and in image space \( \alpha_i, \beta_i, \gamma_i \) — by

\[
dV = -(\alpha_0 dx_0 + \beta_0 dy_0 + \gamma_0 dz_0) n_0 + (\alpha_i dx_i + \beta_i dy_i + \gamma_i dz_i) n_i
\]

(see Fig. 50).

![Fig. 50. The relationship between the portion of the trajectory which lies in image space and the portion which lies in object space.](image-url)
Finally, comparing these two expressions for $dV$, we obtain

$$
\frac{\partial V}{\partial x_0} = -n_0 \alpha_0, \quad \frac{\partial V}{\partial y_0} = -n_0 \beta_0, \quad \frac{\partial V}{\partial z_0} = -n_0 \gamma_0,
$$

$$
\frac{\partial V}{\partial x_i} = n_i \alpha_i, \quad \frac{\partial V}{\partial y_i} = n_i \beta_i, \quad \frac{\partial V}{\partial z_i} = n_i \gamma_i.
$$

(7.23)

Given a point in object space and a point in image space, therefore, the Hamilton eikonal allows us to determine the ray which joins them, by simple operations. Unfortunately, this is neither the kind of data which is available in practical problems, nor the kind of information of which we are particularly appreciative. In consequence, many nineteenth-century mathematicians (Seidel, 1856 and Schwarzschild, 1906 in particular) strove to modify the structure of the eikonal in such a way that it would be better adapted to practical problems, without affecting the ease with which it could be manipulated. Seidel's "mixed eikonal" seems the best adapted to our needs, and it is the eikonal which we shall consider as an example (cf. the general discussion of eikonals in the books by Synge, 1936, by Born, 1933, and by Born and Wolf, 1959).

We select a point $A$ in image space, in principle arbitrary, although it is often instructive to examine particular cases; $A$, for example, might be the centre of the exit pupil, with coordinates $x_a, y_a, z_a$. We examine the function

$$
S = V(M_0, M_i) + \overrightarrow{AM_i} \cdot s_i n_i,
$$

(7.24)

in which $s_i$ is the unit vector tangential to the emergent ray. In the expression for $S$, the coordinates $x_i, y_i, z_i$ are replaced by the slope $\alpha_i, \beta_i, \gamma_i$ of the emergent rays, with the aid of the relations (7.23). Thus

$$
dS = dV + n_i [(x_A - x_i) d\alpha_i + (y_A - y_i) d\beta_i + (z_A - z_i) d\gamma_i],
$$

and substituting the expression for $dV$ already obtained, we arrive at

$$
dS = -n_0 (\alpha_0 dx_0 + \beta_0 dy_0 + \gamma_0 dz_0) + n_i [(x_A - x_i) d\alpha_i + (y_A - y_i) d\beta_i + (z_A - z_i) d\gamma_i],
$$

whence

$$
\frac{\partial S}{\partial x_0} = -n_0 \alpha_0, \quad \frac{\partial S}{\partial y_0} = -n_0 \beta_0, \quad \frac{\partial S}{\partial z_0} = -n_0 \gamma_0,
$$

$$
\frac{\partial S}{\partial \alpha_i} = n_i (x_A - x_i), \quad \frac{\partial S}{\partial \beta_i} = n_i (y_A - y_i), \quad \frac{\partial S}{\partial \gamma_i} = n_i (z_A - z_i).
$$

(7.25)

Given an object point, $(x_0, y_0, z_0)$, and the gradient of a ray in image space, $(\alpha_i, \beta_i, \gamma_i)$, therefore, it is possible to obtain, by straightforward differentiation, the coordinates of a point on the ray. It is with this kind of data that we approach the problem of calculating aberrations, where we express
the variation of these coordinates as a function of the emergent direction. Undeniably, this is an ideal procedure for calculating the aberration coefficients.

From what we have said so far, however, it still seems necessary that we should know the ray path in order to calculate the eikonal. We shall now show, with the aid of an example, that the fact that $S$ is stationary allows us to make certain approximations, which has the effect of resolving this apparent contradiction.

The principle behind the calculation of spherical aberration

We shall attempt to determine the spherical aberration of an electrostatic lens, using the eikonal method. This permits us to consider only meridional rays. The point $M_0$ is situated on the axis, at abscissa $z_0$. The exact ray $r(z)$ will be accurately defined if we know its slope $\beta$ in image space (Fig. 51). First, we calculate the Hamilton eikonal:

$$V = \int_{z_s}^{z_t} \sqrt{\Phi'(r, z) \sqrt{1 + r'^2}} \, dz.$$  

(7.26)

The integrand can be expanded in powers of $r$ and $r'$, thus

$$V = \int_{z_s}^{z_t} \sqrt{\varphi} \left[ 1 + \frac{r'^2}{2} - \frac{1}{8} \frac{\varphi''}{\varphi} r^2 + \frac{1}{64} \left( \frac{\varphi''(iv)}{\varphi} - \frac{1}{2} \left( \frac{\varphi''}{\varphi} \right)^2 \right) r^2 ight. 
\left. - \frac{1}{16} \frac{\Phi}{\Phi} \frac{r'^2}{r^2} - \frac{1}{8} r'^4 \right] \, dz.$$  

Interested as we are only in points lying in the Gaussian image plane with abscissa $z_i$, it is easy to pass over to the mixed eikonal by adding the term $\overrightarrow{AM_i} \cdot s_i n_i$ which takes the form

$$\sqrt{\varphi_i} [z_A - z_i] \cos \beta - r_i \sin \beta$$  

in this particular case.

Much more difficult, however, would be the problem of obtaining an explicit expression for $S$, in which $z_i$ and $r_i$ were eliminated and replaced by $\cos \beta$ and $\sin \beta$, but we have no need to attempt this. For, let us consider first of all only terms of second order in $r$ and $r'$, that is to say, let us make

![Fig. 51. The rotationally symmetric system.](image-url)
the Gaussian approximation. Since the object lies on the axis, the Gaussian image is also axial, and we are led unequivocally to the result,

\[
\frac{\partial S}{\partial (\sin \beta)} = n_i r_i = 0,
\]

if we consider only second order terms.

Now, we include terms up to the fourth order, and evaluate the ordinate of the point at which the ray intersects the Gaussian image plane, which is precisely the radius of the spherical aberration disc. About this radius \( w \), the general eikonal relations state that

\[
n_i(w) = \sqrt{\varphi_i \cdot (w)} = -\frac{\partial S}{\partial \beta_i} = \frac{\partial S}{\partial (\sin \beta)},
\]

so that

\[
\sqrt{\varphi_i \cdot (w)} = \frac{\partial}{\partial (\sin \beta)} \int_{z_0}^{z_i} \sqrt{\varphi} \left[ \frac{1}{64} \left( \frac{q'' (iv)}{q} - \left( \frac{q'''}{q} \right)^2 \right) r^4 - \frac{1}{16} \frac{q'''}{q} r^2 r'^2 - \frac{1}{8} r'^4 \ldots \right] \, dz.
\]

This formula, however, is essentially the expression of a correction term; we can make a good approximation to it, therefore, by replacing the exact trajectory by the Gaussian approximation \( r = \beta G \sqrt{\left( \frac{q_i}{q_0} \right)} X(z) \), following the notation already introduced, which gives

\[
\sqrt{\varphi_i \cdot (w)} \approx \frac{d}{d \beta} \int_{z_0}^{z_i} \beta^4 G^4 \frac{q_i^2}{q_0^2} \sqrt{\varphi} \left[ \frac{1}{64} \left( \frac{q'' (iv)}{q} - \left( \frac{q'''}{q} \right)^2 \right) X^4 - \frac{1}{16} \frac{q'''}{q} X^2 X'^2 - \frac{1}{8} X'^4 \right] \, dz.
\]

We introduce the spherical aberration coefficient \( A \), and relate the angular aperture of the beam in object space, \( \alpha \), to its value in image space, \( \beta \), with the aid of the Lagrange–Helmholtz relation, and arrive finally at

\[
A = \frac{G}{\delta^3} \int_{z_0}^{z_i} \sqrt{\varphi} \cdot X^4 \left[ \frac{1}{32} \frac{q'' (iv)}{q} X'^4 - \frac{1}{32} \left( \frac{q'''}{q} \right)^2 - \frac{1}{4} \frac{q'''}{q} \left( \frac{X'}{X} \right)^2 - \frac{1}{2} \left( \frac{X'}{X} \right)^4 \right] \, dz,
\]

(7.27)

using the notation of the first paragraph of this section; \( \delta \) is the distance between the object and the plane of the entry pupil. This expression is identical to the one obtained in the preceding section. The analogous approximations which are necessary in the general case have been discussed in detail by Glaser (1933), who demonstrated that the method is in every case valid.
The complete results

By introducing non-meridian rays, the calculations for the general case can be effected and the expressions for the other aberration coefficients obtained. For a magnetic lens, the starting point is the optical path-length, which is obtained from the generalized refractive index in the form

\[ V = \int_{z_s}^{z_i} \left[ \sqrt{\Phi(x, y, z)} - \sqrt{\frac{e}{2m}} A(x, y, z) \cos \alpha \right] \, ds. \]  (7.28)

The calculation is set out in detail in the books by Picht (1939), Cosslett (1946) and Zworykin (1945).

In conclusion, we point out one further application of the Hamilton eikonal function. If we leave the coordinates \( x, y, z \) of \( M_i \) unspecified, then the equation

\[ V(x_0, y_0, z_0, x, y, z) = \text{const.} \]

defines a family of wave fronts originating in the point \( M_0 \); since we know that each light ray is a normal to the wave front (Huyghens' principle) it is possible to calculate the effects which will be seen in the image plane from the shape of the wave front. Nijboer (1942) has used this method in glass optics, where he was able to classify in a very logical manner the various aberrations of different orders. In electron optics, however, where only third order aberrations are of appreciable importance, we arrive once again at the results which were obtained in the first paragraph by direct study.

7.3.4 Sturrock's Formulae

Another set of formulae has been obtained by Sturrock (1951a), who started from the reduced equation and the corresponding Gaussian rays. The main feature of interest in the use of the reduced rays arises from the fact that if we consider the two familiar basic Gaussian rays, differentiation is replaced, as it were, by integration since (cf. § 4.2.3) \( R'(z) \) represents the operation†

\[ R'(z) = R'(z_0) - \frac{3}{16} \int_{z_0}^{z} \left( \frac{\varphi'}{\varphi} \right)^2 R(z) \, dz. \]

An integration is far more simple and precise an operation than graphical differentiation.

† It is because this unusual notation is used by Sturrock that we employ it here.
The Sturrock formulae are listed below; an electrostatic lens is described by its "characteristic function" $T = \varphi'/\varphi$, and by two "reduced" Gaussian rays $R(z)$ and $S(z)$, defined in an identical fashion to $X$ and $Y$.

The method with which Sturrock establishes these formulae is of extreme elegance; a Lagrangian, adapted to the reduced variables, is constructed, and after some calculation, for details of which the reader is referred to the original paper, the following expressions are obtained:

\[
A = \sqrt{\varphi_0} \int_{z_0}^{z_i} \frac{R^4}{\sqrt{\varphi}} \left[ g + f \left( \frac{R'}{R} \right)^2 \right] \, dz,
\]

\[
B = \sqrt{\varphi_0} \int_{z_0}^{z_i} \frac{S^2 \ R^2}{\sqrt{\varphi}} \left[ g + f \left( \frac{S'}{S} \right)^2 \frac{R'}{R} \right] \, dz + \int_{z_0}^{z_i} \frac{h}{\sqrt{\varphi}} \, dz - \left[ \frac{a \ S' \ R' + d}{\sqrt{\varphi}} \right]_{z_0}^{z_i},
\]

\[
C = \sqrt{\varphi_0} \int_{z_0}^{z_i} \frac{S^2 \ R}{\sqrt{\varphi}} \left[ g + f \left( \frac{S'}{S} \right)^2 \right] \, dz + \int_{z_0}^{z_i} \frac{f + 2h}{\sqrt{\varphi}} \, dz - \left[ \frac{4 \ a \ R' \ S'}{\sqrt{\varphi}} \right]_{z_0}^{z_i},
\]

\[
D = \sqrt{\varphi_0} \int_{z_0}^{z_i} \frac{S \ R^3}{\sqrt{\varphi}} \left[ g + f \left( \frac{R'}{R} \right)^2 \right] \, dz + \int_{z_0}^{z_i} \frac{f \ R \ R'}{\sqrt{\varphi}} \, dz - \left[ \frac{a \ R'^2}{\sqrt{\varphi}} \right]_{z_0}^{z_i},
\]

\[
F = \sqrt{\varphi_0} \int_{z_0}^{z_i} \frac{S^3 \ R}{\sqrt{\varphi}} \left[ g + f \left( \frac{S'}{S} \right)^2 \frac{R'}{R} \right] \, dz + \int_{z_0}^{z_i} \frac{f S \ S'}{\sqrt{\varphi}} \, dz
\]

\[
+ \left[ \frac{3a \ S'^2 + b \ S' \ S' + c \ S'^2}{\sqrt{\varphi}} \right]_{z_0}^{z_i},
\]

in which

\[
f = \frac{3}{64} T^2, \quad g = \frac{1}{512} (15T^4 - 25T^2T' + 20T'^2), \quad h = -\frac{5}{128} T^2;
\]

\[
a = -\frac{1}{8}, \quad b = -\frac{1}{16} T, \quad c = -\frac{1}{32} (2T^2 + T'), \quad d = -\frac{1}{32} T.
\]

**7.4 Chromatic Aberration**

### 7.4.1 Phenomena which Produce Velocity Dispersion

In electron optics, the refractive index $n$ depends upon the electrostatic potential $\Phi$ or the magnetic field $B$. It is a constant of the apparatus, therefore, and it may at first sight seem strange that we should be able to observe

\* Certain changes have been made in Sturrock's notation to make it accord with the notation of the present work. In his article Sturrock refers to $T = \frac{3}{16}(\varphi')^2$ as the "characteristic function".

\* Certain changes have been made in Sturrock's notation to make it accord with the notation of the present work. In his article Sturrock refers to $T = \frac{3}{16}(\varphi')^2$ as the "characteristic function".

\* Certain changes have been made in Sturrock's notation to make it accord with the notation of the present work. In his article Sturrock refers to $T = \frac{3}{16}(\varphi')^2$ as the "characteristic function".
a phenomenon analogous to the variation with colour of the refractive index of a glass lens.

The function $\Phi$, however, includes the constant $\Phi_0$, the mean kinetic energy of the electrons as they leave the cathode. In fact, the electron energies, like the corresponding velocities, are distributed continuously over a narrow band, and to consider this band as a "line" is only a first approximation. This then is one of the sources of variation of $n$, the effect of which is to replace the image point by a small spot—by analogy with the similar effect in glass optics, this defect is known as chromatic aberration. The fundamental characteristic of this aberration—the arbitrary reduction (as $\Phi_c/\Phi_0$) of its magnitude which can be obtained by increasing the accelerating potential $\Phi_0$—is immediately apparent.

The importance of this effect depends essentially upon the nature of the cathode. When the cathode emits thermal electrons, the velocity spread is very narrow, not departing from a mean value $\Phi_c$ by more than a few tenths of a volt either way. $\Phi_c$ is given accurately enough by $\Phi_c = \frac{T}{11,600}$, so that for oxide-coated cathodes ($T \approx 1100^\circ$), $\Phi_c = 0.1$ V while for tungsten cathodes ($T \approx 2500^\circ$), $\Phi_c = 0.2$ V.

The breadth $\delta \Phi$ of the energy spread, which is the energy such that 50 per cent of the electrons have an energy which lies in the region between $\Phi_c \pm \frac{1}{2} \delta \Phi$, is correspondingly less as the temperature is lower.

Since lens potentials range between 1 and 60 kV, in order of magnitude, the chromatic aberration is in fact slight. For photoelectric emission, it is much larger, for if $\nu$ represents the frequency of the light being used, and $v_0$ is the photoelectric threshold frequency of the cathode, the relation $e \Phi_c = h(\nu - v_0)$ shows that $\Phi_c$ can be as large as a few volts. The aberration becomes most troublesome in secondary emission sources, where the electrons can leave the cathode with any velocity up to that of the primary electrons, and where the energy band may be 25 V wide. The reason why thermal sources are usually employed is vividly illustrated by this comparison.

In the important special case of the microscope, the object itself is a secondary source, since it is simply a thin film through which an electron beam passes, and hence a second way in which dispersion of the electron velocities can be caused appears. On passing through the specimen, the electrons are retarded, not continuously since the phenomenon is quantum mechanical, but in such a way that the electron velocities fall into discrete groups, tracing out a line spectrum; the first "line" corresponds to the primary value, and the subsequent lines appear at intervals of about ten volts (Ruthehann, 1941, 1942, 1948; Möllenstedt, 1949, 1952). The overall spectrum, which is still imperfectly understood despite a very promising recent theory, appears to be characteristic of the substance, and reaches
about a hundred volts for the usual thickness of collodion film which provides the specimen support.

Fortunately, it is usual to work at a high operating potential (between 30 and 100 kV), and a very small contrast diaphragm placed in the focal plane of the objective will eliminate some of the electrons which have suffered scattering. Experiment shows that the overall effect is not too harmful, though why this should be so is still uncertain. Recently, a method of filtering off almost all the electrons which have been retarded has been tested (Boersch, 1953; Möllenstedt, 1955), and found to produce a marked improvement of the image.

### 7.4.2 The Magnitude of the Aberration

We can calculate the size of this defect by a perturbation method, not dissimilar to the method used to calculate spherical aberration. If, in an electrostatic lens, \( q \) increases by a small amount, \( \varepsilon \), the Gaussian equation becomes

\[
r'' + \frac{1}{2} \frac{q'}{q + \varepsilon} r' + \frac{1}{4} \frac{q''}{q + \varepsilon} r = 0.
\]

\( \varepsilon \) being small, we collect up the terms in which it appears onto the right-hand side of the equation:

\[
r'' + \frac{1}{2} \frac{q'}{q} r' + \frac{1}{4} \frac{q''}{q} r = \varepsilon \left[ \frac{1}{2} \frac{q'}{q^2} r' + \frac{1}{4} \frac{q''}{q^2} r \right]. \tag{7.29}
\]

We make the substitution \( r = r_0 + \theta \), in which \( \theta \) is a correction sufficiently small for the product of \( \theta \) and \( \varepsilon \) to be negligible, and \( r_0 \) is the Gaussian trajectory which we can express in terms of the familiar functions \( X(z) \) and \( Y(z) \). The correction term \( \theta \), therefore, satisfies the equation

\[
q'' + \frac{1}{2} \frac{q'}{q} \theta' + \frac{1}{4} \frac{q''}{q} \theta = \varepsilon \left[ \frac{1}{2} \frac{q'}{q^2} r_0' + \frac{1}{4} \frac{q''}{q^2} r_0 \right], \tag{7.30}
\]

which is a linear, inhomogeneous, second order differential equation. Since we know two independent solutions of the corresponding homogeneous equation, it can be solved immediately.

We shall not pursue the calculation further, but simply refer the reader to such original publications as Glaser (1940). Where the full calculation, both for the electrostatic and for the magnetic case, is to be found. The results of these calculations are given below: the effect of a perturbation \( \varepsilon \) is to displace the Gaussian image to a plane with abscissa \( z_t + \Delta z_t \), to alter its magnification from \( G \) to \( G + \Delta G \), and to modify the Larmor
rotation in magnetic lenses by $\Delta \theta$, thus:

$$
\Delta z_i = \frac{G^2 \varepsilon}{Vq_o} \sqrt{\left(\frac{q_1}{q_0}\right)} \int_{z_0}^{z_i} \left[ \frac{1}{2} \frac{q'}{q'^2} \frac{X'}{X} + \frac{1}{4} \frac{q''}{q''^2} + \frac{e B^2}{8 m q^2} \right] X^2 \sqrt{q} \cdot dz,
$$

$$
\Delta G = \frac{G \varepsilon}{Vq_o} \int_{z_0}^{z_i} \left[ \frac{1}{2} \frac{q'}{q'^2} \frac{Y'}{Y} + \frac{1}{4} \frac{q''}{q''^2} + \frac{e B^2}{8 m q^2} \right] X Y \sqrt{q} \cdot dz,
$$

$$
\Delta \theta = \frac{\varepsilon}{4} \sqrt{\frac{e}{2m}} \int_{z_0}^{z_i} \frac{B}{q^{3/2}} dz.
$$

(7.31)

The meaning of these corrections is illustrated in Fig. 52.

In reality, $\varepsilon$ has a continuous range of values, so that a whole series of images is superimposed; the overall effect is very complicated, although in

![Image 52. The perturbation produced by a variation $\varepsilon$ of the potential.](image)

practice the fluorescent screen lies in the image plane which corresponds to $\varepsilon = 0$, and the effect of $\varepsilon$ is to create an aberration disc of radius $r_c$ around this image. The radius $r_c$ of this disc is given by

$$
r_c = \frac{1}{G} \sqrt{\frac{q_0}{q_i}} \Delta z_i \alpha = C_c \alpha \frac{\varepsilon}{q_0},
$$

(7.32)

in which $\alpha$ is the angular aperture of the beam in object space, straightforwardly connected by the Lagrange–Helmholtz relation to the angular aperture in image space which actually appears in the formula. For $\varepsilon$, a mean value is selected; we shall return later to the coefficient $C_c$, introduced here.

7.4.3 The Chromatic Aberration of Magnetic Lenses Used for Beta-ray Spectrography

In the $\beta$-ray spectrograph, we take advantage of the extreme dispersion of electron lenses, that is, of the fact that the position of the image depends strongly upon the emission velocity of the particles. This velocity is inversely
proportional to the de Broglie wavelength of the particles, hence the use of the term "dispersion". The energies which we have in mind range over a very wide interval, between some tens of keV up to 25 MeV. Electrostatic lenses, too feebly convergent in the megavolt region, must therefore be rejected, in favour of magnetic lenses which are effective over the whole energy range.

The diagram below, simplified in the extreme, represents the principle of the arrangement which Kapitza suggested in 1924, and which recalls the optical method of focal separation which R. W. Wood used to study

![Diagram](image)

Fig. 53. The principle of the β-ray spectrograph in which the foci are isolated independently.

the infra-red spectrum. A point source at A (Fig. 53) emits β-rays in all directions; the diaphragm with aperture $D_0$ permits only a beam of angular aperture $\alpha$ to fall upon a lens which focuses the family of rays of energy $\phi_1$ into a point image $A'_1$. Rays corresponding to an energy $\phi_2 > \phi_1$ intersect at a different point $A'_2$, beyond $A'_1$, since the convergence of the lens diminishes as the particles are faster. A small diaphragm $D_S$ with a very small hole placed at $A'_1$, therefore, will separate off the rays of energy $\phi_1$ from all the others, and with the aid of a Geiger counter placed behind the diaphragm the intensity of these rays can be measured.

If the image $A'_1$ were a point image, the selection could be made rigorous, but in fact the lens cannot be perfect, and the image is a small circular spot, of radius $\varrho_s$, given by

$$\varrho_s = A \delta^3 \alpha^3,$$

in which $A$ is the aberration coefficient defined in § 7.2.2, and $\delta$ is the distance between the object and the entry pupil.
The rays which correspond to energy \( \varphi_2 \) will be well separated from those of energy \( \varphi_1 \) only if the former are scattered over a circle of radius \( \varphi_c \) at \( D_s \), such that \( \varphi_c \) is considerably larger than \( \varphi_s \).

Writing \( \varphi_2 = \varphi_1 + \delta \varphi \) and \( \varphi_1 = \varphi \), \( \varphi_c \) will represent the radius of the chromatic aberration spot given by equation (7.32),

\[
\varphi_c = C_c \propto \frac{\delta \varphi}{\varphi}.
\]

An idea of the purity of the spectrum which is obtained is given by the quotient \( \varphi_c/\varphi_s \):

\[
\frac{\varphi_c}{\varphi_s} = \frac{C_c}{A \delta^3} \frac{1}{\alpha^3} \frac{\delta \varphi}{\varphi},
\]

although in fact, too small a value for \( \varphi_c \) has been used. We should rather substitute \( \varphi'_c = \varphi_c + \varphi_s \), since the spherical aberration is produced both at \( A_2 \) and \( A_1 \), so that

\[
\frac{\varphi'_c}{\varphi_s} = 1 + \frac{C_c}{A \delta^3} \frac{1}{\alpha^3} \frac{\delta \varphi}{\varphi}.
\]

(7.33)

Two monochromatic lines of the same intensity with wavelengths \( \lambda_1 \) and \( \lambda_3 \) will be separated, therefore, if \( \varphi'_c/\varphi_s = 2 \), since the transmitted intensity which is proportional to \( \left( \frac{\varphi'_c}{\varphi_s} \right)^2 \) will fall in relative value from 1, when the spectrograph is adjusted perfectly for one of the lines \( \lambda_1 \) or \( \lambda_3 \), to 0.71 when it is adjusted for their mean wavelength \( (\lambda_1 + \lambda_2)/2 \).

Such a calculation is admittedly over-simplified, but in fact it proves to be sufficient to justify selecting the value of the ratio \( \varphi/\delta \varphi \)—or better

\[
P = \frac{\lambda}{\delta \lambda} = 2 \frac{\varphi}{\delta \varphi}
\]

(since this makes the notation conform with the usual optical notation)—which corresponds to \( \varphi_c/\varphi = 1 \) to characterize in the usual way the resolving power of the spectrograph. For \( P \), we have

\[
P = \frac{2C_c}{A \delta^3 \alpha^2}.
\]

(7.34)

This theory can be considerably improved by taking into account the various different designs of instrument (Grivet, 1950b, 1951a). Abundant information on the chromatic aberration of magnetic lenses is naturally available in the literature of \( \beta \)-ray spectrography.

The chromatic aberration of electrostatic lenses too has found a useful application, in the hands of Möllenstedt (1949, 1952); the three-electrode lens combines strong dispersion in the transgaussian region with a high stability in the presence of variations of the source potential. Möllenstedt
took advantage of this to obtain a velocity spectrograph with very high
dispersion, which he used to study the velocity loss of electrons as they
passed through thin films such as those which are used as specimen supports
in transmission microscopy, and to examine the distribution of initial ener-
gies at hot cathodes.

7.5 MECHANICAL DEFECTS

7.5.1 Classification

So far we have been presuming that the electrodes are perfect, in the
sense that they are of exactly the shape and dimensions which we have
ascribed to them in the theory. In practice, however, they are machined
with tools of limited precision and the exactness with which the electrodes are
assembled, too, may leave something to be desired—the planes in which
they lie may be imperfectly parallel, and their symmetry axes may not
quite coincide.

These imperfections have the effect of degrading the definition of the
image; even though the lenses be machined with the most precise tools
available, mechanical defects still produce an appreciable effect. The elec-
tron microscope is the extreme case in this respect, for its optical qualities
depend largely upon the precision with which it has been constructed;
the image quality is an excellent test of the accuracy with which the elec-
trodes have been constructed and assembled. The optical effect of an
ellipticity of a fraction of a micron on the rim of a microscope objective—
with a radius of some millimetres—is perceptible, and the precision has
still further to be improved upon.

In glass optics, the situation is quite different, for this kind of defect is
normally quite negligible. This is due to the ease with which a glass surface
can be ground with extreme precision, and to the simplicity of the shapes
of the lens surfaces, which are always either spherical or planar. If necessary,
the error can be reduced below a tenth of a micron.

We shall segregate these mechanical defects into two groups. On the
one hand, those caused by imperfect alignment, a lack of parallelism, or
of eccentricity in the assembly; and on the other, the defects which are
due to a lack of symmetry of the electrodes themselves.

The theory of defects of the first group is well developed (Cotte, 1938;
Wendt, 1943; and Hachenberg, 1948), and seems to go beyond the needs
of most practical cases. Generally speaking, an adequate and rapid cal-
culation can be made by considering the system as a set of slightly eccentric
lenses, and tracing the rays from one lens to its successor (Bruck, 1947).
We conclude from such a calculation that no appreciable aberration results
from an eccentricity of a hundredth of a millimeter, and that the electrodes can always be set sufficiently far apart for any lack of parallelism of the cross-sectional planes to be significant.

7.5.2 Astigmatism due to Ellipticity

The defects which result from asymmetry in the electrodes themselves are both more important and more difficult to calculate. The simplest method is to consider a model, and we shall suppose that the aperture of the central electrode is slightly elliptical (Fig. 54). The potential $\Phi$, therefore, is no longer axially symmetric but possesses two symmetry planes which contain the axes of the ellipse; it can be expanded in the form

$$
\Phi (r, \theta, z) = \varphi (z) - \frac{1}{4} \varphi''(z) \{1 - \varepsilon(z) \cos 2\theta\} r^2 + \ldots,
$$

(7.35)

where the ellipticity of the electrodes is characterized by the function $\varepsilon(z)$. For an axial point object, therefore, we obtain two slightly different families of rays (to the Gaussian approximation) in the symmetry planes $xOz$ and $yOz$ of the elliptical cross-section. Each family of rays gives a clear sharp image, at one of the two points $A_1$ and $A_2$, while the other rays are such that their projections onto one or other of the two symmetry planes intersect the axis at $A_1$ or $A_2$. The overall effect, therefore, is of two short straight lines, lying in the symmetry planes perpendicular to the axis, at the points $A_1$ and $A_2$.

Every axial point object corresponds to two line foci (Glaser, 1943a), and the best image is to be found at the circle of least confusion midway between them; the radius $d$ of this circle provides a measure of this defect, which is known as “ellipticity astigmatism”; $d$ is given by

$$
d = C_e \alpha,
$$

(7.36)
where \( \alpha \) is the angular semi-aperture of the pencil and \( C_e \) is related to \( \varepsilon(z) \) by the expression

\[
C_e = \frac{1}{4r_0^2} \frac{z_l}{z_o} \int_{z_o}^{z_l} \varepsilon \frac{r^2}{\varphi} \, dz.
\]

(7.37)

If the object is near the focus, an approximate expression for \( C_e \) is

\[
C_e = 1.2 \, f \, \varepsilon
\]

\( \varepsilon \) being the mean value of \( \varepsilon \), presumed to be independent of \( z_0 \) and \( z_l \).

All that remains to be done now is to relate \( \varepsilon \) to the geometrical ellipticity of the opening in the electrode, a rather difficult task. The necessary calculation is to be found in the publications of Bertein (1947a, b, c; 1948a, b) and of Bertein, Bruck and Grivet (1947); confirmation of these results has been obtained by Bruck, Remillon and Romani (1948), who made careful measurements in an electrolytic tank, and by Cotte (1949), who followed an alternative line of reasoning. The result is that \( \varepsilon \) is related to the ellipticity \( \eta \) (if \( a \) is the mean radius of the opening and \( \delta a \) is the difference between the extreme radii, then \( \eta = \frac{\delta a}{2a} \)) by the equation

\[
\varepsilon \varphi'' = \frac{0.88}{a} \frac{E_r \eta}{1 + 0.7 \left( \frac{z}{a} \right)^{2.5/2}} \approx \eta \varphi''
\]

(7.38)

\( E_r \), in this equation, is the radial field at the inside edge of the opening, and \( \varphi'' \) is the value of \( \varphi''(z) \) at the centre of the lens. The relation between cause \( \delta a \) and defect \( C_e \) can therefore be summarized thus:

\[
C_e \sim f \eta \sim \delta a.
\]

(7.39)

This theory provides a qualitative explanation of certain observations in the microscope—a careful scrutiny of the image reveals that it is marred by an astigmatism originating in asymmetries of the electrodes (Hillier and Ramberg, 1947; Bertein and Regenstreif, 1947). Combining this image defect with the effect of diffraction, we obtain the following formulae for the resolving power \( d \) and for the optimum semi-aperture \( \alpha \) of the incident beam (Bruck and Grivet, 1947; Bruck, 1947a):

\[
d = 0.85 \lambda \sqrt{\frac{C_e}{\lambda}},
\]

(7.40)

\[
\alpha = 0.71 \sqrt{\frac{\lambda}{C_e}}.
\]
\( \lambda \), measured in millimicrons, is the electron wavelength, which is related to the accelerating potential \( \Phi_0 \) (V) by the expression

\[
\lambda^2 = \frac{1.24}{\Phi_0}.
\]  
(7.41)

A result which is important for the construction of pole-pieces or of electrodes is that the resolving power depends upon the absolute value of \( \delta a \), and thus upon the absolute precision of the machining. It does not depend upon the diameter of the aperture, which can therefore be as small as is necessary to produce the required focal length.

Good agreement is found between the values of \( d \) and \( \alpha \) given by the formulae above, and their experimental values.

An analogous theory for magnetic lenses has been developed by Hillier and Ramberg (1947) and by Sturrock (1951c), and successfully applied to the magnetic microscope. In this case, however, the defects perhaps originate rather in magnetic inhomogeneities in the pole-pieces than in their mechanical imprecision. Since the methods which we have used to combine mechanical defects and diffraction are rather primitive, this theory only gives the order of magnitude of the various quantities involved. We can appreciate the difficulty of the exact calculation by considering a different problem: the problem of calculating the combined effect of diffraction and spherical aberration. The theory, though it has been to some extent explored, is by no means complete, as the phenomena can only be calculated in the Gaussian image plane, which does not correspond to be best focus (Glaser, 1943a; Bruck, 1947a; Maréchal, 1944, 1947, 1948; Bremmer, 1950).

### 7.6 Measurement of the Aberration Coefficients, and Their Orders of Magnitude

The aberration coefficients which we have defined earlier are in a convenient form for theoretical calculation. The quantities which are obtainable experimentally, however, are: the angular apertures of the beams in object and image space, \( \alpha_0 \) and \( \alpha_i \), and the off-axial distances of the various rays; the longitudinal and transverse aberrations in image space, \( \Delta z \) and \( \Delta r \), which are connected by \( \Delta z = x_i \Delta r \); and the Gaussian focal length \( f_0 \). It is convenient, therefore, to represent the aberration discs of radius \( \Delta r \) by relations of the form

\[
\Delta r = k \lambda i^n,
\]

where

\[
1 \leq n \leq 3.
\]
For third order spherical aberration, for example, we have

$$\Delta r = C_s \alpha^3.$$  \hspace{1cm} (7.42)

If only the angles $\alpha_o$ are known, $\alpha$ can be deduced by a form of the Lagrange-Helmholtz relation modified to include third order terms (Grivet, 1950a).

Experimentally, spherical aberration has received the most attention, for it is present even for axial objects, and restricts both the resolving power, an important quantity in microscopy, and the definition of the spot in cathode ray oscillographs. Rather than introduce a correcting device as complicated as the one designed by Scherzer (see § 7.7.6), experimental workers have tended to reduce $C_s$ to the minimum by choosing the best shapes for their lenses.

The astigmatism which results from ellipticity gives a term of the form $\Delta r = C_d \alpha_o$ (on the axis), which has only very rarely been measured, and which is nowadays easily corrected.

Another aberration is particularly important in projective lenses which are often required to focus rays far from the axis, namely distortion. The effect of this aberration appears not in the resolving power but in the image fidelity. If $\Delta r'$ represents the distance between the actual image point and the corresponding Gaussian image point, $r$ the distance of the corresponding object point from the axis and $G$ the linear magnification of the lens, then

$$\Delta r' = C_d G r^3.$$  \hspace{1cm} (7.43)

In image-converters also, it is necessary to correct the distortion.

7.6.1 Measurement of the Spherical Aberration Constant

The methods which are used were originally modelled on those of geometrical optics, the Hartmann test and the shadow method using the edge of a screen.

The Hartmann test

This method has been used to study experimentally two-cylinder lenses (Epstein, 1936; Klemperer and Wright, 1939) and three-electrode lenses of uncommon design; it gives the optical constants of the lens in question as well as the spherical aberration. A diaphragm $D$, perforated by a number of small holes distributed concentrically about the axis (Fig. 55), is placed in the path of a parallel electron beam. The pencils which are defined by this diaphragm pass through the lens $L$ at distances $r_1, r_2, \ldots, r_i$ from the axis, subsequently intersect it at points $F_1, F_2, \ldots, F_i$, and finally impinge on a mobile fluorescent screen, $E$, which can be moved along the axis; the points of intersection lie at distances $h_1, h_2, \ldots, h_i$ from the axis. With an annular stop we can intercept all but one of the pencils, in object
space; knowing \( h_1, h_2, \ldots, h_i \) for a range of positions of \( E \), therefore, it is a simple matter to determine the positions of the foci \( F_1, F_2, \ldots, F_i \) (and to extrapolate to \( F_0 \), the Gaussian focus), and of the principal planes.

![Fig. 55. The Hartmann test.](image)

If \( \alpha_i \) represents the inclination of the \( i \)-th beam, in image space, and \( \Delta F_i \) the distance \( F_i F_0 \), \( C_s \) is given by

\[
C_s = \frac{\Delta F_i}{\alpha_i^2}.
\]  
(7.44)

The diffraction diagram method

The preceding method is only feasible if the lens being studied has a large aperture, as its accuracy depends upon the holes in the diaphragm \( D \) being small relative to the off-axial distance.

To investigate lenses with a small opening, Mahl and Recknagel (1944) used beams of known inclination, which were produced by diffraction in a thin film of aluminium placed on the specimen support of an electron microscope. In the focal plane of the objective which is being studied, an image

![Fig. 56. The diffraction diagram method.](image)

of the Debye–Scherrer diagram is formed, the outer rings of which are contracted by spherical aberration. This diagram is magnified by a second lens, with a large opening, whose own spherical aberration is negligible since
the beams which emerge from the objective are so feebly inclined. This final diagram is compared with the diagram which is obtained if the objective is suppressed (the specimen lying beneath the second lens only) and the image magnified optically until the first rings of the two diagrams coincide (Fig. 56). The discrepancy $\delta$ between subsequent rings is a function of the various Debye angles at which the beam emerges from the object. In the paper quoted above, $C_s = 80 \text{ mm}$ for one of the three-diaphragm objectives considered, and $65 \text{ mm}$ for the other; the focal lengths were respectively $7 \text{ mm}$ and $61 \text{ mm}$.

**Method using the edge of a screen**

Before arriving at the lens, a parallel beam is partially intercepted by the straight edge of a screen, placed at some minimum distance $\delta_1$ from the axis. Were the lens perfect, the shadow on the screen $E$ would have a straight

![Diagram](image)

**Fig. 57.** The method using the edge of a screen. (a) The overall appearance. (b) The shadow which is seen on the fluorescent screen.

edge. The various points on the edge (Figs. 57a and b) lie at different distances from the axis, however, and the edge of the shadow is convex. If the line $C_0$ represents the ideal Gaussian shadow, and the curve $C_1$ represents the real shadow, then (Liebmann, 1949)

$$\Delta R_l = L \left( \frac{r_L}{f_0} \right)^3 \frac{C_s}{f_0}. \quad (7.45)$$
Methods using the shadows cast by a wire or a grid

A variant upon the preceding method consists of examining the shadow on a screen of a very fine wire (about a micron in diameter) which is inserted into the path of the beam. Dosse (1941b) has studied strongly convergent magnetic objectives in this way, placing the wire before the objec-

![Diagram of a wire shadow](image)

**Fig. 58.** The shadow of a wire situated at the focus.

tive. Castaing (1950 and 1951) has measured the spherical aberration constant of an objective with long focal length by placing the wire in the Gaussian focal plane. The appearance of the shadow is shown in Fig. 58; if \( L \) represents the distance between the wire and the screen, \( r_1 \) and \( r_2 \) the breadth of the image corresponding to the angular apertures \( r_1/L \) and \( R/L \), and \( r \) the radius of the wire, then

\[
C_s = L^3 r \frac{1/r_1 - 1/r_2}{R^2 - r^2}.
\]

(7.46)

Castaing, using a lens with three plane electrodes, found that for \( r_1 = 4.5 \) mm, \( r_2 = 7.35 \) mm, \( R = 9.75 \) mm, \( L = 560 \) mm and \( r = 10^{-3} \) mm, the value of \( C_s \) was 200 mm with \( f_0 = 14 \) mm. Should the objective not be free of ellipticity astigmatism, however, the appearance of the image is very different (see § 7.6.2).

Spangenberg and Field (1942) and Heise and Rang (1949) use two grids: the first, \( G_1 \), is placed in front of the lens \( L \), and defines either the height of the incident pencils if they are parallel, or their inclination if they stem from a point source \( S \); the second, \( G_2 \), has much coarser meshes and is placed just behind the lens. The shadows of the two grids are observed on a fluorescent screen, and the origin for the distances \( h_i \) and \( H_i \) (see Fig. 59) is fixed by marking the bars of \( G_1 \) and \( G_2 \) which pass through the axis in some distinctive way. With \( L \) earthed, \( S \) and the various values of \( \alpha_0 \) are determined; with \( L \) at the desired potential, the values of \( h_i \) and \( \alpha_i \) are measured, and the position of the image \( S_i \) of \( S \) is calculated. If \( S \) is sufficiently distant, \( s_i \Delta S_i \) is effectively the longitudinal aberration. By placing \( S \) in a different
position, we have sufficient data to calculate the focal length and the positions of the principal planes; the curvature of the principal surfaces can also be calculated—they are planar only in the immediate vicinity of the axis.

If the lens is illuminated with a parallel beam, the grid $G_2$ is sufficient to determine $C_2$, and the method is equally applicable to an immersion objective (Septier, 1954b) as Fig. 60 shows in a diagrammatic form. The pencils which are emitted at the surface of the cathode $K$, with median ray parallel to the axis, are focused onto a screen $E$ which is placed at a
distance great in comparison with the focal length $f_0$ of the objective. Despite the shadow of the grid, the image of $K$ can be distinguished on $E$, the magnification $G$ can be measured, and hence $f_0$ is known. Given $L$ and the periodicity of the grid, we can deduce the longitudinal aberration $F_i$.

For openings for which $R = 1\, \text{mm}$, $W$ is $0.7\, \text{mm}$ in thickness, $WA = 2\, \text{mm}$ and $KW = 0.4\, \text{mm}$, $C_s$ is found to be $250\, \text{mm}$. Klemperer and Klinger (1951), using identical objectives but operating them as guns, found that $100\, \text{mm} \leq C_s \leq 600\, \text{mm}$.

The halo test

For certain operating conditions an emission system with an exit diaphragm produces a small spot surrounded by a brilliant ring. This phenomenon can be described only in terms of the spherical aberration of the objective (Fig. 61); only rays which have originated either in the region of the cathode surface between 0 and $r_1$ or between $r_3$ and $r_4$ can pass through the diaphragm $D$—the former produce the central spot, while the latter create the bright ring. The mean radius $R$ and thickness $d$ of the bright ring are the only quantities which need be measured, for a known value of $L$, in order to calculate $C_s$. This method can be extended quite straightforwardly to cover any lens illuminated by parallel radiation.
7.6.2 Measurement of the Other Aberration Coefficients

The coefficient of ellipticity astigmatism

If the lens is weak, the constant \( C_e \) can be obtained rapidly with the shadow method which has already been described (Castaing, 1950). \( C_e \) has a simple geometrical significance, for it represents one-half of the distance between the two line foci (Fig. 62a) \( f_1 \) and \( f_2 \). The form of the shadow of the wire \( D \) varies according as the wire lies beneath, above, or between the line foci; more precisely, the central part of the shadow near the axis rotates through 90° as the wire moves from \( f_1 \) to \( f_2 \). We can obtain \( 2C_e \) directly, therefore, and it is found to be of the order of a few hundredths of a millimetre with \( f_0 = 14 \) mm (and \( C_s = 200 \) mm).

Alternatively, \( C_e \) can be deduced from \( C_s \) for one special position of the wire. If the latter is inclined at 45° to the directions of the focal lines (Fig. 62b), and intersects the axis at the Gaussian focus \( f_0 \) (midway between \( f_1 \) and \( f_2 \)), and if we write \( u_0 \) for the angular aperture which corresponds to the point on the image at which the latter intersects the axis \( X' \), we find

\[
C_e = C_s u_0^2.
\]

Bertein, Grivet and Regenstreif (1949) have suggested yet another method, which depends upon the use of rays far from the axis, which are styled “transgaussian”.

The search for the minimum distortion coefficient

The constant \( C_d \), or the constant \( D_d \) which is derived from it by

\[
D_d = C_d r^2 = \frac{\Delta G}{G},
\]

is measured by forming the images of grids of known dimensions with the lens in question. The quantity

\[
\Delta r' = C_d \cdot G \cdot r^2
\]

is positive or negative according as the distortion is “cushion” or “barrel” — \( G \) is the linear magnification.

The validity of the relation:

\[
\frac{\Delta G}{G} = C_d r^2,
\]

has been verified experimentally (Jacob and Mulvey, 1949) for an electrostatic unipotential lens with three equally large openings, and at the same time, the influence upon \( C_d \) of variations in the thickness \( T \) of the central electrode and in the distance \( S \) between the electrodes was studied; the distortion is zero for various sets of values of the three quantities (\( S = 2D = 2T \) is a particular case).
Heise and Rang (1949) have studied the variation of $C_d$ with the polarization of the central electrode, and find that the distortion is cancelled when the focal length of the lens is a minimum; this is a most important result, for this is the situation in which both the spherical and the chromatic aber-

![Diagram](image)

**Fig. 62.** The shadow of a wire lying perpendicular to the axis of a lens with ellipticity astigmatism. (a) The wave surface and general appearance. (b) The form of the shadow as the wire is moved along the axis.

ration are minimized also. In practice, perfectly corrected lenses can even be achieved without polarization of the central electrode, by altering the thickness or the diameter of this latter in such a way that the lens is operating at its minimum focal length (Regenstreif, 1951a). It is equally possible to reduce the distortion of an immersion objective which is being used in an image converter, by employing a cathode which is concave towards the image side (Morton and Ramberg, 1936).

As for magnetic lenses, Liebmann (1952 and 1953) has shown that the distortion starts off as cushion distortion for feeble excitations, is annulled
at a well-defined value of $B^2/\Phi$, and then becomes barrel-shaped—the critical value is dependent upon the pole-piece geometry, but it is always possible to construct a projective lens free of distortion.

*The study of field curvature and of astigmatism*

Like distortion, field curvature is of interest only in projective lenses and immersion objectives. The field curvature of the objective of an image converter has been studied by Morton and Ramberg (1936) and by Duchesne (1953)—a concave cathode produces partial correction.

The field curvature of the immersion objectives with planar electrodes, operating at a high magnification, which are used in microscopy, has been explored by Johannson (1933) and Septier (1952, 1954b). Briefly, we can conclude that the curvature of the sagittal surface is less pronounced than that of the tangential surface, although both surfaces are always concave on the cathode side; that for any given geometry, the actual curvatures vary markedly with the polarization $\Phi_1/\Phi_2$ which is applied to the focusing electrode of the objective, and pass through a minimum near $\Phi_1/\Phi_2 = 0$; and that at this optimum adjustment, the aberration is virtually independent of the geometrical parameters of the lens.

The field curvature of unipotential three-tube lenses has been measured by Gobrecht (1942), using a point source which can be moved in a plane perpendicular to the axis—the focal surfaces have the same general properties as in the preceding case, and their curvature decreases with the focal length. Using the same method, Becker and Wallraf (1940) have obtained substantially equivalent results for magnetic lenses.

In conclusion, we must point out the fact that the magnitude of distortion, field curvature, and astigmatism depends upon the position of the aperture, as we shall show later in this chapter (§ 7.7.1); they can, therefore, only be specified for a rather fully defined ensemble of lens and diaphragm.

**7.6.3 The Magnitude of Spherical Aberration**

The values of $C_s$ and $C_e$ in some of the common kinds of electrostatic lens have already been mentioned—now, therefore, we shall indicate the laws which govern the behaviour of $C_s$ in terms of the various parameters which characterize electrostatic and magnetic lenses, and quote the smallest values which have so far been attained.

*Electrostatic lenses*

*Two-cylinder lenses* (cf. § 8.1). Lenses with two equal cylinders have been studied by Klemperer and Wright (1939), who show that the spherical aberration, which varies with the ratio of the potentials applied to the two electrodes, $\Phi_1/\Phi_2$, is considerably smaller in retarding lenses than in accele-
rating lenses. Measured in terms of the tube radius as the unit of length, the coefficient \( C_s \) ranges between 300 and 2 when the electrons are being retarded, and \( \Phi_1/\Phi_2 \) is varied between 3 and 15, as opposed to 500 and 25 for the same \( \Phi_1/\Phi_2 \) range when they are being accelerated.

Lenses with two unequal cylinders invariably have a larger spherical aberration than those with two equal cylinders.

*Unipotential lenses* (cf. § 8.2). A number of strongly convergent symmetrical three-electrode lenses, of somewhat original forms, have been studied by Liebmann (1949); for the majority, \( C_s \) obeys the approximate formula:

\[
C_s = K \left( \frac{f_0}{R_1} \right)^2 f_0,
\]

in which \( R_1 \) is the radius of the central opening. \( C_s \) decreases as the radius of the outer openings decreases, when the divergent region of the field extends less and less outside the lens. Again using the radius of the apertures as our length unit, the constant \( C_s \) is of the order of 30 \( R \) for a symmetrical lens with three equal openings, although it falls to 4 \( R \) if the outer electrodes are closed off with grids. Modifications of this kind unfortunately make it virtually impossible to use the lens in practice.

The experimental work of Heise and Rang (1949) has confirmed the existence of a minimum value of \( C_s \) which corresponds to a minimum value of focal length for a given ratio between the potentials; this ratio varies with the geometry of the lens, but in all types of three-diaphragm lens in common use, \( C_s \) is of the order of ten times the focal length.

*Immersion objectives* (cf. § 8.3). Only a little work has been done on such microscope objectives. Septier’s measurements (1954b) show that in many cases,

\[ C_s \approx 50 - 70 f_0 \]

for an objective of standard type with plane electrodes. For this species of lens, however, spherical aberration is usually a negligible factor in calculating the resolving power.

Finally, we would point out that spherical aberration exists equally prominently in the emission systems from which guns are built, but is quite swamped by the aberration which results from space charge effects.

*Mirrors.* In 1948, Regenstreif (unpublished) calculated the spherical aberration coefficients of a number of unipotential lenses used as mirrors; they proved to be some four times greater than the values in the corresponding lenses.

*Magnetic lenses*

\( C_s \) is affected by a host of different parameters—the length of the field, of which the half-width \( a \) of the \( B(z) \) curve (using the bell-shaped distribution) is a convenient measure; the extent to which it is symmetrical; and
the value of the maximum field $B_0$, or better, of the coefficient

$$k^2 = \frac{e}{8m} \frac{B_0^2}{\Phi} a^2.$$

Generally speaking, $C_s$ decreases rapidly as $k^2$ increases (Fig. 63); $C_s$ cannot be decreased indefinitely in this way, however, as $B_0$ possesses an upper limit of about 2.6 weber m$^{-2}$ (26,000 gauss) beyond which the pole-pieces become saturated.

![Fig. 63. The values of $C_s/a$ and $C_\varphi/a$ as functions of $k^2$ in a symmetrical magnetic lens (Dosse, 1941a). The broken curves represent the theoretical values given by a Glaser bell-shaped field distribution (§ 9.2.1).](image)

For given values of $B_0^2/\Phi$, however, $C_s$ passes through a minimum as $a$ is increased (Fig. 64).

Finally, the spherical aberration is diminished if the field distribution is unsymmetrical in the sense that it increases more slowly on the object side (Fig. 65).

Examples of these kinds of variation has been published by Dosse (1941a). For more exact data about these phenomena, the reader is referred to Ruska (1934), von Ardenne (1939) and Liebmann (1950a, b; 1951).

The smallest value of $C_s$ which can be obtained is of the order of

$$C_s = 0.8 f_0$$

for $f_0 \approx 1$ or 2 mm. Objectives are considered very good if

$$C_s = 2 \text{ or } 3 f_0.$$
As far as the smallness of $C_s$ is concerned, it is clear that magnetic lenses are superior to electrostatic ones, the more so in that the risk of breakdown between the electrodes prevents the focal lengths of electrostatic lenses from being as short as those of magnetic lenses. At voltages of some tens of kilovolts, the interelectrode spacing cannot be reduced below about 3 mm, and $f_0$ is then of the order of 3 mm.

**Fig. 64.** $C_s$ as a function of the half-width $a$ of the curve which represents the axial field (cf. Fig. 12; Dosse, 1941a).

**Fig. 65.** $C_s/a$ as a function of $k^2$ for various values of $q = a_1/a_2$; $a = a_1 + a_2$ (Dosse, 1941b).
In Fig. 66, the relative values of the spherical aberration coefficient in a magnetic and a three-electrode electrostatic lens, each of width $2a$, are shown, as calculated by Dosse (1941b). Magnetic lenses also prove to be more satisfactory than electrostatic lenses as regards chromatic aberration $C_e$.

![Graph showing aberration coefficients $C_e$ and $C_m$ for magnetic and electrostatic lenses.]

Fig. 66. A comparison between the aberration coefficients $C_e$ and $C_m$ of electric and magnetic lenses of comparable dimensions (Dosse, 1941b).

## 7.7 Correction of the Aberrations

### 7.7.1 The Influence of the Diaphragm Position on the Field Aberrations

The aberrations of electron optical lenses are far less amenable to correction than their counterparts in glass optics, as the only parameter which is easily varied is the position of the diaphragm. The position of this diaphragm has an effect upon the aberrations which affect only an off-axial
LENS DEFECTS

object point, the coma, distortion, field curvature and astigmatism, but from considerations of symmetry, it affects neither spherical aberration nor the aberrations peculiar to magnetic lenses. The reason why the diaphragm position does affect certain of the aberrations is that the pencil of rays by which the image is produced varies according to where the aperture is placed, and the paths which the rays will have followed through the lens therefore varies too (Fig. 67).

![Fig. 67. The variation of the pencil of rays which produces the image, with the position of the diaphragm.](image)

A detailed quantitative study of magnetic lenses, using the Glaser model, is available (Glaser and Lammel, 1941, 1943) and a parallel study of electrostatic lenses might well be made—this has not as yet been done, however, as the calculation is rather more complicated. In particular, the suppression of distortion has been explored; experimentally, Hillier (1946) has verified the results for a magnetic microscope projective, and Le Rutte (1948) has done the same for an electrostatic projective and for electron mirrors.

### 7.7.2 Spherical Aberration; the Impossibility of Complete Correction

Spherical aberration cannot be perfectly eliminated from electron lenses as we can see if we search for a distribution of refractive index such that all the rays which leave a point object are concurrent in a point image. Gabor (1942) established the requisite distribution, which is identical to that known as “Maxwell’s fish-eye”, but it seems to be impossible to create it, even roughly, with charges on metal electrodes (Glaser, 1948).

Much earlier, however, than the references just quoted and more straightforwardly, Scherzer (1936) had shown that the integrand of the expression for the spherical aberration coefficient can be written as the sum of a number of squared terms, and hence can never be zero. For magnetic lenses, the result is

\[
\frac{|w|}{G} = \frac{e}{16m\Phi_0} \int_{z_0}^{z_f} \left[ \left( B' + B \frac{X'}{X} \right)^2 + B^2 \left( \frac{X'}{X} \right)^2 + \frac{e}{4m} \frac{B^4}{\Phi_0} \right] dz, \tag{7.47}
\]

which is zero only if each of the terms of the integrand is zero, which only happens if there is no lens. In glass optics, the situation is quite dif-
ferent, for the sign of the spherical aberration is different according as the lens is convergent or divergent; by combining two lenses, therefore, the aberration can be cancelled without the overall converging action being lost. In electron optics, however, the spherical aberration is positive whether the lenses be convergent or divergent.

There was a period during which Glaser (1940a) believed Scherzer’s conclusion could be invalidated: by applying certain algebraic transformations, he was able to put equation (7.47) into the form

$$\frac{|w|}{G} = \frac{e}{96m\Phi_0} a^3 \int_{z_a}^{z_1} \left[ \frac{2e}{m\Phi_0} B^4 + 5B'^2 - B B'' \right] X^4 \, dz \tag{7.48}$$

and hence to deduce that a lens whose magnetic field $B(z)$ is a solution of

$$\frac{2e}{m\Phi_0} B^4 + 5B'^2 - B B'' = 0, \tag{7.49}$$

will be strictly free of spherical aberration. Unfortunately, equations (7.47) and (7.48) are equivalent only if the lens which the field describes is very convergent—a lens described by the solution of (7.49), however, is invariably negligibly convergent (Recknagel, 1941; Rebsch, 1940), and the two descriptions are contradictory. Nevertheless, formula (7.48) is of interest, as it can be used to calculate the spherical aberration of the ordinary types of lens which are always powerfully convergent.

### 7.7.3 Partial Correction of Spherical Aberration

It is possible to select the field in such a way as to reduce the spherical aberration; the physical reason for this aberration is that marginal rays are converged more strongly than paraxial rays. This in turn is due to the fact that the longitudinal component of the velocity is less for marginal rays—about this we can do nothing—and to the fact that the longitudinal component of the magnetic field,

$$B_z(r, z) = B(z) - \frac{r^2}{4} B''(z),$$

depends upon the sign of $B''$. The marginal rays pass through regions where $B''$ is negative, where the spherical aberration is accentuated, and regions where $B''$ is positive, in which it is diminished.

A means of producing a magnetic field of the second type (in which the intensity of $B(z)$ is greater in the neighbourhood of the object and the image than at the centre of the lens) has been devised by Siegbahn (1946); using this field, he has built a $\beta$-ray spectrograph of markedly improved brilliance. Marton and Bol (1947) attempted to extend this method to lenses
with the kind of magnification that would be useful in electron microscopy, since Siegbahn's apparatus has unit magnification only; no worthwhile advantage was gained, however, as the clarity of the image depends upon the fourth root of the spherical aberration—a reduction of the latter, therefore, barely affects the resolving power.

If we take into account the fact that Scherzer's formula cannot be applied to the case of mirrors (Recknagel, 1940), we have the possibility of cancelling the positive spherical aberration of a lens with the negative spherical aberration of a mirror, without losing the overall convergence. In a microscope, however, it is difficult to get the mirror sufficiently close to the lens—no feasible arrangement has yet been suggested; if the mirror is not close to the lens, the rays are insufficiently steeply inclined to the axis for the correction to be at all adequate. Ramberg has clarified this situation (1949).

As we shall see later, the attempt to correct mechanical defects has led to the introduction of systems formed from a stigmatic combination of cylindrical lenses.

7.7.4 Correction of Chromatic Aberration

Chromatic aberration likewise cannot be completely corrected, for as Scherzer (1936) has shown, $C_r$ too can be written as the integral of a sum of squared terms, and cannot, therefore, be made zero. In electron optics it is not possible to compensate for the dispersion of the velocities with which the electrons leave the cathode by a judicious choice of lens, unlike the analogous situation in glass optics. As before, we might attempt to correct the aberration by combining lens and mirror action, as mirrors do have a negative chromatic aberration, but the correction is almost always insignificant (Ramberg, 1949). It is interesting, however, to note that the introduction of a composite lens-mirror unit has been envisaged in emission systems (the image-converter).

A closely related defect is the sensitivity of the cardinal elements (the focal lengths, and the position of the principal planes) to accidental variations of the potential $\delta \Phi$ or of the current $\delta i$.

Electrostatic lenses

The harmful effect is a consequence of relativity, when all the potentials are drawn from a single source by means of potentiometers. It can, however, be suppressed, as Le Rutte (1948) discovered, by introducing a comparatively small, fixed polarization potential—provided by a thousand-volt battery, say—into the lead of the central electrode of a three-electrode lens.
**Magnetic lenses**

If the lens is operated under the conditions which correspond to the minimum of the curve which describes $f_i$ as a function of $k^2$ (the parameter which summarizes the form of $B$ and $\Phi$), variations of $B$ and $\Phi$ have little effect, and only a residual effect, upon the distance between the two principal planes, remains.

Alternatively, the construction parameters—the diameter $D$ of the opening between the pole-pieces, the interval $s$ between the poles—can be chosen in such a way as to minimize the chromatic aberration $C_c$. This is particularly desirable in the gun, where it is no easy problem to regulate the potential with precision. Experimentally, Le Poole and van Dorsten (1951) have studied this problem, and, theoretically, Liebmann (1952) has also. Liebmann’s results are summarized in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$D$ mm</th>
<th>$s$ mm</th>
<th>$f_{\text{min}}$ mm</th>
<th>$C_c$ mm</th>
<th>$C_s$ mm</th>
<th>$z_0$ mm</th>
<th>$NI$ ampère turns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment (L.P. and v.D.)</td>
<td>0.5</td>
<td>2</td>
<td>0.85</td>
<td>0.66</td>
<td>1.2</td>
<td>0.43</td>
<td>4,000</td>
</tr>
<tr>
<td>Theor. values giving $f_{\text{min}}$</td>
<td>0.48</td>
<td>1.9</td>
<td>0.83</td>
<td>0.66</td>
<td>1.15</td>
<td>0.32</td>
<td>3,680</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
<td>2.2</td>
<td>0.84</td>
<td>0.66</td>
<td></td>
<td>0.32</td>
<td>4,210</td>
</tr>
</tbody>
</table>

The lenses which were studied were the objectives of a microscope operating at 80kV; between the parallel faces of the pole-pieces, the field may reach some 24,000 gauss, although at the centre it is weaker. From the point of view of constructing microscope objectives, it is very fortunate that the conditions under which $C_s$ and $C_c$ are minimized are so very similar to those under which $f$ is least.

### 7.7.5 Correction of the Mechanical Aberrations

The choice of an elliptic model for these aberrations which we made in §7.5.2 is arbitrary, and at first sight seems highly schematic. In a general theoretical study of this class of aberrations, however, Bertein (1947a, b, c; 1948a, b) has shown to what extent it is valid; the causes of the mechanical aberrations are elucidated in detail, and a great deal of exact information about the whole question is provided. However, the really interesting feature of his theory is that it suggests possible ways by which the aberrations can be corrected. Independently, Hillier and Ramberg (1947) and Scherzer (1947) suggested similar methods. To produce the correction, a weakly convergent cylindrical lens is used which is mechanically extremely asymmetrical (Rang, 1949). This weak lens is placed either just in front of or just behind the lens which is being corrected in such a way that it produces an aber-
roration of the same nature as that of the lens, but of opposite sign. The problem is to find a weak lens which cancels the aberration to be corrected without creating other aberrations—to correct astigmatism without introducing distortion, for example, requires the assembly to be exceedingly accurate. The very simple system of Fig. 68, which is called a stigmatic, possesses this interesting property.

Fig. 68. The stigmatic (Bertein, 1947c).

7.7.6 Cylindrical Lenses and the Correction of Spherical Aberration

However large it may be, the astigmatism produced by ellipticity can always be corrected—we are familiar with the analogous situation in glass optics where two crossed cylindrical lenses have the same effect as a spherical lens, although in practice, three cylindrical lenses have to be used if the image is not to be intolerably distorted.

Scherzer (1947, 1949, 1950) has shown that by placing such a combination of lenses behind an ordinary round lens it should be possible to correct the spherical aberration of the latter. We shall briefly outline the principle upon which this correction is based.

We consider the expression for the potential \( \Phi(x, y, z) \) in an optical system with four radial symmetry planes which intersect along the \( OZ \) axis and are mutually inclined at an angle \( \theta = 45^\circ \). \( \Phi \) can be written in the form:

\[
\Phi(x, y, z) = \Phi_0(x, y, z) + \Phi_2(x, y, z) + \Phi_4(x, y, z) + \ldots,
\]

in which

\[
\Phi_0(x, y, z) = \varphi(0, 0, z) - \frac{1}{4} \varphi''(0, 0, z)(x^2 + y^2)
\]

\[
+ \frac{1}{64} \varphi^{(iv)}(0, 0, z)(x^2 + y^2)^2 - \ldots,
\]

\[
\Phi_2(x, y, z) = \varphi_2(0, 0, z)(x^2 - y^2) - \frac{1}{12} \varphi''_2(0, 0, z)(x^2 - y^2)(x^2 + y^2) + \ldots
\]

\[
\Phi_4(x, y, z) = \varphi_4(0, 0, z)(x^4 - 6x^2y^2 + y^4) - \ldots.
\]
We shall not go beyond terms of the fourth degree in \( x \) and \( y \). \( \Phi_6(x, y, z) \) we recognize as the familiar term associated with axial symmetry; \( \Phi_4(x, y, z) \) represents a term with quadrupole symmetry (see Chapter 10) and \( \Phi_4(x, y, z) \) a term with octupole symmetry. We could obtain this latter by means of eight identical electrodes lying parallel to \( Oz \), distributed regularly round a circle centred on \( Oz \), and supplied with potentials \(+ \Phi_4\) and \(- \Phi_4\) alternately.

\( \varphi'' \) and \( \varphi^{(iv)} \) denote various derivatives of \( \varphi = \Phi (0, 0, z) \) with respect to \( z \). In an axially symmetric electrostatic lens, the radial force is proportional to \( r \) only over an extremely slender axial zone; in reality, this force, which is proportional to the radial field \( E_r \):

\[
E_r = -\frac{1}{2} r \frac{\partial E_z}{\partial r} = \frac{r}{2} \left\{ \varphi(z) - \frac{r^2}{4} \varphi''(z) + \frac{r^4}{64} \varphi^{(iv)}(z) - \cdots \right\},
\]

consists of terms in \( r^{2n+1} \). In the third order approximation, the essential term varies as \( r^3 \), the mean effect of which is a reinforcement of the convergence of the lens for the outer rays. We might hope to be able to introduce a defocusing force proportional to \( r^3 \), with the aid of the term \( \Phi_4(x, y, z) \), but (for the Laplace equation to be satisfied), this potential obligatorily varies periodically with \( \theta \); the force is certainly defocusing in the \( Ox \) and \( Oy \) directions, say, but has then necessarily a focusing effect in the directions at 45° to these. If, therefore, we annul the aberration in the \( Ox \) and \( Oy \) directions, we automatically double it along the bisectrices of the angle \( \angle xOy \).

If, as Scherzer (1947) suggested, we employ a combination of several different systems in which both \( \Phi_2 \) and \( \Phi_4 \) are present, it is effectively possible to correct the third order aberration in every radial direction. The field distributions \( \varphi_2(z) \) and \( \varphi_4(z) \) are not superimposed onto \( \varphi_0(z) \), but succeed it along the optic axis. The potentials \( \varphi_2(z) \) and \( \varphi_4(z) \) determine the paths of the rays through the system to the first order. The beam is axially symmetric as it emerges from the round lens which is to be corrected; it then passes through a quadrupole lens \([\varphi_2(z)]\) which produces a strong astigmatism and results in the formation of two real focal lines, and after this, a second quadrupole lens and a round lens which render the beam again axially symmetric. Generally, we try to obtain the same convergence as that which would have resulted had the lens being corrected been used alone; this determines the potentials to be applied to the different elements.

Various octupole systems create the field distribution \( \varphi_4(z) \) which is necessary for correction.

The equations of motion for the whole system are of the form:

\[
\varphi \, x'' + \frac{1}{2} \varphi' \, x' - \left( \varphi_2 - \frac{1}{4} \varphi'' \right) x = 2 \varphi_4(x^2 - 3 \, x \, y^2) = S_x,
\]

\[
\varphi \, y'' + \frac{1}{2} \varphi' \, y' + \left( \varphi_2 + \frac{1}{4} \varphi'' \right) y = 2 \varphi_4(y^2 - 3 \, x^2 \, y) = S_y.
\]
We calculate two particular trajectories: $x_a$, which lies in the plane $zOx$ in such a way that $x_a = 0$ and $x'_a = 1$ in the object plane, and $y_a$, which lies in $zOy$ and satisfies $y_a = 0$ and $y'_a = 1$ in this same object plane. A real trajectory which originates in an axial object point $A$ at an angle $r'_0$ (the projection of $r'_0$ onto the plane $zOx$ is $\alpha$, and onto $zOy$, $\beta$) can always be represented by

$$x = \alpha x_a, \quad y = \beta y_a.$$  

If we consider all the rays which emerge from $A$ with the same slope $r'_0$ we should obtain a circular aberration figure defined by the equations:

$$\begin{cases} x_b = a \alpha (\alpha^2 + \beta^2) \\ y_b = a \beta (\alpha^2 + \beta^2) \end{cases}$$

and of radius

$$q = a(\alpha^2 + \beta^2)^{3/2};$$

if the system were axially symmetrical the aberration would be completely described by a single coefficient, $a = C_s$.

If axial symmetry is abandoned, however, but nevertheless only quadrupole and octupole field distributions are considered [$q_3(z)$ and $q_4(z)$], we find

$$\begin{cases} x_b = a \alpha^3 + b \alpha \beta^2 \quad \text{(in a stigmatic system)} \\ y_b = c \beta^3 + b \alpha^2 \beta \end{cases}$$

so that here the aberration is characterized by the three coefficients, $a$, $b$ and $c$; a judicious choice of $q_4(z)$ permits all of these to be cancelled.

To do this, we calculate first the trajectories $x_a$ and $y_a$ to the first order, and then the spherical aberration $x_z$ and $y_z$ without taking account of $q_4(z)$. The equations are then completely integrated by considering $S_x$ and $S_y$ as small perturbations—$S_x$ and $S_y$ can be calculated everywhere on the $Oz$ axis by writing $x = \alpha x_a$ and $y = \beta y_a$.

The contributions $(x_b)_c$ and $(y_b)_c$ of the correction term $q_4(z)$ are then obtained in the form of integrals taken from $z_a$ (the object) to $z_b$ (the image):

$$\begin{align*}
(x_b)_c &= \frac{2G}{\sqrt{q_a}} \int_{z_a}^{z_b} \frac{q_4}{\sqrt{q}} (3\alpha \beta^2 x_a^3 y_a^2 - \alpha^3 x_a^3) \, dz, \\
(y_b)_c &= \frac{2G}{\sqrt{q_a}} \int_{z_a}^{z_b} \frac{q_4}{\sqrt{q}} (3\alpha^2 \beta x_a^2 y_a^2 - \beta^3 y_a^2) \, dz.
\end{align*}$$

We next break up $q_4(z)$ into three distinct regions by using three octupole correctors—the first two lie in the planes $P_1$ and $P_2$, with abscissae $z_1$ and $z_2$; in these planes the trajectories pass respectively through the following pairs of points:

$$x_a = 0; \quad y_a \neq 0,$$

$$x_a \neq 0; \quad y_a = 0.$$
We can then choose the amplitudes of the potentials to be applied to the correctors in such a way as to cancel the aberration terms in $\beta^3$ and $\alpha^3$, that is, the coefficients $a$ and $c$.

By placing a third corrector at the exit of the system at a point in which $x_n = y_n$, we can cancel the aberration terms in $\alpha \beta^2$ and $\alpha^2 \beta$ which have been aggravated by the two earlier correctors.

![Fig. 69. The Scherzer correction system.](image)

$R_1$: lens to be corrected. $S$: stigmator. $C_1$, $C_2$: cylindrical lenses (either slit lenses or quadrupoles). $R_2$: intermediate round lens. $O_x$, $O_y$, $O_{45^\circ}$: octupole correcting elements, which act in the directions $Ox$, $Oy$, and along the bisector of $xOy$ respectively. Note the two real line foci which are formed at the centres of $O_x$ and $O_y$.

The complete corrector system, as employed by Seeliger (1949) and Möllenstedt (1956), is shown in Fig. 69. In practice, it is extremely difficult to manipulate as the problems of alignment and adjustment of the potentials of the various elements are most complex.

Seeliger (1951) showed that the third order aberration could indeed be annulled with such an assembly, and hence, for example, larger apertures could be used in microscopy. It is then the fifth order aberration which provides the limit (terms in $\alpha^3$). Recently, Meyer (1961) has calculated the fifth order aberration of this corrector, set out practical alignment procedures, and discussed the residual aberrations which may result from poor alignment.

Many attempts have been made to simplify the assembly to the uttermost (Burfoot, 1953a, 1953b; Archard, 1954a, 1954b, 1955 and 1958; Archard, Mulvey and Petrie, 1960) by collecting the $\varphi_3(z)$ and $\varphi_4(z)$ terms into a single lens, for example. Recently, Deltrap and Cossett (1962) and Deltrap (1964a, b) have studied, first theoretically and then experimentally, a system consisting of a round lens together with four complex magnetic lenses; the aberration coefficients were measured and Deltrap showed that the three coefficients $a$, $b$ and $c$ can be reduced to zero in practice. We might even hope to be able to suppress $\varphi_0(z)$ altogether—to remove the lens under correction that is—and to obtain instead an optical system equivalent to a thick round lens and wholly free of third order spherical aberration.
CHAPTER 8

SOME ELECTROSTATIC LENSES
AND TRIODE GUNS

8.1 TWO COAXIAL CYLINDERS

8.1.1 Principal Studies

Used as they are in cathode ray oscillographs, lenses of this kind received a great deal of attention in the earliest work on electron optics. Epstein (1936) calculated the rays and cardinal elements in such a lens from the potential distribution which he had obtained in an electrolytic tank, and compared these with his experimental results; Spangenberg and Field (1943) completed his experimental work, using the method which has been explained in the chapter on aberrations. We must mention too the work of Gundert (1941) on cylinders with different radii, and of Nicoll (1938), who examined the transition from lens to mirror using the same electrodes.

Not until much later, however, was the theory of this type of lens developed; Goddard (1944) calculated the cardinal elements, by integrating the Gaussian equation with a method of successive approximations. Bernard and Grivet (1951, 1952a) adapted the simple model which Hutter (1943) had already suggested for this form of lens; Glaser and Robl (1951) showed that the axial potential could be described by

\[ \varphi = \varphi_0 + \Delta \varphi \cot^{-1} k z, \]

in which case the trajectories are described by \( J_{1/4}(z) \), although tables of this Bessel function are not easily to be found.

We shall simply outline the method which Grivet and Bernard use; for the experimental results, the reader is referred to the bibliography.

8.1.2 The Grivet–Bernard Model

The two cylinders are assumed to be of the same radius, which we shall use as our unit of length; as a first approximation, we assume also that the gap between the two cylinders is small enough to be neglected (Fig. 70). The axial potential is given by

\[ \varphi = \frac{\Phi_1 + \Phi_2}{2} \left[ 1 + \frac{1 - \gamma}{1 + \gamma} \right] \left[ 1 - 2 \sum_n \frac{e^{-\mu_n |z|}}{\mu_n J_1(\mu_n)} \right], \quad (8.1) \]
in which \( \mu_n \) is the \( n \)th root of \( J_0(\mu) = 0 \). This complicated function is, however, represented very closely by

\[
\varphi = \frac{\Phi_1 + \Phi_2}{2} \left[ 1 + \frac{1 - \gamma}{1 + \gamma} \tanh \omega z \right],
\]

(8.2)
in which \( \Phi_1 \) and \( \Phi_2 \) are the potentials of the two cylinders of unit radius, \( \gamma = \Phi_1/\Phi_2 \) and \( \omega = 1.318 \); Gray (1939) pointed out this happy coincidence.

![Fig. 70](image)

Fig. 70. The form of the electrodes in a two-cylinder lens.

The function \( T(z) \), characteristic of the lens, which is defined by

\[
T(z) = \frac{\varphi'(z)}{\varphi(z)},
\]

(8.3)
is of the form (Fig. 71):

\[
T(z) = \frac{\varphi'(z)}{\varphi(z)} = \frac{2\omega(1 - \gamma)z}{(1 + t)(\gamma + t)}; \quad 2\omega z = \log_e t,
\]

(8.4)
and passes through a maximum, \( T_m \), at \( t_m = \sqrt[4]{\gamma} \), which is given by

\[
T_m = -2\omega \tanh(\omega z_m); \quad z_m = \frac{1}{4\omega} \log_e t.
\]

(8.5)

![Fig. 71](image)

Fig. 71. The form of the functions \( \varphi(z) \) and \( T(z) \).
If we substitute \( Z = z - z_m \), \( s = e^{2\omega z} \) and \( T = y T_m \), it is clear that the two points on the bell-shaped function \( T \) which have reduced ordinate \( y \) are *symmetrical* about the abscissa of the maximum since the equation for the "coordinate" \( s \) of these points is

\[
s^2 - \gamma s \left[ \frac{(1 + \sqrt{\gamma})^2}{2y\sqrt{\gamma}} - \frac{1 + \gamma}{2\sqrt{\gamma}} \right] + 1 = 0, \tag{8.6}
\]

which gives

\[
\frac{(1 + \sqrt{\gamma})^2}{2y\sqrt{\gamma}} - \frac{1 + \gamma}{2\sqrt{\gamma}} = \cosh(2\omega Z), \tag{8.7}
\]

for the abscissae \( Z \). To each value of \( y \), therefore, there correspond two values of \( Z \), since the hyperbolic cosine is an even function.

**Fig. 72.** The cardinal elements of lenses with two equal cylinders.
—-; the Grivet–Bernard model. \( \Delta \Delta \Delta \): Spangenberg’s experimental results.

In terms of these reduced quantities, therefore, the lack of electrical symmetry has vanished, its only effect being to shift the axis of the model relative to the gap between the cylinders by an amount \( z_m \). If \( F_0 \) and \( F_i \) are the focal lengths of the reduced lens, defined by:

\[
F_0 = \frac{R(+\infty)}{R'(-\infty)}, \quad F_i = \frac{R(-\infty)}{R'(+\infty)},
\]

then the focal lengths of the real lens will be

\[
f_0 = F_0 \left[ \frac{\Phi_2}{\Phi_1} \right]^{\frac{1}{4}}, \quad f_i = F_i \left[ \frac{\Phi_2}{\Phi_1} \right]^{\frac{1}{2}}. \tag{8.8}
\]
We can now replace the actual bell-shaped function \( T \) by the Glaser expression for such functions, \( T = \frac{T_m}{1 + (\frac{Z}{a})^2} \), having the same maximum \( T_m \), and a half-width \( a \) such that the \( T^2 \) curves, real and model, have the same area—using formula (4.15), we see that this ensures that the focal lengths are identical in the extreme case of weak convergence. For \( a \) we obtain the expression:

\[
a = -\left(\frac{1}{\pi \omega}\right)\left(\frac{1 + \sqrt{\gamma^2}}{1 - \sqrt{\gamma^2}}\right)\left[\frac{1 + \gamma}{1 - \sqrt{\gamma}} \log_e \gamma + 2\right].
\]

(8.9)

The values of focal length which have been calculated in this way are accurate to within two per cent for very convergent lenses in the region which is used in practice: \( 0.1 < \gamma < 10 \); an excellent agreement. The numerical values are obtained from the formulae:

\[
f_i = a K^{\frac{1}{4}} \sin(K \pi), \quad f_0 = a K^{\frac{1}{4}} \sin(K \pi),
\]

\[
z_{R_i} = a K \cot(K \pi) - \frac{T_m a^2}{4},
\]

\[
z_{R_e} = -a K \cot(K \pi) - \frac{T_m a^2}{4},
\]

\[
K = \sqrt{1 + \left(\frac{3}{16}\right) a^2 T_m^2},
\]

(8.10)

\[
a = 0.483 \coth^2\left(\frac{x}{2}\right) (x \coth x - 1),
\]

\[
T_m = 2.636 \tanh\left(\frac{x}{2}\right), \quad x = -\log_e V\gamma.
\]

\( K \) is always very close to unity, and the formulae approximate to:

\[
V\gamma = 1 + \varepsilon, \quad K = 1 + 0.0675 \varepsilon^2 \left(\frac{1}{V\gamma}\right),
\]

(8.11)

\[
a = 0.644 \left(1 + \frac{\varepsilon^2}{10}\right), \quad f_0 = 3.036 \frac{y^{3/4}}{\varepsilon^2}.
\]

The reduced ray can be written in terms of sinusoidal functions.

In Fig. 72, the closeness of the agreement between the values of the focal length calculated in this way and the values obtained experimentally by Spangenberg and Field (1943) is displayed—the other curve in the same figure represents the behaviour of the abscissa of the image focal plane.

The curve \( T(Z) \) and the cardinal elements are even more accurately represented if we write

\[
T = T_0 \text{sech}\left(\frac{Z}{b}\right)
\]

(8.12)
(Fig. 73); the reader is referred to the article by Bernard and Grivet (1952a) for further details. In Fig. 74, the results furnished by this improved model, the theoretical results of Goddard (1944), and the experimental results of Spangenberg and Field (1943) are all presented simultaneously for comparison.

So far we have been making the assumption that the gap between the cylinders is thin; if this gap is in fact of width 2\(d\), however, the axial potential is found to be well represented by

\[
\varphi = \frac{\Phi_1 + \Phi_2}{2} \left[ 1 + \frac{1 - \gamma}{1 + \gamma} \frac{1}{2\omega d} \log_e \frac{\cosh \omega(z + d)}{\cosh \omega(z - d)} \right],
\]

so that the formulae listed earlier are still valid provided \(\omega\) is replaced by \(\omega_1\), where

\[
\omega_1 = \frac{\tanh(\omega d)}{d}, \quad \omega = 1.318.
\]

These same formulae also describe quite accurately (to within a few percent) the lens which is formed by two diaphragms in which two equal circular holes have been cut.

8.2 LENSES FORMED FROM THREE DIAPHRAGMS

8.2.1 The Potential

The three-electrode, or "einzel" lens is frequently used as the second lens of a cathode ray oscillograph, and is also the standard type of lens for an electrostatic electron microscope. As the bibliography for this paragraph shows, it has often been studied; we shall describe the properties of the einzel lens in terms of the very complete, though analytically simple, theory which Regenstreif (1951a) has proposed. We have already stated that the potential around a single plate with a circular hole, lying between two distant parallel plates without holes, can be calculated; on the axis, it takes the form

\[
\Psi(z) = a + bz + c z \tan^{-1} \left( \frac{z}{R} \right).
\]

Regenstreif, and later Bertein (1952b), have shown that the potential on the axis of the three-electrode lens can be represented approximately by a linear superposition of three potential functions of the form (8.15). We try, therefore, to use a function of the form

\[
\varphi(z) = A \Psi(z - z_0) + B \Psi(z) + C \Psi(z + z_0).
\]

A function constructed in this way satisfies Laplace's equation exactly, but can only be approximately matched to the boundary conditions.
Fig. 73. Approximations to the characteristic function.

--- : the actual form of $T^2(Z)$. △△△ : the model $T^2(Z) = \frac{T_0^2}{1 + (Z/a)^2}$.
--- : the model $T^2(Z) = T_0^2 \text{sech}^2(Z/b)$.

Fig. 74. The cardinal elements of lenses with two equal cylinders.

--- : results obtained with the second approximation (Grivet–Bernard).
△△△ : experimental results. --- : Goddard’s theoretical results.
Regenstreif selected the values of $A$, $B$, $C$, $a$, $b$, $c$ in such a way that $\varphi(z)$ would represent the actual field most faithfully at the centre of the lens where the electrons move most slowly, and are hence most susceptible to deviation. This produces

$$
\varphi(z) = a + b \left[ (z + z_0) \tan^{-1}\left(\frac{z + z_0}{R_2}\right) + (z - z_0) \tan^{-1}\left(\frac{z - z_0}{R_2}\right) - 2z \tan^{-1}\left(\frac{z}{R_1}\right) \right],
$$

$$
a = \Phi_1 - 2b \left[ R_1 + z_0 \tan^{-1}\left(\frac{z_0}{R_2}\right) \right],
$$

$$
b = \frac{\Phi_2 - \Phi_1}{2R_1 + 2z_0 \tan^{-1}\left(\frac{z_0}{R_1}\right)},
$$

(8.17)

in which $R_1$ is the radius of the opening in the central electrode, which is held at potential $\Phi_1$; $R_2$ is the radius of the opening in each of the outer electrodes, which are held at potential $\Phi_2$; and $z_0$ is the distance between the central electrode and each of the other two. This notation is shown in Fig. 75a; in Figs. 75b and 75c a cross-section of the actual lens and a photograph of the actual component parts are to be seen. A potential function of the form (8.17) is not difficult to handle, and a comparison between electrolytic tank measurements and this mathematical model shows that the latter provides a very good description of the lens. We should, however, point out that a considerably less accurate model is necessary.

![Fig. 75a. The lens formed from three thin diaphragms.](image-url)
Fig. 75b. The cross-section of a three-electrode lens (scale: $\frac{1}{4}$).

casing

Fig. 75c. Some typical electrodes.

In this case than in the case of the immersion objective which is discussed later; an error of a few per cent is of little importance, as the electron velocity never drops below some 20 per cent of the maximum velocity, whereas near a cathode, the corresponding velocity might be less than one-thousandth of its value at the anode.
In a unipotential lens with all three openings the same size, and with a thick central electrode, there is a different method of approximating to the real situation which gives a better expression for \( \varphi(z) \) in the neighbourhood of the centre of the lens, particularly if the holes are large in diameter with respect to the inter-electrode distance (a situation in which Regenstreif’s formulae can no longer be applied).

The general solution of the Laplace equation can be put into the form

\[
\Phi(z, r) = \int_0^\infty [A(k) \sin kz + B(k) \cos kz] I_0(kr) \, dk. \tag{8.18}
\]

If the potential is known over the surface of a cylinder of radius \( R \), \( A(k) \) and \( B(k) \) can be determined and the axial potential \( \varphi(z) \) deduced. We suppose, therefore, that the potential over the cylinder on which the peripheries of the three apertures lie varies in some simple fashion; we suppose, for example, that the potential, which is constant over the electrode surfaces, changes linearly between them, an assumption fully justified by measurements in the electrolytic tank.

With a potential \( \Phi_2 \) applied to the outer electrodes and \( \Phi_1 \) to the central one, the formula which is finally obtained for \( \varphi(z) \) (for \( z \geq 0 \)) is

\[
\varphi(z) = \Phi_1 - \frac{\Phi_2 - \Phi_1}{2\omega(z_2 - z_1)} \log \left[ \frac{\cosh \omega(z + z_2) \cosh \omega(z - z_2)}{\cosh \omega(z + z_1) \cosh \omega(z - z_1)} \right] \tag{8.19}
\]

(see Septier (1953), for example), in which \( \omega = 1.318 \), \( z_1 \) is the abscissa of the face of the central electrode, and \( z_2 \) that of the inner face of the outer electrode. The origin \( z = 0 \) is at the centre of the lens, and the radius of the openings is taken as the unit of length.

The case \( z_1 = z_2 \) corresponds to the limiting situation of a three-cylinder lens, with an infinitely small gap between each pair of cylinders. The potential reduces to

\[
\varphi(z) = \Phi_1 - \frac{1}{2} (\Phi_2 - \Phi_1) [\tanh \omega(z + z_2) - \tanh \omega(z - z_2)]. \tag{8.20}
\]

The situation in which the diaphragms have different radii has been studied by Ehinger and Bernard (1954).

### 8.2.2 The Cardinal Elements

The formulae of the preceding section are too complicated to be substituted into the ray equation with any hope of successfully solving it. Regenstreif, following an idea of Rüdenberg (1948), represented the curve \( \varphi(z) \) by three smoothly joined parabolic arcs. The Gaussian or transgaussian equation of motion is soluble in terms of hyperbolic functions (with arguments which are circular functions) or of circular functions (with hyperbolic functions as arguments). Each arc is joined continuously
and with the same gradient at the frontier between each parabolic region: although the calculation is long, it is straightforward and possesses the advantage that the resulting formulae are particularly tractable in that

$$\frac{\varphi(0)}{\varphi(z_0)} = \frac{\Phi_2 - \Phi_1}{1 + \frac{z_0}{R_1} \tan^{-1}\left(\frac{z_0}{R_1}\right)}$$

$$x = \frac{\Phi_1 + \frac{\Phi_2 - \Phi_1}{1 + \frac{z_0}{R_1} \tan^{-1}\left(\frac{z_0}{R_1}\right)}}{\Phi_1 + (\Phi_2 - \Phi_1) \left[ 1 - \frac{R_2}{2R_1} \left( 1 + \frac{z_0}{R_1} \tan^{-1}\left(\frac{z_0}{R_1}\right) \right) \right]}$$

Fig. 76. The focal length of a weak lens as a function of $x$.

for a whole range of microscope lenses, the physical data only appear in the form of a single, and thus most convenient, parameter $x$, given by

$x$ is simply the ratio of the axial potential at the centre, $\varphi(0)$, to the potential on the axis in the plane of one of the outer electrodes. From among Regenstreif's numerous results on this series of lenses, we list five important examples:

(i) The focal length of a lens of weak or medium convergence is given by

$$\frac{f}{z_0} = \frac{8}{3} \frac{x}{(1 - x)^2}$$

(see Fig. 76); the principal planes then coincide with the central electrode.
(ii) The focal length $f$ of strong lenses and the position $z_F$ of their foci are given by

$$\frac{f}{z_0} = \frac{0.72}{\sin(0.707 \log_e x + 0.355)} ,$$

$$\frac{z_F}{z_0} = 1 + 0.764 \frac{\sin(0.707 \log_e x - 0.887)}{\sin(0.707 \log_e x + 0.355)}$$

(8.23)

(see Fig. 77).

Fig. 77. The behaviour of the focal length of a strong lens.

(iii) The critical potential $\Phi_c$ of the central electrode at which the lens is transformed into a mirror is given by

$$\frac{\Phi_c}{\Phi_2} = -\frac{1}{\frac{z_0}{R_1} \tan^{-1} \left( \frac{z_0}{R_1} \right)}$$

(8.24)

(see Fig. 78).

(iv) The focal length of convergent or divergent mirrors is given by

$$\frac{f}{z_0} = \frac{0.72}{\sin \left[ 0.707 \log_e (-x) + 0.355 \right]}$$

(8.25)

(see Fig. 79).
(v) The cardinal elements of lenses in which the central electrode is positive and accelerating in its action are given by Regenstreif, and the behaviour of the focal length is illustrated in Fig. 80.

The formulae which we have at our disposal, therefore, are adequate to describe very fully indeed the behaviour of this family of lenses. These theoretical predictions are in satisfactory agreement with the measurements of Heise and Rang (1949).

![Fig. 78. The repulsive potential \( \Phi_e/\Phi_2 \) as a function of \( z_0/R_1 \).](image)

![Fig. 79. The focal lengths of mirrors.](image)
Fig. 80. The focal lengths of lenses with a positive central electrode.

Fig. 81. The appearance of a lens with a thick central electrode.
Finally, we mention the case in which the central electrode is very thick, for which the formulae are more complicated but no more difficult—the potential is represented by a wide “plateau” in the centre of the lens, and to the parabolic arcs, we must add a fourth section, a horizontal straight line; the results are again in good agreement with experiment.

In Fig. 81, the lens is shown schematically; in Fig. 82, the values of the focal length are plotted. The parameter \( v \) is defined by \( v = z_1/z_2 \).

![Graph showing the focal length of a lens with a thick central electrode as a function of \( x \), for various thicknesses.](image)

**Fig. 82.** The focal length of a lens with a thick central electrode as a function of \( x \), for various thicknesses.

### 8.2.3 Aberrations; Ellipticity Astigmatism

The results of the preceding section have a special relevance when we consider aberrations, since the focal length, distortion, and spherical aberration all pass through their minima simultaneously (Bruck and Romani, 1944; Heise, 1949). With the aid of Regenstreif’s formulae, therefore, we can easily establish the conditions for these minima which correspond to the value \( x_m = 5.8 \times 10^{-2} \) \((f_m = 0.7730)\) when the central electrode is thin. Only with difficulty can the aberration coefficients \( C_z \) and \( C_c \) of the lens be determined using this approximate method; we need to represent \( \Phi(z) \) by a single, well-chosen, analytic function, which would enable us to integrate the equation of the paraxial rays. Glaser (1952) and Glaser
and Schiske (1954, 1955) have suggested the following model:

$$\Phi_0(z) = \Phi_0 - \frac{\Phi(O)}{1 + (z/d)^2} = \Phi_0 \left(1 - \frac{k^2}{1 + (z/d)^2}\right),$$

with

$$k^2 = \frac{\Phi(O)}{\Phi_0}.$$

$\Phi_0$ denotes the accelerating voltage and $d$ is the half-width of the bell-shaped curve representing $\Phi(z)$; $\Phi_0(O) = \Phi_0 - \Phi(O)$ is the potential at $z = 0$ (Fig. 82a).

**Fig. 82a.** The potential distribution in an einzel (unipotential) lens, the Glaser-Schiske model; $d$ denotes the half-width at half the maximum height.

The factor $k^2$ can be calculated by Regenstreif’s method, recently improved by Kanaya et al. (1966), for each type of lens and each value of $\Phi_1$. The following table gives the values of $k^2$ and $d$ for various lenses, when the central electrode is held at cathode potential ($\Phi_1 = 0$).

<table>
<thead>
<tr>
<th>$R_1/R_2$</th>
<th>$z_1/R_1$</th>
<th>$z_2/R_1$</th>
<th>$k^2$</th>
<th>$d/z_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.550</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.740</td>
<td>0.677</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.860</td>
<td>0.590</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.720</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.720</td>
<td>0.581</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.855</td>
<td>0.581</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.950</td>
<td>0.700</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.950</td>
<td>0.700</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
<td>3.5</td>
<td>0.980</td>
<td>0.705</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>4.5</td>
<td>0.988</td>
<td>0.659</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.996</td>
<td>0.750</td>
<td></td>
</tr>
</tbody>
</table>
Integration of the paraxial equations of motion leads to well-tabulated elliptic integrals; the cardinal elements and the aberration constants were then calculated by Glaser and Schiske. Kanaya et al. have cast these into a simplified form. Throughout the range $0 < k^2 < 1$, the following formulae are accurate to within a few per cent:

\[
\frac{f}{d} = \frac{64}{3\pi} \left( \frac{1-k^2}{k^4} \right), \quad \frac{C_s}{d} = 2 \left( \frac{f}{d} \right) \left( \frac{1+k^2/4}{1-3k^2/4} \right), \quad \frac{C_t}{d} = \frac{1}{2k^2} \left( \frac{f}{d} \right)^3.
\]

The values of $f/d$ (the "immersion" focal length), $f'/d$ (the asymptotic focal length), $z_H/d$ (the absolute value of the position of the principal plane), $C_s/f$ and $C_s/d$ are shown in Fig. 82b.

In theory, we should be able to achieve very small values of the aberration coefficient $C_s$ for $k^2$ close to 1 ($C_s < d$). In fact, however, the object must be placed at the focus in object space in an electron microscope objective, outside the field (or, at least, in a region where $\Phi(z) = \Phi_0$); under these circumstances, we must have $f \approx f'$ and hence $k^2 < 0.8$ which implies $C_s/f > 20$ ($C_s/d > 40$) as Fig. 82b shows. These values agree well with those measured on real objectives.

The stronger excitations will only be usable when the object is virtual, therefore (projective lenses).

The astigmatism which results from an ellipticity of the aperture also passes through a minimum at the same point as the focal length. Intuitively,
this is reasonable, as the focal lengths of two lenses with exactly circular
openings, one of radius equal to the major axis of the actual lens, the other
of radius equal to the length of the minor axis, are in fact the same. To
prove rigorously that the minima occur simultaneously is less easy. Regen-
streif has succeeded in demonstrating that it would be so in a lens of which
the central opening is only slightly elliptical (1951b), a restriction which
allowed him to make certain approximations.

Regenstreif (1951c) has even gone so far as to consider strongly
elliptical openings which might be very useful, as we have already seen,
in the correction of spherical and chromatic aberration, where we need
lenses with a pronounced ellipticity.

8.2.4 Application to the Electrostatic Electron Microscope

The results of the previous sections have proved to be very valuable
as guides towards possible improvements of the electrostatic electron
microscope.

The objective

Our formulae suggest that we can construct a remarkable type of ob-
djective, in which the focus is within a few tenths of a millimetre from the
face of the lens at which the electrons enter, and the "working distance"
is thus a minimum; further, the focal length corresponds to the minimum
of the curve which represents \( f \) as a function of \( x \)—this, in turn, means
that the objective is very stable and completely insensitive to small variations
of the potential such as those which can only with difficulty be suppressed
since they originate in stray currents; finally, by choosing this value of the
focal length, the spherical aberration and ellipticity astigmatism are auto-
matically minimized.

This kind of objective can be built as a unipotential lens in which the
potential of the central electrode and that of the cathode are the same;
the central electrode, therefore, will be quite thick. There is, however,
advantage to be gained by taking into account the insensitivity to variations
in the potential, and polarizing the central electrode negative with respect
to the cathode (by a few kV, provided by an autobias resistance). Under
these conditions, for the same focal length, the ellipticity astigmatism is
reduced for machining of the same precision, since the central opening
can be made perceptibly larger.

Projective lenses

With the aid of the formulae quoted above, a whole range of projective
lenses free of distortion can be designed—it proves always to be valuable
to use a thick central electrode, but in this case, we are far less restricted
in our choice of the various dimensions.
8.2.5 Other Studies

A new and simple calculation procedure has been suggested by Bernard (1951a). Useful information is also to be found in Mahl (1940), Dosse (1941c), Chanson (1947), Plass (1942) and Ramberg (1942), Lippert and Pohlitz (1952), see § 17.3.2. Mirrors, which have been studied in particular by Regenstreif (1951a), have been used by Mahl and Pendzich (1943) in a microscope at magnification 5000. More recent work, which has appeared since the first edition of this volume, and which contains copious theoretical and experimental results on the symmetrical three-electrode einzel lens, has been collected together into a special section (17.3).

8.3 "MESH" OR "GRID" LENSES

(cf. § 4.1.3)

Lenses of this kind were used first—and very brilliantly—to improve the clarity of the mass spectrograph of J. J. Thompson (Cartan, 1937); nowadays, their principal application is in linear ion accelerators, and Bernard (1951a, 1953a) has recently established the full exact theory of their mode of operation.

The only difference between this sort of lens and the familiar three-electrode lens lies in the fact that the central electrode is replaced by a network of fine mesh, held in a suitable support (Fig. 83a). This grid is held at a potential \( \Phi_G \), and the outer electrodes at \( \Phi_0 \); we define \( \sigma \) to be the quantity \( \sigma = \frac{\Phi_G - \Phi_0}{\Phi_0} \). The axial potential can be calculated in the form of a Fourier integral, which results in a complicated expression which is, however, well represented by the function

\[
\varphi(z) = \Phi_0 \left[ 1 + \sigma - \frac{\sigma}{2\omega d} \log \frac{\cosh \omega \frac{z + d}{R}}{\cosh \omega \frac{z - d}{R}} \right], \quad (8.26)
\]

\( \omega = 1.318 \).

The distribution of potential on the axis is symmetrical with respect to the plane of the grid, and is shown in Fig. 83b; the reason why this particular design of lens is of interest is evident, for the potential curve is always concave in the same sense. The sign of \( \varphi''(z) \) and the sense of the radial electric field, therefore, remain unchanged throughout the whole length of the lens. Lenses of this kind are thus either uniformly divergent or uniformly convergent. The differential effect from which the usual type of lens suffers has been eliminated.
An approximate representation of the potential enables us to calculate the cardinal elements, just as in the case of the two-cylinder lens (§ 8.1.2). Picht's reduced form of the Gaussian equation is used, with a characteristic function \( T = \varphi' / \varphi \) in the form of two of the bell-shaped fields of Glaser, one for each half of the lens.

Fig. 83. The form of the electrodes and the distribution \( \varphi(z) \) in a lens with a mesh electrode.

\[
T(z) = \frac{T_0}{1 + \left( \frac{z - z_0}{a} \right)^2}, \quad z < 0,
\]

\[
T(z) = \frac{T_0}{1 + \left( \frac{z + z_0}{a} \right)^2}, \quad z > 0.
\]

The parameters \( T_0 \), \( z_0 \) and \( a \) are chosen in such a way as to minimize the systematic error which is introduced by this approximate representation. The following results are obtained:

\[
\frac{1}{f_i} = - \frac{1}{f_0} = - \frac{2 \sin^2(K \alpha)}{K^2 \alpha} \left[ K \cot(K \alpha) - \cot \alpha + \frac{1}{4} \frac{T_1 \alpha}{\sin^2 \alpha} \right],
\]

\[
z_{i1} = -z_{i0} = a K \cot(K \alpha) + f_i,
\]
in which

\[ K = \sqrt{1 + \frac{3}{16} T_0^2 a^2}, \]  

(8.28)

and

\[ \alpha = - \cot^{-1} \left( \frac{z_0}{a} \right), \]  

(8.29)

and \( T_1 \) is the value of the characteristic function at \( z = 0 \).

The numerical values which have been calculated from this formula prove to be in good agreement with the experimental values obtained by Knoll and Weichart (1938). The comparison between the convergence of a grid lens and the corresponding three-electrode lens in which the grid is replaced by an electrode with an opening of radius \( \frac{1}{2} R \) is shown on the same scale, in Fig. 84. At the same potential, the grid lens is always far more convergent than a normal lens, above all for small values of \( \sigma \). Grid lenses are likely to be useful, in consequence, for focusing fast particles (the velocity of which corresponds to a high value of the potential \( \Phi_0 \)) when one can apply only \( \Phi_G - \Phi_0 \), small compared with \( \Phi_0 \), as focusing potential; this is just the situation in linear accelerators.

![Fig. 84. A comparison between the convergence of a lens with a grid and the corresponding three-electrode lens.](image-url)
It has been assumed, in the preceding discussion, that the grid behaves like a continuous membrane. In fact, it is only rarely possible to construct metallic membranes which combine mechanical robustness with the extreme transparency which is necessary if the majority of the electrons are to pass through unretarded. In practice, a fine-mesh network is used, the holes of which produce a slight perturbation of the potential in their vicinity—the ideal trajectories are slightly modified, and the Gaussian image, whose appearance can be calculated from the cardinal elements, is slightly blurred; it suffers, we might say, from "grid aberration". By regarding each hole in the grid as a weak lens, Bertiein (1951) has established a simple expression, with the aid of the parametric representation of the Gaussian trajectories, for the radius $r$ of the small aberration disc by which the ideal point is replaced:

$$r = \frac{\sigma}{(1 + \sigma)^{3/4}} \cdot \frac{f_i}{2d} \cdot \frac{\sin K x}{K \sin x} \cdot \tanh \left( \frac{\omega d}{R} \right); \quad (8.30)$$

$\sigma$ is the radius of the holes in the mesh. Further calculation shows that the coefficient of $\sigma$ is always less than unity, and thus that the radius of the spot into which the electrons are spread is less than that of the hole itself.

The most finely meshed grid at present commercially available seems to be that which is known as "electroformed" in nickel, with holes of some 15 $\mu$ across. The construction of this type of gauze is described in detail in an article by Wilder (1949). The aberration is not, therefore, a sufficiently important defect to prohibit the use of this type of lens.

The geometrical aberrations inherent in all lenses are no less apparent in grid lenses; although spherical aberration is still present, Scherzer's theorem according to which the spherical aberration is always positive (§ 7.7.2) can no longer be proven, as a result of the discontinuity in $E_z$. Rather, as the full calculation by Bernard (1952b) shows, grid lenses may possess either positive or negative spherical aberration. We can conceive, therefore, of an ensemble of stigmatic systems in which a grid lens cancels the spherical aberration of the other lenses (Gianola, 1950; Seman, 1952).

That the grid always produces its own aberration must not, however, be forgotten—such a combination of mesh lens and ordinary lenses, therefore, may well be capable of a full correction of spherical aberration only when it is possible to replace the grid by a continuous membrane, virtually transparent to the particles which are being used. For ions, such a goal seems indeed far off, but for electrons, it is already possible to construct beryllium membranes about 10 m$\mu$ thick (Hast, 1948) which are quite transparent to electrons, as the following results of Möllenstedt (1950) show

Kivolovts 30 40 50; Transparency 0.53 0.3 0.15.

Transparency is defined to be the ratio of the number of electrons which pass through the thin foil and are slowed down (losing a few eV) to the number which are scattered elastically.
8.4 THE IMMERSION OBJECTIVE

8.4.1 The Optical Properties of an Accelerating Region of Constant Field

The simplest type of immersion objective is the two-electrode version—it consists of a plane cathode, \( K \), and a thin plane anode, \( A \), in which a hole of small diameter has been punched (Fig. 85a). The electrons are emitted by \( K \) with a mean velocity fixed by \( \Phi_0 \); they then cross the space between \( K \) and \( A \), of length \( a \), in which the electric field is assumed to be constant and given by

\[
E = -\frac{\Phi_a - \Phi_0}{a}.
\]  

(8.31)

The electrons finally enter the field-free region beyond \( A \) \((E_z = 0)\), after being refracted as they pass through the anode.

---

![Diagram](image)

**Fig. 85a.** The two-electrode immersion objective \((K = \text{cathode}; A = \text{anode})\) with two particular trajectories.

(i) The trajectories in the inter-electrode region

We have

\[
\varphi(z) = \Phi_0 + E_z = (\Phi_a - \Phi_0) \frac{z}{a} + \Phi_0,
\]

and hence

\[
\varphi'(z) = \frac{\Phi_a - \Phi_0}{a},
\]

and

\[
\varphi''(z) = 0.
\]
The general equation of motion becomes

$$r'' + r' \frac{\varphi'}{2 \varphi} = 0$$

(8.32)

in the region $0 \leq z \leq a$ so that on integrating,

$$r' \sqrt{\varphi} = r_0 \sqrt{\Phi_0} = \text{const.},$$

(8.33)

whence

$$r(z) = r_0 + \int_0^z \frac{r_0'}{\sqrt{\varphi(z)}} \, dz,$$

and finally

$$r(z) = r_0 + 2r_0' \sqrt{\frac{\Phi_0}{\Phi_a}} \frac{\varphi(z)}{(\varphi(z) - \Phi_0)z}.$$

(8.34)

At $z = a$ (where the electrons arrive at the anode),

$$r(a) \simeq r_0 + 2r_0' \sqrt{\frac{\Phi_0}{\Phi_a}} a,$$

since

$$\frac{\Phi_0}{\Phi_a} \ll 1,$$

and

$$r'(a) = r_0 \sqrt{\frac{\Phi_0}{\Phi_a}}.$$

(ii) The passage through the anode

We assume that the anode is thin enough for the electric field to vary suddenly from $E_1$ on the left to $E_2$ on the right over an infinitely short interval between $z = a^-$ and $z = a^+$. The value of $\varphi''$ becomes infinite, but $\varphi$, $\varphi'$, $r$ and $r'$ remain finite. We obtain, on integrating,

$$\int_{a^-}^{a^+} r'' \, dz = - \int_{a^-}^{a^+} \left( \frac{\varphi'}{2 \varphi} r' + \frac{\varphi''}{4 \varphi} r \right) \, dz,$$

or

$$r_{a^+} - r_{a^-} = \frac{\Phi_{a^+}' - \Phi_{a^-}'}{4 \Phi_a} = \frac{E_2 - E_1}{4 \Phi_a} r_a.$$

(8.35)

The anode is clearly equivalent to a thin lens with convergence

$$\frac{1}{f} = - \frac{\Delta r'}{r_a} = - \frac{E_1 - E_2}{4 \Phi_a}.$$

(8.36)

When $E_2 = 0$, we have simply

$$- \frac{1}{f} = + \frac{E_1}{4 \Phi_a} \simeq \frac{1}{4a},$$

(8.36')

or $f = - 4a$. 
The lens is divergent; on passing through such a lens, the slope of a trajectory is altered by

$$\Delta r'_a = \frac{r_a}{4a}. \tag{8.37}$$

(iii) The position of the image of $K$; the magnification

By calculating a trajectory such that $r_0 = 0$ and $r'_0 = 1$, we can determine the position, $z_{K'}$, of the image of $K$.

When the electrons emerge from the anode,

$$r_a = 2a \sqrt{\frac{\Phi_0}{\Phi_a}},$$

$$r'_a = \frac{3}{2} \sqrt{\frac{\Phi_0}{\Phi_a}},$$

so that

$$z_{K'} = a - \frac{r_a}{r'_a} = -\frac{a}{3}. \tag{8.38}$$

A second trajectory, parallel to the axis, such that $r_0 = 1$ and $r'_0 = 0$ produces $r_a = 1$ and $r'_a = \frac{1}{4a}$. This trajectory intersects the plane $z = -\frac{a}{3}$ at $r\left(-\frac{a}{3}\right) = \frac{2}{3}$. The image, $K'$, of $K$ is thus virtual, upright, lies at $z = -\frac{a}{3}$ and is linearly magnified, $G = +2/3$.

The simple two-electrode objective cannot, therefore, give by itself a real magnified image of $K$. It must be used in conjunction with a lens $L$ placed beyond $A$, which produces a real image of $K'$ on the final screen; this combination has been employed effectively in emission microscopes ($L$ may be either electrostatic or magnetic).

8.4.2 The Three-electrode Objective: the Complexity of the Problem

This type of lens is used for two quite distinct purposes. In the electron guns of microscopes and cathode ray oscillographs, it serves to concentrate as intense a beam of electrons as possible into the smallest possible spot, or more generally, it produces an electron beam of some given shape and intensity. In emission microscopy and image conversion, on the other hand, the immersion objective produces a magnified image of the emissive cathode on the fluorescent screen (Fig. 85b).

These two uses are totally dissimilar; in the first, the beam of electrons is intense, and space charge effects have often to be considered—it is for the sake of the very fine crossover that this kind of system is studied. In the other application, space charge is quite negligible, as the intensities are very feeble—a few $\mu$A for high tensions of the order of tens of kilovolts.
Further, it is not the crossover that is of interest, but the real image of the emitting surface. Only a first approximation to the real situation in electron guns is provided by results obtained with, for example, the emission microscope (see § 8.4.3).

![Fig. 85b. The electron trajectories in an immersion objective.](image)

### 8.4.3 Properties of the Three-electrode Immersion Objective

Two complications appear in this situation: firstly, the curve \( \varphi(z) \) has a most complex shape, as is shown by the electrolytic tank measurements which have been made on objectives with plane electrodes by Septier (1954b)—see Fig. 86a—and Duchesne (1953)—see Fig. 86b, c; the notation is explained by Fig. 87. Septier’s lens was designed for a microscope, and hence was to provide strongly magnified images; Duchesne’s, on the other hand, had only to produce the unit magnification required in an “electron telescope”. The result is that the ray equation cannot be handled directly, and the reduced equation is no advantage, as \( T(z) = \frac{\varphi'(z)}{\varphi(z)} \) goes to infinity at the origin.

The other reason why this lens is not easy to study arises from the fact that the ray paths are peculiarly susceptible to small variations in the lens power near the cathode, where the electrons are still moving rather slowly; such variations can occur as a result of small variations in the potential applied to the control electrode or Wehnelt (the electrode adjoining the cathode), or of slight changes in the shape of the control electrode, or even the cathode itself. Trifling alterations in the potential distribution produce considerable repercussions on the focusing and magnification of the image.
Fig. 86. The form of the distribution $\varphi(z)$. (a) In a very convergent objective (Septier). (b) In an objective with a weak magnification. (c) In an objective with a large aperture (Duchesne).

Fig. 87. The appearance of an objective with thick electrodes.
Only with extreme care, therefore, can the optical properties of this type of lens be calculated with methods similar to those we have been using for other kinds of lens, namely straightforward calculation using a convenient but approximate model for \( \varphi(z) \) or for \( T(z) \). In particular, such models must be chosen with the utmost care in the region around the cathode.

A recent study which Septier (1954b) has made of an immersion objective with thick plane electrodes shows that it is possible to calculate the axial potential in such a lens quite accurately; a step-by-step calculation then gives the cardinal elements (focal length, focus, and principal plane, on the image side) to within one or two per cent.

There are two formulae with which the potential \( \varphi(z) \) in an objective with thick plane electrodes (see Fig. 87) can be calculated quite accurately; they overestimate and underestimate the potential near the cathode (which is chosen to be the origin of potential) respectively. The first, which can only be applied to lenses with all openings of the same size, has the form

\[
\varphi(z) = \frac{\sigma}{2\omega a} \log \frac{\cosh \omega(z + a)}{\cosh \omega(z - a)} + \frac{1 - \sigma}{2\omega c} \log \frac{\cosh \omega[z + (a + b + c)]}{\cosh \omega[z - (a + b)]} \frac{\cosh \omega[z - (a + b)]}{\cosh \omega[z + (a + b)]},
\]

(8.39)
in which \( R_1 = R_2 = 1 \), \( \Phi_1 / \Phi_2 = \sigma \), and \( \omega = 1.318 \).

The second, valid for any situation, is given by

\[
\varphi(z) = \frac{\sigma}{a \pi} \left[ (z + a) \tan^{-1} \left( \frac{z + a}{R_1} \right) - (z - a) \tan^{-1} \left( \frac{z - a}{R_1} \right) \right] + \frac{1 - \sigma}{c \pi} \left[ (z + \beta) \tan^{-1} \left( \frac{z + \beta}{R_2} \right) - (z - \beta) \tan^{-1} \left( \frac{z - \beta}{R_2} \right) \right] - (z + \alpha) \tan^{-1} \left( \frac{z - \alpha}{R_1} \right) + (z - \alpha) \tan^{-1} \left( \frac{z - \alpha}{R_1} \right),
\]

(8.40)
in which \( \alpha = a + b \), and \( \beta = a + b + c \).

A comparison between the results of these two formulae and the experimental results for a particular lens is shown in Fig. 86a.

A great deal of theoretical work has been devoted to the calculation of the resolving power of this kind of system, which can always be written in the form

\[
\delta = k \frac{\Phi_c}{E_0},
\]

(8.41)
in which \( \Phi_c \) is the most probable energy with which the electrons are emitted (measured in volts), and \( E_0 \) is the electric field at the cathode; \( k \) is a constant, to which different authors ascribe values which vary between 4 and 1.2—the latter seems the most likely. (For further details, see Langmuir, 1937; Recknagel, 1941; Artcimovitch, 1944; and Septier, 1954a.) If, however,
instead of calculating the diameter of the electron beam which has originated at an object point in or close to the Gaussian image plane, we determine the electron intensity distribution in the same cross-section, and if we define the resolving power as the diameter of the circle which contains about 90 per cent of the electrons of the beam, it becomes apparent that in the neighbourhood of the point of the caustic of the beam which lies in front of the Gaussian image plane, the value of \( k \) can be lowered to about 0.2 or 0.3, and theoretical values of \( d \) can be obtained which are far lower than those which are normally admitted (Septier, 1955). We can also explain the fact that the resolving powers which have been measured by certain workers have been smaller than the theoretical resolving power which is given by the formulae quoted above, with \( k = 1.2 \).

If we now stop down the elementary beams which emerge from different points on the object, the angular aperture of these beams can be reduced and the resolving power improved. The actual mechanism of this improvement will be examined in \( \S \) 13.2.2.

The first experimental exploration of the objective with very thin plane electrodes was performed by Johansson (1933, 1934), while more recently, Septier (1952, 1953a, b, 1954b) has examined the same type of objective with much thicker electrodes. We shall now list the main features of the Gaussian optics of such a lens: in the work of Duchesne (1953), the reader will find a study of a similar objective used at low magnification as an image converter, where rays steeply inclined to the axis, originating in a large cathode, are considered. In addition Soa (1959) has recently studied several three-electrode immersion objectives; his results, which show the way in which the optical characteristics vary with the various geometrical parameters, complement the experimental data outlined below.

The focal curve and the focal length at high magnification

The parameters upon which the optical properties depend are \( a, \sigma, R_1, R_2, b \) and \( c \) (see Fig. 87); \( \sigma \) is the relative polarization of the control electrode.

Once the parameters which describe the actual construction of the lens have been fixed (\( R_1, R_2 \) and \( c \)), the image is focused onto some fixed point \( D \) for an unlimited family of pairs of values \( (a, \sigma) \). The curve \( a = f(\sigma) \) is known as the focal curve, and is characteristic of the objective. To obtain this curve, \( a \) is set at a number of known values, and at each the image is focused by varying \( \sigma \). For each \( (a, \sigma) \) pair, the magnification can be measured either photographically, or by measuring the lateral displacement of the image which corresponds to a known displacement of the object. As the objective is far from the screen \( (D >> a + b + c) \), the focal length at the focus can be calculated from the formula \( f = \frac{D}{G} \). In Fig. 88a, the
values of \( a \) and \( f \) are shown (as functions of \( \sigma \)) for an objective defined by \( R_1 = R_2 = 1, c = 2 \) and \( b = 0.7 \). For all the objectives which were studied, \( f \) proved to be an increasing function of \( \sigma \), which could be regarded as a hyperbola in the usual region of operation \((-0.05 < \sigma < 0.05)\),

\[
f = \frac{A}{\sigma - B}.
\]

\( G \), correspondingly, varies linearly with \( \sigma \); for \( \sigma = 0 \), the immersion objective is strongly convergent (3 mm < \( f < 4.5 \) mm).

Figure 88b (Soa, 1959) shows how the positions of the focus, the principal plane and the image vary when \( \sigma \) is varied but \( a \) is held constant together with the variation of \( f \) as a function of \( \sigma \) for the same objective. It can be seen that the condition \( f = f_{\text{min}} \) corresponds to an image located very close to the anode; if we are to obtain a high magnification image, we shall automatically have \( f > f_{\text{min}} \). This phenomenon is common to all lenses of this type.

**The influence of the geometrical parameters**

*The control electrode–anode distance.* This distance, \( c \), has to be kept within a rather narrow band of values around \( c = 2 \) mm. Below this, there is a danger of electrical breakdown with \( \Phi_2 = 25 \) or 30 kV; above, the electric field is inadequate. An increase in \( c \) is equivalent to a decrease in \( \Phi_2 \).

The focal curve can be deduced from the curve corresponding to \( c = 2 \) mm, through their similarity parallel to the \( \sigma \)-axis.

*The radius of the anode.* No new effect is produced by varying \( R_2 \), as an increase in this radius corresponds to a decrease in \( \Phi_2 \). To transfer from the focal curve for an objective with \( R_2 = 1 \) to an objective with \( R_2 = 2.5 \), for example, we simply multiply the values of the abscissa \( \sigma \) by 0.8. The focal length increases with \( R_2 \), visible evidence of the progressively stronger
divergent action of the opening in $A$. For negative polarizations, however, this effect is slight, a few per cent only between apertures with radii in the ratio 2.5:1.

*Control electrode thickness.* An increase $\Delta b$ in $b$ shifts the focal curve along the ordinate axis $a$ towards the origin. Defining the angular aperture $\alpha$ of the

---

**Fig. 88b.** The convergence (—) and the positions of the focus (—), principal plane (· · ·) and image (+++ ) as functions of the polarization $\sigma$.

**Fig. 89.** The variation of $f$ with the various geometrical parameters.
control electrode by the expression \( \tan \alpha = \frac{R_1}{a + b} \), \( \alpha \) is effectively constant near \( \sigma = 0 \) when \( b \) alone is varied. For \( \sigma = 0 \), the focal length \( f \) remains practically constant; for \( \sigma < 0 \), \( f \) increases slightly as \( b \) decreases, although never more than 3 per cent for variations in \( b \) between 0.2 and 0.4 mm. The change is in the opposite sense if \( \sigma \) is positive.

**Control electrode radius.** The focal curves are again shifted parallel to the \( Oa \) axis, but the effect of a change in \( R_1 \) is more pronounced than a change in \( b \), and is in the opposite sense, as increasing \( R_1 \) corresponds to increasing \( a \). The aperture of \( W \) decreases when \( R_1 \) increases, irrespective of the polarization. \( R_1 \) proves to be the vital parameter of the objective, for the focal length varies rapidly with \( R_1 \) — for example, with \( R_2 = 1, b = 0.7 \) and \( c = 2, f_\sigma = 0 \) changes from 4 to 6 mm when \( R_1 \) increases from 1 to 1.5 mm.

Figure 89 summarizes the effects upon \( f \) of the various parameters, in a few numerical examples.

**The field \( E_0 \) at the surface of the cathode**

The resolving power in the central region is restricted only by the spherical aberration of the component pencils, which in turn is a decreasing function of the field \( E_0 \) which prevails in the neighbourhood of \( K \) if the objective is unstopped. It is most important, therefore, to know the magnitude of this field. Knowing the focal curves for various combinations of the geometrical parameters, the values and variations of \( E_0 \) for various values of the objective parameters \( \sigma, R_1, R_2 \) and \( b \) can be studied in the electrolytic tank. \( E_0 \) will be measured in volts per centimetre if we regard the anode potential \( \Phi_a \) as one volt.

For any given geometrical arrangement, \( E_0 \) is unaffected by wide variations in \( a \) (see Fig. 90), and begins slowly to decrease only for large positive values of \( \sigma \) (and hence large values of \( a \)). (See Septier, 1954b; Soa, 1959.)

![Figure 90](image_url)  
**Fig. 90.** The field \( E_0 \) as a function of the distance between the cathode and the grid.
Increasing the anode radius $R_2$ produces a decrease in $E_0$, and is thus equivalent to reducing $\Phi_2$. The effect of decreasing the control electrode thickness $b$ becomes progressively more pronounced the larger the opening in the control electrode, and is a slight increase in $E_0$; $E_0$ falls rapidly, however, if the radius of the hole in the control electrode is increased. In every case, therefore, this latter proves to be the most important of the parameters. In the following table, a few numerical examples of these changes are given ($c = 2$ mm throughout).

<table>
<thead>
<tr>
<th>$R_1$ (mm)</th>
<th>$R_2$ (mm)</th>
<th>$b$ (mm)</th>
<th>$E_0$ (V cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.7</td>
<td>0.63</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>0.7</td>
<td>0.48</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.7</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.62</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.2</td>
<td>0.64</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>0.7</td>
<td>0.62</td>
</tr>
<tr>
<td>1.2</td>
<td>1</td>
<td>0.7</td>
<td>0.54</td>
</tr>
<tr>
<td>1.5</td>
<td>1</td>
<td>0.7</td>
<td>0.45</td>
</tr>
</tbody>
</table>

There would be an advantage, therefore, in using a very thin control electrode with an extremely small hole, but in fact, it is not feasible to work other than with $a \geq 0.2$ mm, $b \geq 0.2$ mm, and $R_1 \geq 0.5$ mm (the opening in the control grid inevitably becomes dirty, and if this hole is too small, the ensuing perturbations are excessive). For $R_1 = 0.5$ mm, $R_2 = 0.25$ mm, $b = 0.2$ mm, $E_{\text{max}}$ is of the order of 1 V cm$^{-1}$; if, therefore, we take

$$\Phi_c = 0.1 \text{ V}, \quad \Phi_2 = 30 \text{ kV}, \quad \text{and} \quad \delta_{\text{min}} = 1.2 \frac{\Phi_c}{E_0},$$

the resolving power can never exceed

$$\delta_{\text{min}} = 40 \text{ m}\mu.$$

**Field curvature**

Off the axis, the resolving power deteriorates rapidly as a result of field curvature. For given $a$, the effect of increasing the polarization $\sigma$ from the value at which the centre of the image is in focus is to produce an annular region of increasing mean diameter $d_\text{m}$ in which the image is in focus.

Just how important is the curvature of the image surface is shown by plotting the curve $d_\text{m} = f(\Delta \sigma)$ for a few values of $\sigma$; the limiting case of a plane image surface corresponds to $d_\text{m} = \infty$ for $\Delta \sigma = 0$.

For a given geometrical design, this curvature is smallest, though not zero, near $\sigma = 0$, increases slowly for $\sigma < 0$, but tends rapidly towards infinity for $\sigma > +0.1$. If we select some mean value $\delta_\text{M}$ for the resolving power, we can measure the diameter $d$ of the sharply imaged zone, on the edges of which $\delta = \delta_\text{M}$; $d$, therefore, is greatest for $\sigma = 0$, and if the point
of observation is at a distance \( L = 66 \text{ cm} \), \( d_{\text{max}} \approx 5 \text{ cm} \). Under these operating conditions, \( d_{\text{max}} \) is substantially the same for all the objectives; decreasing \( R_3 \) has a slightly adverse effect upon the image quality; \( b \) has no appreciable influence. If \( R_1 \) is very small, the image can be focused only for positive values of \( \sigma \), and below \( R_1 = 0.5 \text{ mm} \), the field curvature, which decreases with \( R_1 \) (as the "aperture" of \( W \) increases), again becomes troublesome.

The optimum electrical and geometrical parameters for an immersion objective with plane electrodes suitable for electron microscopy will be, therefore, \( \sigma = 0 \), \( R_1 = 0.5 \) to 0.6 mm, \( b = 0.2 \) to 0.3 mm and \( c = 2 \text{ mm} \) (this means that \( 0.2 \text{ mm} < a < 0.4 \text{ mm} \)); \( R_2 \) can be taken equal to 0.25 mm.

**A new type of objective**

In order to increase \( E_0 \), and hence the resolving power, we might construct a system consisting of an accelerating unit and a weak three-electrode lens \( L \). A thin diaphragm \( D \) with a large opening has, however, to be inserted between the cathode and the accelerating anode (Fig. 91) to reduce the aberration. It is tied to the cathode, and although this lowers the field \( E_0 \), this latter is still considerably larger than in the objectives described above (Septier, 1952, 1954b). If the electrode \( A \) is very thin, and contains an extremely small opening, the diaphragm \( D \) is unnecessary, and we can then obtain \( E_0 \approx 150 \text{ kV cm}^{-1} \) (Düker and Illenberger, 1962).

- \( \phi = 0 \)
- \( O \)
- \( \hat{\phi}_1 \)
- \( \hat{\phi}_2 \)
- \( K \)
- \( D \)
- \( A \)
- \( L \)
- \( z \)

**Fig. 91.** An objective with a large field \( E_0 \).

**Theoretical work on the immersion objective**

In an article by Hahn (1958) is to be found a full study of the trajectories in an immersion objective, setting out from a series expansion of the poten-
tial $\varphi(z)$ and the first order equation of motion for the Gaussian optical properties of an axially symmetrical system; the trajectories are also obtained in series form. The cardinal elements and the third order aberrations are then studied, and in the original paper is to be found a table of the various coefficients (transverse and longitudinal chromatic aberration, spherical aberration, coma, astigmatism, distortion and field curvature); these are primarily of theoretical interest.

8.5 THE TRIODE ELECTRON GUN

8.5.1 Introduction

The three-electrode immersion objective is very often used as an electron source, under the name of triode gun. It consists of an emissive cathode, a focusing electrode or "grid" (or Wehnelt in the German literature) and an accelerating electrode (or anode).

If we characterize the gun by the quantity $P = \frac{I}{V_{a}^{3/2}}$, which is measured in $AV^{-3/2}$ and is known as the perveance, guns can be classified into two broad categories ($I$ represents the total intensity and $V$ the potential of the anode).

(i) The first category comprises guns with a very low perveance ($P < 10^{-9}$), in which the role of the space charge of the beam is negligible (these are the guns used in electron microscopes and instruments of a similar nature) or of little importance (as in oscilloscope guns where the only effect of the space charge is a slight widening of the spot—see Chapter 12). The electron trajectories are identical to those in the immersion objective which we have already discussed, and in particular, they cross the axis to form a zone of minimum diameter known as the "crossover".

(ii) The second category contains all guns with a large perveance ($P > 10^{-7}$), in which we are trying to extract the maximum useful current while applying a potential which is sometimes very small. This is the case with numerous modern ultra-high-frequency generators (klystrons and travelling-wave tubes, for example) and devices which make use of the energy in an electron beam for melting, cutting and welding the refractory metals in vacuo. The best electrostatic guns at present lie in the range $10^{-6} < P < 5\cdot10^{-6}$ (on this topic, the survey article by Süsskind, 1956, should be consulted).

Guns with a very high perveance are distinguished from those in the categories above by two very special characteristics: the cathode emission is limited by the space charge (there is no electric field at the cathode) and the trajectories, subjected both to the converging action of the three-electrode system and to the diverging action of the space charge, no longer cross the axis. The beam has only a zone where its cross-section shrinks
to a minimum (or "throat") which lies at the exit of the gun; subsequently it diverges very rapidly. We make a special effort, therefore, to intensify the convergence of the triode system to the utmost, by choosing special shapes for the electrodes.

We shall first of all review the different types of gun with very low perveance, essentially those to be found in electron microscopes and other such devices; after this, we shall describe succinctly the methods of calculation which are used in connection with very high perveance guns of simple construction.

8.5.2 Electron Microscope Guns

For any given electrode system, the electrolytic tank provides the potential distribution \( \varphi(z) \) and we can thus calculate the optical properties of the gun—the emissive surface area of the cathode, the position of the crossover, and the aperture of the beam where it leaves the anode. Ellis (1947) has made a simplified analysis in which \( \varphi(z) \) is regarded as being composed of two straight lines, and integration of the equations of motion is thus straightforward. Jacob (1948, 1950a, 1950b) too has performed calculations for such a system in which \( \varphi(z) \) is described by a function of the form

\[
\varphi_1(z) = A_1 \sinh(kz)
\]

near the cathode, and by another function

\[
\varphi_2(z) = A_2 e^{kz}
\]

far from the cathode; again, the equations of motion can be integrated exactly.

A more realistic but more complicated theory has been worked out by Maloff and Epstein (1938) and by Plocke (1951, 1952). In these articles, the possible modifications which space charge could make to \( \varphi(z) \) are taken into account. The effect of space charge is not only to reduce the emissive zone of the cathode and of the electric field at the cathode, and to move the crossover further away from the cathode, but also to decrease the angular aperture of the beam. All these theoretical predictions agree well with the observations which have been made upon oscillographs. Review articles on cathode ray tube guns have been published by Morton (1946), Moss (1945), and Knoll and Thiele (1941).

Just as in an ordinary immersion objective, the potential distribution in the region near the cathode plays an overwhelming role; the following three parameters are therefore of great importance:

- the distance between cathode and control electrode, or grid, \( h \);
- the radius of the opening in the grid, \( R \), and its thickness;
- the polarization of the grid \( (-\sigma) \).

The emissive area of the cathode is restricted to the interior of the equi-
potential surface $\Phi = 0$. The trajectories which originate in this zone intersect the axis at a distance from the cathode which lengthens as the convergence is reduced; it is this "crossover" which is regarded as the effective electron source. For high voltage guns, the crossover lies between the grid and the anode. As the polarization $\sigma$ is reduced, the crossover approaches the anode, then leaves the gun and moves farther and farther away (the same effects occur if the cathode is brought nearer to the grid). This axial displacement is accompanied by a variation in the diameter—the further the crossover penetrates into the interior of the gun, the smaller is its diameter $d_c$. (If the real crossover lies between the grid and the anode, rays which emerge from the anode will seem to have come from a virtual crossover, namely, the image of the real crossover formed by the optical system which consists of the region between the real crossover and the anode. The values of $d_c$ which are measured refer to this virtual crossover.)

We can distinguish between two kinds of gun, according to the way in which this system is used:

"Short-focus" guns in which the crossover is formed internally, and is of very small diameter (3 to $5 \times 10^{-2}$ mm); these guns are used together with an additional (condenser) lens, which forms the image of this crossover on the preparation.

"Long-focus" guns which form the crossover externally, and are used without a condenser. By varying $\sigma$, the position of the crossover is varied in the neighbourhood of the specimen (placed at a distance from the anode which may vary between 5 and 15 cm). Even in the best conditions, however, $0.3 < d_c < 0.5$ mm; nevertheless, the aperture of the beam $\alpha$ is smaller than in the preceding case, as the Lagrange–Helmholtz relation shows.

At first sight, it may seem somewhat artificial to make this distinction between the two "kinds" of gun, as the same gun can be used in both cases. In reality, however, every electrode system has an optimum mode of operation, which gives the highest brightness, and if this optimum is to be achieved, the gun must be operated in the right conditions.

The vital feature of these guns is neither the intensity which is emitted, nor the current density, but the brightness, measured in some cross-section of the beam, the crossover in particular. The brightness is defined as the current per unit area per unit solid angle (see § 4.3.5). In theory, this quantity depends only upon the current density $j_0$ emitted at the cathode, the working temperature $T^\circ\text{K}$ and the accelerating potential $\Phi_0$, thus:

$$B_{th} = j_0 \frac{e}{\pi k T} \Phi_0 \text{ A cm}^{-2} \text{ sterad}^{-1}. \quad (8.42)$$

The current density at the crossover is given by

$$j_c(r) = j_c(0) \exp \left\{ - \left( \frac{r}{d_c} \right)^2 \right\}, \quad (8.43)$$
in which \( j_0(0) = B \pi \alpha^2 \) (\( \alpha \) being the natural semi-aperture of the beam), and \( r_c \) is the "natural half-width" of the bell-shaped curve above, which is related to the radius \( r_0 \) of the emissive surface of the cathode by the expression

\[
r_c = \frac{r_0}{\alpha} \sqrt{\frac{kT}{e\Phi_0}}.
\] (8.44)

To increase \( B_{th} \) when \( \Phi_0 \) is fixed, the only possibility is to search for cathodes with a high specific emission \( j_0 \), working at a low temperature. From this point of view, oxide-coated cathodes would be superior to tungsten ones, but their lifetime is too short in the vacuum which obtains in a microscope, which is relatively poor (10\(^{-4}\) to 10\(^{-6}\) mm Hg) and often dirty (containing oil or grease molecules, and water vapour).

It is better to use a tungsten filament bent into a "V", of which only the point, which is placed on the axis, emits in normal conditions. This point is then equivalent to a spherical cap of very small radius (\( \sim 0.1 \) mm) and the gun produces an axially symmetric beam. This is no longer true if the grid potential is too small; the two branches of the filament or the part behind the point emit electrons which are then "accepted" by the gun but which follow different trajectories. This gives the illusion that the main beam is broken up into several parts.

For a filament of about \( 1/10 \) mm diameter, the lifetime \( t \) is given effectively by

\[
t \text{ (hours)} = \frac{32}{j_0} \quad (j_0 \text{ in A cm}^{-2})
\] (8.45)

(Bloomer, 1957). For \( j_0 = 1 \) A cm\(^{-2}\), the working temperature is of the order of 2650°C and \( t \sim 32 \) hours. Under these conditions, the value of \( B_{th} \) is of the order of \( 5 \times 10^4 \) A cm\(^{-2}\) sterad\(^{-1}\). A number of authors have suggested that the filament should be replaced by a genuine point, with a radius of curvature of the order of a fraction of a micron, to which an electric field of the order of 10\(^7\) V cm\(^{-1}\) is applied. The current density may reach 10\(^8\) A cm\(^{-2}\), but this field emission is unstable and the lifetime of the source is very short if an extremely high vacuum is attained (10\(^{-7}\) mm Hg). Recent work has shown that it is possible to obtain sources of this type with a useful lifetime by employing mixed or "thermal-field" emission (thermionic emission, \( T \sim 2000^\circ \)K, reinforced by the Schottky effect. Maruse and Sakaki, 1958; Sakaki and Maruse, 1960; Swift, 1960; Drechsler, Cosslett and Nixon, 1958; and Pilod and Sonier, 1961). The brightness which is obtained is 1000 or 1500 times better than the brightness which would have been produced by simple thermionic emission (at \( T = 2000^\circ \)K, \( j_0 \) would be of the order of 2\( \times 10^{-3} \) A cm\(^{-2}\)) and about ten times better than we should obtain if a filament of the ordinary kind were used; \( B \sim 3 \times 10^6 \) A cm\(^{-2}\) sterad\(^{-1}\). The stability of the emission from a heated point has been studied by several workers (Elésin, 1958; Trolan and Dyke, 1958). The total current emitted by the point is only of the order
of 1 to 10 $\mu$A, as opposed to the 50 to 200 $\mu$A emitted by other types of filament.

When the crossover is real, and lies beyond the anode, we can obtain a source having practically the theoretical value of the brightness by stopping down the peripheral part of the crossover; if, on the contrary, we use the whole crossover as the source, the mean measured brightness falls to about 10 or 20 per cent of the theoretical brightness. To measure $B$, the following method may be followed. A magnified image of the crossover is formed on a screen containing a very small hole, with the aid of a lens $L$ of known magnification; a deflection system moves the image across the hole. By measuring the current which is collected in a Faraday cage, we can plot the curve $j_i(r)$ across a diameter of the image and hence calculate the distribution $j_c(r)$ over the crossover. The lens $L$ is then removed and the natural divergence of the beam $\alpha$ measured; with these data, $B$ can be calculated.

If $B$ is to be compared with $B_{th}$, the temperature $T$ must be known very accurately as must the emissivity $j_0$; the former is measured pyrometrically, while the latter is either taken from the tables or measured using experimental diodes. Experiment shows (Haine and Einstein, 1952) that in a triode gun with plane electrodes and a tungsten filament, designed for use with an internal crossover, the two important parameters in the adjustment are the cathode-grid distance $h$ and the (negative) grid potential ($-\sigma$). When this potential is reduced, starting at cut-off, the brightness increases rapidly as a result of the increase in the emissive area of the cathode, passes through a maximum, and finally diminishes very gradually, doubtless influenced by the aberrations of the system which spread out the crossover (Fig. 92 A2); $h$ is meanwhile held constant. For each value of $h$, there is thus an optimum potential $\sigma_{opt}$ which becomes greater as $h$ is reduced (see Fig. 92 A1 where the cathode is behind the grid).

If we always select the potential $\sigma_{opt}$, the brightness remains practically constant for a given temperature $T$ when $h$ is varied; there is even so a rather ill-defined optimum value of $h$, $h_{opt}$.

![Fig. 92 A1. Short-focus gun with plane electrodes (Haine and Einstein).]
When $\sigma$ is reduced from the cut-off value, and $h$ is kept fixed, $\alpha$ increases (since the emissive surface of the cathode increases); this increase is very rapid when $h$ is large, less so when $h$ is small. The diameter of the crossover, on the other hand, remains virtually constant (see Fig. 92 A2).

If $h$ is now fixed and $T$ increased, the value of $\sigma_{\text{opt}}$ decreases slightly; $B$ increases very rapidly with $T$, since $j_0$ varies exponentially with $T$. In this way, values of $B$ of the order of $10^4 \text{ A cm}^{-2} \text{ sterad}^{-1}$ have been measured on a gun with plane electrodes, at a temperature $T$ of $2700^\circ \text{K}$ and accelerating potential $\Phi_0 = 50 \text{ kV}$; this represents a value which is about 95 per cent of $B_{\text{th}}$. The aperture $\alpha$ is then of the order of 6 milliradians and $d_c \simeq 0.04 \text{ mm}$. The total current $I_c$ which leaves the cathode when $\sigma = \sigma_{\text{opt}}$ decreases steadily when $\sigma_{\text{opt}}$ is increased, when $h$ is increased and when the diameter of the opening in the grid is diminished. These are all factors which we have already seen to be important in the preceding section—they all have the effect of reducing the emissive area of the cathode. The theoretical brightness $B_{\text{th}}$ is independent of the geometry of the gun, but in practice the maximum brightness attainable, $B_{\text{max}}$ (which is close to $B_{\text{th}}$), is obtained for different values of the aperture $\alpha$ and the total current $I_c$ in each different type of gun. The best gun will be the one which yields $B_{\text{max}}$ with the smallest possible values of $\alpha$ and $I_c$.

Another gun with a conical grid has also been studied by Haine and

---

**Fig. 92 A2.** Curves showing the brightness $B$, the diameter $d_c$ and the angular aperture $\alpha_c$ of the crossover as functions of the polarization of the grid ($W$) for a given distance between the filament ($F$) and $W$, for the case of the gun illustrated in Fig. 92 A1 (after Haine, Einstein and Borcherds. 1958). $B$ varies with temperature, $T$ ($T_1 = 2,585^\circ \text{K}$; $T_2 = 2,650^\circ \text{K}$; $T_3 = 2,720^\circ \text{K}$ and $T_4 = 2,800^\circ \text{K}$), but $\alpha_c$ and $d_c$ are virtually independent of $T$. 
Einstein (1958; see Fig. 92 A3). The results obtained resemble those outlined above, but the values of \( h_{opt} \) and \( \sigma_{opt} \) are clearly different. We should also mention the work of Hillier and Baker (1945), Dosse (1940), Ellis (1947), von Borries (1948) and more recently of Boersch and Born (1958) whose results confirm those of Haine.

A great deal of effort has gone into the experimental determination of the forms which triode guns should take if the crossover is to be formed outside ("long-focus guns") and the brightness is to reach its maximum value. Bricka and Bruck (1948) were the first to find a satisfactory solution to this problem; their gun, shown schematically in Fig. 92 A4, has spherical electrodes and the crossover always lies beneath the hole in the anode. Bricka and Bruck quote a maximum measured current density of 25 mA cm\(^{-2}\) at the centre of the crossover which lies 150 mm away from the anode, as opposed to the 3 mA cm\(^{-2}\) which would be obtained with a gun with plane electrodes used in the same conditions. The aperture was \( 8 \times 10^{-3} \) and hence the brightness \( B \) of the order of 1 or \( 2 \times 10^4 \) A cm\(^{-2}\) sterad\(^{-1}\) which is about 20 per cent of the theoretical brightness.

Castaing (1951, 1960) has shown that it is possible to obtain a brightness close to the theoretical brightness with this gun, working under the same conditions as a near-focus gun, by stopping down the crossover \( (B = 5.8 \times 10^4 \) A cm\(^{-2}\) sterad\(^{-1}\) for \( j_0 = 2 \) A cm\(^{-2}\), \( T = 2700^\circ \text{K} \) and \( \Phi_0 = 30 \) kV or \( B \sim 0.7 B_{th} \)). The crossover is, however, very close to the anode.

---

Fig. 92 A4. The Bruck and Bricka long-focus gun (with spherical electrodes).
More recently, another solution has been proposed by Steigerwald (1949); this is the *Fernfokus*, shown schematically in Fig. 92 A5. In normal operation, the region near the point of the grid is divergent; the polarization is low and the current emitted is in consequence intense. The beam then converges in the grid–anode region, but the overall magnification is small; at distances between 15 and 50 cm from the anode, the crossover is of the order of 0.3 to 0.5 mm in diameter.

A thorough study of the gun has been made by Braucks (1958, 1959) in which the influence of all the geometrical parameters and the polarization upon the performance of the gun (spot diameter, aperture, and brightness) is analysed; Braucks also examines the effect of placing an additional electrode within the ordinary grid. With the ordinary three-electrode gun, very high current densities (~50 or 60 mA cm\(^{-2}\)) can be obtained in a crossover 23 cm from the anode with an aperture \(a\) of the order of \(3 \times 10^{-3}\) radians. \(B\) is then of the order of \(4 \times 10^4\) A cm\(^{-2}\) sterad\(^{-1}\) for \(\Phi_0 = 30\) kV, and the total current leaving the gun is 100 \(\mu\)A.

A long-focus gun based on the *Fernfokus* has been suggested by Bas (1954) and Bas and Gaydou (1959) which is designed to be used in an X-ray tube with a fine focus. The cathode is now the flat end of a tungsten cylinder about 0.5 mm in diameter, heated by lateral electron bombardment (Fig. 92 A6). The current density measured at the centre of the cross-
over, which lies 5 mm beneath the anode, leads to a maximum value of
the brightness of the order of 60 per cent of $B_{th}$. The difference between
measured and theoretical brightness is due to a broadening of the cross-
over; this in turn is a consequence of the crater which is produced at the
centre of the anode by the impacts of ions created in the residual gas and
accelerated in the sense opposite to that of the electrons. With this gun, a
crossover 0.2 mm in diameter can be obtained at a great distance (150 to
300 mm); further, this gun is capable of giving very large currents if the
diameter of the cathode is increased.

Other long-focus guns have been studied by Ehrenberg and Spear (1951)
and Spear (1952).

In conclusion, it is worth considering the effect of sending electrons
through a microscope with a high magnification in the opposite direction—
the linear magnification will be the inverse of its usual value, and we shall
obtain a spot of very small diameter, the minimum being simply the limit of
resolution in the same instrument when used in the normal way. The pro-
duction of these ultrafine spots, which are usually known as "electron
probes", has been studied in detail by Castaing (1951). They have such
important applications as shadow microscopy, scanning microscopy and
microanalysis. Castaing's probe instruments are described in § 21.5.

8.5.3 Electrostatic Guns with High Pervenance

When the beams of electrons (or ions) are very intense but have fairly
low energies, the role of the space charge becomes dominant within the
gun. In the beam itself, the particles are subjected to a transverse defocusing
force which tends to make them diverge rapidly. To obviate this defocusing,
a transverse field must be created between the cathode and the anode which
opposes the defocusing force; this can be done if the shape of the electrodes
which are placed around the beam is suitably chosen.

This shape can only be calculated rigorously in a very limited number
of cases, and even then, certain simplifying hypotheses must be made. We
shall outline the principle of this calculation rapidly.

![Diagram](image_url)

**Fig. 92 S1.** A thin beam, with plane symmetry. The cathode lies in the plane $z = 0$. 

Within the beam, the potential satisfies Poisson's equation:

\[ \nabla^2 \Phi = \frac{\rho}{\varepsilon_0} \quad \text{(R.M.K.S. units)}, \]  

(8.46)

in which \(\rho\) represents the charge density, which is assumed to be constant in the transverse cross-section; it is related to the total current \(I\), the longitudinal component \(v_z\) of the velocity and the cross-section \(S(z)\) by

\[ \rho = \frac{I}{S(z) v_z}. \]  

(8.47)

We consider first of all a system with translation symmetry, infinite in the Ox direction. The width of the cathode which is plane is \(2 Y_0\); the cathode lies in \(z = 0\) and is symmetrical with respect to the \(yOz\) plane. We denote the marginal trajectory of the beam by \(Y(z)\); this is the trajectory which emerges from the edge of the cathode, \(y = Y_0, z = 0\), on the side where the values of \(y\) are positive (see Fig. 92 S1).

The potential \(V(z, y)\) in the neighbourhood of the axis can, by virtue of the symmetry, be expanded in the form:

\[ V(z, y) = V_0(z) + \frac{1}{2} y^2 \left( \frac{\partial^2 V(z, y)}{\partial y^2} \right)_0, \]  

(8.48)

in which \(V_0(z)\) is the potential on \(Oz\).

To a first approximation, the motion of the particles is described by the two equations

\[ \frac{d^2 z}{dt^2} = \frac{e}{m} \frac{\partial V(z, y)}{\partial z} = \frac{e}{m} \frac{dV_0(z)}{dz}, \]  

(8.49a)  

\[ \frac{d^2 y}{dt^2} = \frac{e}{m} \frac{\partial V(z, y)}{\partial y} = \frac{e}{m} y \left( \frac{\partial^2 V(z, y)}{\partial y^2} \right)_0. \]  

(8.49b)

Equation (8.49a) is readily integrable, and we obtain

\[ v_z^2 = \frac{2e}{m} V_0(z). \]  

(8.50)

In the transverse direction, the equation is linear in \(y\): if \(y = Y(z)\) is a trajectory, therefore, \(y = c Y(z)\) is also a possible trajectory, with \(0 < c < 1\); the flow is laminar (this is true only on the assumption that the transverse emission velocities of the particles can be neglected).

On eliminating the time between equations (8.50) and (8.49b), we obtain

\[ \sqrt{V_0(z)} \frac{d}{dz} \left[ \sqrt{V_0(z)} \frac{dy}{dz} \right] = \frac{1}{2} y \left( \frac{\partial^2 V(z, y)}{\partial y^2} \right)_0. \]  

(8.51)

The right-hand side of equation (8.51) can be evaluated with the aid of equation (8.46). If \(I\) is the current emitted per unit length of the cathode in the \(Ox\) direction, we have
\[ q = \frac{I}{2Y(z)\sqrt{\frac{2e}{m}V_0(z)}} \quad (8.52) \]

and

\[ \left( \frac{\partial^2 V(z, y)}{\partial y^2} \right)_0 = \frac{q}{\varepsilon_0} - \frac{d^2 V_0(z)}{dz^2}. \quad (8.53) \]

Finally, therefore, on substituting from equation (8.53) into equation (8.51), we obtain

\[ \sqrt{V_0(z)} \frac{d}{dz} \left[ \sqrt{V_0(z)} \frac{dy}{dz} \right] + \frac{1}{2} V''_0(z)y = \frac{I}{4\varepsilon_0 \sqrt{\frac{2e}{m}V_0(z)}} \cdot \frac{y}{Y(z)}. \quad (8.54) \]

By hypothesis, however, \( Y(z) \) is a particular trajectory, a solution of (8.54). If \( y \) is replaced by \( Y(z) \), we obtain

\[ Y(z) V''_0(z) + Y'(z) V'_0(z) + 2 Y''(z) V_0(z) = \frac{I}{2\varepsilon_0 \sqrt{\frac{2e}{m}V_0(z)}}. \quad (8.55) \]

**First special case**

Here, we try to obtain a beam of constant width parallel to the axis, \( Y(z) = Y_0 \). When the emission at the cathode is space charge limited, the condition

\[ V'_0(z) = 0 \quad (8.56) \]

is obtained in \( z = 0 \). If we take the zero of potential at the cathode, equation (8.55) simplifies to

\[ V_0(z) = \left( \frac{9}{4} k_1 I \right)^{2/3} z^{4/3} = A_1 z^{4/3}. \quad (8.58) \]

On substituting from equations (8.58), (8.53), and (8.52) into (8.48) it becomes apparent that at the edge of the beam \( y = \pm Y_0 \), the distribution \( V(z, y) \) is such that

\[ \sqrt{V_0(z)} V''_0(z) = \frac{I}{2\varepsilon_0 Y_0 \sqrt{\frac{2e}{m}}} = k_1 I, \quad (8.57) \]

in which \( k_1 \) is a constant. The solution of this equation gives the axial potential in the form

\[ V(z, Y_0) = V_0(z), \]
which entails
\[ \frac{\partial V(z, Y_0)}{\partial y} = 0, \quad (8.59a) \]
and
\[ \frac{\partial^2 V(z, Y_0)}{\partial y^2} = 0. \quad (8.59b) \]
Along an arbitrary trajectory within the beam, \( y(z) = cY_0 \), we find the same law:
\[ V(z, cY_0) = V_0(z). \]
The equipotential surfaces are therefore planes perpendicular to \( Oz \).
At the boundaries, the potential within the beam and the external potential \( V_{\text{ext}} \) must coincide; \( V_{\text{ext}} \) satisfies the Laplace equation
\[ \frac{\partial^2 V_{\text{ext}}}{\partial z^2} + \frac{\partial^2 V_{\text{ext}}}{\partial y^2} = 0, \quad (8.60) \]
and must satisfy the condition:
\[ \frac{\partial^2 V_{\text{ext}}}{\partial y^2} = 0 \text{ at } y = Y_0, \]
to conform with equation (8.59a). In the upper half-plane, we select a new axis \( O_1z \) which coincides with the edge of the beam \( y = Y_0 \) (see Fig. 92 S2). From the general theory of potential (Durand, 1952), we know that both the real and the imaginary parts of any analytic function of a

![Fig. 92 S2. A flat beam with constant cross-section: the change of coordinates.](image)

complex variable, \( u = z + jy \), are possible solutions of the Laplace equation. If, in equation (8.58), \( z \) is replaced by \( z + jy \), a solution of the Laplace equation for which \( V = f(z) \) for \( y = 0 \) is \( V = \text{Re}[f(u)] \). We can thus obtain an expression for \( V(z, y) \) which satisfies equation (8.60) [see Pierce, 1949, in which we find \( V(z, y) = \text{Re}(A_1u^{4/3}) \)]; in polar coordinates \((r, \theta)\), we should have
\[ V_{\text{ext}}(r, \theta) = \left( \frac{9}{4} k_1 \sqrt{I} \right)^{2/3} r^{4/3} \cos \frac{4}{3} \theta \quad \text{for} \quad 0 \leq \theta \leq \frac{3\pi}{2}. \quad (8.61) \]
By writing $V_{\text{ext}} = \text{const}$, we can trace the equipotentials outside the beam. If in particular $V = 0$, we shall obtain a \textit{straight line}, defined by $\theta_0 = \frac{3\pi}{8} = 67^\circ 30'$. The appearance of the equipotentials which are obtained in this way is shown in Fig. 92 S3, in which the accelerating potential has been put arbitrarily equal to unity. In a real system, for an accelerating potential $V_0$, we should have

$$V_0(z) = V_0 \quad \text{at} \quad z_0 = \left(\frac{V_0}{A_1}\right)^{3/4}.$$ 

Alternatively by fixing $V_0$ and $z_0$, we can calculate the maximum current $I$ which leaves the cathode, and the perveance $P = I/V_0^{3/2}$ of the gun.

Certain of the equipotential surfaces are materialized in the form of metal electrodes. The cathode $K$ lies in the plane $z = 0$, and is $2Y_0$ in width. In general, the grid $W$ follows the equipotential $V = 0$ and is connected electrically to the cathode, but we may use a negative electrode which will be placed behind $K$. The anode and the equipotential $V = V_0$ coincide.

![Fig. 92 S3. Equipotentials outside a flat beam of constant cross-section, when the cathode emission is limited by space charge. The divergence which occurs during the passage through the anode is neglected.](image)

After passing through the anode, the beam enters a region free of electric field; the opening in the anode behaves like a divergent lens, the focal length of which we can calculate (see § 8.3.1), and further, the beam diverges under the action of its own space charge.

If we are to obtain a convergent beam at the exit of the gun, the shape of the electrodes must be altered.
Second special case: the convergent beam with a plane of symmetry

By using the solution of Langmuir and Blodgett (1923) for the linear flow of charge in the space between two concentric cylindrical electrodes when the conditions are space charge limited, the shapes of the electrodes of the gun can still be calculated by Pierce's method (1949), but only in an approximate fashion. Alternatively, we can set off from equation (8.54), in which we write

$$Y(z) \approx Y_0(1 - z/a)$$  \hspace{1cm} (8.62)

(the trajectory produced backwards intersects the axis again behind the anode at $z = a$).

Equation (8.54) then becomes

$$\frac{d}{dz} \left[ \left(1 - \frac{z}{a} \right) \frac{dV}{dz} \right] = k \frac{I}{Y_0 \sqrt{V}}$$  \hspace{1cm} (8.63)

(Wakerling and Helmholz, 1949), in which $I$ represents the total intensity emitted per unit length of the cathode in the $Oz$ direction.

We try to find a solution of the form

$$V(z) = A z^{4/3} F(z),$$  \hspace{1cm} (8.64)

which satisfies the two conditions imposed by the space charge limited emission in the plane $z = 0$, namely

$$V(z) = 0 \quad \text{and} \quad V'(z) = 0.$$

The function $F(z)$ is obtained in an approximate manner in series form; in the plane $z = 0$, $F(0) = 1$. If we revert to the coordinate system shown in Fig.92 S2, where $Oz$ is the edge of the beam, we find:
Fig. 92 S5. The case of a convergent beam with plane symmetry and straight edges. The anode corresponds to the equipotential \( \varphi_0 \); the real beam diverges after passing through this electrode.

\[
V(r, \theta) = A \left( r^{4/3} \cos \frac{4}{3} \theta + \frac{8}{15} r^{7/3} \cos \frac{7}{3} \theta + \cdots \right). \tag{8.65}
\]

The shape of the grid is obtained by putting \( V = 0 \). For \( r = 0 \), we see that as before, the electrode leaves the cathode at an angle of \( \frac{3\pi}{8} \) with the edge of the beam, but that this electrode is no longer plane. An approximate calculation is normally made using the two terms above, and a suitable form for the electrodes which will lead to the distribution of equation (8.64) is then sought with the aid of an electrolytic tank. The method used is as follows: the electrodes are built of thin metal sheets, and placed in a flat-bottomed tank. The edges of the beam are simulated by two insulating strips, so that the condition \( \partial V / \partial n = 0 \) at the edge of the beam is imposed (that is, the normal field is zero). The potential distribution along a strip is measured, outside the beam, and the grid and anode are then bent and shaped until the desired distribution is obtained.

The general arrangement of the tank in the case of an axially symmetric beam is illustrated in Fig. 92 S4.

The form of the equipotentials obtained in this way is illustrated in Fig. 92 S5.

**Third special case: the thin beam with arbitrary cross-section**

\( Y(z) \) is now an arbitrary function; it might be part of a hyperbola, for example (Wakerling and Helmholtz, 1949), of the form
Fig. 92 S6. A beam bounded by two hyperbolic arcs, with asymptotes $\Delta$ and $\Delta'$.

\[ Y(z) = (Y_0 - Y_m) \frac{\frac{x - x_m}{x_m}}{1 + \frac{Y_0 - Y_m}{x_m} \frac{x}{x_m} \cot \alpha}. \]  
(8.66)

The meanings of $Y_0$, $Y_m$, $x_m$ and $\alpha$ are shown in Fig. 92 S6.

The case $\alpha = \pi/2$ corresponds to a parabolic trajectory. Equation (8.54) is solved with the aid of an electronic computer, which furnishes the function $V_0(z)$. We can then determine the potential outside the beam (which is assumed to be thin enough for a series expansion of the potential to be adequately represented by terms up to those in $y^2$) by using the equation

\[ V(z, y) = V_0(z) - \frac{1}{2} y^2 \frac{d^2 V_0(z)}{dz^2} + \frac{1}{2} y \frac{k I}{\sqrt{V_0(z)}}, \]  
(8.67)

in which we write $V(z, y) = \text{constant}$. Again using an electronic computer, we solve for $y$ and thus obtain the equation $y(z)$ for each of the different equipotentials.

Figure 92 S7 is an example of such a distribution, in a particular case. We see that it is possible to obtain the desired beam with the aid of electrodes which are wholly outside the beam, and that it is even possible to slow the beam down after it has been accelerated; this can be extremely useful in certain research in which intense beams of electrons or ions are in use.

Fourth special case: the axially symmetrical beam

The problem now becomes even more complicated, as the equation which gives $V_0(z)$ is still less tractable than equation (8.54), save when the
cross-section of the beam remains constant. We find, in fact:

\[
R(z) \frac{d^2 V(z)}{dz'^2} + 2R'(z) \frac{dV(z)}{dz} + 4R''(z) V(z) = \frac{I}{2\pi \varepsilon_0 \sqrt{\frac{2e}{m} V_0(z)}} \cdot \frac{1}{R(z)}.
\]  

(8.68)

If \(R(z) = R_0\), we can again find a solution of the form

\[V(z) = A z^{\lambda/3}.\]  

(8.69)

To a first approximation, and when the transverse dimensions of the beam are very slight, the distribution on the edge of the beam resembles that of equation (8.69) and the shapes of the electrodes which are necessary can be determined with the aid of an electrolytic tank with a sloping bottom. The family of equipotentials which is obtained is represented in Fig. 92 S8.

The case of a convergent beam with a rectilinear border (emitted by a cathode in the shape of a spherical cup) has also been solved approximately, with the aid of the results of Langmuir and Blodgett (1924) on rectilinear particle flow between two concentric spheres. The shape of the electrodes is shown in Fig. 92 S9.

Equation (8.68) could obviously be solved with the aid of a computer by means of a method resembling that described in the preceding sections. In practice, however, axially symmetric beams are far more common than plane beams and a great deal of attention has thus been devoted to the former; both the approximate calculation of the electrode shape, and the experimental determination of this shape have been carried out, when a beam with given characteristics is to be produced.

Requirements which are frequently encountered are that the beam should converge as it leaves the anode, and pass through a point of minimum cross-section at a distance \(z_c\) from the anode as far from the latter as possible
without sacrificing the highest possible intensity. In this connection, we mention the calculations of Guénard (1945) and Huber (1949) and the experimental work of Helm, Spangenberg and Field (1947), Samuel (1945), Heil and Ebers (1950) and Müller (1956); the articles by Süsskind (1956), devoted to guns descended from the "Pierce gun", and by Ivey (1954) which is concerned with the focusing of intense beams, may also be consulted with profit.

We shall mention yet one further method, suggested by Charles and Septier (1948), which depends only upon the chart of the potential $\Phi(r, z)$ which is obtained in the electrolytic tank. It allows us to calculate the shape of the beam and the maximum intensity which the cathode can emit at a given potential when the gun is of arbitrary shape and is working in space

![Diagram](image-url)

**Fig. 92S8.** Theoretical chart of the equipotentials for a beam of circular cross-section, parallel to the axis.

![Diagram](image-url)

**Fig. 92S9.** An example of the electrodes with which we can produce a convergent axially symmetric beam which has straight edges between $C$ and $A$. The electrodes $C$ and $W$ are connected electrically to one another. The beam cross-section passes through a minimum in the opening in the anode, and subsequently diverges rapidly.
charge limited conditions \( \frac{\partial V}{\partial z} = 0 \) on \( K \). This method has in particular been used for klystron guns.

In conclusion, it should be stressed that calculation can provide only approximate solutions, which represent a first approximation to the shape of the electrodes. Further, none of the preceding calculations is valid unless the influence of the opening in the anode and above all, the transverse components of the particle velocities at the cathode, are neglected. Finally, experiment shows that electrode systems with forms as simple as those illustrated in Figs. 92 S5 and 92 S9, while always ensuring that the potential at the edges of the beam is distributed in a manner closely resembling equation (8.69), give very good results.
CHAPTER 9

MAGNETIC LENS MODELS

9.1 THE LONG SOLENOID

For a detailed study of the uniform magnetic field which is produced by a long solenoid, the reader is referred to the standard textbooks; a full description of the application of the solenoid in β-ray spectrographs is to be found in an article by Grivet (1950b, 1951a), together with an extensive bibliography. Here, we shall attempt no more than to summarize the basic formulae. In a rotating meridian plane (cf. § 5.1) the trajectory is of the form

\[ r = 2q \sin \alpha \sin \left( \frac{z}{2q \cos \alpha} \right), \quad (9.1) \]

in which \( \alpha \) is the angle between the initial velocity at the object point (which lies within the field) and the direction of the axis. \( Oz \) is an axis which is parallel to the field at the object point through which it passes. The distance between the object \( A \) and the image \( A' \) is given by

\[ AA' = 2\pi q. \quad (9.2) \]

Expanding \( r \) in powers of \( \alpha \) about \( A' \), the coefficients of spherical aberration \( C_s \) and chromatic aberration \( C_c \) prove to be equal, thus

\[ C_s = \pi q, \]
\[ C_c = \pi q. \quad (9.3) \]

The magnification \( G \) in the rotating meridian plane is \(-1\), but since this plane rotates through just 180° between the object and its image, the actual magnification is \(+1\): image and object are equal. The number of ampère-turns per unit length which corresponds to some given value of \( q \) is given by

\[ n I = \frac{5.366 \sqrt{\Phi (1 + 0.979 \times 10^{-6} \Phi)} \Phi^*}{2q} = \frac{5.366 \sqrt{\Phi^*}}{2q} \quad (A, V, m) \quad (9.4) \]

for electrons of energy \( \Phi \) eV; the relativistically corrected value of the accelerating potential for electrons is \( \Phi^* \), but for ions \( \Phi^* \approx \Phi \).
9.2 MODELS REPRESENTING SHORT SYSTEMS

In the majority of magnetic lenses, the axial field \( B(z) \) is bell-shaped in distribution. Step-by-step integration is feasible, but it is more attractive to try to find a simple model to represent this bell-shaped field, of such a kind that the trajectory equation is integrable in terms of tabulated functions.

9.2.1 The Glaser Model (1941b)

In a symmetric lens, the bell-shaped curve \( B(z) \) can be closely described by the function

\[
B(z) = \frac{B_e}{1 + (z/a)^2},
\]

(9.5)

in which \( B_e \) is the field at the centre of the lens, and \( a \) is the distance between the centre of the lens and the points at which the longitudinal component of the field has fallen to \( B_e/2 \).

If we write

\[
x = z/a, \quad y = r/a, \quad k^2 = \frac{eB_e^2 r^2}{8m_0 \Phi^*}, \quad p = \sqrt{1 + k^2},
\]

the trajectories can be expressed in terms of elementary functions. (For electrons, \( k = 1.48 \times 10^5 \frac{aB_e}{\sqrt{\Phi^*}} \) in m.k.s. units.) This can be seen as follows. The equation of the Gaussian trajectories is of the form

\[
\frac{d^2 y}{dx^2} + \frac{k^2 y}{(1 + x^2)^2} = 0
\]

(9.6a)

and after a new change of variable,

\[
x = \cot \phi \quad (\pi > \phi > 0),
\]

which gives

\[
dx = -\frac{d\phi}{\sin^2 \phi} \quad \text{and} \quad 1 + x^2 = \frac{1}{\sin^2 \phi},
\]

we obtain

\[
\frac{d^2 y}{d\phi^2} + 2 \cot \phi \frac{dy}{d\phi} + k^2 y = 0.
\]

(9.6b)

Figure 93a shows how the variables \( z \) and \( \phi \) can be visualized simply. We write

\[
y(\phi) = \frac{v(\phi)}{\sin \phi} = v(\phi) \sqrt{1 + x^2}
\]

and equation (9.6b) becomes

\[
v''(\phi) + p^2 v(\phi) = 0,
\]

(9.6c)
Fig. 93a. The Glaser model: definition of the variable $\Phi$; the classical cardinal elements ($F, H, f$) and the immersion elements ($G, P$ and $g$).

which gives the general solution

$$\frac{r}{a} = y = \frac{1}{\sin \phi} \left( C_1 \sin p\phi + C_2 \cos p\phi \right) = C_1 \frac{\sin p(\phi + C_2)}{\sin \phi}, \quad (9.6d)$$

in which

$$\phi = \cot^{-1}(z/a).$$

Let us now consider a trajectory $y_1(\phi)$, parallel to the axis for $z = +\infty$ ($\phi = 0$) and traversing the lens from right to left (Fig. 93a). It remains at a finite distance from the axis ($y = 1$ for $\phi = 0$), so that

$$C_2 = 0 \quad \text{and} \quad y = C_1 \frac{\sin p\phi}{\sin \phi} \quad \text{with} \quad C_1 = 1/p. \quad (9.7)$$

The slope of this trajectory at any point is given by

$$y' = \frac{dy}{d\phi} \frac{d\phi}{dx} = \frac{1}{p} \left( \sin p\phi \cos \phi - p \sin \phi \cos p\phi \right),$$

which gives $y' = 0$ for $\phi = 0$, as required. At $x = -\infty$ ($\phi = \pi$), we have

$$y'(-\infty) = \frac{\sin p\pi}{p}.$$ 

This trajectory intersects the axis at points where $\sin p\phi = 0$, that is where $\phi_n = n\pi/p$, in the range $0 < \phi < \pi$. For high values of $p$, and hence of the excitation parameter $k^2$, we can thus have several real foci. For $p \approx 2$ (or $k^2 \approx 3$), there is only a single focus; the lens is more commonly used in this range of excitations.
It is to be noticed that the focus defined in this way is “immersed” in the field (see § 4.3.4); it is not the “asymptotic” focus that we have used hitherto to ascertain the behaviour of the rays in a lens.

Glaser’s bell-shaped field was the first example known of “Newtonian” fields, fields which can lead to two sets of cardinal elements—“classical” cardinal elements and “immersion” elements—for both of which the relations for conjugacy between object and image remain valid.

*The classical cardinal points* (Fig. 93a). The asymptote to the trajectory of equation (9.7) is

\[ y = -\cos p\pi + x \frac{\sin p\pi}{p}, \]  

(9.8a)

and the usual lens constants (focal length, position of the focus and principal plane) for a symmetrical lens are, on the image side,

\[ \frac{f}{a} = \frac{p}{\sin p\pi}, \quad \frac{z_F}{a} = p \cot (p\pi), \quad \frac{z_H}{a} = p \cot \left( \frac{p\pi}{2} \right). \]  

(9.8b)

The image rotation between the object point at infinity and the image \( z_F \) is given by

\[ \theta_{oi} = \pi k. \]

These “asymptotic” elements are used whenever we wish to form a real image at very high magnification \((z_i \to \infty)\) with a virtual object; this is the situation at the projective lens in an optical system containing several lenses. In object space, the rays converge towards a point near the

---

**Fig. 93b.** Position of the immersion foci and principal planes (object and image) as a function of \( k^2 \).
classical object focus $F_0$ (Fig. 93a); the final magnification $G$ is given by

$$G = \frac{A'B'}{AB} = \frac{z_0}{f_0}.$$

It can be shown (see Glaser, 1952) that, under these conditions, the magnification passes through a maximum for $k^2 = 1.05$; the minimum focal length $f_{\text{omin}} = 1.5a$ and $z_{f0} = 0.8a$, so that the virtual object is then immersed a relatively short distance. This is the arrangement that is generally used in electron microscopy with real objects.

The immersion cardinal elements (Fig. 93b). Using equation (9.7) directly, with an immersed object, we find that on the image side,

$$\frac{g}{a} = \frac{1}{\sin \frac{\pi}{p}}, \quad \frac{z_0}{a} = \cot \frac{\pi}{p},$$

and the rotation is given by

$$\theta_{ot} = \pi k/p.$$ 

If we consider two particular trajectories,

$$u(\phi) = \frac{\sin p\phi}{\sin \phi} \quad \text{and} \quad v(\phi) = \frac{\cos p\phi}{\sin \phi}$$

and an object plane defined by $\phi = \phi_o$ with the corresponding image plane $\phi = \phi_i$, the magnification $G$ will be given by

$$G = \frac{u(\phi_i)}{u(\phi_o)} = \frac{v(\phi_i)}{v(\phi_o)}.$$

This implies

$$\sin p(\phi_i - \phi_o) = 0$$

or, for $n = 1$

$$\phi_i - \phi_o = \frac{\pi}{p}. \quad (9.9b)$$

(This is a property common to all Newtonian fields.)

The image will again be rotated through an angle $\theta = k\pi/p$ with respect to the object.

In Fig. 93b, the variation of the positions of the foci and the principal planes with $k^2$ is shown, and in Fig. 93c the behaviour of the reduced convergence $C = a/g$ is illustrated.

For $k^2 = 3(p = 2)$, the immersion foci coincide at the centre of the lens and we have $g = g_{\text{omin}} = a$. In this situation, we have (for electrons)

$$g_{\text{omin}} = 1.17 \times 10^{-5}\sqrt{\frac{\Phi^*}{B_e}}. \quad (9.10a)$$

($g$ in metres, $\Phi^*$ in volts and $B_e$ in tesla).
For a given value of the ratio $\sqrt{\Phi^* / B_c}$, we can derive the value of $a$ that yields the absolute minimum value of $g$. We regard $a$ as a function of $k$:

$$a = \sqrt{\frac{8m_0}{e}} \cdot \frac{\sqrt{\Phi^*}}{B_c} \cdot k,$$

and seek the minimum of

$$g(k) = \frac{a(k)}{\sin \frac{\pi}{\sqrt{1+k^2}}}. $$

We find that $g = g_{\text{min}}$ for $k^2 = 0.8$, which yields

$$a_{\text{min}} = 0.54 \frac{\sqrt{\Phi^*}}{B_c} \times 10^{-5},$$

(9.10b)

$$g_{\text{min}} = 0.84 \frac{\sqrt{\Phi^*}}{B_c} \times 10^{-5}$$

(in the same units as (9.10a)).

A magnetic lens differs from an electrostatic lens in that a real (non-magnetic) object can be placed near the immersion focus and hence a highly magnified image will be obtained. In electron microscopy, however, the object is “illuminated” with an electron beam from a gun, accelerated through a potential difference $\Phi$; these electrons will therefore experience the action of a large part of the field $B(z)$ before they reach the object. It will thus be difficult to ascertain the exact conditions of illumination at the object and this is why an external (or only slightly immersed) focus has, for so long, always been used. (In reality, all the real foci will be immersed with Glaser’s mathematical model, since the field extends from $-\infty$ to
Asymmetrical distributions. It is always possible to represent an asymmetrical bell-shaped curve by two functions of the foregoing type (Dosse, 1941):

\[
B(z) = \frac{B_c}{1 + (z/a_1)^2} \text{ for } z \leq 0; \quad z = a_1 \cot \phi; \quad p_1 = \sqrt{1 + k_1^2},
\]

\[
B(z) = \frac{B_c}{1 + (z/a_2)^2} \text{ for } z \geq 0; \quad z = a_2 \cot \phi; \quad p_2 = \sqrt{1 + k_2^2}.
\]

A trajectory is calculated in one of the regions \((z > 0, \text{ for example})\); from its characteristics \((y \text{ and } y')\) at \(z = 0\), we can determine the constants \(C_1\) and \(C_2\) that appear in the general expression for the same trajectory in the second field region \((z < 0)\). Object and image are related thus (Glaser, 1952):

\[
\tan \left\{ \frac{p_2}{a_2} \left( \frac{\pi}{2} - \phi_1 \right) \right\} = \frac{a_1}{a_2} \frac{p_2}{p_1} \tan \left\{ \frac{p_1}{a_1} \left( \frac{\pi}{2} - \phi_0 \right) \right\}.
\]

If we characterize the degree of asymmetry by the ratio \(q = a_1/a_2\), we can show that the cardinal elements do not change much as \(q\) is varied from 0 to infinity. We notice that the focal lengths on the object and image sides remain equal, but the foci are no longer symmetrical about \(z = 0\).

Aberration coefficients. Glaser and Lammel (1943) succeeded in putting the aberration coefficients of their lens in a form that could be calculated directly.

For a lens with very high magnification (object situated near the object focus), the spherical aberration coefficient (referred to the object plane) is given by the general expression

\[
\frac{C_s}{a} = \frac{1}{\sin^4 \phi_0} \left\{ \frac{\pi k^2}{4p^3} + \frac{1}{8} \frac{4k^2 - 3}{4k^2 + 3} (\sin 2\phi_1 - \sin 2\phi_0) \right\}, \quad (9.11)
\]

which becomes

\[
\left( \frac{C_s}{a} \right)_{\phi_0 = 0} = \left( \frac{\pi k^2}{4p^3} - \frac{1}{8} \frac{4k^2 - 3}{4k^2 + 3} \sin \frac{2\pi}{p} \right) \frac{1}{\sin^4 (\pi/p)}. \quad (9.12)
\]

When \(k^2\) is increased, \(C_s\) initially decreases steadily (see Fig. 93d); for \(k^2 = 3\), in particular, we have \(C_s = 0\cdot3a\). Between \(k^2 = 3\cdot5\) and \(k^2 = 7\) it remains virtually constant and then begins to increase very slowly. The minimum \((C_s = 0\cdot25a)\) occurs in a range of excitations that are too high to be used in practice.
If $\Phi^*$ and $B_c$ are fixed, we can vary $a$ to give $C_s$ its minimum value. We find that now

$$k^2_{\text{min}} = 2.8, \quad a_{\text{min}} = \frac{113\sqrt{\Phi^*}}{B_c}$$

and

$$(C_s)_{\text{min}} = 0.30a_{\text{min}}$$

($a$ in mm, $\Phi^*$ in volts and $B_c$ in gauss).

We see, therefore, that for an objective lens with $k^2 = 3$ and the minimum focal length ($f_{\text{min}} = a$), we shall have the focus at $z = 0$ and very nearly minimum aberration if we choose a lens for which $a$ is given by the foregoing formula.

The chromatic aberration constant is given by

$$\frac{C_c}{a} = \frac{\pi k^2}{2p^3} \frac{1}{\sin^2 \phi_0},$$

or, for high magnification (object at the focus),

$$\frac{C_c}{a} = \frac{\pi k^2}{2p^3} \frac{1}{\sin^2 (\pi/p)}.$$

$C_c$ passes through a minimum near $k^2 = 4$ and we have

$$\left(\frac{C_c}{a}\right)_{\text{min}} \approx 0.58.$$

If we try to extract the maximum magnification from the objective, we have for $k^2 = 3$ ($p = 2$)
MAGNETIC LENS MODELS

\[
\frac{C_c}{a} = 1.8 = 3\left(\frac{C_c}{a}\right)_{\text{min}}
\]

\[
\frac{C_s}{a} = 0.3.
\]

The dependence of \(C_s\) and \(C_c\) upon \(k^2\) is illustrated in Fig. 93d.

**Correction**

This Glaser model is capable of describing to within a few per cent the properties of magnetic lenses with a wide variety of pole-pieces, and under very diverse conditions of excitation; it is very well adapted to the case of iron-free coils, which may well be used in electron microscopes or diffraction devices working at very high voltages, now that superconductors with high critical field have become available (§9.5). There do exist, however, cases in which the experimental \(B(\varepsilon)\) curve differs markedly from the Glaser model, in particular when the iron is far from saturation. In such cases, however, the precision with which the model describes the actual situation can be considerably improved by choosing the half-width \(a\) of the “bell” suitably. Normally, the model curve is given the same half-width and height as the actual curve (measured, generally speaking) for the lens, and the agreement elsewhere is found to be sufficiently good. In this situation, on the other hand, a satisfactory representation of the real characteristic is not possible; we no longer attempt, therefore, to match the two curves, other than by making them of the same height, but rather choose a half-width, \(a_1 = ma\), for the model curve (\(a\) being the real half-width) in such a way that the error in one of the cardinal elements—the focal length \(f_0\) or \(f_1\) is usually selected—is minimized. Glaser (1950a) has given an exact solution of this problem, but often a less rigorous procedure, suggested by Grivet (1951b, 1952a) proves to be both considerably more simple and adequately precise. If \(T_1\) and \(T_2\) are the equations of the two bell-shaped curves, we simply write

\[
\int_{-\infty}^{\infty} (T_1^2 - T_2^2) \, dz = 0.
\]

(9.13)

**9.2.2 The Grivet–Lenz Model**

Glaser’s model curve is very satisfactory at the centre of the lens, less so at its two edges—in fact, the field falls off very rapidly, following, in general, an exponential law. This exponential law seems to be characteristic of the way in which the field decreases whenever this decrease is produced as a result of deliberate “screening”. Glaser’s curve, however, decreases very slowly, much too slowly, and it was this fact that led Grivet (1951b, 1952a, b) and Lenz (1951) to consider a model less unsatisfactory in this
respect. The new model which was suggested takes the form:

\[ B(z) = B_0 \text{sech}(z/b), \]  

(9.14)

in which

\[ b = 0.7593a, \]  

(9.15)

where \( a \) is the half-width of the "bell" at one-half of its maximum height.

The Gaussian rays

If \( b \) is chosen as the unit of length, and \( x = z/b, R = r/b \), the Gaussian ray equation is of the form

\[ R'' \cosh^2 x + h^2 R = 0, \]  

(9.16)

in which \( h = 0.7593k \), and \( k \) is the usual parameter of magnetic lenses defined by

\[ k^2 = \frac{e B^2}{8m\Phi_0} a^2. \]

A change of variable, \( u = \tanh x \), transforms this equation into the Legendre equation

\[ (1 - u^2) \frac{d^2 R}{du^2} - 2u \frac{d R}{du} + \nu(\nu + 1) R = 0, \]  

(9.17)

so that a general ray can be written in the form

\[ R = A P_\nu(\tanh x) + B P_\nu(-\tanh x), \]  

(9.18)

in which \( P_\nu(u) \) is the Legendre function of order \( \nu \), with \( \nu(\nu + 1) = h^2 \). Some values of this function, which has been tabulated by Schelkinoff, are reproduced in Table 4 (see pp. xx--xxi). The cases for which \( \nu \) is an integer, \( n \), are physically significant and mathematically special—the function \( P_\nu(u) \) goes over into the polynomial \( P_n(u) \), while \( P_n(-u) \) has to be replaced by \( Q_n(n) \), a Legendre function of the second kind.

The particular values \( \nu = n \) (integral)

Whether or not \( \nu \) be an integer, the ray \( r = P_\nu(\tanh x) \) is always a principal ray, parallel to the axis and unit distance away from it as \( x \to \infty \), since \( P_\nu(1) = 1 \) is always true, and \( P_\nu'(1) \) is finite; the relation \( u = \tanh x \) ensures that the asymptotic behaviour is acceptable. If, however, \( \nu \) takes some integral value \( n \), this principal ray emerges parallel to the axis for \( x \to -\infty \) as well, since \( P_n(-1) = 1 \) if \( n \) is an integer; in the classical sense of the word, the lens is afoveal. Inside the lens, the ray intersects the axis \( n \) times, giving \( n \) immersion foci, \( G_n \), symmetrically placed about the centre of the lens which itself is a focus if \( n \) is odd—if \( n = 1 \), it is the only focus, and in this case

\[ r = A \tanh x + B x \tanh x. \]
The values \( v = n \) for which the system is afocal separate the values of \( v \) which correspond to convergent systems \((0 < v < 1; \ 3 < v < 4; \ 5 < v < 6; \ldots)\) from those which correspond to divergent systems \((2 < v < 3; \ 4 < v < 5; \ldots)\). In actual instruments, multiple foci are always avoided, and we need only study the region \( 0 < v \leq 1 \), that is \( 0 < h^2 \leq 2 \).

**The classical cardinal elements**

These elements define the homographic correspondence between the two asymptotes of each ray; the expansion of \( P_r(u) \) in the neighbourhood of \( u = -1 \) is known, and after a certain amount of transformation, we obtain

\[
P_r(\tanh x) \rightarrow \frac{\sin \nu \pi}{\pi} [2x + 2\psi(v) + C] + \pi \cot \pi \nu,
\]

in which \( C = 0.5772 \) is Euler’s constant, and \( \psi(v) \) is the derivative of \( \log(v!) \).

For the position of the focus \( z_F \), the focal length \( f \), and the image rotation \( \varphi \) we eventually obtain

\[
\frac{z_F}{b} = \psi(v) + C + \frac{\pi}{2} \cot(\pi v), \quad (9.19)
\]

\[
\frac{f}{b} = \frac{\pi}{2\sin(v\pi)}, \quad (9.20)
\]

\[\varphi = \pi v.\]

When the lens is weak, the disparity between the results obtained with this model and those obtained with Glaser’s model is slight; it becomes important, however, when the convergence is strong, and the most powerfully convergent projective lens proves to be defined not by \( k = 1, f = 1.5a \) but by \( v = 1/2, k = 1.14, f = 1.19a \).

The function \( \psi(v) \) is well tabulated, in the tables of Jahnke and Emde in particular; it varies only between 0 and 1 when \( v \) itself varies from 0 to 1. The quantities \( z_F \) and \( f \) are plotted as functions of \( k \) in Fig. 94a.

**Immersion**

When either the object or the image (or both) lies inside the field, there is no longer a rigorous homographic correspondence between the two as there is in Glaser’s model field; cardinal elements valid for every object and image no longer exist, as can easily be verified for the simple case \( v = 1 \). The cardinal elements defined by Glaser—which we shall simply call “immersion”—are, however, useful to describe the connexion between the object and its image in the neighbourhood of one particular pair of conjugate points, namely a focus and infinity. The focus \( G \) lies at the point at which
\[ P_v(u) = 0, \quad u_v = \tanh x_G, \quad z_G = b x_G, \]  \hspace{1cm} (9.21)

in which \( u \) is one of the \( n \) real roots of \( P_v(u) = 0 \), and \( n \) is the integer nearest to and larger than \( v \); the root is unique if \( 0 < v \leq 1 \). The immersion focal length, \( g \), is given by

\[ \frac{b}{g} = \frac{P'_v(u_v)}{\cosh^2 x_G} = \frac{2\sin(v \pi)}{\pi P_v(-u_v)}, \]  \hspace{1cm} (9.22)

the second form being convenient when \( v \) is small. The more \( P_v(-u_v) \) differs from unity, the greater is the difference between \( g \) and its classical counterpart \( f \); for \( v \leq 0.3 \) (so that \( P_v \geq 0.98 \), the immersion is negligible, and for small \( v \), the immersion focus \( G \) tends towards the familiar focus \( F(x_F = x_G = 1/2v) \); see Fig. 94b. The image rotation varies with the position of the object, but between infinity and a focus it is given by

\[ \varphi = \left\{ \frac{\pi}{2} - \tan^{-1} (\sinh x_G) \right\} h. \]  \hspace{1cm} (9.23)

The value \( g_0 \) of \( g \) for which the focus \( G \) is situated at the centre of the lens characterizes an important type of microscope objective. The numerical value proves to be \( g_0 = 0.7593a \) for \( v = 1 \) and \( k^2 = 2 \), which is appreciably
different from the value furnished by the Glaser model, namely, \( g_0 = a \), \( k^2 = 3 \), and which is in fact closer to reality than the latter; in particular, the relation between ampère-turns \( n I \), half-width \( a \) (millimetres), and the field at the centre of the lens \( B_0 \) (kilogauss) is

\[
B_0 a = 5.27 n I, \tag{9.24}
\]

for the Grivet–Lenz model, while for Glaser's field the numerical coefficient is 4. Experimentally, the value of the coefficient ranges between 4.4 and 5.8.

Another point of agreement with reality is that \( g_0 \) is not the minimum value of \( g \); the cardinal elements of this model are tabulated numerically below.

### 9.2.3 Generalization of Glaser's Formulae

A number of authors have considered the properties of the more general type of bell-shaped field:

\[
B(z) = \frac{B_0}{[1 + (z/a)^2]^\nu}, \tag{9.25}
\]

for various values of \( \nu \) using numerical integration, (Glaser and Lenz, 1951; Lenz, 1950). Sturrock (1951b), on the other hand, has used a method due to Regenstreif to extend Glaser’s results to fields of the form
The Cardinal Elements in the Grivet–Lenz Model

<table>
<thead>
<tr>
<th>( v )</th>
<th>( k^2 )</th>
<th>( \frac{z_G}{a} )</th>
<th>( \frac{g}{a} )</th>
<th>( \frac{f}{a} )</th>
<th>( \frac{z_F}{a} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1908</td>
<td>3.787</td>
<td>3.86</td>
<td>3.850</td>
<td>3.787</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4163</td>
<td>1.861</td>
<td>2.025</td>
<td>2.029</td>
<td>1.860</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6765</td>
<td>1.190</td>
<td>1.450</td>
<td>1.474</td>
<td>1.176</td>
</tr>
<tr>
<td>0.4</td>
<td>0.9539</td>
<td>0.8302</td>
<td>1.180</td>
<td>1.254</td>
<td>0.779</td>
</tr>
<tr>
<td>0.5</td>
<td>1.301</td>
<td>0.5914</td>
<td>1.0276</td>
<td>1.193</td>
<td>0.466</td>
</tr>
<tr>
<td>0.6</td>
<td>1.659</td>
<td>0.4200</td>
<td>0.932</td>
<td>1.254</td>
<td>0.1464</td>
</tr>
<tr>
<td>0.7</td>
<td>2.065</td>
<td>0.2847</td>
<td>0.864</td>
<td>1.193</td>
<td>0.2700</td>
</tr>
<tr>
<td>0.8</td>
<td>2.496</td>
<td>0.1745</td>
<td>0.820</td>
<td>1.029</td>
<td>0.987</td>
</tr>
<tr>
<td>0.9</td>
<td>2.965</td>
<td>0.0810</td>
<td>0.786</td>
<td>0.860</td>
<td>2.962</td>
</tr>
<tr>
<td>0.95</td>
<td>3.211</td>
<td>–</td>
<td>–</td>
<td>7.525</td>
<td>9.796</td>
</tr>
<tr>
<td>1</td>
<td>3.469</td>
<td>0</td>
<td>0.7593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>4.007</td>
<td>0.0717</td>
<td>0.739</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>4.579</td>
<td>0.1359</td>
<td>0.722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>5.186</td>
<td>0.1933</td>
<td>0.708</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>5.828</td>
<td>0.2477</td>
<td>0.697</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>6.504</td>
<td>0.2968</td>
<td>0.688</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10.407</td>
<td>0.500</td>
<td>0.658</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ B(z) = \frac{B_0}{1 - (z/a)^2}, \]  

(9.26)

\[ B(z) = \frac{B_0}{(z/a)^2 - 1}, \]

or, by writing \( a^* = j a \), to

\[ B(z) = \frac{B_0}{1 + (z/a^*)^2}. \]  

(9.26')

\( B \) is clearly infinite at \( z = \pm a \).

For \(|z| < a\), we replace the quantities \( u, a, k, p \) of § 9.2.1 by

\[ \frac{\pi}{2} + j \varphi, \ j a, \ j k, \ \omega \sqrt{k^2 - 1}, \]

and finally obtain

\[ y = \text{sech}(C_1 \sin \varphi + C_2 \cos \varphi), \]

\[ z = a \tanh \omega \varphi. \]

The image is defined by

\[ \varphi_1 - \varphi_0 = \frac{\pi}{\omega}, \]  

(9.27)
so that

\[-z_{f_0} = z_{f_i} = a \coth \frac{\pi}{\omega},\]

\[-f_0 = f_i = a \cosech \frac{\pi}{\omega},\]

\[-z_{H_0} = z_{H_i} = a \coth \frac{\pi}{2\omega},\]

and the principal planes, which are not crossed, lie outside the lens. The spherical and chromatic aberration coefficients are given by

\[C_s = a \left[ \frac{\pi k^2}{4\omega^3} + \frac{1}{8} \frac{4k^2 + 3}{4k^2 - 3} (\sinh 2\varphi_i - \sinh 2\varphi_0) \right] \sech^4 \varphi_0, \quad (9.28)\]

\[C_c = a \frac{\pi k^2}{2\omega^3} \sech^2 \varphi_0. \quad (9.28')\]

The β-ray spectrograph of Siegbahn and Släta (1949) is an example of a case where \(\omega\) is small.

If \(|z| > a\), however, the Glaser symbols \(u, a, B_0, k\) and \(\mu\) are replaced by \(j\varphi, j\alpha, -B_0, -jk\) and \(\omega = \sqrt{1 - k^2}\) respectively, and we find

\[y = \cosech\varphi [C_1 \sin \omega \varphi + C_2 \cos \omega \varphi],\]

\[z = a \coth \varphi.\]

Since the image is defined by

\[\varphi_0 - \varphi_i = \pi/\omega,\]

the cardinal elements are

\[z_{f_0} = -z_{f_i} = a \coth \pi/\omega, \quad (9.30)\]

\[z_{H_0} = -z_{H_i} = a \coth \pi/2\omega,\]

\[f_0 = -f_i = a \cosech \pi/\omega,\]

and

\[C_s = a \left[ -\frac{\pi k^2}{4\omega^3} + \frac{1}{8} \frac{4k^2 + 3}{4k^2 - 3} (\sinh 2\varphi_0 - \sinh 2\varphi_i) \right] \cosech^4 \varphi_0, \quad (9.31)\]

\[C_c = a \frac{\pi k^2}{2\omega^3} \cosech^2 \varphi_0. \quad (9.32)\]

The case for which \(\omega\) is small corresponds to a microscope objective.

9.3 THE UNSHIELDED SHORT COIL

9.3.1 The Field

Lyle (1902) made a detailed study of the fields around short toroidal coils, and gave very simple formulae for the fields along their axes.
Torus of square cross-section

The distribution of the longitudinal component $B_z$ of the magnetic field along the $z$-axis is only very slightly different from that which is produced by a single turn carrying the same current $i$, lying coaxially in the symmetry plane of the coil. Its radius $a_2$ is related to the mean radius $a_1$ of the coil by the expression

$$a_2 = a_1 \left(1 + \frac{b^2}{24a_1^2}\right),$$

(9.33)

in which $b$ is the length of a side of the square cross-section (Fig. 95); $a_2$ will be referred to as the mean corrected radius.

Torus of rectangular cross-section

We shall assume that the radial thickness of the coil $b$ is greater than its axial extent $c$; the field component $B_z$ is then the same as that which would be produced by two circular coplanar turns each carrying a current $\frac{i}{2}$, of radii $a_2 + e$, $a_2 - e$, where

$$a_2 = a_1 \left[1 + \frac{c^2}{24a_1^2}\right] \quad \text{and} \quad e^2 = \frac{b^2 - c^2}{12}.$$  

(9.34)

The formulae applicable to a number of other possible cases are to be found in the article by Lyle already mentioned; there seems little point in listing them here, as the results given above are the only ones as yet to have found practical application. Readers who are interested in more rigorous methods of calculating the induction, $B(z)$, produced by coils of rectangular cross-section should consult the article by Garrett (1951).
9.3.2 Adjustment of Glaser's Model

The field on the axis of a circular current-carrying loop of radius \( r_a \) is given by

\[
B(z) = \frac{\mu_0 i}{2r_a} \left( 1 + \left( \frac{z}{r_a} \right)^2 \right)^{1/2},
\]

which is not in a very convenient form for calculating the optical properties. We replace this exact expression, therefore, by a Glaser bell-shaped field, and optimize the value of \( a \) by Grivet's procedure.

(i) Square cross-section

Starting from

\[
B_c = \frac{\mu_0 i}{2r_a} \quad \text{(weber m}\textsuperscript{-2}, \text{At, m}),
\]

\( a \) is adjusted to satisfy

\[
\rho_z = 1.333a.
\]

(ii) Rectangular cross-section

With

\[
B_c = \frac{\mu_0 i}{2r_a} \left( 1 - \frac{e^2}{r_a^2} \right)^{-1},
\]

\( a \) is given by

\[
\rho_z = 1.333a \left[ 1 + 3.84 \left( \frac{e}{r_a} \right)^2 \right].
\]

This expression is partially empirical, as after arranging that the bell-shaped fields are equally wide, the coefficient 1.3 is replaced by 1.33, so that the expression reduces to the value chosen in the previous paragraph when \( e = 0 \).

An old formula of Ruska (1933) is still useful to give a fist approximation:

\[
\frac{d}{f} = \frac{1}{(220)^2} \frac{(n I)^2}{\Phi_0}, \quad n I = 220 \sqrt{\Phi_0 \frac{d}{f}} \quad (9.36)
\]

\((n I)\) is the number of ampère-turns on the coil, \( \Phi_0 \text{ kV} \) is the accelerating potential, \( f \) is the focal length, and \( d \) is the mean diameter of the coil. Experimental data are to be found in profusion in studies of the cathode ray tube.
9.4 SHIELDED COILS

When a coil is shielded save for a narrow slit $s$ around the bore, the field on the axis is sharply concentrated, and much higher for any given number of ampère-turns than its unshielded counterpart. Iron-clad coils are, in this sense, more economical than unshielded coils.

Ruska's formula still gives an order of magnitude estimate of the focal length, when the gap between the iron is narrow, provided $d$ is regarded as the interior diameter of the opening.

9.4.1 The Axial Field

A precise theory can be established by assimilating the magnetic circuit into the simple model shown in Fig. 96 studied experimentally by van Ments and Le Poole (1947). We shall consider only the case in which the permeability of the iron is sufficiently large throughout the pole pieces to be legitimately regarded as infinite. Ampère's circuitual theorem expresses the difference between the scalar magnetic potential at the two poles, $\Psi_2 - \Psi_1$, in terms of the ampère-turns $nI$, thus:

$$\Psi_2 - \Psi_1 = nI = \frac{1}{\mu_0} \int B_z \, dr.$$  

The formula established by Bertram (1940, 1942)

$$B = \frac{\mu_0 n I}{2R} \cdot \frac{\sinh \omega \varepsilon}{\varepsilon \cosh \omega (z + \frac{\varepsilon}{2}) \cosh \omega (z - \frac{\varepsilon}{2})},$$  

(9.37)

can now be applied with $\varepsilon = s/R$, $s$ being the size of the gap and $\omega = 1.318$.

Fig. 96. Cross-section of the pole-pieces of a shielded magnetic lens.
Introducing the maximum value $B_m$ of the axial field, which obtains at the centre of the lens, we find

$$B = B_m \frac{\cosh^2 \frac{\omega \varepsilon}{2}}{\cosh \omega \left(z + \frac{\varepsilon}{2}\right) \cosh \omega \left(z - \frac{\varepsilon}{2}\right)}$$  \hspace{1cm} (9.38)

where $B_m$ is given by

$$B_m = \frac{\mu_0 n I}{R \varepsilon} \tanh \frac{\omega \varepsilon}{2}.$$  \hspace{1cm} (9.39)

Comparing $B_m$ with the value of the field in the gap, $B_0$, which is given by

$$B_0 s = \mu_0 n I,$$

we find

$$\frac{B_m}{B_0} = \tanh \frac{\omega \varepsilon}{2}.$$  \hspace{1cm} (9.40)

For the half-width $a$ of the real curve, at the point at which it has fallen to half its maximum height, we obtain

$$\cosh \left(2 \omega \frac{a}{R}\right) = 2 + \cosh \left(\omega \frac{s}{R}\right).$$  \hspace{1cm} (9.41)

The quantities $a/R$ and $B_m/B_0$ are plotted in Fig. 97, and prove to be in excellent agreement with the experimental curve obtained by Le Poole.

![Fig. 97. The magnetic characteristics of a shielded lens—a comparison between theory and experiment.](image-url)
The agreement is very good, too, with the curves calculated precisely by Liebmann (1951), Liebmann and Grad (1951) and Lenz (1950), using the relaxation method.

9.4.2 Adapting the Model

Adapting Glaser’s model by Grivet’s procedure, to calculate the cardinal elements, we find

\[
\frac{a'}{R} = \frac{1}{2} \left( \coth \frac{\omega \varepsilon}{2} \right)^2 \left( \frac{\omega \varepsilon}{\tanh \omega \varepsilon} - 1 \right).
\]  (9.42)

That \( a' \) and \( a \) prove, on evaluation, to be almost identical shows that the model is a good one (for \( s = 0 \), for example, \( a_0 = 0.667 \) while \( a'_0 = 0.669 \)). The formulae of §9.2.1 yield values which approximate closely to the values of \( f, z_F, g, \) and \( z_G \) which are observed in practice.

In particular, if we follow Le Poole and write

\[
K = \frac{(n I)^2}{\Phi}
\]

(in which \( n I \) is the number of ampère-turns, and \( \Phi \) is the accelerating potential of the electrons), we find that

\[
K = 28.796 \left( \frac{\varepsilon}{\tanh \frac{\omega \varepsilon}{2}} \right)^2 \frac{R^2}{a^2} k^2,
\]  (9.43)

which agrees very well with experiment. This formula also gives a very satisfactory value \( K_{\text{min}} \) of \( K \), at which a projective lens has its minimum focal length, namely \( K_{\text{min}} = 1.3 \); more precision can be attained by distinguishing between \( a' \) of the model and \( a \) of the actual lens, in which case we find

\[
K_{\text{min}} = 115.184 \varepsilon^2 \tan \omega \varepsilon^2 \frac{1.3}{2 (\omega \varepsilon \coth \omega \varepsilon - 1)^2}.
\]  (9.44)

The predicted value \( f_{\text{min}} \) of \( f \) is also accurate, and if we make the parallel calculation for the immersion focus, determining the value \( K_0 \) of \( K \) for which \( G \) lies at the centre of the lens and the associated focal length \( g_0 \), the agreement with experiment is again close; a comparison of the following table with Le Poole’s curves shows this and suggests that the greatest disparity appears when the induction is largest.

In the following table, a few numerical values, calculated with the aid of the preceding formulae, are listed; together with \( K/k^2 \), the values \( K_0 \) of \( K \) for which the immersion focus \( G \) lies at the centre of the lens are tabulated, as are the corresponding focal lengths \( g_0 \); likewise, the values \( K_{\text{min}} \) of \( K \) for which the ordinary focal length \( f \) is a minimum, \( f_{\text{min}} \).
9.4.3 Magnetic Microscope Lenses

These lenses have been extensively studied; the basic problem is to design a lens with a very short focal length, of the order of millimetres, so that the microscope is as compact as possible. The objective is the most difficult, as it must also satisfy the further conditions of minimum spherical aberration.

![Diagram of magnetic lens models](image)

**Fig. 98.** Cross-sections of the pole-pieces of the various magnetic lenses built by Ruska, together with the field distributions on the axes of the systems.
and ellipticity astigmatism. Another difference between the design of objective and projective lenses stems from the fact that the former must leave enough room for the specimen-holder, which means that the focus must not lie too near the centre of the objective if the advantage of working at the minimum focal length is not to be sacrificed. The focal length $f$, therefore, must be 8–10 mm rather than 3 mm, and it is often useful to add another lens to the microscope in order to keep its overall length within acceptable bounds. In practice, the specimen-holder is often so cumbersome that the focal length of the objective has to be longer than that of the projective (3 mm instead of 1.5, for example). At the projective lens, however, only distortion need be considered, as the beam is very fine.

The pole-pieces in both kinds of lens may be made of a cobalt alloy with a very high saturation induction (26,000 gauss), in order to produce a very high field density in a restricted volume.

With the aid of Glaser's model, the aberrations can be calculated (Dosse, 1941c, d; Glaser and Lammel, 1943)—the results have been verified experimentally for the cases of spherical aberration and distortion. This latter aberration, though, can be eliminated as Hillier (1946) has demonstrated. Mechanical defects and imperfections in the magnetic homogeneity of the objective can also be corrected (Hillier and Ramberg, 1947), and in consequence, a resolution of better than 1 mμ is now attainable.

Saturation of the field implies that the field distribution along the axis is flatter than a Glaser bell-shaped curve, and this effect becomes the more pronounced as the opening in the pole-pieces is reduced or, equivalently,
Fig. 100. The focal length as a function of the accelerating potential of the electrons, in the various objectives studied by Ruska.

Fig. 101. The focal length as a function of the maximum axial field, for two of Ruska's objectives.
as \( f \) is shortened. The effect of saturation may become important when the magnetizing current is high, but it does not alter the general behaviour of the lens; only the numerical values of the lens constants are modified, as the simple theory outlined above is no longer adequate. In particular, if we study the variation of the focal lengths \( f \) and \( g \) with the parameter \( k^2 \), defined by \( k^2 = \frac{e a^2 B^2}{8 m \Phi_0} \), we find that both \( f \) and \( g \) reach a minimum. In this region, the lens is insensitive to small variations of \( k^2 \) produced by the magnetizing current, perhaps, or by the high tension at the electron gun; it is, in other words, achromatic. Although the effect is more marked for \( f \) (the projective lens) than for \( g \) (the objective), it is perceptible and can be made use of even in the latter case. For the corresponding value of \( k^2 \), the pole-pieces are nearly saturated, and the effects of magnetic heterogeneities less pronounced. For these reasons, these are the operating conditions under which objectives are designed to work. In Fig. 98, the various stages in the construction of the magnetic objective designed by Ruska (1944) are shown; the values of the maximum field are plotted as a function of the number of magnetizing ampère-turns in Fig. 99, while Figs. 100 and 101 summarize the optical properties.

Only two parameters are required to characterize the behaviour of the symmetrical projective lens shown in Fig. 102: the angle of the cone and the ratio of the width of the gap to the diameter of the part which is bored away. The cone angle, always of the order of 70° or 75°, is of very little use as a means of reducing \( f \); in order to make the instrument less bulky. The optimum choice for the width of the gap depends upon the velocity of the electrons, the diameter of the bore and the magnetizing current. Figure 102 shows the variation of \( B(z) \) in a typical lens, while the optical properties are given by the theory of § 9.2.

**Fig. 102a.** Cross-section of the pole-pieces of a magnetic projective lens, and the axial field distribution.
The cardinal elements prove to be described very satisfactorily by the formulae which we have obtained earlier, when compared with the experimental results obtained by Liebmann and Le Poole.

Figure 103, in conclusion, shows a cross-section of a complete magnetic lens.

The very thorough studies which have been made by Durandeau and Fert (1957) are to be found summarized in § 17.4. They comprise a complete account of the properties—magnetic as well as optical—of these lenses; with their aid, any desired type of objective, projective or condenser lens can easily be constructed.

9.5 SUPERCONDUCTING COILS

For some years now, superconducting metal compounds have been known that retain their superconducting property in the presence of very high magnetic fields. The value of the magnetic field which, at a given temperature, destroys this zero-resistance state \((\rho = 0)\) and introduces an ohmic resistance is known as the "second critical field", \(H_{c2}\).

The "first critical field", \(H_{c1}\), is the value of the field at which the lines of force begin to penetrate the superconductor; for \(H < H_{c1}\), there is a "total Meissner effect" and no lines of force penetrate the superconduc-
tor, which behaves as a perfectly diamagnetic material \((B = 0\) inside the metal). For \(H_{c1} < H < H_{c2}\), the magnetic field does penetrate the superconductor, but this does not destroy the "superconductivity" and the resistance of the metal remains zero. For \(H_{c2} < H\), the resistance reappears, the superconductivity is destroyed and the metal (or alloy) behaves like an ordinary conductor with a normal Joule effect. It ceases to be suitable for the application we are considering.

The following table gives the values of the critical temperature \(T_c\) \((\rho = 0\) if \(T < T_c\)) and critical field \(H_{c2}\) for the three materials that are at present manufactured commercially.

<table>
<thead>
<tr>
<th></th>
<th>(T_c (^\circ \text{K}))</th>
<th>(B_c (\text{kG}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alloys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nb Zr</td>
<td>12</td>
<td>60–70</td>
</tr>
<tr>
<td>Nb Ti</td>
<td>13</td>
<td>100–150</td>
</tr>
<tr>
<td>Compound</td>
<td>Nb3 Sn</td>
<td>18</td>
</tr>
</tbody>
</table>

Figure 103 shows the variations of the maximum current density that can be obtained in the superconducting wire or ribbon. (In fact, if the coil is to behave in a stable fashion, the superconducting material must be covered with a thick layer of copper or aluminium and the current densities shown in Fig. 103 must then be roughly halved for small coils and divided by about 10 or 20 for very big coils (Donadiou, 1965).

We see that by using these materials, it will be possible to reduce the cross-section of the magnetizing coil of ordinary microscope lenses considerably, but their main virtue arises from the possibility of obtaining very high fields with iron-free coils, unlimited by saturation phenomena. We can always reduce the longitudinal extent of the curve \(B(z)\), (i) by using an outer screen of cobalt steel (which will saturate only close to the coil) or (ii) by introducing cylindrical superconducting tubes close to the axis. (A superconducting substance provides a perfect magnetic screen so long as the magnetic field does not exceed \(H_{c1}\).) Finally (iii) we might use the new materials that saturate only at high fields and which are ferromagnetic at liquid helium temperature (these are the rare earth metals).

The use of these lenses involves us in delicate technological problems: the coils are permanently immersed in a liquid helium bath \((T_c = 4·2^\circ \text{K})\), but offers new opportunities for focusing beams of very high energy particles (electrons between 1 and 5 MeV, for example, see Chapter 17).

With the aid of formulae (9.33) or (9.34), or the formula below, we can calculate the field \(B_c\) at the centre of a coil of length \(2b\), inner radius \(a_1\) and outer radius \(a_2\), in which a mean current density \(j (\text{A/m}^2)\) is flowing:
Fig. 103. Curves showing the maximum current density as a function of the induction $B$ for various superconductors.

$$B_c = \mu_0 j b \log \frac{a_0}{a_1} + \sqrt{\left(\frac{a_0}{a_1}\right)^2 + \left(\frac{b}{a_1}\right)^2}.$$  (9.45)

With a coil made of niobium–tin ribbon 12 mm wide and 0.03 mm thick carrying 500 amperes, a field of 50 kG could be created without difficulty, with a half-width $a$ of the bell-shaped field equal to 6 mm; treating the distribution as a Glaser bell-shaped curve, we could then obtain the following values of the various parameters:

<table>
<thead>
<tr>
<th>$\Phi$ (MV)</th>
<th>$\Phi^*$ (MV)</th>
<th>$k^2$</th>
<th>$g$ (mm)</th>
<th>$C_s$ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>5.8</td>
<td>6.3</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>2.9</td>
<td>6.5</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>1.15</td>
<td>7.5</td>
<td>3.6</td>
</tr>
</tbody>
</table>
With particle energies of a few hundred keV, and $a < 2$ mm, we could achieve short focal lengths and aberration constants $C_2$ of a few tenths of a millimetre.

Nevertheless, great precautions will have to be taken in constructing the coils, to ensure that the field $B(z)$ has perfect rotational symmetry. The wires will need to be positioned with the utmost care, and deformation of the conductors during cooling and under the action of the intense field will have to be avoided.†

†Work on microscope lenses (by Fernández-Morán, among others) will be described in Chapter 17.
CHAPTER 10

STRONG FOCUSING LENSES

10.1 THE STRONG FOCUSING PRINCIPLE

10.1.1 The Inability of Ordinary Lenses to Focus High Energy Particles

During the last ten years, the range of applications of corpuscular optics has widened considerably, and with the construction of large particle accelerators a number of fresh problems have appeared; the basic problem is to focus high energy ions or electrons, or even just roughly to channel them over large distances.

In an electrostatic lens, the convergence diminishes rapidly when the accelerating potential $\Phi_0$ is increased, then passes through a minimum, and subsequently increases regularly, as we have seen in § 4.4; for $\Phi_0 > 5 \times 10^8$ V, the convergence is comparable with its value at low energies. This, however, requires the ratio $k = \frac{\Phi_1}{\Phi_0}$ to be held constant ($\Phi_1$ represents the potential applied to the lens), and although it may be easy to speak of holding $k$ constant in theory, in practice such an operation is not possible; breakdown phenomena, which affect the optical properties adversely, restrict $\Phi_1$ absolutely to about 100 kV. Further, for a given value of $\Phi_1$, the convergence of a lens varies effectively as $(\Phi_1/\Phi_0)^2$. Electrostatic lenses formed either from tubes or from diaphragms are therefore out of the question where the focusing of particles with energies higher than 1 MeV is concerned ($\Phi_0 = 10^6$ V).

The domain in which electrostatic lenses can be used is somewhat extended by using grid lenses (§ 8.4), the convergence of which varies as $(\Phi_1/\Phi_0)$; a certain number of particles are intercepted by the grid, however, and the worst drawback is not the loss of current but the secondary emission which is provoked by the impacts of the primary particles at the wires of the grid and which can be extremely troublesome.

Magnetic lenses too are inadequate at high energies. Their convergence falls steadily when $\Phi_0$ is increased, and further, it is not at present possible to increase the induction on the axis, $B_z$, beyond the limiting value of 2·5 or 2·6 tesla which is imposed by saturation of the iron in the magnetic circuit. Another consideration is that even for fairly low energies of a few tens or hundreds of keV, these lenses focus beams of heavy ions inefficiently.
These difficulties have been surmounted in a few special cases by using lenses without iron; very short current pulses, several hundreds or even thousands of ampères in amplitude, are fed into specially constructed coils and an induction which may reach several hundred kilogauss is created. A pulsed supply of this kind is absolutely vital if excessive Joule heating is to be avoided. In this way, it has been possible to focus intense beams of protons with energies of a few hundred keV; here, the principal cause of the divergence is the mutual repulsion of the ions due to space charge.

We may conclude, therefore, that the normal types of lens which have been described earlier can no longer be used with electrons when their energy is more than a few hundred MeV, nor with protons beyond a few hundred keV. The reason for this is to be found in an analysis of the convergence mechanism itself; both in electrostatic and magnetic lenses, all the energy which is expended goes into creating a principal field (either electric or magnetic) parallel to the direction of motion of the particles, and therefore without direct focusing action.

We consider, first of all, the case of the unipotential three-electrode lens—the particles are slowed down in the first half of the lens, and accelerated in the second. The variation of $E_z$ in the $Oz$ direction gives rise to a radial field $E_r$ which produces defocusing at the two extremities and focusing in the central zone where the particles move most slowly. The final convergence is thus a differential effect.

In a magnetic lens, the radial field $B_r$ which exists at the point where the particles enter the lens endows the particles, which had been travelling parallel to the axis, with a transverse velocity $v_y$, and it is only when this deviation takes effect that the particles are usefully affected by the principal field $B_z$. The action of $B_r$ is of the opposite sign at the exit, however, so that here again the overall effect is a secondary effect, proportional to $B^3$.

A field which is perpendicular to the direction of motion (the field in the deflector of an oscillograph or in a magnetic prism, for example) will, on the contrary, act directly upon the convergence, as the force exerted upon the particles is proportional to the field strength $B$. If the system is to have the properties of a lens—if, namely, it is to be capable of concentrating the particles which are emitted by an axial point source into another axial (image) point—this force must everywhere be proportional to the distance from the axis, and hence the field itself must be proportional to this distance. The lenses which we are about to describe, which were proposed by Courant, Livingston and Snyder (1952), only partially satisfy this requirement that the field should be transversal, and proportional in strength to the distance from the axis. Here, too, we have a differential phenomenon, but far more favourable than in ordinary lenses. These are "strong convergence" lenses, which are also known as quadrupole lenses on account of the particular way in which the electrodes or pole-pieces of which they are composed are disposed around the axis.
10.1.2 "Strong Focusing" or "Quadrupole" Lenses†

These lenses consist of four cylindrical electrodes (or pole-pieces), parallel to the \( Oz \) axis; the lenses have four symmetry planes which intersect along this axis, and which are separated by angles of \( \pi/4 \). They produce an electric field \( E \) or a magnetic field \( H \), the intensity of either of which at any point \( M \) is proportional to the distance between \( M \) and the axis \( Oz \); the gradient of the field in the radial direction is constant throughout the whole useful region of space. The theoretical interest of these field distributions was pointed out by Melkikh in 1944 in a general study of cylindrical systems, but their practical application as lenses dates only from 1952. Since then, several theoretical and experimental studies have been devoted to elucidating their properties; a very complete bibliography relating to this question is to be found in the articles by Grivet and Septier (1960) and Septier (1961a).

We consider a transverse section of such a lens (Fig. 104) and we suppose first of all that the electrodes are infinitely long in the \( Oz \) direction. One pair of diametrically opposed electrodes (or pole-pieces), 1 and \( 1' \), is held at a potential (which may be either electrostatic or scalar magnetic) \( +\Phi_1 \), with respect to the earth and the other pair, 2 and \( 2' \), at a potential \( -\Phi_1 \). The aperture of the lens is defined by the radius \( a \) of the circular channel, which is tangential to the four electrodes. The two symmetry planes which do not intersect the electrodes we designate \( zOX \) and \( zOY \); the other two symmetry planes which do pass through the electrodes we label \( zOx \) and \( zOy \).

First, let us consider the electrostatic case. If a positive particle is incident parallel to the axis in the plane \( zOX \), it will experience a repulsion due

†Quadrupole optics is reviewed in Hawkes (1966).
to the electrode 1 (or 1'), but will not be affected by the presence of 2 and 2' as a result of the symmetry. The particle will remain in the plane \( zOx \), and will converge towards the axis. In the plane \( zOy \), the trajectory will also be planar, but the particle will be attracted by 2 (or 2') and will diverge away from the axis. Particles which are incident at the lens other than in the planes \( zOx \) and \( zOy \) will follow skew trajectories, approaching \( Oz \) in the \( OX \) direction, but moving away from \( Oz \) in the \( OY \) direction. In one direction the effect is of convergence, therefore, and in the other, of divergence.

In the magnetic case, the Biot–Savart law shows that the force which acts is perpendicular to the field, so that the privileged planes in which the trajectories remain planar are \( zOX \) and \( zOY \).

If we want finally to obtain convergence in every meridian plane, two of these lenses, \( \mathcal{Q}_1 \) and \( \mathcal{Q}_2 \), of length \( l \), must be placed end to end, differently orientated about the axis; the second lens is identical to the first save for a rotation about the axis of \( 90^\circ \). The divergent plane of \( \mathcal{Q}_2 \) corresponds to the convergent plane of \( \mathcal{Q}_1 \), and vice versa. The two lenses should also be as close as possible to one another to prevent the particles which diverge away from \( \mathcal{Q}_1 \) getting too far from the axis before being subjected to the focusing action of \( \mathcal{Q}_2 \). If \( f_{1x} \) is the focal length in the convergent plane of \( \mathcal{Q}_1 \), \( f_{2y} \) the focal length in the divergent plane of \( \mathcal{Q}_2 \), and \( e \) the width of the gap between the two lenses, the overall effect can be calculated roughly by assuming that each lens is a thin lens. From the standard formula for combinations of lenses, we can show immediately that even for \( f_{2y} = -f_{1x} \), the whole ensemble is convergent; the power is given by

\[
\frac{1}{f} = \frac{1}{f_{1x}} + \frac{1}{f_{2y}} - \frac{e}{f_{1x} f_{2y}},
\]

and if \( f_{2y} = -f_{1x} \),

\[
\frac{1}{f} = + \frac{e}{f_{1x}^2}.
\]

The effect of such a doublet is thus again a differential one, but the immense progress with respect to ordinary lenses resides in the fact that by lengthening the lenses, we can increase the convergence as much as we like without altering the electric or magnetic excitations. It is physically impossible to produce the same effect—convergence (or divergence) in every radial plane—with a single quadrupole lens; the potential is obliged to satisfy the Laplace equation within the useful region, and we can demonstrate that in consequence the lens must behave in the manner just described. This is proved in the course of the analysis of the field distribution which is given in the following paragraph; the higher terms of the general expansion are of a different symmetry from that of the basic quadrupole component, and cannot counteract the effect of the latter.
10.2 THE FIELD DISTRIBUTION

10.2.1 The Potential

We shall discuss the scalar electrostatic or magnetic potential, first of all making the assumption that the electrodes are infinitely long in the \( O_z \) direction; \( \Phi \) is then a function of the two variables \( x \) and \( y \) only (or \( X \) and \( Y \) according to the coordinate system chosen; one is appropriate in the electrostatic case, the other in the magnetic case). If we write

\[
s = X + i Y, \quad W(s) = V(X, Y) + i \Phi(X, Y),
\]

and take the symmetry into account, we obtain the following expression for \( W(s) \):

\[
W(s) = \frac{1}{2} h_2 s^2 + \frac{1}{6} h_6 s^6 + \frac{1}{10} h_{10} s^{10} + \cdots.
\]

The equipotential surfaces can be extracted thus:

\[
\Phi(X, Y) = \text{Im}\{W(s)\}
\]

\[
= h_2 X Y + h_6 X Y \left( X^4 - \frac{10}{3} X^2 Y^2 + Y^4 \right) + \cdots.
\]

With respect to the axes \( Ox, Oy \) we find

\[
\Phi(x, y) = \text{Re}\{W(s)\}
\]

\[
= \frac{1}{2} h_2 (x^2 - y^2) + \frac{1}{6} h_6 \{x^6 - 15x^4 y^2(x^2 - y^2) - y^6\} + \cdots.
\]

On the electrodes, at a distance \( a \) from \( O_z \), the potential is \( \pm \Phi_1 \), so that for magnetic lenses:

\[
\frac{\Phi(X, Y)}{\Phi_1} = \frac{2K_2}{a^2} X Y + \frac{2K_6}{a^6} X Y \left( X^6 - \frac{10}{3} X^2 Y^2 + Y^6 \right) + \cdots,
\]

and for electrostatic lenses:

\[
\frac{\Phi(x, y)}{\Phi_1} = \frac{K_2}{a^2} \frac{x^2 - y^2}{a^2} + \frac{K_6}{a^6} \frac{x^6 - 15x^4 y^2(x^2 - y^2) - y^6}{a^6} + \cdots.
\]

The coefficients \( K_n \) are constants which depend only upon the form of the cross-section of the electrodes.

The field components \( E_x \) and \( E_y \), \( B_x \) and \( B_y \) are given by:

\[
E_x = -\frac{\partial \Phi(x, y)}{\partial x}, \quad E_y = -\frac{\partial \Phi(x, y)}{\partial y},
\]

\[
B_x = -\frac{\partial \Phi(X, Y)}{\partial X}, \quad B_y = -\frac{\partial \Phi(X, Y)}{\partial Y}.
\]
If the gradient of the field is to be constant, the coefficients $K_{2n}$ must be zero for $n > 1$, so that

$$\Phi(x, y) = K_2 \frac{\Phi_1}{a^2} (x^2 - y^2).$$

The electrodes must then coincide with the equipotential surfaces $\pm \Phi_1$, the equation of which has the form

$$\frac{x^2 - y^2}{a^2} = 1 \quad \text{or} \quad X Y = \frac{1}{2} a^2,$$

which implies $K_2 = 1$.

![Diagrams showing various electrode shapes](image)

**Fig. 105.** Various electrode shapes. (a) Hyperbolic segments. (b) Plane electrodes. (c) Concave circular electrodes. (d) Convex circular electrodes.

The ideal distribution for which we are searching would be produced, therefore, by four electrodes the cross-sections of which are of the form of a pair of rectangular hyperbolae with asymptotes $OX$ and $OY$. In practice, however, it is not possible to build electrodes which are infinitely long in these two directions and even when the electrodes consist of portions of hyperbolic cylinders (Fig. 105a), higher order terms reappear in the expression for the potential; these latter are fortunately of little importance as we shall see later. The coefficients $K_2, K_6, \ldots$ can be calculated in a
few simple cases (plane electrodes—see Fig. 105 b; right cylinders formed from a segment of a circle with an angular separation of $2 \gamma$—see Fig. 105 c). The terms $K_2$ and $K_6$ in the expansion \[ \frac{\Phi(X, Y)}{\Phi_1} \] then take the following values:

<table>
<thead>
<tr>
<th></th>
<th>$K_2$</th>
<th>$K_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infinite hyperbolae</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Plane electrodes</td>
<td>1.037</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Concave circular electrodes. \[ 1.273 \cdot \frac{\sin 2\gamma}{2\gamma} \quad 0.042 \cdot \frac{\sin 6\gamma}{6\gamma} \]

In the final case, we can clearly reduce $K_6$ to zero by choosing $\gamma = \pi / 6 = 30^\circ$; this means that over an extended region in the neighbourhood of the axis, the gradient is accurately constant.

The sharp transition from hyperbolae to planes modifies the ideal distribution relatively little; we might consider taking advantage of this insensitivity by constructing electrodes which are circular in section (Fig. 105 d) rather than hyperbolic, as the former are considerably simpler to machine. Some measurements of the coefficients in magnetic lenses have produced the following results:

**Hyperbolic pole-pieces**

Width $E = 2a$ \[ K_6 = 0.5 \times 10^{-2} \]

**Circular pole-pieces**

Width $E = 2.5a$

Radius $R = 1.25a$ \[ K_6 = 2 \times 10^{-2} \quad K_{10} = -1.24 \times 10^{-2} \]

$R = 1.125a$ \[ K_6 = -0.5 \times 10^{-2} \quad K_{10} = -1.25 \times 10^{-2} \]

With circular pole-pieces of width $E = 2.5a$ and radius $R = 1.15a$, $K_6$ is effectively zero.

The potential distribution established above is only valid in a real lens far from the extremities of the electrodes and in particular, at the centre of the lens (where $z = 0$) provided the mechanical length $l$ is long in comparison with $a$. Near the longitudinal extremities, the potential and field become functions of all three variables—this is the region of leakage or overlap fields. There are still symmetry planes, however, and we can express $\Phi(x, y, z)$ in a form in which the simplifications which result from the symmetry have been made; the general expression in which only the symmetry conditions have been inserted can be written

\[ \frac{\Phi(x, y, z)}{\Phi_1} = \alpha \cdot \frac{x^2 - y^2}{a^2} + \beta \cdot \frac{x^4 - y^4}{a^4} + \gamma \cdot \frac{x^6 - y^6}{a^6} + \delta \cdot \frac{x^2 y^2 (x^2 - y^2)}{a^6}. \]
Φ has also to satisfy the Laplace equation
\[ \nabla^2 \Phi(x, y, z) = 0, \]
which implies
\[ \beta = -\frac{1}{12} \frac{d^3 \alpha}{dz^3} a^2, \]
\[ \delta = -15 \gamma + \frac{1}{24} \frac{d^4 \alpha}{dz^4} a^4, \]
so that finally
\[
\frac{\Phi(x, y, z)}{\Phi_1} = \frac{k_2(z)}{a^2} (x^2 - y^2) - \frac{1}{12} \frac{d^2 k_2(z)}{dz^2} \frac{x^4 - y^4}{a^4} + k_6(z) \frac{x^6 - y^6}{a^6} \\
- \left( 15k_6 - \frac{1}{24} \frac{d^4 k_2}{dz^4} a^4 \right) \frac{x^2 y^2 (x^2 - y^2)}{a^6}.
\]
At the centre \((z = 0)\):
\[ k_2(z) = K_2 \quad \text{and} \quad k_6(z) = K_6. \]
If we consider only the fundamental term, the function \(k_2(z)\) which is equal to unity at the centre falls away to zero far from the lens as shown in Fig. 106. We shall write
\[ k_2(z) = k(z). \]

![Fig. 106. The appearance of the characteristic function \(k(z)\) for a lens of mechanical length \(l\); the "equivalent length" \(L\) is greater than \(l\).](image)

The behaviour of \(k(z)\), which is the "characteristic function" of the lens, can only be obtained experimentally; at present, the function can only be measured easily in the magnetic case.

We have performed the various calculations relative to the coordinate system \(xOy\) only; in the various special cases considered later, we change to the axes \(XOY\) by a rotation through \(\pi/4\).
10.2.2 The Transverse Gradient

To obtain the first order optical properties, we assume that the potential distribution in the lens is given by the terms of lowest degree in the expansion of $\Phi$, namely

$$\frac{\Phi(x, y, z)}{\Phi_1} = \frac{k(z)(x^2 - y^2)}{a^2},$$

or

$$\frac{\Phi(X, Y, z)}{\Phi_1} = 2 \frac{k(z) X Y}{a^2}.$$

In the magnetic case, we use the second expression, and we obtain

$$B_x = - \frac{2\Phi_1}{a^2} X k(z),$$

$$B_y = - \frac{2\Phi_1}{a^2} X K(z),$$

$$B_z = - \frac{2\Phi_1}{a^2} X Y \frac{dk(z)}{dz}.$$

If we assume that the surface of each pole is a magnetic equipotential (and experiment shows that this is always true, since the effects of saturation always begin at a point of the yoke far from the poles) and that the consumption of ampère-turns in the iron is negligible, we can write:

$$\Phi_1 = \mu_0 n I,$$

where $n I$ represents the number of ampère-turns per pole.

The gradient, therefore, is given by

$$K(z) = \left| \frac{\partial B_x}{\partial Y} \right| = \left| \frac{\partial B_y}{\partial X} \right| = \frac{2\mu_0 n I}{a^2} k(z).$$

At $z = 0$, $K(0) = \frac{2\mu_0 n I}{a^2}$, $k(0) = 1$.

If the permeability of the iron is not infinite, we can always measure the potential difference, $2 \Phi_1$, between two adjacent poles, and we shall then have

$$K(0) = \frac{2\Phi_1}{a^2}.$$

In the electrostatic case, we find in the same way

$$K(0) = \frac{2\Phi_1}{a^2} \quad \text{and} \quad K(z) = K(0) k(z).$$
10.2.3 The Equivalent Length

As in all optical systems with a transverse field, the lens action can be characterized by the integral

\[ I = \int_{-\infty}^{\infty} B_r \, dz, \]

calculated along a line parallel to \( Oz \); \( B_r \) is the amplitude of the transverse component of \( B \), a distance \( r \) from \( Oz \). Along these lines parallel to the \( z \) axis, the amplitude of \( B_r \) is given to a first approximation by

\[ B_r(z) = B_r(0) k(z) = r K(0) k(z). \]

We can, therefore, define a hypothetical lens which is equivalent to the real lens by supposing that a field \( B_r(0) \) acts over a distance \( L \), and that for the two lenses the integral \( I \) takes the same value. This is equivalent to replacing the \( k(z) \) curve by a rectangle of unit height and length \( L \) in the direction of the lines parallel to \( Oz \). It is in this way that the "equivalent length" \( L \) is defined:

\[ L = \frac{1}{B_r(0)} \int_{-\infty}^{\infty} B_r(z) \, dz. \]

This length \( L \) is greater than the mechanical length \( l \) of the electrodes (Fig. 106), and measurement shows that to all intents and purposes,

\[ L \approx l + 1.1 \, a. \]

10.3 THE FIRST ORDER OPTICAL PROPERTIES

10.3.1 The Equations of Motion

The relativistically correct forms of the basic equations:

\[ \frac{dp}{dt} = \frac{d(mv)}{dt} = e E, \quad \text{(electrostatic case)} \]

or

\[ \frac{d(mv)}{dt} = -e(v \times B), \quad \text{(magnetic case)} \]

lead to simple expressions which have the same form in both cases provided we refer to the axes \( Ox, Oy \) for \( E \) and \( OX, OY \) for \( B \). If \( s \) is a unit vector, tangent to the trajectory, and \( ds \) is an element of the trajectory, we have

\[ ds = dz \left[ 1 + \left( \frac{dX}{dz} \right)^2 + \left( \frac{dY}{dz} \right)^2 \right]^{1/2} = dz \left[ 1 + \left( \frac{dx}{dz} \right)^2 + \left( \frac{dy}{dz} \right)^2 \right]^{1/2}. \]
We can eliminate the time $t$ from the equations above, by writing
\[
v = \frac{ds}{dt} = v \frac{ds}{ds} \quad \text{or} \quad \frac{d}{dt} = v \frac{d}{ds}.
\]

Further, we know that
\[
p = \{2 \, e \, m_0 \, \Phi_0 (1 + \varepsilon \, \Phi_0)\}^{1/2},
\]
and therefore
\[
v = \left( \frac{2e \, \Phi_0}{m_0} \right)^{1/2} \frac{1 + \varepsilon \, \Phi_0}{1 + 2 \varepsilon \, \Phi_0}, \quad \text{with} \quad \varepsilon = \frac{e}{2m_0 c^2},
\]
where $\Phi_0$ represents the total accelerating potential of the particles.

Finally we obtain
\[
\frac{d}{dz} \left[ \left( \frac{\Phi_0(1 + \varepsilon \, \Phi_0)}{1 + x'^2 + y'^2} \right)^{1/2} \frac{ds}{dz} \right] = \left( \frac{e}{2m_0} \right)^{1/2} \frac{ds}{dz} \times B,
\]
or
\[
\frac{d}{dz} \left[ \left( \frac{\Phi_0(1 + \varepsilon \, \Phi_0)}{1 + x'^2 + y'^2} \right)^{1/2} \frac{ds}{dz} \right] = \frac{1}{2} \left( 1 + 2 \varepsilon \, \Phi_0 \right) \left[ \frac{1 + x'^2 + y'^2}{\Phi_0(1 + \varepsilon \, \Phi_0)} \right] E.
\]

To a first approximation, for trajectories which are only slightly inclined to the axis (where $ds \approx dz$), and neglecting the variations in the velocity due to the local potential $\Phi(x, y, z)$ which is always small in comparison with $\Phi_0$ we arrive at the following equations if we consider an ideal lens in which the radial gradient is constant; these equations are valid for any velocity, as they are relativistically correct.

(i) The magnetic case

\[
X'' + \left( \frac{e}{2m_0 \Phi_0(1 + \varepsilon \, \Phi_0)} \right)^{1/2} K(z) X = 0,
\]
\[
Y'' - \left( \frac{e}{2m_0 \Phi_0(1 + \varepsilon \, \Phi_0)} \right)^{1/2} K(z) Y = 0,
\]
or, expressed in a different form

\[
X'' + \frac{K(z)}{B \varrho} X = 0, \quad Y'' - \frac{K(z)}{B \varrho} Y = 0.
\]

$B \varrho$ is the “magnetic stiffness” or “rigidity” of the particle, defined by
\[
B \varrho = \frac{mv}{e} = p/e.
\]

If the particles are slow enough for Newtonian mechanics to be a valid approximation, which occurs when $\varepsilon$ is small, we find

\[
X'' + \sqrt{\frac{e}{2m_0 \Phi_0}} K(z) X = 0,
\]
\[
Y'' - \sqrt{\frac{e}{2m_0 \Phi_0}} K(z) Y = 0,
\]
or
\[ X'' + \beta^2 k(z) X = 0, \]
\[ Y'' - \beta^2 k(z) Y = 0, \]
where
\[ \beta^2 = K(0) \sqrt{\frac{e}{2m_0 \Phi_0}} = \frac{\mu_0 n I}{a^2} \sqrt{\frac{2e}{m_0 \Phi_0}}. \]

\( K(0) \) is the gradient at \( z = 0 \).

(ii) The general relativistic electrostatic case
\[ x'' + \frac{1 + 2\varepsilon \Phi_0}{2\Phi_0(1 + \varepsilon \Phi_0)} K(z) x = 0, \]
\[ y'' - \frac{1 + 2\varepsilon \Phi_0}{2\Phi_0(1 + \varepsilon \Phi_0)} K(z) y = 0. \]

Where classical mechanics is adequate (\( \varepsilon \ll 1 \)),
\[ x'' + \beta^2 k(z) x = 0, \]
\[ y'' - \beta^2 k(z) y = 0, \]
with
\[ \beta^2 = \frac{K(0)}{2\Phi_0} = \frac{1}{a^2} \frac{\Phi_1}{\Phi_0}. \]

We have assumed implicitly that the convergent plane of the lens is \( OX \) (or \( Ox \)) and the divergent plane, \( OY \) (or \( Oy \)).

10.3.2 Integration of the Equations

If \( k(z) \) is replaced by the equivalent rectangular distribution of width \( L \), \( k(z) = 1 \) everywhere within the lens and \( k(z) = 0 \) outside; the equations above give the values of \( X \) and \( Y \) (or \( x \) and \( y \)). If, for example, we take the origin of the \( Oz \) axis in the entry plane of the lens
\[ X = X_0 \cos \beta z + \frac{X'_0}{\beta} \sin \beta z, \]
\[ Y = Y_0 \cosh \beta z + \frac{Y'_0}{\beta} \sinh \beta z. \]
\( X_0, Y_0, X'_0 \) and \( Y'_0 \) represent the initial conditions in the plane \( z = 0 \); these are the coordinates of the point at which the trajectory in question is incident, and the projections of its slope onto the \( XOz \) and \( YOz \) planes.

The equations of motion can equally well be integrated for other forms of the function \( k(z) \), but the calculations become more complicated. The approximation of the rectangular model gives results which are satisfactory to within a few per cent; in the domain of quadrupole lenses, the accuracy and simplicity of this model make it a basic tool.
The relations which give $X_s$, $Y_s$ and the slopes $X'_s$ and $Y'_s$ at the exit of a lens of length $L$ are of the form

$$\begin{align*}
X_s &= X_0 \cos \beta L + X'_0 \frac{\sin \beta L}{\beta}, \\
X'_s &= -\beta X_0 \sin \beta L + X'_0 \cos \beta L, \\
Y_s &= Y_0 \cosh \beta L + Y'_0 \frac{\sinh \beta L}{\beta}, \\
Y'_s &= \beta Y_0 \sinh \beta L + Y'_0 \cosh \beta L.
\end{align*}$$

These equations are in a form in which a matrix method can well be employed, and this method proves to be particularly fruitful when we are studying combinations of lenses, as we shall see later on (§ 10.4.3).

Any lens can be characterized by a pair of transfer matrices, one $\| T_C \|$ in the convergent plane $XOz$ and the other $\| T_D \|$ in the divergent plane $YOz$; we can then write:

$$\begin{align*}
\begin{pmatrix} X_s \\ X'_s \end{pmatrix} &= \| T_C \| \begin{pmatrix} X_0 \\ X'_0 \end{pmatrix}, \\
\begin{pmatrix} Y_s \\ Y'_s \end{pmatrix} &= \| T_D \| \begin{pmatrix} Y_0 \\ Y'_0 \end{pmatrix},
\end{align*}$$

with

$$\| T_C \| = \begin{pmatrix} \cos \beta L & \frac{1}{\beta} \sin \beta L \\ -\beta \sin \beta L & \cos \beta L \end{pmatrix},$$

and

$$\| T_D \| = \begin{pmatrix} \cosh \beta L & \frac{1}{\beta} \sinh \beta L \\ \beta \sinh \beta L & \cosh \beta L \end{pmatrix}.$$

10.3.3 The Cardinal Elements

If we consider a particular incident ray which is parallel to the axis $Oz$, the conditions at the origin are defined by $X_0$, $Y_0$ and $X'_0 = Y'_0 = 0$. We find, therefore, that

$$\begin{align*}
\begin{pmatrix} X_s \\ X'_s \end{pmatrix} &= \| T_C \| \begin{pmatrix} X_0 \\ 0 \end{pmatrix}, \\
\begin{pmatrix} Y_s \\ Y'_s \end{pmatrix} &= \| T_D \| \begin{pmatrix} Y_0 \\ 0 \end{pmatrix}.
\end{align*}$$

The cardinal elements are defined by the emergent rays, and are given by the following relations:

$$\begin{align*}
f_x &= -\frac{X_0}{X'_s}, \\
f_y &= -\frac{Y_0}{Y'_s}, \\
u_x &= \frac{X_s}{X'_s}, \\
u_y &= -\frac{Y_s}{Y'_s}, \\
v_x &= u_x - f_x, \\
v_y &= u_y - f_y;
\end{align*}$$
$f_x$ and $f_y$ represent the image focal lengths in the two planes $XOz$ and $YOz$. The abscissae of the foci, $u_x$ and $u_y$, and those of the principal planes $v_x$ and $v_y$, are measured from the exit face (see Figs. 107a and 107b). Finally,

$$
\begin{align*}
  f_x &= \frac{1}{\beta \sin \beta L}, \\
  u_x &= \frac{\cos \beta L}{\beta \sin \beta L}, \\
  v_x &= \frac{\cos \beta L - 1}{\beta \sin \beta L},
\end{align*}
$$

$$
\begin{align*}
  f_y &= -\frac{1}{\beta \sinh \beta L}, \\
  u_y &= -\frac{\cosh \beta L}{\beta \sinh \beta L}, \\
  v_y &= \frac{1 - \cosh \beta L}{\beta \sinh \beta L}.
\end{align*}
$$

Fig. 107. The cardinal elements of a lens. (a) In the convergent plane $XOz$. (b) In the divergent plane $YOz$.

Fig. 108. The variation of the reduced convergence $L/f$ and of the abscissa of the image principal plane $v/L$ as functions of the excitation $\beta L$, in the two planes $XOz$ and $YOz$ for a single lens.
The "object" and "image" elements are symmetrically placed with respect to the centre of the lens. If the excitations are sufficiently slight, or if the particles are moving at very high energies, the factor $\beta$ is small and we may write

$$\sin \beta L \approx \sinh \beta L \approx \beta L,$$

$$\cos \beta L \approx 1 - \left(\frac{\beta L}{2}\right)^2, \quad \cosh \beta L \approx 1 + \left(\frac{\beta L}{2}\right)^2,$$

so that for weak lenses

$$f_x = -f_y = \frac{1}{\beta^2 L},$$

and

$$v_x = v_y = -\frac{1}{2} L.$$

The two focal lengths are equal in absolute magnitude, and the principal planes coincide at the centre of the lens. This is only a valid approximation when $\beta L < 0.2$, however; when $\beta L > 0.2$, the ratio $f_x/f_y$ increases rapidly and the principal planes move away from the centre (Fig. 108).

10.4 COMBINATIONS OF LENSES

10.4.1 The Simple Doublet

A doublet consists of two lenses $Q_1$ and $Q_2$, of lengths $L_1$ and $L_2$ and excitations $\beta_1$ and $\beta_2$ respectively, separated by a region of field-free space of length $D$. The convergent plane of $Q_1$ corresponds to the divergent plane of $Q_2$. The optical properties and in particular the cardinal elements can easily be calculated with the aid of the matrix formalism.

Two transfer matrices correspond to each lens—one $\|T_C\|$ for the convergent plane, the other $\|T_D\|$ for the divergent plane. A region of space of length $D$ in which only the position of the current point is altered, and not the slope of the trajectory, is represented by the matrix

$$\|T_C\| = \begin{pmatrix} 1 & D \\ 0 & 1 \end{pmatrix}.$$  

The transfer matrices of the doublet we shall denote by $\|M_x\|$ in the $XOz$ plane (or the "convergent–divergent" plane, since $Q_1$ is convergent and $Q_2$ divergent in this plane) and by $\|M_y\|$ in the $YOz$ plane (the "divergent-convergent" plane); we obtain these matrices by forming the product of the individual matrices, written in the order opposite to that in which the particles encounter the corresponding elements. For example,

$$\|M_x\| = \|T_D\| \cdot \|T_C\| \cdot \|T_C\|.$$
or

\[
||M_x|| = \begin{pmatrix}
\cosh \beta_2 L_2 & \sinh \beta_2 L_2 \\
\beta_2 \sinh \beta_2 L_2 & \cosh \beta_2 L_2
\end{pmatrix} \begin{pmatrix}
1 & D \\
0 & 1
\end{pmatrix} \begin{pmatrix}
\cos \beta_1 L_1 & \sin \beta_1 L_1 \\
-\beta_1 \sin \beta_1 L_1 & \cos \beta_1 L_1
\end{pmatrix},
\]

and similarly

\[
||M_x|| = \begin{pmatrix}
\cos \beta_2 L_2 & \sin \beta_2 L_2 \\
-\beta_2 \sin \beta_2 L_2 & \cos \beta_2 L_2
\end{pmatrix} \begin{pmatrix}
1 & D \\
0 & 1
\end{pmatrix} \begin{pmatrix}
\cos \beta_1 L_1 & \sinh \beta_1 L_1 \\
\beta_1 \sin \beta_1 L_1 & \cosh \beta_1 L_1
\end{pmatrix}.
\]

And once again, in the exit plane of \(Q_2\)

\[
\begin{pmatrix}
X_s \\
X'_s
\end{pmatrix} = ||M_x|| \begin{pmatrix}
X_0 \\
X'_0
\end{pmatrix},
\]

\[
\begin{pmatrix}
Y_s \\
Y'_s
\end{pmatrix} = ||M_y|| \begin{pmatrix}
Y_0 \\
Y'_0
\end{pmatrix}.
\]

For \(X'_0 = Y'_0 = 0\), we obtain the focal lengths \(f_x\) and \(f_y\) and the positions of the foci. The expressions for these quantities are very complicated in the general case, but are considerably simplified when the doublet is symmetrical (\(\beta_1 = \beta_2, L_1 = L_2\)) and for example

\[f_x = f_y = [\beta (\sin \beta L \cosh \beta L - \cos \beta L \sin \beta L + \beta D \sin \beta L \sinh \beta L)]^{-1}.
\]

The relative positions of the cardinal elements are shown in Fig. 109. The relations which express the correspondence between the object and the image are calculated by the same method, but it must be admitted that the

[Fig. 109. The positions of the cardinal elements of a doublet. \(Q_1\) and \(Q_2\) have here the same excitation; the two focal lengths, \(f_x\) and \(f_y\), are equal.]

full calculations for an optical system adapted to a particular problem are extremely complicated. In most cases, we are obliged to resort to an electronic computer. A great deal of data is to be found, however, in recent publications, both graphical (Enge, 1959) and in the form of tables of calculated values of the cardinal elements (Blewett, 1959).
10.4.2 The Pseudo-stigmatic Doublet and the Stigmatic Doublet

In general, a doublet produces two distinct images of an axial point source which lie in the planes $zOX$ and $zOY$. Each image is a line focus so that the system has innate astigmatism. Nevertheless, by exciting $Q_1$ and $Q_2$ differently, the two focal lines can be brought into the same image plane; the combination is then called "pseudo-stigmatic", as the linear magnifications along $OX$ and $OY$ are different (see Fig. 110a). This mode of operation is used very often to concentrate the maximum number of particles which emerge from a source into a region around a point.

At very high excitations, we can demonstrate theoretically—and confirm experimentally—that a symmetrical doublet can behave like an axially symmetrical lens; both the foci $F_X$ and $F_Y$ and the principal planes $H_X$ and $H_Y$ coincide. When $D = 0$, the doublet behaves in this way for $\beta L = \pi$. The lens can then be used in an electron microscope to form a faithful image (Septier, 1958); the asymptotic foci of the doublet lie, however, inside the lens so that the strongly convergent lens which is obtained in this way is only suitable as a projective lens.
10.4.3 The Triplet

The doublet discussed above possesses one marked drawback: the optical centre of the equivalent lens is not the same in the two planes $zOX$ (or $zOx$) and $zOY$ (or $zOy$), and moves when the excitations of the lenses are varied—this complicates the adjustment especially.

The present tendency is to replace the doublet by a symmetrical triplet, a diagram of which is given in Fig. 111. The two outer lenses $Q_1$ and $Q_3$ are identical and have the same excitation $\beta_1$; the central lens $Q_2$ is turned through $90^\circ$ with respect to $Q_1$ and $Q_3$, and has an independent excitation $\beta_2$. The principal planes are different in $zOX$ and $zOY$, but remain symmetrical with respect to the centre of symmetry $O$. Moreover, for weak lenses, the triplet can be regarded as a pair of thin convergent cylindrical lenses situated at $O$.

The calculations associated with such a system are extremely complicated in the case of strongly convergent lenses. The cardinal elements of a few triplets are tabulated in Blewett (1959) and Enge (1961).

Under certain conditions of excitation, the symmetric triplet is stigmatic for a pair of conjugate planes, $P$ and $P'$, which are themselves symmetrically placed with respect to the centre, $O$. The magnification is then equal to unity. This particular system has been studied in detail by Septier and Dhuicq (1965), Septier (1966), and has recently been employed in electron microscopy by Bauer (1965–6) together with a symmetric doublet equivalent to a round lens. Images of good quality have been obtained.

10.4.4 The Quadruplet

Work by Dhuicq and Septier (1959) and Dhuicq (1960) has shown that it is possible to design an optical system with a very high convergence that is equivalent to a round lens for all pairs of conjugate points on the axis $Oz$, using four identical quadrupoles, symmetrical in pairs about the
optical centre, \( z = 0 \) (Fig. 111a). The excitations will be symmetric when
the scalar potentials at the electrodes or pole-pieces are the same in the
two central lenses and are likewise the same in the two outermost lenses.
The plane \( zOx \) is then “convergent–divergent–divergent–convergent”
(the \( CDDC \) arrangement).

![Diagram](image)

**Fig. 111a.** Quadruplets, electrically symmetrical (a) and antisymmetrical (b).

In this situation, the asymptotic foci will always be situated inside the
outermost quadrupoles so that the lens can only be used as a projective
lens (the real immersed foci do not coincide in the two symmetry planes
and one cannot therefore place a real object at them).

In contrast, we can find working regions for which the foci lie beyond
the outside quadrupoles and the focal length is very short by exciting
the lens of such a system so that the \( zOx \) plane becomes \( CDCA \) (Dymnikov
and Yavor, 1963; Dymnikov et al., 1965a; Dhucq and Septier,
1966a). These lenses are equivalent to round lenses, and could be used
with high energy particles (electrons of several MeV, or even ions of a
few hundred keV); the focal length would remain less than or equal to
one centimetre.

For the case \( l_2/l_1 = 2, l = 0 \), the curves \( D = d/l_1 = \text{const.} \)
\( F = F_x/l_1 = F_y/l_1 = \text{const} \) and \( Z_{Fx}/l_1 = Z_{Fy}/l_1 = \text{const} \) are shown in Fig. 111b as
functions of the excitations, \( U \) and \( V \), of the two pairs of lenses \( Q_1, Q_4 \) and
\( Q_2, Q_3 \) respectively (\( U = \beta_x^1 l_1 \) and \( V = \beta_x^2 l_2 \)).

Thus, for \( U = 1.95, V = 4.5 \) for example, the central gap is \( d/l_1 = 0.155 \)
and the focal length is \( f = 0.041 l_1 \). The focus is located at \( Z_F = 0.267 l_1 \)
from the terminal face. If \( l_1 = 20 \text{ mm}, l_2 = 40 \text{ mm} \) and \( a = 4 \text{ mm} \) (\( a \) is the
bore radius), we have \( d = 3.1 \text{ mm}, f = 0.8 \text{ mm} \) and \( p = 6.4 \text{ mm} \).
For 3 MeV electrons ($\Phi^*_0 = 12$ MeV), we should need field gradients in the lenses of the order of $K_1 = 11,000 \text{ G/cm}$ and $k_2 = 15,000 \text{ G/cm}$ and these values are attainable. The superiority of this type of lens over ordinary lenses so far as convergence is concerned is clear. The aberration coefficients are much higher, however.

![Diagram](image)

**Fig. 111b.** The optical elements of an antisymmetric quadruplet equivalent to a round lens, when $M = l_2/l_1 = 2$ and $L = l/l_1 = 0$. $P = p/l_1$ denotes the distance from the common focus to the terminal face of the fourth lens; $D = d/l_1$ is the distance between the two central lenses. The focal length is given by $F = f/l_1$.

It is also possible to obtain a similar solution with a symmetrical quadruplet. It possesses the desired properties only for a single pair of conjugate points on the axis $Oz$, which severely limits its usefulness.
10.5 A CHAIN OF $N$ IDENTICAL CROSSED QUADRUPOLE LENSES

10.5.1 Introduction

If we want to channel a beam of high energy particles over a long distance, we can employ a series of $N$ quadrupole lenses, provided a lateral stability condition which is identical to the condition which is imposed on the optical system of an ordinary periscope is fulfilled. The transverse dimensions of the beam must always remain smaller than the apertures of the lenses, and hence smaller than the radius of the bore, $a$. The same problem crops up in all periodic optical systems, whether they be formed from $N$ identical convergent lenses or, as in this case, from $\frac{1}{2}N$ convergent lenses with $\frac{1}{2}N$ divergent lenses interleaved between them. We shall only consider the case in which the lenses are all identical (both in length and excitation). A thorough examination of more general cases shows that, to within the degree of complication considered, the conclusions are unaltered. Figure 112 shows the evolution of a beam which is incident parallel to the axis in a poorly chosen system; the particles are lost at the walls after the fifth lens.

To obtain the condition for lateral stability, we shall argue first of all in terms of an optical system consisting of thick convergent lenses, and use the formalism of elementary geometrical optics; afterwards, we shall introduce the matrix formalism which proves to be particularly fruitful. The thick lens systems which we have in mind may be either ordinary convergent lenses or quadrupole doublets. In the latter case, the positions of the principal planes and the foci, measured relative to the beginning of the
channel, will be different in the two planes $XOz$ and $YOz$; if we suppose
the succession to be infinitely long, however, we can consider that the basic
unit is identical in the two planes.

10.5.2 The Behaviour of a Ray through a Characteristic Section of the Chain

The regularly repeated cell of the pattern will consist of a thick convergent
system which is defined by its object and image principal planes, $P_0$ and
$P_i$, and by its object and image foci, $F_0$ and $F_i$; the constants of the system
will thus be
\[ \overrightarrow{P_0 F_0} = f_0, \quad \overrightarrow{P_i F_i} = f_i, \quad \overrightarrow{P_0 P_i} = \alpha. \]
We have $f_0 = -f_i$, and for convergent lenses, $f_i > 0$; we write $f = f_i = |f_0|$.
The position of one cell or "element" relative to its successor will be fixed
by the distance $a + b$ between the focus $F_{i,m}$ and the focus $F_{0,m+1}$, so that
the length of the element (the geometrical period) is (see Fig. 113):
\[ L = a + b + 2f + d. \]

![Fig. 113. The paths of two particular rays in an elementary cell of the $m$-th rank
(the cell is equivalent to a thick lens with the same foci and principal planes).](image)

Each element is bounded by two planes $H_m$ and $H_{m+1}$ which are at a distance
$a$ from $F_i$ and $b$ from $F_0$ respectively.
\[ \overrightarrow{H_m F_{0,m}} = b, \quad \overrightarrow{F_{i,m+1} H_{m+1}} = a. \]
We suppose that the distance of the ray from the axis, $r_m$, and its slope $r'_m$
at the entry to the $m$-th element are known; the corresponding quantities at
the exit, in the plane $H_{m+1}$, will be linearly related to the values at the entry:
\[ r_{m+1} = \alpha r_m + \beta r'_m, \]
\[ r'_{m+1} = \gamma r_m + \delta r'_m. \]
It is easy to relate $\alpha$, $\beta$, $\gamma$ and $\delta$ to the known optical quantities $f$, $d$, $a$ and $b$ by applying these equations to the two rays usually employed in optical constructions.

The first is parallel to the axis at $H_m$, and a distance $r_m = 1$ from it; it leaves the element with a slope $r_{m+1} = -\frac{1}{f}$, passing through the image focus $F_i,m$. The second ray is symmetrical to the first with respect to the symmetry plane of the thick lens—it passes through $F_{0,m}$ and has a slope $r'_m = +\frac{1}{f}$ at the entry; it leaves parallel to the axis and a distance $r_{m+1} = 1$ from it (Fig. 112).

We have therefore

(i) $r'_m = 0$, $r_m = 1$, $r'_{m+1} = \frac{1}{f}$, $r_{m+1} = -\frac{a}{f}$,

(ii) $r'_m = \frac{1}{f}$, $r_m = \frac{b}{f}$, $r'_{m+1} = 0$, $r_{m+1} = 1$,

so that

$\alpha = -\frac{a}{f}$, $\beta = \frac{f^2 - ab}{f}$, $\gamma = -\frac{1}{f}$, $\delta = -\frac{b}{f}$.

The transfer relations thus become:

$$r_{m+1} = -\frac{a}{f} r_m + \frac{f^2 - ab}{f} r'_m,$$

$$r'_{m+1} = -\frac{1}{f} r_m - \frac{b}{f} r'_m.$$

We eliminate the slopes $r'$, and obtain the following recurrence relation:

$$r_{m+2} - \frac{a + b}{f} r_{m+1} + r_m = 0,$$

the general solution of which is a linear combination, with constant coefficients, of the two fundamental solutions $e^{m\omega}$ and $e^{-m\omega}$:

$$r_m = A e^{m\omega} + B e^{-m\omega}.$$

The constants $A$ and $B$ are determined by the initial conditions $r_0$ and $r'_0$ at the beginning of the chain of lenses. From the recurrence relation we can calculate $\omega$; by substituting for $r_m$, and dividing by $e^m$, we obtain

$$e^{2\omega} - \frac{a + b}{f} e^\omega + 1 = 0,$$

or

$$\frac{e^\omega + e^{-\omega}}{2} = \frac{a + b}{2f},$$
the solution of which is
\[
\cosh \omega = \frac{a + b}{2f} \quad \text{if} \quad \frac{a + b}{2f} > 1,
\]
or
\[
\cos \mu = \frac{a + b}{2f} \quad \text{if} \quad \frac{a + b}{2f} \leq 1 \quad \text{with} \quad \omega = j \mu.
\]

The outermost trajectory of the beam defined by \( r_m \) will be stable only when \( \omega \) is imaginary, so that
\[
\begin{align*}
\quad r_m & = A \cos m \mu + B \sin m \mu. \\
\end{align*}
\]

\( r_m \) is then always bounded, irrespective of the values of \( A \) and \( B \). Otherwise, when \( \omega \) takes real values, \( r_m \) increases with \( m \) without limit for every set of initial conditions for which \( A \neq 0 \).

We therefore obtain a simple relation between the value of the focal length and the distance between the two foci \( F_1 \) and \( F_0 \):
\[
a + b < 2f.
\]

We have thus expressed the condition in terms of the basic lens parameters.

**10.5.3 The Matrix Representation**

The form of the stability condition obtained above is highly suggestive, but nevertheless, it is preferable to obtain a more abstract but still more general expression by employing the matrix representation. The method which we shall use allows us to find the stability condition in cases where the individual units of the structure consist of several different elements—convergent or divergent lenses, regions of field-free space, etc. For a motif or "cell" which consists of three separate parts, for example, we have
\[
||M|| = ||T_3|| \cdot ||T_2|| \cdot ||T_1||,
\]
so that
\[
\begin{pmatrix} r_{m+1} \\ r'_{m+1} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} r_m \\ r'_m \end{pmatrix},
\]
where
\[
||M|| = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.
\]

The expressions given above for \( \alpha, \beta, \gamma \) and \( \delta \) show that
\[
|M| = \alpha \delta - \beta \gamma = 1.
\]
(This is a perfectly general property of all the elementary matrices \( ||T_n|| \) which we use.)

The stability condition \( a + b < 2f \) here takes the form
\[
\alpha + \delta < 2,
\]
or
\[
\text{Trace } ||M|| < 2.
\]
(The trace of a matrix is the sum of the elements on the principal diagonal.)
10.5.4 The Case of \( N \) Quadrupole Lenses

The stability condition takes a particularly simple form when the lenses (represented by their equivalent rectangular model) are in contact. (If the distance \( d \) between the lenses is not zero, the expression which is obtained is more complicated, but is still obtained by writing Trace \( ||M|| < 2. || \).

The motif of length \( 2L \) is now composed of a convergent lens and a divergent lens each of length \( L \). The corresponding elementary matrices are then respectively:

\[
||T_c|| = \begin{pmatrix}
\cos \beta L & -\frac{1}{\beta} \sin \beta L \\
-\beta \sin \beta L & \cos \beta L
\end{pmatrix},
\]

\[
||T_d|| = \begin{pmatrix}
\cosh \beta L & \frac{1}{\beta} \sinh \beta L \\
\beta \sinh \beta L & \cosh \beta L
\end{pmatrix}.
\]

The transfer matrix for a region of field-free space \( D \) is, we recall,

\[
||T_G|| = \begin{pmatrix} 1 & D \\ 0 & 1 \end{pmatrix}.
\]

Writing \( \beta L = x \), the matrix \( ||M|| \) is therefore given by

\[
||M|| = \begin{pmatrix}
\cosh x & 1 \\
\beta \sinh x & \cosh x
\end{pmatrix} \begin{pmatrix}
\cos x & 1 \\
-\beta \sin x & \cos x
\end{pmatrix},
\]

or

\[
||M|| = \begin{pmatrix}
\cosh x \cos x - \sinh x \sin x & \frac{1}{\beta} (\cosh x \sin x + \sinh x \cos x) \\
\beta (\sinh x \cos x - \sin x \cosh x) & \sinh x \sin x + \cosh x \cos x
\end{pmatrix}.
\]

The periscopic stability condition is thus

\[ 2 \cosh x \cos x \leq 2, \]

or

\[ \cosh x \leq \sec x, \]

which gives

\[ 0 < \beta L \leq 1.873. \]

When \( \beta L > 1.873 \), the system is over-focused and the beam is rapidly lost at the walls.

Within the channel, the trajectory will display a certain periodicity, of length \( A \), which is equal to twice the distance between the two successive points at which \( r = 0 \); the shortest value of \( A \) will be obtained when \( \beta L = 1.873 \), when it is equal to twice the period, \( p = 2L \), of the structure itself.
10.6 APPLICATIONS OF QUADRUPOLE LENSES

We have already mentioned two common uses of these lenses, namely, to focus into an image plane the particles which are emitted at an object plane, which is usually the exit diaphragm of an accelerator or the plane of a target which is being bombarded by some primary radiation, and to channel a beam over a long distance, the only requirement being that the trajectories shall be stable; here it is not a faithful image that interests us.

The same channelling principle has been applied to the circular accelerators which have recently been constructed; the principle was described in 1952 by Courant, Livingstone and Snyder for strong-focusing synchrotons. Each elementary magnet of the synchroton plays two roles; the trajectory has to be curved in such a way that the mean path is circular, and the particles have to be focused in the horizontal and the vertical planes. These two operations are combined by employing hyperbolic pole-pieces, between which the radial gradient is constant.

Another application of a similar kind, though in theory more simple, was first suggested by Blewett (1952) and later studied by Bernard (1953a). Here, small quadrupole lenses are inserted into the tubes which separate the accelerating gaps in a linear ion accelerator in order to cancel the defocusing effect of the high frequency accelerating field. We can briefly analyse the stability conditions of the whole system by supposing that each accelerating gap behaves like a divergent thin lens, the focal length of which is known as soon as the conditions which produce acceleration have been fixed (the field, the phase angle, the value of the high frequency). If we assume that the velocity of the particles varies very little as they pass through a gap, and if we place one lens in each tube, we obtain a certain periodicity $p$, which may be produced, for example, by:

— the second half of a convergent quadrupole lens $Q_1$ in the tube preceding the first gap;
— a divergent lens $L_1$ in the first gap;
— a divergent quadrupole lens $Q_2$ which is identical to $Q_1$ in the tube following the first gap;
— a lens $L_2$ identical to $L_1$;
— and, finally, the first half of a convergent quadrupole lens $Q_3$ identical to $Q_1$ (see Fig. 114a).

The stability condition provides the value of the minimum gradient which is necessary if the accelerator is to function well (see, for example, the study by Teng, 1954). The lower the energy of the particles, the greater must the convergence of the lenses be; we cannot even consider using electrostatic lenses, as the potentials involved would be too high. In a magnetic version which has recently been constructed in Geneva at CERN in a linear proton accelerator, gradients of the order of 10 kilogauss cm$^{-1}$ are obtained.
at the point of injection (with an initial energy of 500 keV) in pulsed operation; when the energy lies between 10 and 50 MeV, the gradient which is necessary varies between about 2000 and 500 G cm⁻¹, and the quadrupole lenses can be supplied steadily as the heat which is dissipated in the windings is now sufficiently low.

**Fig. 114a.** Focusing in a linear ion accelerator. (a) Tubes in which the motion of the particles is left undisturbed, save by a single quadrupole lens in each tube, alternately convergent and divergent. (b) The optical elements of the period; \( L_1 \) and \( L_2 \) are fictitious divergent lenses which lie in the gaps.

In conclusion, we should point out that quadrupole lenses can equally well be used to focus molecules and atoms which possess a magnetic moment proportional to the applied field, as Vauthier discovered (1949). The focusing force is then proportional to the distance from the axis in every radial direction, and the lens behaves like an axially symmetrical lens (Keller, 1957; Bennewitz and Paul, 1954; and Bennewitz, Paul and Schlier, 1955), and finally, several authors have shown that electric quadrupole lenses supplied with a high frequency potential provide a mass filter with an extremely good resolution (Paul and Raether, 1955; and Paul, Reinhard and von Zahn, 1958).

### 10.7 The Aberrations of Quadrupole Lenses

#### 10.7.1 Aperture Aberrations

The vital aberration here is the aperture aberration. In the Gaussian image plane, the aberration figure depends upon the symmetry of the system; in the case of stigmatic operation (Burfoot, 1954), it may be (Fig. 114b):
(a) a circle, as in axially symmetrical lenses;
(b) a star with four points;
(c) a rosette;
(d) or, more commonly, a composite figure formed from a superposition of the three elementary aberrations. Three coefficients are necessary to describe the aberration (whereas one alone, $C_s$, is sufficient when the lenses are axially symmetric).

When the system is pseudo-stigmatic, we again obtain these same shapes, but with different magnifications in the $X$ and $Y$ directions (Septier, 1960). Finally, when the system is astigmatic, the effect of the aberrations on the focal lines has been studied in detail both for magnetic and electrostatic lenses (Septier, 1958; Septier and van Acker, 1960. See Fig. 6b).

Four coefficients are required to describe the transverse aberration in the Gaussian image plane. We write

\[
\Delta x = C_{xx} x_a^2 + C_{xy} x_a y_a^2,
\]
\[
\Delta y = C_{yx} y_a^2 + C_{yy} x_a y_a^2.
\]

(10.A)

(For a stigmatic or pseudo-stigmatic system, we should find that $C_{xy} = C_{yx}$.) Here, $x_a$ and $y_a$ denote the distance from the axis in the plane $z = z_a$ of the entrance pupil and $x(z)$, $y(z)$ are trajectories travelling in the two orthogonal planes of symmetry.

These trajectories $x(z)$ and $y(z)$ can be expressed in terms of two particular trajectories $g(z)$ and $h(z)$, defined by

\[
g_x(z_0) = g_y(z_0) = h_x(z_a) = h_y(z_a) = 1
\]
and

\[
g_x(z_a) = g_y(z_a) = h_x(z_0) = h_y(z_0) = 0.
\]

We find (Fig. 114c)

\[
\begin{align*}
x(z) &= x_0 g_x(z) + x_a h_x(z) \\
y(z) &= y_0 g_y(z) + y_a h_y(z),
\end{align*}
\]

where $z = z_0$ denotes the object plane and $z = z_a$, the aperture plane. At $z = z_i$, we have

\[
x'(z_i) = x_a h'_x(z_i) = \alpha, \\
y'(z_i) = y_a h'_y(z_i) = \beta,
\]

if we consider a particle from an axial point source $S$. We can also express $\Delta x$ and $\Delta y$ in terms of $\alpha$ and $\beta$, which are easier to measure experimentally:
\[ \Delta x = C'_x \alpha^2 + C'_{xy} \alpha \beta^2, \quad \Delta y = C'_y \beta^2 + C'_{yx} \beta \alpha^2. \]

If \( S'_x \) and \( S'_y \) coincide, \( C'_{xy} = C'_{yx} \) even if the magnifications \( M_x \) and \( M_y \) are different.

As in the case of rotationally symmetric systems, the aberrations may be calculated by either of two methods: the trajectory method or the characteristic function method, developed and applied by Hawkes (1963, 1965) in the general case of systems with two symmetry planes (rectilinear orthogonal systems).

Hawkes' theory also allows us to calculate all the aberration coefficients and we do indeed recover, as a special case, the formulae listed in Chapter 7 for rotationally symmetric systems.

### 10.7.2 The Trajectory Method

Into the equations of motion we now introduce firstly the transverse components of the velocity and secondly the higher order terms in the field expansion in the fringe field zone (Septier, 1961a). In the plane that is initially convergent, we obtain the following equations:

**Magnetic case**

\[
X'' + \beta^2 kX = -\beta^2 \left\{ kX \frac{3X'^2 + Y'^2}{2} - kX'Y'Y - k'XYY' \right\} - \frac{k''}{12} X (X^2 + 3Y^2) = -\beta^2 P(z). \tag{10.B}
\]

**Electrostatic case**

\[
x'' + \beta^2 kx = -\beta^2 \left\{ kx(x'^2 + y'^2) + \beta^2 k^2 x(x^2 - y^2) \right\} - \frac{k'x'}{2} (x^2 - y^2) - \frac{k''x^3}{6} = -\beta^2 Q(z). \]
In the initially divergent plane, we obtain identical equations if the following exchanges are performed:

\[ \beta^2 \rightarrow -\beta^2 \quad X \Leftrightarrow Y \quad X' \Leftrightarrow Y' \]

\[ x \Leftrightarrow y \quad x' \Leftrightarrow y' \]

For a given system, the function \( \beta^2 k(z) \) is known everywhere on the axis \( Oz \), and can be represented by simple analytic functions, usually bell-shaped fields (Bernard and Hue, 1956; Grivet and Septier, 1960; Septier, 1961a). With the aid of an electronic computer, we can then either (i) integrate the so-called “third order equations” and calculate the final discrepancy between their solutions and those of the “first order equations” (obtained by setting the right hand sides equal to zero) or alternatively (ii) we can calculate the differences, \( \varepsilon(z) \) and \( \eta(z) \), between the Gaussian and “third order” trajectories throughout the whole system. For this, we replace such quantities as \( X, Y, X' \) and \( Y' \) by the first order solutions, obtained in a preliminary calculation. \( P(z) \) and \( Q(z) \) are then known functions.

If the rectangular model (\( K = \text{const.} \)), which yields very simple first order solutions, is used, the aberrations cannot at first sight be calculated by this method, for the terms in \( k'(z) \) and \( k''(z) \) have to be neglected in the equations. Efforts have been made, nevertheless, to take them into account (Bernard and Hue, 1956). It seems, however, that by carrying out partial integrations on the general equations, expressions for the aperture aberration coefficients not containing \( k' \) and \( k'' \) can be obtained (Dymnikov et al., 1965c); these can then be integrated directly using the simple solutions afforded by the rectangular model.

We can also calculate the aberration coefficients of each lens and hence the total aberration coefficients for a system consisting of a number of lenses (Deltap, 1964a, 1964b).

Using this trajectory method, Septier (1961a), Strashkevich (1964) and Dymnikov et al. (1964a) have succeeded in determining the aperture aberrations of electrostatic and magnetic single lenses and doublets. More recently, Meads (1966) has written a programme for the IBM 7094 computer with which the aberration coefficients can be calculated and an optimum system designed (by varying the geometrical and electrical parameters). Dhuiqcq and Septier (1966b) have calculated the aperture aberration coefficients of a quadruplet equivalent to a round lens (the characteristics of which have been given in § 10.4.4): for the magnetic case,

\[ \frac{C_x}{f} = 39, \quad \frac{C_{xy}}{f} = \frac{C_{ux}}{f} = 72, \quad \frac{C_y}{f} = 1.4 \times 10^5 \]
and for the electrostatic case
\[
\frac{C_x}{f} = 12, \quad \frac{C_{xy}}{f} = \frac{C_{yx}}{f} = 550, \quad \frac{C_y}{f} = 0.5 \times 10^5
\]
for
\[f = 0.8 \text{ mm.}\]

They have shown that it is possible to cancel these aberrations completely by employing three octopole lenses, judiciously disposed within the system.

Deltrap (1964a, 1964b) has measured the aberration coefficients of a system consisting of four mixed magnetic lenses (each a combined quadrupole and octopole) together with a round lens and has shown that it is possible to reduce the aberration coefficients considerably. Septier and Dhuiq (1964) have performed similar experiments using a symmetric triplet of unit magnification. These two latter systems are stigmatic only for a privileged pair of conjugate points.

10.7.3 The Characteristic Function Method (Hawkes, 1963, 1965a, 1965b)

Setting out from the expression for Fermat’s principle, as applied in electron optics,
\[
\delta \int n ds = 0,
\]
in which \(n\) denotes the refractive index
\[n = \Phi^{1/2} - \sqrt{\frac{e}{2m_0}} A \cdot s,
\]
Hawkes transforms the variational equation into
\[
\delta \int m dz = 0,
\]
in which
\[m = \{ \Phi^{1/2} (1 + x'^2 + y'^2)^{1/2} - \eta (A_x x' + A_y y' + A_z) \}, \quad (10.\text{C})\]
where
\[\eta = \sqrt{\frac{e}{2m_0}}.
\]
The function \(\Phi(x, y, z)\) and the components of the vector \(A(x, y, z)\) can be expressed as power series in \(x\) and \(y\), while the variable \(z\) figures only in the coefficients. On substituting these series into \(m\) and collecting up, we obtain
\[m = m^{(0)} + m^{(2)} + m^{(4)} + \cdots + m^{(2n)} + \cdots, \quad (10.\text{D})\]
in which \(m^{(2n)}\) contains all the terms of degree \(2n\).
If \( S \) denotes the optical path-length between two points in the planes \( z_1 \) and \( z_2 \), and \( S^* \) an infinitesimally close optical path, we have
\[
S = \int_{z_1}^{z_2} m \, dz,
\]
\[
S^* = S + \delta S.
\]
If the two rays begin and end at the same points, we have
\[
\delta S = 0,
\]
which implies
\[
\frac{dp}{dz} = \frac{\delta m}{\delta x}, \quad \frac{dq}{dz} = \frac{\delta m}{\delta y},
\]
where we have written \( p = \delta m/\delta x', q = \delta m/\delta y' \).

The function \( V_{1,2} = \int_{z_1}^{z_2} m \, dz \) is Hamilton's point characteristic function, and we have
\[
\delta V_{1,2} = p_2 \delta x_2 + q_2 \delta y_2 - p_1 \delta x_1 - q_1 \delta x_1 \left\{ \begin{array}{l}
 p_1 = -\frac{\partial V_{1,2}}{\partial x_1} \\
 q_1 = -\frac{\partial V_{1,2}}{\partial y_1} \\
 p_2 = \frac{\partial V_{1,2}}{\partial x_2} \\
 q_2 = \frac{\partial V_{1,2}}{\partial y_2}.
\end{array} \right. (10.E)
\]

If \( m \) is replaced by a perturbed function, \( m^* \),
\[
m^* = m + m^{(p)},
\]
we obtain (Sturrock, 1951)
\[
V_{1,2}^{(p)} = \int_{z_1}^{z_2} m^{(p)} \, dz
\]
and if in addition we impose the conditions
\[
x_1^{(p)} = y_1^{(p)} = x_2^{(p)} = y_2^{(p)} = 0,
\]
we can deduce the perturbations of \( p \) and \( q \) (\( p^{(p)} \) and \( q^{(p)} \)) from formulae resembling (10.E) except that the perturbation \( V_{1,2}^{(p)} \) replaces \( V_{1,2} \).

The Gaussian (first order) properties of a system will be described by the term
\[
m = m^{(2)},
\]
the third order geometrical aberrations by
\[
m^{(6)} = m^{(4)},
\]
and the chromatic aberration by

\[ m^{(p)} = \frac{\partial m^{(2)}}{\partial \Phi}. \]

We obtain \( V^{(p)}_{1,2} \) by integrating \( m^{(p)} \) with respect to \( z \); to be able to perform this integration, we replace the quantities \( x, y, x' \) and \( y' \) by their unperturbed first order values, obtained from \( m = m^{(2)} \). We can then calculate the quantities \( p^{(p)}_c, q^{(p)}_c, x^{(p)}_c \) and \( y^{(p)}_c \) in an arbitrary current plane of abscissa \( z = z_c \) in terms of \( V^{(p)}_{ac} \) and \( V^{(p)}_{ac} \) (these two quantities correspond to integrations from the object point \( z_0 \) to \( z_c \) and from the aperture \( z_a \) to \( z_c \), respectively).

In a system of quadrupoles possessing two planes in which the trajectories remain plane, we have

\[ \begin{align*}
    m^{(2)} &= \frac{1}{2} C(x'^2 + y'^2) + \frac{1}{2}(\tilde{C}_x x^2 + \tilde{C}_y y^2) \\
    p &= C x', \quad q = C y'
\end{align*} \tag{10.F} \]

(\( C, \tilde{C}_x \) and \( \tilde{C}_y \) are functions of \( z_c \).) The equations of motion are now

\[ \frac{d}{dz}(C x') = \tilde{C}_x x, \quad \frac{d}{dz}(C y') = \tilde{C}_y y, \tag{10.G} \]

and have general solutions

\[ \begin{align*}
    x(z) &= x_0 g_x(z) + x_a h_x(z), \\
    y(z) &= y_0 g_y(z) + y_a h_y(z).
\end{align*} \tag{10.H} \]

It is easy to show from equations (10.G) that the quantities

\[ \begin{align*}
    k_x &= C(g_x h'_x - g'_x h_x) \\
    k_y &= C(g_y h'_y - g'_y h_y)
\end{align*} \]

are constants. We then find

\[ \begin{align*}
    k_x x^{(p)}_c &= h_x \frac{\partial V^{(p)}_{ac}}{\partial x_0} - g_{xc} \frac{\partial V^{(p)}_{ac}}{\partial x_a}, \\
    k_x y^{(p)}_c &= C \left( h'_x \frac{\partial V^{(p)}_{ac}}{\partial x_0} - g'_{xc} \frac{\partial V^{(p)}_{ac}}{\partial x_a} \right), \tag{10.1}
\end{align*} \]

with two analogous expressions for \( y^{(p)}_c \) and \( q^{(p)}_c \).

Setting out from equations (10.H) and (10.1), and taking the symmetry into account, we can express \( x_c \) and \( y_c \) as functions of \( x_0, y_0, x_a \) and \( y_a \), and the corresponding slopes \( x'_0, y'_0, x'_a \) and \( y'_a \); this enables us to calculate the cross-section of the beam at every plane \( z = z_c \) (and in particular, at the Gaussian image plane) and to ascertain the various aberration coefficients.

In order to calculate the aperture aberration coefficients, Hawkes obtains the expressions for \( m^{(2)} \) and \( m^{(4)} \), using the expansions of \( \Phi(x, y, z) \)
(scalar electrostatic potential) and of $A_x$, $A_y$ and $A_z$ given by Glaser (1956); we recall that the expression for the vector potential $A$ chosen by Glaser differs from the (more common) one that has been given in Chapter 3.

In an orthogonal system with a straight axis, consisting of round electrostatic lenses together with (electrostatic or magnetic) quadrupoles and octopoles, oriented in such a way that there are two privileged planes $zOx$ and $zOy$ in which the trajectories remain plane, round magnetic lenses may be included; the function $B(z)$ must, however, not overlap any of the other functions of the type $\Phi(z)$. In a magnetic lens, the equations for $x$ and $y$ are independent only if we use a rotating coordinate system, the rate of rotation of which is given (see Chapter 5) by

$$\frac{d\theta}{dz} = \eta \frac{B(z)}{2 \{\Phi^*(z)\}^{1/2}},$$

the trajectories are then "uncoupled" over two curved orthogonal surfaces defined by $\theta(z)$ and $\theta(z) + \pi/2$. If the round magnetic lens is placed between two series of systems with plane orthogonal surfaces, the planes behind the magnetic lens will have to be twisted through an angle

$$\theta_{1,2} = \int_{z_1}^{z_2} \frac{d\theta}{dz} dz$$

with respect to the planes in front of the magnetic lens. This situation never arises in practice, for $\theta_{1,2}$ would vary with the excitation of the round lens. A system consisting of a round lens followed by quadrupoles and octopoles has, on the other hand, been widely used with a view to correcting the spherical aberration of a round electron microscope objective (see §7.7).

Hawkes' major contribution in this domain is to have put the expressions for the aperture aberration coefficients into forms that can be used directly, in several special cases:

(i) systems of magnetic and electrostatic quadrupoles and octopoles (without round lenses);

(ii) systems containing only one type of quadrupole and octopole lens (magnetic or electrostatic);

(iii) systems with rotational symmetry.

Full expressions for all these coefficients are to be found in Hawkes (1965a, b). We give as an example only those relating to a purely electrostatic stigmatic system (a quadruplet, for example), using the rectangular model approximation ($k(z) = 1$ inside the lenses). Each lens, labelled $i$, will be represented by a function $D(z)$:

$$D(z) = D_i = \text{const}.$$
(or with the notation of §10.2, we should have \( D_i = 4\Phi_i / a^2 \)) and to obtain \( C_x, C_y \) and \( C_{xy} \) in the Gaussian image plane, we must calculate \( g_x, g_y, h_x \) and \( h_y \); in the Gaussian image plane, \( z = z_G \), we have \( g_x = g_y = g \) and \( h_x = h_y = 0 \). If \( z_{ia} \) and \( z_{ib} \) denote the abscissae of the entry and exit planes of the \( i\)-th lens, we have

\[
kC_x = \left( \frac{g}{12\Phi_0^{3/2}} \right) \sum_i \int_{z_{ia}}^{z_{ib}} D_i^2 h_x^4 \, dz - \left( \frac{g}{24\Phi_0^{1/2}} \right) \sum_i \left[ D_i h_x^3 h_x' \right] z_{ib}^2 z_{ia}.
\]

\[
kC_y = \left( \frac{g}{12\Phi_0^{3/2}} \right) \sum_i \int_{z_{ia}}^{z_{ib}} D_i^2 h_y^4 \, dz + \left( \frac{g}{24\Phi_0^{1/2}} \right) \sum_i \left[ D_i h_y^3 h_y' \right] z_{ib}^2 z_{ia}.
\]

\[
kC_{yx} = kC_{xy} = - \left( \frac{g}{8\Phi_0^{3/2}} \right) \sum_i \int_{z_{ia}}^{z_{ib}} \frac{D_i^2}{h_x^2 h_y^2} \, dz + \left( \frac{g}{32\Phi_0^{1/2}} \right) \sum_i 3 D_i \{ h_x^2 (h_y^2)' - (h_x^2)' h_y^2 \} z_{ia} z_{ib}
\]

\[
- \{ h_x^2 (h_y^2)' - (h_x^2)' h_y^2 \} z_{ib}^2 z_{ia}.
\]

We recall that in the Gaussian image plane, \( z_G, k_x = k_y = k \):

\[
k = \sqrt{\Phi_0} (gh' - g'h) = gh' \quad \text{at} \quad z = z_G \quad \text{(no distortion)}.
\]

In conclusion, it is possible to show that the aperture aberration coefficients of a system of quadrupoles and octopoles, or of a round lens followed by such a system, can be cancelled by a judicious choice of the electrical and geometrical parameters.

### 10.7.4 Chromatic Aberration

*The achromatic system.* By superimposing an electrostatic potential \( \Phi(x, y, z) \) and a magnetic potential \( \Phi_m(x, y, z) \) within a single lens, in which the electrodes are turned through 45° with respect to the magnetic poles (to make the system orthogonal), we can design an achromatic mixed lens, as first Kel’man and Yavor (1961, 1963) and later Septier (1963) have shown; the converging power suffers, however. This interesting property has been demonstrated experimentally (Kel’man *et al.*, 1963; Yavor *et al.*, 1964). The presence of the electrostatic lens means that such systems can only be used for relatively low energies.

We reconsider the equations of motion in a mixed quadrupole system (in the non-relativistic case), using the rectangular approximation:

\[
x'' + \left( \frac{K_M}{2\Phi_0} + \frac{eK_M}{mv} \right)x = 0.
\]
The electric field gradient is denoted by $K_E$ and the magnetic field gradient by $K_M$; the equivalent lengths of the magnetic and electrostatic contributions are supposed identical. We may rewrite this

$$x'' + \left( \frac{eK_E}{mv^2} + \frac{eK_M}{mv} \right) x = 0$$

or

$$x'' + A(v) x = 0,$$

so that the system will be achromatic if $\partial A(v)/\partial v = 0$, which implies

$$2 \frac{eK_E}{mv^2} + \frac{eK_M}{mv} = 0.$$

Since $\beta_k^2 = eK_E/mv^2$ and $\beta_M^2 = eK_M/mv$, the condition to be fulfilled is

$$\beta_k^2 = -\frac{1}{2} \beta_M^2$$

and the equations of motion will then be of the form

$$x'' + \left( \frac{1}{2} \beta_M^2 \right) x = 0,$$

$$y'' - \left( \frac{1}{2} \beta_M^2 \right) y = 0.$$

The convergence is thus half as strong as that of the magnetic lens acting alone.

Using the method set out in the preceding section, Hawkes (1965c) has found it possible to calculate the chromatic aberration of rectilinear orthogonal systems (and in particular, of these mixed lenses) in a completely general way, together with the aperture aberration of achromatic lenses. In addition, Septier (1963) and later Dymnikov et al. (1964b, 1965b) have shown that the aperture aberration remains of the same order of magnitude as that of electrostatic or magnetic lenses.
CHAPTER 10A

PRISM OPTICS

The foregoing chapters have been devoted to the study of optical systems with straight axes that have the same effect on charged particles as glass lenses have on photons. We have, moreover, considered only beams that consist of a single type of particle (electrons or identical ions) and are monochromatic (or quasi-monochromatic). Any small differences between the energies or velocities of the various particles are embraced within the "chromatic aberration" of the lenses.

In particle optics, there is a whole range of situations in which we have to analyse the different components of a heterogeneous beam. We then use deflector systems with dispersive properties and by analogy with glass optics, we call such systems "prisms". We shall find that these prisms also possess valuable focusing properties; in fact they form another class of thick optical systems, for which asymptotic cardinal elements can be defined. The straight optic axis of the electron lens is here replaced by a curved axis coinciding with a particular "mean trajectory" (or "equilibrium trajectory") which lies in a plane of symmetry or antisymmetry of the system.

Prisms consist of a pair of electrodes (in the electrostatic case) or of poles (in the magnetic case), which create an electric field $E$ or a magnetic field $B$ directed in such a way that the force exerted on a particle is always normal to the direction in which the particle is travelling, at least in the privileged ("median") plane containing the curved axis.

In its simplest form, a prism consists of a limited region of space containing a homogeneous electric or magnetic field: a plane condenser or a pair of parallel poles very close together. We should use such a system to obtain fairly weak angular deflexions (the beam of electrons in an oscilloscope is deflected in this way for example, see Chapter 12). In order to obtain larger deflexions and a high dispersion, it is better to use long systems with a large radius of curvature in which—for obvious reasons of constructional simplicity and ease of calculation—the mean trajectory (or curved axis) is a segment of a circle centred at $O$ and of radius $R$.

Electrostatic prisms will thus consist of portions of cylindrical, spherical (or even toroidal) condensers and magnetic prisms of circular sector
magnets. The angle at the centre of the prism, $\phi$, is then equal to the deflexion imposed on the mean trajectory of the beam in the median plane.

This median plane, which contains the curved optic axis, is a plane of symmetry in the electrostatic prism and of antisymmetry in the magnetic prism; the scalar potentials in the two cases therefore have different expressions. We shall study the two types of prisms in turn, first examining in detail the first order optical properties (Gaussian optics) and then briefly considering the aberrations.

10A.1 MAGNETIC PRISMS

10A.1.1 Potential and Field

We consider a magnet consisting of two identical poles in the form of circular sectors of angle $\phi$, centre $C$ and mean radius $r_0$, terminated by the plane faces $AB$ and $A'B'$ (Fig. 115a(a)). The gap has a plane of mechanical symmetry (Fig. 115a(b)). The magnetic field is constant along an arbitrary circle $(C)$ of radius $r$ lying in this symmetry plane (the "median" plane)—on $(C_0)$ we have $B = B_0$—and we assume to begin with that $B = 0$ outside the end faces and that $B = \text{constant along } (C)$ for $0 \leq \theta \leq \phi$

![Diagram of a magnetic prism. (a) The mean trajectory $(C_0)$ in the median plane. (b) Section by a vertical plane. (c) Rectangular model: real pole faces at $P$ and $P'$, magnetic virtual faces at $S$ and $S'$ (equal areas).](image)
This is the rectangular model approximation that has already been used for quadrupole lenses. For reasons of symmetry, the field $B$ is parallel to the axis $CZ$ in the median plane. In reality, the fringing fields at the edges extend beyond the poles, and the rectangular model obtained by setting the areas under the curves equal:

$$B_0L = r_0 \int_{-\infty}^{\infty} B_z(\theta) \, d\theta,$$

is slightly longer than the circular arc $(C_0)$:

$$L > r_0 \phi_{\text{mech}},$$

where $\phi_{\text{mech}}$ denotes the angle of the real sectors. We must therefore replace the real sector by a magnetic sector of angle $\phi = L/r_0 > \phi_{\text{mech}}$; the difference $\Delta \phi = \phi - \phi_{\text{mech}}$ depends upon the width of the gap, $D$, and on the geometry of the poles and coils, but it is always true that

$$0.5D < r_0 \Delta \phi < 0.7D.$$

We shall assume henceforward that the faces $AB, A'B'$ are the hypothetical magnetic faces corresponding to the angle $\phi = L/r_0$. It is easy to determine $L$ from measured values of the magnetic field.

Instead of the standard cylindrical coordinates $r, \theta, Z$, with axis $CZ$, it is more convenient to use the reduced dimensionless coordinates $x, z, \theta$, defined by

$$x = \frac{r - r_0}{r_0}, \quad z = \frac{Z}{r_0}, \quad \theta = \frac{s}{r_0}. \quad (10A.1)$$

$(C_0)$ then plays the part of a curved axis, which extends outside the sector as two straight lines $T_0$ and $T_S$ (see Fig. 115a(a)); $ds$ is an element of length along this axis, $(C_0)$.

In the gap, the scalar magnetic potential $\Phi_m$ corresponding to the induction $B$ satisfies Laplace's equation,

$$\nabla^2 \Phi_m = 0. \quad (10A.2)$$

Taking the rotational symmetry into account ($\partial/\partial \theta = 0$) and introducing the new variables $x$ and $z$, equation (10A.2) becomes

$$\frac{\partial^2 \Phi_m}{\partial x^2} + \frac{1}{1 + x} \frac{\partial^2 \Phi_m}{\partial x} + \frac{\partial^2 \Phi_m}{\partial z^2} = 0. \quad (10A.3)$$

In the symmetry plane of the system, which we shall call the “horizontal plane” $H$, $\Phi_m$ is a function of $x$ only. We can therefore write it in the form

$$\Phi(0, x) = \Phi_0(x) = \Phi_0(1 - nx + bx^2 - cx^3 + dx^4 - \cdots). \quad (10A.4)$$

Outside the plane $H$, $\Phi_m$ can be expanded as a polynomial containing only
odd powers of \( z \):

\[
\frac{\Phi_m}{\Phi_0} = - \sum_{n=0}^{\infty} A_{2n+1}(x)z^{2n+1}.
\]

(10A.5)

Substituting equation (10A.5) into (10A.3), we obtain a recurrence relation between the terms \( A_{2n+1}(x) \) and their \( x \)-derivatives, from which we can calculate the first few terms of (10A.5). We obtain, for example,

\[
A_3(x) = - \frac{1}{6} \left\{ (2b - n) + (n + 2b - 6c)x \right. \\
- \left. (n + 2b + 3c - 12d)x^2 + \cdots \right\}.
\]

(10A.6)

The first term, \( A_1(x) \), corresponds to \( n = 0 \) and is given by equation (10A.4). We finally obtain, up to terms of fourth order in \( x \) and \( z \),

\[
\frac{\Phi_m}{\Phi_0} = (-1 + nx - bx^2 + cx^3)z - \{(n - 2b) - (n + 2b - 6c)x\} \frac{z^3}{6} + \cdots.
\]

(10A.7)

Writing

\[
B_z = -\frac{\partial \Phi_m}{\partial Z} = -\frac{1}{r_0} \frac{\partial \Phi_m}{\partial z},
\]

\[
B_x = -\frac{\partial \Phi_m}{\partial r} = -\frac{1}{r_0} \frac{\partial \Phi_m}{\partial x},
\]

(10A.8)

we arrive at the expressions for the two components of the field \( B \):

\[
\frac{B_z}{B_0} = (1 - nx + bx^2 - cx^3) + \frac{1}{2} z^2 \{(n - 2b) - x(n + 2b - 6c)\} + \cdots,
\]

\[
\frac{B_x}{B_0} = (-n + 2bx - 3cx^2)z - (n + 2b - 6c) \frac{z^3}{6} + \cdots
\]

(10A.9)

Only the first terms in these expansions are retained when we study the “first order” optical properties, so that

\[
B_z = (1 - nx)B_0 = B_0 \left(1 - n \frac{r - r_0}{r_0}\right),
\]

(10A.10)

\[
B_x = -B_0 nz = -B_0 n \frac{Z}{r_0}.
\]

The coefficient \( n \) is called the “index” of the field; it is proportional to the transverse gradient of \( B_z \) in the vicinity of \( r_0 \):

\[
n = -\frac{r_0}{B_0} \frac{\partial B_z}{\partial r} = -\frac{1}{B_0} \frac{\partial B_z}{\partial x}.
\]

(10A.11)

When \( n = 0 \), we have \( B_z = B_0 \) and \( B_x = 0 \) so that the magnetic field is homogeneous throughout the sector.
10A.1.2 First Order Trajectory Equations (Gaussian Optics)

We first observe that one of the zero order trajectories is the circle \((C_0)\); this curve is the path of a particle of charge \(e\) and momentum \(p_0 = mv_0 = \sqrt{2em_0V_0^*}\) (with \(V_0^* = V_0(1 + eV_0/2m_0c^2)\): the relativistic potential) chosen such that

\[B_0r_0 = p_0/e, \quad (10A.12)\]

injected along \(T_0\), normal to the entry face \(AB\) at \(S\). For non-relativistic particles (electrons at a few keV or ions up to a few hundred keV), we have \(V_0^* = V_0\), and so

\[B_0r_0 = \sqrt{\frac{2m_0}{e}}V_0. \quad (10A.13)\]

Particles of momentum \(p_1 \neq p_0\) can then also follow circular paths \((C)\) of radius \(r_1 = r_0(1 + x_1)\), satisfying the equation

\[B(x_1)r_1 = p_1/e.\]

We now study the trajectories of particles of momentum \(p_0\), injected near to the curved axis \((C_0)\), \(x, z \ll 1\), with small slopes \((x'^2, y'^2 \ll 1)\). We know that in a static magnetic field, the initial energy of the particle does not change, since the magnetic force

\[f_m = ev \times B \quad (10A.14)\]

is always normal to the direction of the trajectory. We thus deduce that \(m = \text{const.}\) and \(v = \text{const.} = v_0\). Furthermore, since \(v_z\) and \(v_x\) are both much smaller than \(v_0\), we may write

\[v_\theta = v_0 \left\{1 - \left(\frac{v_x}{v_0}\right)^2 - \left(\frac{v_r}{v_0}\right)^2\right\}^{-1/2} \approx v_0,\]

and so

\[v_\theta = r \frac{d\theta}{dt} = r_0 \left(\frac{d\theta}{dt}\right)_0 = v_0\]

or

\[\frac{d\theta}{dt} = \frac{1}{1 + x} \left(\frac{d\theta}{dt}\right)_0 = \frac{\omega}{1 + x}. \quad (10A.15)\]

\(\omega\) is known as the “cyclotron frequency” of the particle in the field \(B_0\). From equation \((10A.12)\), we have

\[\omega = eB_0/m. \quad (10A.16)\]

In the particular coordinate system that we have selected (Fig. 115a), the set of vectors \((r, Z, s)\) are mutually orthogonal. The equations of motion for \(r\) and \(Z\) then take the following form, when we project onto the corre-
sponding axes:

\[
\frac{d^2r}{dt^2} - r \left( \frac{d\theta}{dt} \right)^2 = -\frac{e}{m} r B_z \frac{d\theta}{dt},
\]

\[
\frac{d^2Z}{dt^2} = \frac{e}{m} r B_r \frac{d\theta}{dt}.
\]

Replacing \( B_z \) and \( B_r \) by the expressions given in equation (10A.10), we obtain

\[
\frac{d^2x}{dt^2} + \omega^2(1-n)x = 0,
\]

\[
\frac{d^2Z}{dt^2} + \omega^2 nz = 0.
\]

We can eliminate time with the aid of the formula \( dt = d\theta/\omega \), and we find

\[
\frac{d^2x}{d\theta^2} + (1-n)x = 0,
\]

\[
\frac{d^2Z}{d\theta^2} + nz = 0.
\]

For \( 0 < n < 1 \), the solutions of equations (10A.19) and (10A.20) can be expressed in terms of circular functions: the prism is convergent in both the \( x \)- and \( z \)-directions (that is, in both the horizontal plane \( H \) in which the deflexion occurs and in the plane that is customarily called the "vertical plane", \( V \), containing the axes \( O_z \) and \( O_s \)). If \( n = 0 \) (homogeneous field), there is no focusing in \( V \), but there is again a focusing effect in \( H \). When \( n < 0 \), the plane \( V \) will be a divergent plane and \( H \) will be a convergent plane. We have already encountered such a situation in the chapter on strong-focusing lenses. Finally, we observe that the two equations (10A.19) and (10A.20) are identical when \( n = \frac{1}{2} \), so that the convergence is the same in \( H \) and \( V \).

10A.1.3 Optical Properties of the Prism with Entry and Exit Faces Normal to the Mean Trajectory (\( C_o \))

10A.1.3.1 The general case

The solutions of equations (10A.19) and (10A.20) are of the form

\[
x = A_H \cos (\sqrt{1-n} \theta) + B_H \sin (\sqrt{1-n} \theta),
\]

\[
z = A_V \cos (\sqrt{n} \theta) + B_V \sin (\sqrt{n} \theta).
\]
The constants $A$ and $B$ can easily be expressed as a function of the initial conditions at the entry face, and we obtain the coordinates (of position and slope) of the trajectory in the plane of the exit face, after traversing an angular path of length $\theta = \phi$, in the form

\[
\begin{align*}
x_S &= x_0 \cos (\sqrt{1-n \phi}) + \frac{x_0'}{\sqrt{1-n}} \sin (\sqrt{1-n \phi}), \\
x'_S &= -x_0 \sqrt{1-n} \sin (\sqrt{1-n \phi}) + x_0' \cos (\sqrt{1-n \phi}), \\
z_S &= z_0 (\cos \sqrt{n \phi}) + \frac{z_0'}{\sqrt{n}} \sin (\sqrt{n \phi}), \\
z'_S &= -z_0 \sqrt{n} \sin (\sqrt{n \phi}) + z_0' \cos (\sqrt{n \phi}),
\end{align*}
\]

(10A.21)

(10A.22)

where $x' = dx/d\theta = dr/ds$ and $z' = dz/d\theta = dZ/ds$. From equations (10A.21) and (10A.22), we can define the transfer matrices of the prism, $\|H\|$ and $\|V\|$, in the planes $H$ and $V$ respectively. We write

\[
\begin{pmatrix} x_S \\ x'_S \end{pmatrix} = \|H\| \begin{pmatrix} x_0 \\ x'_0 \end{pmatrix},
\]

with

\[
\|H\| = \begin{pmatrix} \cos (\sqrt{1-n \phi}) & \frac{1}{\sqrt{1-n}} \sin (\sqrt{1-n \phi}) \\ -\sqrt{1-n} \sin (\sqrt{1-n \phi}) & \cos (\sqrt{1-n \phi}) \end{pmatrix}
\]

(10A.23)

\[
= \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}
\]

and

\[
\begin{pmatrix} z_S \\ z'_S \end{pmatrix} = \|V\| \begin{pmatrix} z_0 \\ z'_0 \end{pmatrix},
\]

with

\[
\|V\| = \begin{pmatrix} \cos (\sqrt{n \phi}) & \frac{1}{\sqrt{n}} \sin (\sqrt{n \phi}) \\ -\sqrt{n} \sin (\sqrt{n \phi}) & \cos (\sqrt{n \phi}) \end{pmatrix}
\]

(10A.24)

\[
= \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.
\]

We recall that all distances in real space are measured in units of $r_0$. From $\|H\|$ and $\|V\|$, we can determine the focal lengths, $f_H$ and $f_V$ and the positions of the foci and principal planes (Fig. 115b). The cardinal elements are symmetrical with respect to the plane $\theta_1 = \phi/2$. 

Vol. 1–M
Returning to the unscaled variables, we can easily show that

$$ f'_H = -f_H = \frac{r_0}{\sqrt{1 - n \sin(\sqrt{1 - n \phi})}}; $$

$$ g'_H = -g_H = \frac{r_0}{\sqrt{1 - n \tan(\sqrt{1 - n \phi})}}; \tag{10A.25} $$

$$ s'_H = -s_H = \frac{r_0}{\sqrt{1 - n}} \tan \left( \frac{1}{2} \sqrt{1 - n \phi} \right); $$

$$ f'_V = -f_V = \frac{r_0}{\sqrt{n \sin(\sqrt{n \phi})}}; $$

$$ g'_V = -g_V = \frac{r_0}{\sqrt{n \tan(\sqrt{n \phi})}}; \tag{10A.26} $$

$$ s'_V = -s_V = \frac{r_0}{\sqrt{n}} \tan \left( \frac{1}{2} \sqrt{n \phi} \right). $$

10A.1.3.2 First special case: $n = 0$

The matrices $\|H\|$ and $\|V\|$ now take the simple form

$$ \|H\| = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}, \quad \|V\| = \begin{pmatrix} 1 & \phi \\ 0 & 1 \end{pmatrix}. \tag{10A.27} $$

In the plane $V$, the prism is equivalent to a drift space of length $L = r_0 \phi$, and has no effect on the trajectories. In the plane $H$, it behaves like a thick lens, convergent if $\phi < \pi$ and divergent if $\pi < \phi < 2\pi$. We then have
\[ f'_H = -f_H = r_0 \csc \phi, \]
\[ g'_H = -g_H = r_0 \cot \phi, \]
\[ s'_H = -s_H = r_0 \tan (\phi/2). \]

The diagram of Fig. 115c(a) shows that the principal planes lie at \( P \), the point at which \( T_0 \) and \( T_S \) intersect; we can therefore replace the prism by a thin lens of focal length of \( f = r_0 \csc \phi \) at \( P \) and two drift spaces \( SP = PS' = r_0 \tan (\phi/2) \). The length of this optical system is greater than that of the real trajectory \( L = r_\phi \), however, and it is thus better to consider the thick lens of length \( L \) with crossed principal planes (Fig. 115c(b)). For \( \phi = 90^\circ \), the foci lie at \( S \) and \( S' \) and we have \( f'_H = r_0 \).

**THE CORRESPONDENCE BETWEEN OBJECT AND IMAGE.** We consider an object \( P \), a distance \( p \) from \( S \), and its image \( Q \), distant \( S'Q = q \) from \( S' \). In reduced coordinates, \( p/r_0 \) and \( q/r_0 \), the complete transfer matrix between \( P \) and \( Q \) becomes

\[
\begin{pmatrix} T_11 & T_12 \\ T_21 & T_22 \end{pmatrix} = \begin{pmatrix} 1 & q/R \\ 0 & 1 \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} 1 & p/R \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}.
\]

(10A.29)

For a ray leaving \( P \) \( (x_p = 0) \) at slope \( x'_p \neq 0 \), we see that at \( Q \)

\[ x_Q = x'_p T_{12} = 0, \]

so that

\[ T_{12} = 0. \]

(10A.30a)

![Diagram](a)

![Diagram](b)

**Fig. 115c.** The special case of the homogeneous field prism (faces normal to \((C_0)\)). (a) Thin lens representation. (b) Representation in terms of thick centred systems.
Equation (10A.30a) is the condition that $P$ and $Q$ be conjugate points. $T_{12} = 0$ implies that

$$\tan \phi = \frac{(p+q)/r_0}{1-pq/r_0^2}.$$  \hspace{1cm} (10A.30b)

From Fig. 115d, we see that since $CS = CS' = r_0$, we have

$$\tan \phi = -\frac{\tan \phi_P + \tan \phi_Q}{1 - \tan \phi_P \tan \phi_Q}$$

$$= -\tan (\phi_P + \phi_Q),$$

and hence

$$\phi + \phi_P + \phi_Q = \pi.$$  \hspace{1cm} (10A.31)

Thus the points $P$, $C$ and $Q$ are collinear (Barber's rule). Once $P$ has been selected, and knowing $T_0$ and $T_s$, we have only to produce $PC$ to obtain $Q$. *Stigmatic operation* is obtained when

$$p = q = r_0 \cot (\phi/2).$$

![Fig. 115d. Correspondence between object and image in a homogeneous field prism (faces normal to $(C_0)$).](image)

**Linear Magnification.** Consider now a trajectory for which $x_p \neq 0$. Knowing that $T_{12} = 0$, we have

$$x_q = T_{11} x_p.$$  

The linear magnification, $G_H$, is given by

$$G_H = T_{11} = \cos \phi - q/R \sin \phi.$$  \hspace{1cm} (10A.32a)

Using equation (10A.31), we can easily show that

$$G_H = -\frac{\cos \phi_P}{\cos \phi_Q}$$

$$= -\frac{CQ}{CP}.$$  \hspace{1cm} (10A.32b)

If $p = q$, we have $G_H = -1$. 
10A.1.3.3 Second special case: \( n = \frac{1}{2} \)

For this value of \( n \), equations (10A.19) and (10A.20) are identical and the prism is equivalent to a stigmatic, centred, rotationally symmetric system (apart from the deflexion through an angle \( \phi \)), for particles of the same velocity \( v \). We now have

\[
\begin{align*}
  f' &= -f = r_0 \sqrt{2} \csc(\phi/\sqrt{2}), \\
  g' &= -g = r_0 \sqrt{2} \cot(\phi/\sqrt{2}).
\end{align*}
\]

(10A.33)

If \( \phi = \pi/\sqrt{2} \approx 127^\circ \), the foci are situated at \( S \) and \( S' \), and \( f' = r_0 \sqrt{2} \). Symmetrical operation \( (p = q) \) is obtained for \( |p-g| = |f| \), or

\[
  p = q = r_0 \sqrt{2} \cot\left(\frac{1}{2} \frac{\phi}{\sqrt{2}}\right).
\]

(10A.34)

We then have \( G_H = G_V = -1 \).

In the general case \( (p \neq q) \), the relation between conjugates takes the form

\[
\tan(\phi/\sqrt{2}) = -\frac{(p+q)/r_0 \sqrt{2}}{1-pq/2r_0^2}
\]

or

\[
q/r_0 = \frac{(p/r_0) \sqrt{2} + 2 \tan(\phi/\sqrt{2})}{(p/r_0) \tan(\phi/\sqrt{2}) - \sqrt{2}}.
\]

(10A.35)

Barber’s rule still applies, provided we consider a fictitious system of angle \( \phi' = \phi/\sqrt{2} \), of the same radius \( r_0 \), and divide \( p \) and \( q \) by \( \sqrt{2} \). The corresponding points \( P' \) and \( Q' \) will be collinear with \( C \).

For a given angle \( \phi \) and the same radius of curvature \( r_0 \), a prism of index \( n = \frac{1}{2} \) is less convergent in the \( H \) plane than the corresponding prism with homogeneous field, but has the advantage of convergence in the \( V \) plane.

10A.1.4 Method of Obtaining a Field of Index \( n \)

If the material of which the poles are composed has a very high permeability \( (\mu \rightarrow \infty) \), the surfaces of the poles are equipotentials. If each of the poles is excited by a coil of \( N \) turns carrying a current \( I \), the magnetic potential will be given by \( \pm \Phi_1 = \pm \mu_0 NI \). When the width of the gap is much less than \( r_0 \), we can use equation (10A.9) and write down the equation in \( x \) and \( z \) that gives the profile of the pole held at potential \( \Phi_1 \):

\[
-\frac{\Phi_1}{\Phi_{om}} = (1-nx+bx^2-cx^3)z + \{(n-2b)-(n+2b-6c)x\} \frac{z^3}{3}.
\]
At \( x = 0 \), the reduced width of the gap is \( 2h/r_0 \). To first order, therefore,

\[
-\Phi_1 = \frac{h}{r_0} \Phi_0 = B_0 h = \mu_0 NI
\]  

(10A.36)

and

\[
z = \frac{h}{r_0} (1 - nx)^{-1}. \tag{10A.37}
\]

In a first approximation, the field with index \( n \) will therefore be obtained with poles of hyperbolic cross-section, defined by equation (10A.37). In order to simplify their mechanical construction, the hyperbola is usually replaced by its tangent at the point \( (x = 0, z = h) \); the pole pieces are then portions of cones of revolution, with meridian

\[
z = \frac{h}{r_0} (1 + nx). \tag{10A.38}
\]

When \( n = \frac{1}{2} \),

\[
z_{(1/2)} = \frac{h}{r_0} (1 + x/2) \tag{10A.39}
\]

and the apex of the cone, \( C' \), is at \( x = -2 \), that is at a point distant \( 2r_0 \) from \( (C_0) \) (Fig. 115e). The slope of the tangent is then

\[
\tan \alpha = h/2r_0. \tag{10A.40}
\]

The index \( n = \frac{1}{2} \) is only obtained in a narrow zone on either side of \( (C_0) \) and the prism must only be used with beams of small diameter.

In the special case for which \( n = 2b \) (for example, \( n = \frac{1}{2} \) and \( b = \frac{1}{4} \)), a rigorous solution of Laplace's equation, \( \nabla^2 \Phi_m = 0 \), can be found (Bosi, 1965, 1967):

\[
-\frac{\Phi_m}{\Phi_0} = z\{1 - n \log (1 + x)\}. \tag{10A.41}
\]

The equation of the pole is then given by

\[
z = -\frac{\Phi_1/\Phi_0}{1 - n \log (1 + x)} = -\frac{h}{1 - n \log (1 + x)}. \tag{10A.42}
\]

Fig. 115e. The pole cross-section that gives \( n = \frac{1}{2} \) \((H, H')\): hyperbolae; \( T, T' \): tangents at \( A, A' \).
This expression coincides with equation (10A.37) in first order, and at \( x = 0 \) equation (10A.36) is satisfied exactly.

10A.1.5 Properties of the Prism with Inclined Entry and Exit Faces

So far, we have assumed the mean trajectory \( T_0T_S \) to be normal to the entry and exit faces of the prism; the same is true to a first approximation of the other trajectories, in Gaussian optics. Let us now suppose that the entry and exit faces are turned (about \( S \) and \( S' \)) through angles \( \alpha \) and \( \beta \) respectively, with respect to their initial orientation. By convention, the angle of rotation is taken to be positive when the normal to the tilted face is outside the mean trajectory \( T_0(C_0)T_S \) (Fig. 115f). For simplicity, we only consider here the case of the prism with homogeneous magnetic field \( (n = 0) \), and it is, in fact, extremely rare for prisms to be constructed with tilted faces and \( n \neq 0 \). To a first approximation, however, the results are true for any value of \( n \).

![Fig. 115f. Tilted faces: definition of the signs of \( \alpha \) and \( \beta \).](image)

10A.1.5.1 Trajectories in the median plane

Trajectories that arrive at the inclined entry face at a distance \( r = r_0(1 + x) \) with \( x > 0 \) ("external" trajectories) will experience the action of \( B_z \) over a distance \( L' \) which is less than \( L = r_0\phi \) (when \( \alpha > 0 \)); they will therefore be deflected through an angle \( \phi_{\text{ext}} < \phi \). For "internal" trajectories \( (x < 0) \), on the other hand, we have \( L' > L \) and hence \( \phi_{\text{int}} > \phi \). The rotation of the faces will thus modify the convergence in the \( H \) plane. In order to calculate the effect of a rotation of the entry face through an angle \( \alpha \), we can reason as follows (Enge, 1967): when the face \( AB \) is rotated to \( A_1B_1 \), a small prism \( SAA_1 \) of angle \( \alpha \) (Fig. 115g) is added to the original prism \( (P) \), and in it the field \( B_z \) is in the same sense as in \( (P) \); simultaneously, a small prism \( SSB_1 \), also of angle \( \alpha \), is removed from \( (P) \). The new prism with tilted face is thus equivalent to the prism \( (P) \) provided that we add to the latter a double prism of angle \( \alpha \) consisting of two symmetrical parts. \( SAA_1 \) and \( SBB_1 \). In \( SAA_1 \), the field \( B_z \) has the same ampli-
tude and sense as the main field in \((P)\), but in \(SBB_1\), the field has the same amplitude as that in \((P)\) but is in the opposite direction.

If \(x\) is always much smaller than unity, the paths of the beams in the small prisms will be short, so that to a first approximation the double prism will be equivalent to a thin divergent lens, which causes only a refraction (that is, a change of slope) \(\epsilon\), without affecting the position \(x\). A trajectory \(T_1\), parallel to \(T_0\) (Fig. 115h), travels a distance

\[
\text{d}l = r_0x \tan \alpha
\]

in the prism. The radius of curvature of the trajectory is equal to \(r_0\), since \(|B| = B_0\), and hence

\[
r_0\epsilon = r_0x \tan \alpha,
\]

so that

\[
\epsilon = x \tan \alpha.
\]

The trajectory \(T_2\) experiences an equal and opposite refraction, \(-\epsilon\). The double prism is thus equivalent to a thin divergent lens of focal length

\[
f_{EH} = \frac{-r_0x}{\epsilon} = -r_0 \cot \alpha.
\]

If \(\alpha < 0\), the lens will be convergent.

Similarly the exit double prism will be equivalent to a lens of focal length

\[
f_{SH} = -r_0 \cot \beta \quad (\beta > 0 \text{ or } < 0).
\]

The transfer matrix of a thin lens of convergence \(c = 1/f\) is of the form

\[
\|T\| = \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix}.
\]

The total transfer matrix of the prism \((P')\) with inclined end-faces is therefore given in the \(H\) plane by the product of three matrices:
PRISM OPTICS

\[\|H\| = \begin{pmatrix} 1 & 0 \\ \tan \beta & 1 \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \tan \alpha & 1 \end{pmatrix},\]

which yields

\[\|H\| = \begin{pmatrix} \frac{\cos (\phi - \alpha)}{\cos \alpha} & \sin \phi \\ -\frac{\sin \{\phi - (\alpha + \beta)\}}{\cos \alpha \cos \beta} & \frac{\cos (\phi - \beta)}{\cos \beta} \end{pmatrix}.\] (10A.46)

In the symmetrical situation for which \(\alpha = \beta\), \(\|H\|\) simplifies to

\[\|H\| = \begin{pmatrix} \frac{\cos (\phi - \alpha)}{\cos \alpha} & \sin \phi \\ -\frac{\sin (\phi - 2\alpha)}{\cos^2 \alpha} & \frac{\cos (\phi - \alpha)}{\cos \alpha} \end{pmatrix}.\] (10A.47)

When \(\phi = \alpha + \beta\), the entry and exit planes are parallel; the prism \((P')\) then behaves as a "parallel plate" and in the \(H\) plane, its convergence is zero, since \(H_{21} = 0\). If \(\alpha + \beta < \phi\), the prism is convergent in the \(H\) plane, and we have

\[f'_H = -f_H = \frac{r_0 \cos \alpha \cos \beta}{\sin \{\phi - (\alpha + \beta)\}},\]

\[g'_H = r_0 \frac{\cos \beta \cos (\phi - \alpha)}{\sin \{\phi - (\alpha + \beta)\}},\]

\[g_H = -r_0 \frac{\cos \alpha \cos (\phi - \beta)}{\sin \{\phi - (\alpha + \beta)\}}.\] (10A.48)

\(g_H = -g'_H\) when \(\alpha\) and \(\beta\) are equal.

10A.1.5.2 Trajectories in the vertical plane

In the prism \((P)\), where \(\alpha = \beta = 0\), there are zones in the vicinity of the faces in which the field falls rapidly from \(B = B_0\) to \(B = 0\); the field lines become curved outside the prism (Fig. 115i). A component \(B_\theta \neq 0\) appears, which has virtually no effect on the particles since \(B_\theta\) is parallel to the velocity up to terms in second order (it must be taken into account in calculating the aberrations). Because of the symmetry, we still have \(B_r = 0\), but if the faces are tilted, the field lines remain plane and normal to the surfaces of the poles; we thus find that \(B_r\) no longer vanishes. We thus write (in unscaled coordinates \(Z\) and \(s\)
Fig. 115i. The appearance of the lines of force of $B$ in the fringing field zone near the entry face. $F$: magnetic virtual face.

\[
\frac{d^2 Z}{dt^2} = \frac{e}{m} vB_r \quad \text{with} \quad v = \frac{ds}{dt}. \tag{10A.49}
\]

\[
\frac{d^2 Z}{ds^2} = -\frac{B_r}{r_0B_0}.
\]

We do not know the expression for $B_r$, but we do know that \( \text{curl} \ B = 0 \), so that

\[
\frac{\partial B_r}{\partial Z} = \frac{\partial B_z}{\partial r}. \tag{10A.50}
\]

Integrating over the fringing field region, we obtain

\[
\frac{\partial}{\partial Z} \int B_r \, ds = \frac{\partial}{\partial r} \int B_z \, ds.
\]

In the \((r, s)\) plane, we have characterized the action of this fringing field by that of the equivalent double prism containing the homogeneous field $B_o$ and at the entry, we have

\[
\int B_z \, ds = B_o r_0 \tan \alpha \tag{10A.51}
\]

according to equation (10A.43). Since $\partial r = r_0 \partial x$ and $\partial Z = r_0 \partial z$, we obtain

\[
\frac{\partial}{\partial z} \int B_r \, ds = B_o \tan \alpha
\]

or

\[
\int B_r \, ds = B_o z \tan \alpha. \tag{10A.52}
\]

In the vertical plane, an incident ray is thus refracted by an amount

\[
\frac{dZ}{ds} = -\frac{1}{r_0 B_0} \int B_r \, ds = -\frac{Z}{r_0} \tan \alpha. \tag{10A.53}
\]

The double prism is therefore equivalent to a thin lens in the $V$ plane, of
focal length

\[ f_{EV} = -\frac{Z}{dZ/ds} = r_0 \cot \alpha = -f_{EH}. \tag{10A.54} \]

Similarly, at the exit plane we have

\[ f_{SV} = r_0 \cot \beta = -f_{SH}. \tag{10A.55} \]

The relations (10A.44, 10A.45) and (10A.54, 10A.55) remain valid to a first approximation when \( n \neq 0 \).

A tilted face thus has the same effect as a thin strong-focusing lens: divergent in the \( H \) plane and convergent in the \( V \) plane (or conversely, depending on the sign of \( \alpha \) or \( \beta \)). This brings out the reason for this modification to the end faces when \( n = 0 \): the prism with homogeneous field, which is much the easiest to construct mechanically, becomes convergent in the \( V \) plane and remains convergent in the \( H \) plane. These features were foreseen by Cotte as early as 1938. The transfer matrix \( \| \mathbf{V} \| \) therefore becomes

\[
\begin{pmatrix}
1 & 0 & 1 & \phi \\
-\tan \beta & 1 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
\phi \tan \alpha & 1
\end{pmatrix}
\begin{pmatrix}
1 - \phi \tan \alpha & \phi \\
\phi \tan \alpha \tan \beta - (\tan \alpha + \tan \beta) & 1 - \phi \tan \beta
\end{pmatrix}.
\tag{10A.56}
\]

We therefore have

\[ f' = -f_{SV} = \frac{r_0}{\tan \alpha + \tan \beta - \phi \tan \alpha \tan \beta}, \]

\[ g' = \frac{r_0 (1 - \phi \tan \alpha)}{\tan \alpha + \tan \beta - \phi \tan \alpha \tan \beta}, \tag{10A.57} \]

\[ g = -\frac{r_0 (1 - \phi \tan \beta)}{\tan \alpha + \tan \beta - \phi \tan \alpha \tan \beta}. \]

10A.1.5.3 Stigmatic operation

With a homogeneous field prism with tilted faces, we can achieve stigmatic operation but, unlike the case \( n = \frac{1}{3} \), the system is stigmatic for only one particular pair of conjugate planes, since the cardinal elements (\( f \) and \( g \)) are different in the planes \( H \) and \( V \).

In the case of a symmetric system (\( \alpha = \beta, p = q \)), we can easily calculate the values of \( \tan \alpha \) and \( p/r_0 \) for which this type of operation obtains.
We establish the overall transfer matrices $\|T_H\|$ and $\|T_V\|$ (see equation 10A.29) and then write down the conjugacy relation for each:

$$T_{H12} = 0,$$

$$T_{V12} = 0.$$  

(10A.58)  

(10A.59)

From equation (10A.59), we derive

$$\tan \alpha = \frac{r_0}{p}.$$  

(10A.60)

Substituting (10A.60) into (10A.58), we finally obtain

$$\frac{p}{r_0} = 2 \cot (\phi/2),$$

$$\tan \alpha = \frac{1}{2} \tan (\phi/2).$$  

(10A.61)

The magnification $G$ is $-1$ in both planes.

**Example:** The prism with deflexion $\pi/2$. The matrices $\|H\|$ and $\|V\|$ take the form

$$\|H\| = \begin{pmatrix} \tan \alpha & 1 \\ \frac{1 - \tan^2 \alpha}{2} \tan \alpha \end{pmatrix},$$

$$\|V\| = \begin{pmatrix} 1 - \frac{\pi}{2} \tan \alpha \\ -\tan \alpha (2 - \frac{\pi}{2} \tan \alpha) \\ 1 - \frac{\pi}{2} \tan \alpha \end{pmatrix}.$$

(10A.62)

and equation (10A.61) gives

$$\tan \alpha = \frac{1}{2} \quad (\text{or} \quad \alpha = 26^\circ34'),$$

$$p = 2r_0.$$  

(10A.63)

A device incorporating a stigmatic prism of this type ($\phi = \pi/2$) has been successfully used by Castaing and Slodzian (1962a, b) to transport the image in ion microscopy and to act as a mass selector simultaneously.

### 10A.1.6 The Effect of the Fringing Fields

#### 10A.1.6.1 In the $H$ plane

In a real magnet, the trajectories do not enter the region containing the field $B_z$ abruptly; a particle of momentum $p_0$ is progressively deflected, and the radius of curvature of the trajectory decreases from $r = \infty$ to $r = r_0$ (if $B = B_0$). The trajectory that follows the circle ($C_0$) inside the prism does not coincide with $T_0$ and $T_S$ outside, but corresponds instead to the trajectories $T_0'$ and $T_S'$ (Fig. 115j). A more detailed analysis shows
Fig. 115j. The effect of the fringing field in the $H$ plane. (a) Shift of the trajectories. $A_1B_1$ and $A'B'_1$ are the mechanical faces and $AB, A'B'$ are the magnetic faces that form the edges of the rectangular model; in reality, the fringing fields extend to $PP'$. (b) Curvature of the magnetic face. (c): misorientation of the magnetic virtual face when the mechanical face is tilted, $\beta < \beta_1$.

that $T'_0$ and $T'_S$ are parallel to $T_0$ and $T_S$ respectively, but are slightly shifted towards the centre of curvature $C$ by an amount $\Delta x_1$. The effect can be corrected by slightly altering the relative positions of the prism and the ideal path selected initially. With the new curved axis $T'_0(C_0)T'_S$, the formulae derived above are directly applicable between the virtual magnetic faces $AB$ and $A'B'$. Furthermore, the virtual magnetic faces obtained from measurements of the magnetic field distribution by the method suggested in §10A.1.1 are slightly curved, since the dimensions of the poles in the radial direction are finite, and are convex towards the outside of the prism (Fig. 115k). This curvature introduces aberrations.

10A.1.6.2 In the vertical plane

When the entry and exit faces $(AB, A'B')$, which are assumed to be of infinite extent in the radial direction, are inclined $(\alpha, \beta \neq 0)$, it can be shown by calculating the trajectories rigorously (Enge, 1964) that the convergence of the prism in the $V$ plane is slightly weaker than that predicted by the formulae given above. This error can be compensated by increasing the angles $\alpha, \beta$ of the mechanical faces to values $\alpha', \beta'$ slightly
larger than those given by the theory. When the width of the poles is small, on the other hand (Fig. 115k), the rotation of the magnetic face is less than that of the mechanical face, and this effect becomes greater as $r_0$ is made smaller. The angle $\beta_1$ of the mechanical faces must therefore be adjusted in such a way that the desired rotation $\beta$ of the magnetic faces is achieved. Accurate magnetic measurements are necessary when designing a prism with tilted faces. Alternatively, the magnet may be equipped with rotatable mechanical faces; this is particularly convenient when rigorously stigmatic operation is required.

10A.1.7 The Dispersive Properties of a Magnetic Prism

10A.1.7.1 Trajectories and transfer matrices for particles of momentum $p_0 + \Delta p$

Hitherto, we have been considering particles of charge $e$, all having the same momentum $p_0 = mv_0 = eB_0r_0$. We now consider a particle incident along the trajectory $T_0$, entering the prism at $S$, but having momentum

$$p = p_0 + \Delta p = p_0(1 + \Delta p/p_0) = p_0(1 + \delta),$$

where $\delta \ll 1$. In a prism of arbitrary index $n$, the circular equilibrium trajectory corresponding to this momentum will be a circle $(C)$ of radius $r = r_0 + dr$, such that

$$(r_0 + dr)B_z = p_0(1 + \delta)/e$$

and

$$B_0r_0(1 + x)(1 - nx) = p_0(1 + \delta)/e,$$

$$x = \frac{\delta}{1 - n}.$$  \hspace{1cm} (10A.65)

With respect to the new curved axis $(C)$, defined by these conditions, the initial values where the particle enters the prism are as follows:

$$x_0 = -\delta/(1 - n), \quad x_0' = 0.$$  \hspace{1cm} (10A.64)

The transfer matrix $|H|$ of the prism is again given by equation (10A.23). The trajectory is therefore of the form

$$x_{(C)} = x_0 \cos \left(\sqrt{1 - n} \theta\right),$$
$$x'_{(C)} = -x_0 \sqrt{1 - n} \sin \left(\sqrt{1 - n} \theta\right).$$ \hspace{1cm} (10A.66)

If we now measure the abscissa, $x$, with respect to the curved axis $(C_0)$ corresponding to momentum $p_0$, we find at the exit of the prism

$$x_s = \left\{1 - \cos \left(\sqrt{1 - n} \phi\right)\right\} \frac{\delta}{1 - n},$$
$$x'_s = \frac{\delta}{\sqrt{1 - n}} \sin \left(\sqrt{1 - n} \phi\right).$$ \hspace{1cm} (10A.67)
We can thus extend the matrix $\|H\|$ to include the effect on $x_s$ and $x_s'$ of a variation $\Delta p/p_0 = \delta$: for $\alpha = \beta = 0$, we may write

$$
\begin{pmatrix}
x_s' \\
x_s'' \\
\delta
\end{pmatrix}
= \|H\| \begin{pmatrix}
x_0 \\
x_0' \\
\delta
\end{pmatrix}
= \begin{pmatrix}
H_{11} & H_{12} & H_{13} \\
H_{21} & H_{22} & H_{23} \\
H_{31} & H_{32} & H_{33}
\end{pmatrix}
\begin{pmatrix}
x_0 \\
x_0' \\
\delta
\end{pmatrix},
$$

(10A.68)

with

$$
H_{11} = \cos (\sqrt{1 - n} \phi), \quad H_{12} = \frac{1}{\sqrt{1 - n}} \sin (\sqrt{1 - n} \phi)
$$

$$
H_{13} = \frac{1}{1 - n} \{ 1 - \cos (\sqrt{1 - n} \phi) \}
$$

$$
H_{21} = -\sqrt{1 - n} \sin (\sqrt{1 - n} \phi), \quad H_{22} = \cos (\sqrt{1 - n} \phi)
$$

(10A.69)

$$
H_{23} = \frac{1}{\sqrt{1 - n}} \sin (\sqrt{1 - n} \phi)
$$

$$
H_{31} = H_{32} = 0, \quad H_{33} = 1
$$

When the entry and exit faces are tilted ($\alpha, \beta \neq 0$), we find

$$
H_{11} = \cos (\sqrt{1 - n} \phi) + \frac{1}{\sqrt{1 - n}} \tan \alpha \sin (\sqrt{1 - n} \phi)
$$

$$
H_{21} = \cos (\sqrt{1 - n} \phi) (\tan \alpha + \tan \beta)
$$

$$
-\sqrt{1 - n} \sin (\sqrt{1 - n} \phi) \left( 1 - \frac{1}{1 - n} \tan \alpha \tan \beta \right)
$$

(10A.70)

$$
H_{22} = \cos (\sqrt{1 - n} \phi) + \frac{1}{\sqrt{1 - n}} \tan \beta \sin (\sqrt{1 - n} \phi)
$$

$$
H_{23} = \frac{1}{\sqrt{1 - n}} \sin (\sqrt{1 - n} \phi) + \frac{1}{1 - n} \{ 1 - \cos (\sqrt{1 - n} \phi) \}.
$$

$H_{12}, H_{13}, H_{31}, H_{32}$ and $H_{33}$ are unchanged (see 10A.69).

When $n = 0$ (homogeneous field), $\|H\|$ simplifies:

$$
\begin{pmatrix}
\|H\| \\
(n = 0)
\end{pmatrix}
= \begin{pmatrix}
\cos (\phi - \alpha) & \sin \phi & 1 - \cos \phi \\
\sin \{ \phi - (\alpha + \beta) \} & \cos (\phi - \beta) & \sin \phi + (1 - \cos \phi) \tan \beta \\
\cos \alpha \cos \beta & \cos \beta & 0
\end{pmatrix}
$$

(10A.71)

The transfer matrix in the $V$ plane is unaltered, so far as the first order properties are concerned at least.
10A.1.7.2 Dispersion

10A.1.7.2.1 Definitions. The dispersive power of a prism—its ability to separate (in the image plane) particles of slightly different momenta originating in the same object point on the axis \((x_0 = 0)\) by an amount \(x_i\) (Fig. 115k)—is characterized by a dimensionless quantity \(\mathcal{D}_p\):

\[
\mathcal{D}_p = x_i \frac{p_0}{\Delta p} = \frac{x_i}{\delta}.
\]

(10A.72)

Fig. 115k. Dispersion of the prism.

Writing \(SP = p\) and \(SQ = q\), the total transfer matrix between \(P\) and \(Q\) takes the form

\[
\|T\| = \begin{pmatrix}
1 & q/r_0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
1 & p/r_0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

\[
= \begin{pmatrix}
T_{11} & T_{12} & T_{13} \\
T_{21} & T_{22} & T_{23} \\
T_{31} & T_{32} & T_{33}
\end{pmatrix}.
\]

Since \(P\) and \(Q\) are conjugates,

\[T_{12} = 0,
\]

and hence

\[
\frac{p}{r_0} H_{11} + \frac{q}{r_0} H_{22} + \frac{pq}{r_0^2} H_{21} + H_{12} = 0,
\]

(10A.73)

where \(H_{ij}\) are the elements of \(\|H\|\). In the image plane \(Q\), we have

\[x_i = x_0 T_{11} + x_0 T_{12} + \delta T_{13},
\]

which reduces to

\[x_i = \delta T_{13}.
\]

We therefore find

\[
\mathcal{D}_p = T_{13} = H_{13} + \frac{q}{r_0} H_{23}.
\]

(10A.74)
We recall that the linear magnification in the $H$ plane is given by

$$G = T_{11} = H_{11} + \frac{q}{r_0} H_{21}. \quad (10A.75)$$

Examining the matrices $\|H\|$ given in equation (10A.69), corresponding to systems with no inclination of the entry and exit faces, we see that

$$H_{13} = \frac{1}{1-n} (1-H_{11}), \quad H_{23} = H_{12} = -\frac{1}{1-n} H_{21},$$

and hence

$$\mathcal{D}_p = H_{13} + \frac{q}{r_0} H_{23} = \frac{1}{1-n} \left\{ 1 - \left( H_{11} + \frac{q}{r_0} H_{21} \right) \right\}$$

or

$$\mathcal{D}_p = \frac{1}{1-n} (1-G_H). \quad (10A.76)$$

This extremely useful relation was proved analytically by H. Bruck (1958).

10A.1.7.2.2 Special cases. We assume symmetric conditions ($p = q$, $G = -1$).

(i) Homogeneous field prism ($n = 0$). When the faces are normal to $T_0$ ($\alpha = \beta = 0$), we have $p = q = r_0 \cot(\phi/2)$ and we find that

$$\mathcal{D}_p = 2. \quad (10A.77)$$

When the faces are tilted and $\alpha = \beta$, $\tan\alpha = \frac{1}{2} \tan(\phi/2)$ and $p = q = 2 \cot(\phi/2)$. We find

$$\mathcal{D}_p = 4. \quad (10A.78)$$

The dispersion is thus doubled by the use of inclined faces.

(ii) Prisms of non-zero index. For $\alpha = \beta = 0$, we obtain

$$\mathcal{D}_p = \frac{2}{1-n},$$

so that for $n = 1/2$,

$$\mathcal{D}_p = 4. \quad (10A.79)$$

In all these special cases, $\mathcal{D}_p$ is independent of $\phi$.

10A.1.7.2.3 The non-dispersive system. When we wish simply to deflect a beam consisting of particles of different momenta and focus it bodily without dispersion, a prism (or a more complicated optical system) for which the term $T_{13}$ of the total transfer matrix vanishes must be employed. With $n = 0$ and $\alpha = \beta = 0$, this condition implies $\phi = 2\pi$ so that there would be no overall deflection. With $\alpha \neq \beta (n \neq 0)$, however, non-dispersive prisms can be obtained. Figure 1151, for example, shows
a solution to this problem (net deflection of 90°) given by Enge (1967): a beam incident parallel to $T_0$ is focused in both the $H$ and the $V$ planes, and also in energy, at the image focus $F'$. The parameters of the system are then $\phi = 270^\circ, \alpha = 45^\circ, \beta = 32^\circ40', p = \infty$ and $q = 2.74 \, r_0$.

10A.1.7.2.4 Mass dispersion. When the incident beam consists of ions of different masses, all carrying the same charge $e$ and accelerated through the same potential $V_0$, the different masses can be separated with the aid of a prism. A mass dispersion, $\mathcal{D}_m$, may be defined by the relation

$$\mathcal{D}_m = \frac{m_0}{\delta m},$$

where $\delta m$ is the difference in mass between two ions of similar mass in the vicinity of the mass $m_0$. If $\delta m \ll m_0$, the relation $mv = p = \sqrt{2emV_0}$ implies

$$\frac{\delta p}{p_0} = \frac{1}{2m_0} \delta m,$$

and hence

$$\mathcal{D}_m = \frac{1}{2} \mathcal{D}_p.$$

10A.1.7.2.5 Theoretical resolving powers $R_p$ and $R_m$. In a prism free of aberrations, all particles of the same momentum $p_0$, from a slit source (object) of reduced width $a_0/r_0$ in the $H$ plane and situated at $P$, will pass through the exit slit of reduced width $a/r_0$ situated at $Q$ (the image of $P$);
we have

\[ a_i = |G_H|a_0. \]

In a first approximation, the particles of momentum \( p_i = p_0 + \Delta p \) will be focused over an image of the same width \( a_i \) but situated at a distance \( x_i = D_p(\Delta p/p_0) \) from the axis (Fig. 115m).

\[ \begin{align*}
  &\text{Object} \quad \underline{a_0} \\
  &\text{Images} \quad \underline{a_1} \\
  &S \quad p_0 \quad \Delta p \\
  &S' \quad p_0 + \Delta p \\
  &x_i \\
\end{align*} \]

**Fig. 115m.** Images when the particles have different momenta.

If we wish to record particles of momentum \( p_0 \) and those of momentum \( p_0 + \Delta p_i \) simultaneously on a double collector, we can establish the minimum value of \( \Delta p \), \( \Delta p_{\text{min}} \), such that no particle of momentum \( p < p_0 - \Delta p_{\text{min}} \) or \( p > p_0 + \Delta p_{\text{min}} \) passes through the image slit of width \( a_i \) lying on the axis. We then have \( a_i = x_i \) and

\[ D_p \frac{\Delta p_{\text{min}}}{p_0} = |G_H|a_0. \]

The theoretical resolving power, in the first order approximation, is then defined by

\[ R_p = \frac{p_0}{\Delta p_{\text{min}}} = \frac{D_p}{|G_H|a_0}. \] (10A.82)

If, on the other hand, we have a single collector behind a slit of reduced width \( s_0 = s/r_0 \) placed on the axis (and this case occurs more frequently), we may pass particles of different momenta across this slit (by varying \( B_0 \)); particles of momentum \( p_0 \), focused into an image of width \( a_i \), cannot all be completely separated from particles of momentum \( p_0 + \Delta p \) unless (Fig. 115m)

\[ x_i = a_i + s_0 \]

or

\[ D_p \frac{\Delta p_i}{p_0} = |G_H|a_0 + s_0. \]

Hence,

\[ R_p = \frac{D_p}{|G_H|a_0 + s_0} \] (10A.83)
since the electric signal recorded by the collector C is trapezoidal and its width is \( a_t + s_0 \).

When a source of given width \( a \) is used, it will be advantageous to increase \( D_p \), the ratio \( D_p/|G_H| \) and the radius \( r_0 \) and to reduce \( s \), since in real space we have

\[
R_p = \frac{r_0 D_p}{|G_H|a + s} \tag{10A.84}
\]

(where \( s = r_0 s_0 \) and \( a = r_0 a_0 \)).

Formulae (10A.83) and (10A.84) are used to calculate the theoretical resolving power in first order (which is valid, as we shall see, when the angular aperture of the beam is very small: less than \( 10^{-2} \), say).

**Examples.** For a homogeneous field prism, in the symmetrical case (\( |G| = 1 \)), we find:

\[
\alpha = \beta = 0 \quad R_p = \frac{2r_0}{a + s},
\]

\[
\alpha = \beta \neq 0 \quad R_p = \frac{4r_0}{a + s}.
\]

For \( n = \frac{1}{2}, \alpha = \beta = 0 \), in the symmetrical case:

\[
R_p = \frac{4r_0}{a + s}.
\]

When the prism is used for mass separation, we can define a corresponding resolving power:

\[
R_m = \frac{m_0}{\Delta m}
\]

and, from equation (10A.81), we have

\[
R_m = \frac{1}{2} R_p. \tag{10A.85}
\]

In reality, the width of the image in the \( H \) plane is not equal to \( a_t = G_H a_0 \), for the aberrations of the prism must be taken into account, especially the second order aberrations, which are the most important.

10A.1.8 **Aberrations of Magnetic Prisms**

We have obtained the first order optical properties with the aid of a rectangular field model (taking edge effects into account when necessary). We have, however, neglected

(i) terms of higher order in the expansions of \( B_z \) and \( B_x \);

(ii) transverse velocities \( v_z \) and \( v_x \);
(iii) the effect of the component $B_0$ that is present in the stray fields and which affects $v_z$ and $v_x$;
(iv) the curvature of the virtual magnetic faces.

The dominant aberration terms that appear when we take all these factors into account are of second order, and they have been extensively studied both for the homogeneous field prism \((n = 0)\) with normal and tilted faces and for the case where the index \(n \neq 0\). Just as for lenses, the most important aberration, the defect that limits the resolving power of the prism, is the aperture aberration, or more exactly the aperture aberration in the \(H\) plane. Unlike round lenses, however, for which the (third order) spherical aberration cannot be removed, the second order aperture aberration of magnetic prisms can be eliminated in a relatively simple way. Some attention has also been devoted to the aberrations of higher order.

The aperture aberration in the \(H\) plane can be calculated easily in the simple case of the \(180^\circ\) prism (Fig. 115n). The trajectory \((C_0)\), centred on \(C\) and of radius \(r_0\), arrives at \(A'\) (the Gaussian image). The trajectories \((C_1)\) and \((C_2)\) corresponding to particles of the same momentum \(p_0\) but initially inclined to \((C_0)\) at angles \(\pm \gamma_0\) intersect \(CA\) at \(A''\), between \(C\) and \(A'\). We have

\[
AA'' = 2r_0 \cos \gamma_0 = 2r_0(1 - \gamma_0^2/2).
\]

![Fig. 115n. Second order aberration in a homogeneous field prism with \(\phi = 180^\circ\).](image)

In this special case, therefore, the aberration \(A'A''\) is given by

\[
A'A'' = r_0 \gamma_0^2.
\]  (10A.86)

The reader will find a detailed study of the second order aberrations in the articles by Hintenberger (1948, 1949, 1951), Kerwin (1949), Ikegami (1958), Tasman and Boerboom (1959), Wachsmuth, Tasman and Boerboom (1959), Tasman, Boerboom and Wachsmuth (1959), Bretscher (1960), Brown, Belbeco and Bounin (1964) and Enge (1967). Here, we confine
ourselves to the principal results, usually obtained by applying a perturbation method to the trajectories; we have already used this technique in the sections on round lens aberrations.

We first consider an ideal prism of angle $\phi$, limited at $S$ and $S'$ by two plane faces $F$ and $F'$ (the virtual magnetic faces), in which the mean trajectory is $(C_0)$. A second trajectory falls on the entry face $F$ with initial conditions $x_1, x_1', z_1, z_1', \delta$. The corresponding quantities in the plane $F'$ of the exit face can be obtained to a first order approximation by means of the transfer matrices $\|H\|$ and $\|V\|$. In reality, however, the trajectory intersects $F'$ at a point $x_2 + \Delta x_2, z_2 + \Delta z_2$ with slope $x_2' + \Delta x_2', z_2' + \Delta z_2'$ (Fig. 115o). The differences $\Delta x_2 \ldots \Delta z_2'$ may be expanded, and if we consider only terms of second order, we obtain

\[
\Delta x_2 = a_1 x_1^2 + a_2 x_1 x_1' + a_3 x_1'^2 + a_4 z_1^2 \\
+ a_5 z_1 z_1' + a_6 z_1'^2 + a_7 x_1 \delta + a_8 x_1' \delta + a_9 \delta^2,
\]

\[
\Delta z_2 = c_1 x_1 z_1 + c_2 x_1 x_1' + c_3 x_1' z_1' + c_4 z_1'^2 \\
+ c_5 z_1 \delta + c_6 z_1' \delta.
\]

\[(10A.87)\]

**Fig. 115o.** Definition of the notation: ideal prism limited at $F_E$ and $F'_E$.

The expansions for $\Delta x_2'$ and $\Delta z_2'$ are similar to those for $\Delta x_2$ and $\Delta z_2$ except that the coefficients $a_1 \ldots a_9$ and $c_1 \ldots c_6$ are replaced by $b_1 \ldots b_9$ and $d_1 \ldots d_6$ respectively. A consequence of the symmetry about the median plane is that there are only thirty second order aberration coefficients; the terms independent of $\delta$ are the geometrical aberrations (20 coefficients).

For a flat monochromatic beam ($z_1 = z_1' = \delta = 0$) emerging from the point $P$ on $T_0$, the aperture aberration term is given by $\Delta x_2 = a_4 x_1'^2$ in the exit plane of the prism.

All the coefficients $a_k, b_k, c_k$ and $d_k$ can be calculated by integrating the equations of motion, in which second order terms arising from $B_z, B_x$ and the velocities $v_x$ and $v_z$ have been retained.

As the second stage, we calculate the trajectories in the entry and exit zones more accurately. For this, we assume that the magnetic entry and exit faces passing through $S$ and $S'$ (the $F_E$ and $F'_E$) are curved (with radii of curvature $R_1$ and $R_2$) and tilted at angles $\alpha$ and $\beta$ (these are the
angles at \( S \) and \( S' \)). The real ray (Fig. 115p(a)), which impinges on the plane of the flat face \( F \) at \( x_F \) with slope \( x_F' \) crosses \( F_E \) at \( x_E \) (with slope \( x_E' = x_F' \)); it then follows a curved path \( (C) \). If the prism were in fact bounded by the face \( F \), the trajectory \( (C) \) would correspond to a particle emerging from \( P' \) and impinging on \( F \) at \( x_i \) with slope \( x_i' \). The parameters \( x_i \) and \( x_i' \) are then calculated as a function of \( x_E, x_E', \alpha \) and \( R_1 \).

When the particle leaves the magnet (Fig. 115p(b)), the trajectory intersects the real face \( F_E' \) at \( x_s \) (slope \( x_s' \)) and then continues in a straight line to \( F' \), which it crosses at \( x_3 \) with slope \( x_3' \) (Fig. 115p(b)). If the prism ended at \( F'' \), the curve \( (C) \) would cut this face at \( x_2 \) (with slope \( x_2' \)). It is possible to calculate \( x_s \) and \( x_s' \) as a function of \( x_2 \) and \( x_2' \).

The first calculation, on the prism with normal plane faces \( FF' \), allows us to relate the entry parameters, \( x_1 \) and \( x_1' \), to the exit parameters, \( x_2 \) and \( x_2' \), however; it is therefore possible to express \( x_s \) and \( x_s' \) as a function of \( x_E \) and \( x_E' \). The same procedure can be followed in the vertical plane for \( z_s \) and \( z_s' \). In first order, we recover the transfer matrices \( \|H\| \) and \( \|V\| \) that we have met above (these will contain supplementary terms involving the curvatures \( 1/R_1 \) and \( 1/R_2 \) of the entry and exit faces).

The values in the Gaussian image plane, situated at a distance \( q \) from \( S \), are then calculated from \( x_s, x_s', z_s, \) and \( z_s' \). We know that

\[
x_E = x_0 + \frac{p}{r_0} x_0', \quad x_E' = x_0',
\]

\[
z_E = z_0 + \frac{p}{r_0} z_0', \quad z_E' = z_0'
\]

(the suffix 0 corresponds to the object at \( P \), distance \( p \) from the magnet) and at \( Q \) we have

\[
x_t = x_s + \frac{q}{r_0} x_s'.
\]
Equation (10A.89) may be put in the form
\[ x_i = G_{\mu}x_0 + M_1x'_0 + M_2\delta + \Delta x_i, \]
\[ z_i = G_{\nu}z_0 + N_1z'_0 + \Delta z_i, \] (10A.90)
where \( \Delta x_i \) and \( \Delta z_i \) are second order terms.

In the Gaussian image plane, the relation expressing the fact that \( P \) and \( Q \) are conjugates in the \( H \) plane implies that \( M_1 = 0 \). When there is double focusing in first order, we have \( M_1 = 0 \) and \( N_1 = 0 \) simultaneously. The dispersion is given by \( M_2 \). The term \( \Delta x_i \) is of the form
\[ \Delta x_i = M_{11}x_0^2 + M_{12}x_0x'_0 + M_{13}x'_0^2 + M_{14}z_0^2 + M_{15}z_0z'_0 \]
\[ + M_{16}z'_0^2 + N_{11}x_0\delta + N_{12}x'_0\delta + N_{13}\delta^2. \] (10A.91)
The \( M_{1k} \) correspond to geometrical aberrations and the \( N_{1j} \) to chromatic aberrations.

In devices (such as mass or velocity spectrographs) in which a high resolving power is required in the \( H \) plane, sources (objects) that are very narrow in the \( x \)-direction are employed; the terms in \( x_0^2 \) and \( x_0x'_0 \) can then be neglected. Furthermore, the optics can be designed so that \( z_0 \) is also small, and that the incident beam is parallel to \( T_0 \) in the \( V \) plane \((z'_0 = 0)\); the principal aberration term is then given by
\[ \Delta x_i = M_{13}x'_0^2. \] (10A.92)
The coefficient \( M_{13} \) depends on the radii of curvature, \( R_1 \) and \( R_2 \), of the faces of the prism; these curvatures can be selected in such a way that the coefficient vanishes (Hintenberger, 1949). In the case of the prism with homogeneous field, we arrive at an expression of the form
\[ M_{13} = r_0\left(\frac{\gamma_1}{R_1} + \frac{\gamma_2}{R_2}\right) - (C_1 + C_2) \]
with
\[ \gamma_1 = \frac{1}{\cos^3\alpha \left\{1 + (r_0/p + \tan \alpha)^2\right\}^{3/2}}, \]
\[ \gamma_2 = \frac{1}{\cos^3\beta \left\{1 + (r_0/q + \tan \beta)^2\right\}^{3/2}}, \] (10A.93)
\[ C_1 = \frac{r_0^2}{p^2} \frac{r_0/p + 3 \tan \alpha}{\left\{1 + (r_0/p + \tan \alpha)^2\right\}^{3/2}}, \]
\[ C_2 = \frac{r_0^2}{q^2} \frac{r_0/p + 3 \tan \beta}{\left\{1 + (r_0/q + \tan \beta)^2\right\}^{3/2}}; \]
\( p \) and \( q \) are related by the first order conjugacy condition. If the entry and
exit faces are plane \((R_1 = R_2 = \infty)\), we cannot achieve \(M_{13} = 0\) except for particular values of \(p\) and \(q\):

\[
\frac{r_0}{p} = -3 \tan \alpha \quad \text{and} \quad \frac{r_0}{q} = -3 \tan \beta. \tag{10A.94}
\]

In the symmetric case \((p = q, \alpha = \beta)\), we should then have

\[
\tan \alpha = -\frac{1}{2} \tan (\phi/2),
\]

\[
\frac{p}{r_0} = \frac{2}{3} \cot (\phi/2). \tag{10A.95}
\]

The faces must thus be tilted towards the interior of the prism, which increases the horizontal convergence in \(H\) but renders the \(V\) plane divergent. (We recall that a doubly convergent system is obtained for \(\tan \alpha = \frac{1}{2} \tan (\phi/2), \quad p/r_0 = 2 \cot (\phi/2)\).)

More generally, with \(\alpha, \beta \neq 0\) but again with \(p = q\), we obtain

\[
r_0 \left( \frac{1}{R_1} + \frac{1}{R_2} \right) = 2 \frac{r_0^2}{p^2} \left( \frac{r_0}{p} + 3 \tan \alpha \right) \cos^3 \alpha. \tag{10A.96}
\]

The radii \(R_1\) and \(R_2\) and the tilt \(\alpha\) can then be chosen arbitrarily, subject to this condition.

For prisms in which \(n \neq 0\), relations of the form (10A.93) or (10A.96) are again obtained, and it is possible to remove \(M_{13}\) as before.

If we now consider rays from a point object \(P\) lying in the median plane \((x_0 = z_0 = 0)\), we have

\[
\Delta x_i = M_{13} x_0'^2 + M_{16} z_0'^2 \tag{10A.97}
\]

(neglecting the chromatic terms). The first term represents the aperture aberration (broadening of the image), which is independent of the height \(z_i\) of the point at which the trajectory falls on the Gaussian image plane. The second term, on the other hand, vanishes when \(z_i = 0\), and is only present outside the median plane \(H\); it corresponds to curvature of the image (which in first order would be a segment of an infinitesimally thin straight line). In first order, we have

\[
z_i = K z_0',
\]

so that

\[
\Delta x_i \propto z_i^2.
\]

The image becomes a parabolic segment, symmetrical about the median plane, and \(M_{13} x_0'^2\) wide.

\textbf{Resolving Power.} If we take only the principal (aperture) aberration term into account, the resolving power will be given by
\[ R_p = \frac{\Phi_p}{|G|a_0 + s_0 + |M_{13}|x_0'^2} \]  

(10A.98)

and

\[ R_m = \frac{1}{2} R_p. \]

Example: The prism with deflection \( \pi/2 \). The matrices \( |H| \) and \( |V| \) take the form

\[ R_p = \frac{2r_0}{a + s + r_0 x_0'^2}. \]

Unsymmetrical conditions \( (G_H \neq -1) \). It can be shown (Bruck, 1958) that

\[ |M_{13}| = \frac{1}{2} (|G_H| + 1/G_H^2), \]

(10A.99)

and hence

\[ R_p = \frac{r_0 (1 + |G_H|)}{a + s + \frac{1}{2} r_0 (|G_H| + 1/G_H^2) x_0'^2}. \]

(10A.100)

### 10A.2 Electrostatic Prisms

#### 10A.2.1 Description

An electrostatic prism (Herzog, 1934) consists of a portion of a condenser with two metal electrodes, having an axis of revolution \( Z \) and a plane of mechanical and electrical symmetry. This symmetry plane will be called the horizontal or \( H \) plane, as before. The electrodes coincide with part of the surface of a torus of revolution with axis \( C \) and they therefore have a double curvature: in the \( H \) plane and in a vertical plane \( (r, Z) \). The angle of the prism (the angle of deflexion) is \( \phi \) and the mean radius of curvature in the \( H \) plane is \( r_0 \).

Figure 115q shows a section through the prism by the \( H \) plane, and by the radial plane \( \theta = 0 \). The radii of curvature of the electrodes in the \( H \) plane are denoted by \( r_i \) and \( r_e \) and those in the radial plane by \( R_i \) and \( R_e \). We restrict the discussion to the simple case in which the signs of \( R_i \) and \( R_e \) are the same and the corresponding centres of curvature are coincident at \( C_1 \). We also assume that the separation of the electrodes, \( \Delta r = R_e - R_i = r_e - r_i \), is small in comparison with \( r_i \) and \( R_i \).

The electrodes are held at potentials \( \pm V_1 \), symmetric with respect to earth \( (V = 0) \); the mean equipotential surface, which intersects \( H \) along a circle \( (C_0) \) of radius \( r_0 \), corresponds to \( V = 0 \). If we take the origin of potential at the source which is emitting accelerated particles at voltage \( \Phi_0 \), the potential on \( (C_0) \) will be equal to \( \Phi_0 \) (and the electrode voltages
PRISM OPTICS

Fig. 115q. Electrostatic prism. (a) Section by a horizontal plane $H$. (b) Section by a radial vertical plane.

will be $\Phi_0 \pm V_1)$. With our assumption that $(r_e - r_i)/r_0 \ll 1$, we have $r_0 = (r_e + r_i)/2$ and $R_0 = (R_e + R_i)/2$.

As a first approximation, we assume that the electric field vanishes outside the prism and is constant within it, for $0 < \theta < \phi$, when we follow one of the family of circles $(C)$ of radius $r$ centred on $\hat{CZ}$. In the $H$ plane, this field is wholly radial. We shall thus be using a rectangular model, whereby the prism is bounded not by the edges of the real electrodes but by virtual electric faces situated outside the prism. In order to obtain the distance $l$ between the edges of the electrodes and this virtual face (Fig. 115r), the real distribution of $E_r$ in the $H$ plane must be established, either by calculation or by means of an electrolytic tank. When (earthed) metal slits or diaphragms are placed at the entry and exit of the prism, the stray fields will be altered; the distribution $E_r(\theta)$ must be determined for the complete

Fig. 115r. Entry zone of an electrostatic prism with a diaphragm. (a) Position of the electric face $F$ with respect to the electrodes and the diaphragm (separation $l$). (b) Determination of the limit of the rectangular model (equal areas).
system of electrodes and diaphragms, therefore. The distance \( l \) satisfies the empirical expression

\[
l = 0.4b + 0.2d + 0.5s \tag{10A.101}
\]

(Wollnik and Ewald, 1965), in which \( 2b \) represents the separation of the electrodes, \( d \) the distance between the electrodes and the diaphragm and \( 2s \) the width of the opening in the diaphragm in the \( H \) plane. This formula is valid for \( d \geq 2b \) and for plane diaphragms.

If the curvatures \( r_0 \) and \( R_0 \) are equal, the condenser is called a spherical condenser; if \( R_0 = R_i = R_e = \infty \), the condenser is cylindrical, with \( \vec{CZ} \) as axis.

We shall first consider only particles travelling in the vicinity of \( (C_0) \) with small slopes \( dr/ds \) or \( dZ/ds \) (Gaussian optics), as we did in the case of magnetic prisms; we also limit the discussion to prisms whose entry and exit faces lie in planes intersecting along \( \vec{CZ} \).

### 10A.2.2 Potential and Field

In the ideal prism (rectangular model), these quantities are independent of \( \theta \) in the range \( 0 < \theta < \phi \).

We adopt as before the system of reduced coordinates

\[
x = \frac{r - r_0}{r_0}, \quad z = \frac{Z}{r_0}, \quad \theta = \frac{s}{r_0},
\]

where \( s \) is the arc-length along the curved axis \( (C_0) \). In the neighbourhood of \( (C_0) \) in the \( H \) plane, the radial field \( E_r \) may be written in the form

\[
E_r = E_0(1 - nx + bx^2 - \cdots), \tag{10A.102}
\]

where

\[
n = \frac{r_0}{E_0} \left( \frac{\partial E_x}{\partial r} \right)_{z=0}
\]

is the index of the field and

\[
b = \frac{1}{2} \frac{r_0^2}{E_0} \left( \frac{\partial^2 E_x}{\partial r^2} \right)_{z=0}.
\]

Taking \( \Phi = \Phi_0 \) on \( (C_0) \), the potential is given by

\[
\Phi(x, 0) - \Phi_0 = -E_0r_0(x - \frac{1}{2} nx^2 + \frac{1}{3} bx^3 - \cdots). \tag{10A.103}
\]

The coefficients \( n \) and \( b \) depend on the radii of curvature, \( r_0 \) and \( R_0 \), of the equipotential surface \( \Phi(x, z) = \Phi_0 \). Since we know that \( \Phi(x, z) \) must satisfy Laplace's equation and that, because of the symmetry about the plane \( z = 0 \), it can be expanded in the form
we can show (Ewald and Liebl, 1955; Albrecht, 1956) that
\[
\frac{\Phi(x, z) - \Phi_0}{E_0r_0} = -\left\{ x - \frac{1}{2} nx^2 + \frac{1}{3} bx^3 \right. \\
\left. -\frac{1}{2} (1-n)z^2 + \frac{1}{2} (1+n-2b)xz^2 + \cdots \right\}.
\]

For the field components, we obtain
\[
E_r(x, z) = E_0 \left\{ 1 - nx + bx^2 + \frac{1}{2} (1+n-2b)z^2 + \cdots \right\},
\]
\[
E_z(x, z) = E_0 \left\{ - (1-n)z + (1+n-2b)xz + \cdots \right\},
\]
where
\[
n = 1 + r_0/R_0.
\]
\[
b = 1 + \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1/R_0').
\]
\[
R_0' = \left( \frac{dR'}{dr} \right)_{r=r_0, z=0}.
\]

Halting the field expansions at second order terms, we find
\[
\Phi(x, z) = \Phi_0 - E_0r_0 \left[ x - \frac{1}{2} \left( 1 + \frac{r_0}{R_0} \right) x^2 \\
+ \frac{1}{3} \left( 1 + \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1 + R_0') \right) x^3 + \frac{1}{2} \frac{r_0}{R_0} z^2 \\
- \frac{1}{2} \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1 + R_0') \right] xz^2 \\
\right],
\]
\[
E_r(x, z) = E_0 \left[ 1 - \left( 1 + \frac{r_0}{R_0} \right) x + \left\{ 1 + \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1 + R_0') \right\} x^2 \\
- \left\{ \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1 + R_0') \right\} \frac{z^2}{2} \right],
\]
\[
E_z(x, z) = E_0 \left[ \frac{r_0}{R_0} z - \frac{r_0}{R_0} + \frac{r_0^2}{2R_0^2} (1 + R_0') \right] xz \\
\right].
\]

For the study of the first order properties, we simply write
\[
E_r(x, z) = E_0(n - 1)x = E_0 \left\{ 1 - \left( 1 + \frac{r_0}{R_0} \right) x \right\},
\]
\[
E_z(x, z) = E_0(n - 1)z = E_0 \frac{r_0}{R_0} z.
\]
A positive field $E_0$ is created by applying a potential difference $2V_1$ across the electrodes in such a way that the outer electrode is negative with respect to $\Phi_0$. In two special cases (the spherical condenser and the cylindrical condenser), $\Phi(r, z)$ is given for every point in space by simple analytical expressions. For the spherical prism ($n = 2$):

(a) \[ V_1 = \Phi_0 \frac{r_0}{r} \left( \frac{1}{r_i} - 1 \right) = \Phi_0 \frac{r_e - r_i}{r_e r_i}, \]

(b) \[ r_0 = \frac{2r_i r_e}{r_i + r_e}, \]

(c) \[ \Phi(0, r) = \Phi_0 \left( \frac{2r_0}{r} - 1 \right), \]

and for the cylindrical prism ($n = 1$):

(a) \[ V_1 = \Phi_0 \log \left( \frac{r_e}{r_i} \right), \]

(b) \[ r_0 = r_i r_e, \]

(c) \[ \Phi(0, r) = \Phi_0 \left\{ 1 - 2 \log \left( \frac{r}{r_0} \right) \right\}. \]

If $r_e - r_i \ll r_0$, we can regard $r_0$ as approximately equal to $(r_i + r_e)/2$ in both cases.

**10A.2.3 Trajectories**

We consider only the non-relativistic situation; this implies

\[ m = m_0, \quad \frac{dm}{dt} = 0, \quad v_0 = \sqrt{\frac{2e\Phi_0}{m}} \]

(where $m_0$ is the rest mass of the particle). When a particle of mass $m$, charge $e$ and energy $W_0 = e\Phi_0 = \frac{1}{2}mv_0^2$, travelling in the $H$ plane, enters the prism at $S$, it will follow the circle $(C_0)$ if the value of $E_0$ is such that

\[ \frac{mv_0^2}{r_0} = -eE_0, \]

and hence

\[ E_0 r_0 = -2\Phi_0. \]

The velocity of the particle will not remain equal to $v_0$ except on $(C_0)$, since $\Phi(r, z) \neq \Phi_0$. More generally, a particle incident with velocity

\[ v = v_0(1 + \beta), \quad \beta \ll 1 \]

will have a velocity

\[ v = v_0 \left\{ 1 + \beta - x + \frac{1}{2} \frac{r_0}{K} (x^2 - z^2) + x\beta \right\} \]
at a point \((x, z)\); in a first order approximation, this reduces to

\[
v = v_0(1 + \beta - x).
\]  
(10A.116)

In cylindrical coordinates, the equations of motion take the form

\[
\begin{align*}
\frac{d^2 r}{dt^2} &= m r \left( \frac{d\theta}{dt} \right)^2 + e E_r, \\
\frac{d^2 z}{dt^2} &= e E_z, \\
\frac{d}{dt} \left( r^2 \frac{d\theta}{dt} \right) &= 0.
\end{align*}
\]  
(10A.117)

Equation (10A.117c) may be integrated immediately, and gives an expression for \(d\theta/dt\)

\[
\frac{d}{dt} \left( r_0^2 (1 + x)^2 \frac{d\theta}{dt} \right) = \frac{d}{dt} \left[ r_0 (1 + x) v_\theta \right] = 0
\]

\((v_\theta)\) denotes the azimuthal velocity). Hence

\[
r_0^2 (1 + x)^2 \frac{d\theta}{dt} = r_0 (1 + x) v_\theta = A = \text{const.} 
\]  
(10A.118)

The constant \(A\) is easily obtained from the initial conditions \((x_1, x'_1, z_1, z'_1)\) at \(\theta = 0\):

\[
A = r_0 (1 + x_1)(v_\theta)_1, 
\]  
(10A.119)

and

\[
(v_\theta)_1 = v_1 \left\{ 1 - (x_1'^2 + z_1'^2) \right\}^{1/2} = v_1 \left\{ 1 - \frac{1}{2}(x_1'^2 + z_1'^2) \right\}. 
\]  
(10A.120)

In first order

\[
(v_\theta)_1 = v_1 = v_0(1 + \beta - x_1) \]

so that

\[
A = r_0 v_0 (1 + x_1)(1 + \beta - x_1) \approx r_0 v_0 (1 + \beta). 
\]  
(10A.121)

\(A\) is independent of \(x_1\) and hence, on combining equations (10A.118) and (10A.121), we obtain

\[
\frac{d\theta}{dt} = \frac{v_0}{r_0} (1 + \beta - 2x). 
\]  
(10A.122)

To the second order approximation, we should find (Ewald and Liebl, 1965)
\[
\frac{d\theta}{dt} = \frac{v_0}{r_0} \left\{ 1 + \beta - 2x - 2x\beta + 3x^2 - x_i^2 + 2\beta x_i \right\} - \frac{1}{2}(x_i'^2 + z_i'^2) + \frac{1}{2} \frac{r_0}{R_0^2} (x_i^2 - z_i^2) \right\}. 
\]

(10A.123)

Substituting for \( E_r \) and for \( d\theta/dt \) from equation (10A.122) into equation (10A.117a) gives

\[
\frac{d^2r}{dt^2} = \frac{v_0^2}{r_0} (1 + \beta - 2x)^2 (1 + x) + \frac{e}{m} E_0 (1 - nx),
\]

so that with the aid of equation (10A.112), we find

\[
\frac{d^2x}{dt^2} = \frac{v_0^2}{r_0^2} \{(1 + x)(1 + \beta - 2x)^2 - (1 - nx)\}
\]

or

\[
\frac{d^2x}{dt^2} = \frac{v_0^2}{r_0^2} \{(n - 3)x + 2\beta\}. 
\]

(10A.124)

Finally, from equation (10A.117b) we arrive at

\[
\frac{d^2z}{dt^2} = \frac{e}{m} E_0 (n - 1)z = -\frac{v_0^2}{r_0} (n - 1)z,
\]

and hence

\[
\frac{d^2z}{dt^2} = \frac{v_0^2}{r_0^2} (1 - n)z. 
\]

(10A.125)

On eliminating the time from equations (10A.123) and (10A.124), we obtain the equations of the paraxial trajectories in reduced coordinates:

(a) \[
\frac{d^2x}{d\theta^2} + (3 - n)x = 2\beta,
\]

(10A.126a)

(b) \[
\frac{d^2z}{d\theta^2} + (n - 1)z = 0.
\]

(10A.126b)

For a monochromatic beam of particles, these equations simplify still further to

(a) \[
x'' + (3 - n)x = 0,
\]

(b) \[
z'' + (n - 1)z = 0.
\]

(10A.126b)

**Remark.** It can be shown (Bruck, 1958) that if the particle velocities are relativistic, the trajectory equations become

\[
\frac{d^2x}{d\theta^2} + \left(3 - n - \frac{v^2}{c^2}\right)x = \frac{d^2x}{d\theta^2} + \left(2 - \frac{r_0}{R_0} - \frac{v^2}{c^2}\right)x = 0,
\]

\[
\frac{d^2z}{d\theta^2} + (n - 1)z = \frac{d^2z}{d\theta^2} + \frac{r_0}{R_0} z = 0.
\]

(10A.127)
We note that \( 2 \beta = 2dv/v_0 = d\Phi/\Phi_0 \) is the relative energy variation in the beam.

Equations (10A.126) show that the particles oscillate about the curved axis \((C_0)\), provided that \(3 - n > 0 \) and \(n - 1 > 0\). The motion in the \(H\) plane and that in the vertical or \(V\) plane \((Z, s)\) are uncoupled, exactly as in the case of the magnetic prism.

We set

\[
\begin{align*}
\omega_H^2 &= 3 - n, \\
\omega_V^2 &= n - 1
\end{align*}
\]  

(\(\omega_H\) and \(\omega_V\) may be imaginary); the solutions of equations (10A.126) then take the following form, expressed in terms of the initial conditions, at \(\theta = 0\):

\[
\begin{align*}
x &= x_1 \cos (\omega_H \theta) + \frac{x_1'}{\omega_H} \sin (\omega_H \theta), \\
z &= z_1 \cos (\omega_V \theta) + \frac{z_1'}{\omega_V} \sin (\omega_V \theta).
\end{align*}
\]  

The values of \(n, \omega_H\) and \(\omega_V\) for the three types of prism commonly used are tabulated below, for non-relativistic particles; the relativistic values are related to these by

\[
\begin{align*}
(\omega_H^2)_{\text{rel}} &= \omega_H^2 - v^2/c^2, \\
(\omega_V^2)_{\text{rel}} &= \omega_V^2.
\end{align*}
\]  

<table>
<thead>
<tr>
<th>Condenser</th>
<th>Toroidal</th>
<th>Spherical</th>
<th>Cylindrical</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>(1 + r_0/R_0)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(\omega_H)</td>
<td>((2 - r_0/R_0)^{1/2})</td>
<td>1</td>
<td>(\sqrt{2})</td>
</tr>
<tr>
<td>(\omega_V)</td>
<td>((r_0/R_0)^{1/2})</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

10A.2.4 First Order Optical Properties; Transfer Matrices

When the incident beam is not monoenergetic, the equation of motion in the \(H\) plane can be obtained by integrating the first of equations (10A.126a); the solution when the right hand side is zero (10A.126b) is already known, and hence the full solution is

\[
x = A \cos (\omega_H \theta) + B \sin (\omega_H \theta) + \frac{1}{3 - n} \frac{\Delta \Phi}{\Phi_0}.
\]  

(10A.131)
Introducing the initial conditions at $\theta = 0$, we find

$$x = x_1 \cos (\omega_H \theta) + \frac{x_1'}{\omega_H} \sin (\omega_H \theta) + \frac{1}{\omega_H} \left\{ 1 - \cos (\omega_H \theta) \right\} \frac{\Delta \Phi}{\Phi_0},$$

$$x' = -x_1 \omega_H \sin (\omega_H \theta) + x_1' \cos (\omega_H \theta) + \frac{1}{\omega_H} \sin (\omega_H \theta)$$

and

$$z = z_1 \cos (\omega_V \theta) + \frac{z_1'}{\omega_V} \sin (\omega_V \theta),$$

$$z' = -z_1 \omega_V \sin (\omega_V \theta) + z_1' \cos (\omega_V \theta).$$

These equations enable us to write down the transfer matrices $\|H\|$ and $\|V\|$ and hence to express the trajectory parameters at $\theta = \phi$ in terms of those at $\theta = 0$:

$$\begin{pmatrix} x_2 \\ x_2' \\ \Delta \Phi \\ \Phi_0 \end{pmatrix} = \|H\| \begin{pmatrix} x_1 \\ x_1' \\ \Delta \Phi \\ \Phi_0 \end{pmatrix},$$

$$\begin{pmatrix} z_2 \\ z_2' \end{pmatrix} = \|V\| \begin{pmatrix} z_1 \\ z_1' \end{pmatrix},$$

with

$$\|H\| = \begin{pmatrix} \cos (\omega_H \phi) & 1 \omega_H^{-1} \sin (\omega_H \phi) & \frac{1 - \cos (\omega_H \phi)}{\omega_H} \\ -\omega_H \sin (\omega_H \phi) & \cos (\omega_H \phi) & \frac{\omega_H}{\sin (\omega_H \phi)} \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\|V\| = \begin{pmatrix} \cos (\omega_V \phi) & \frac{\sin (\omega_V \phi)}{\omega_V} \\ -\omega_V \sin (\omega_V \phi) & \cos (\omega_V \phi) \end{pmatrix}.$$

Unlike the case of magnetic prisms, the mass of the particles does not occur in the foregoing equations; we shall therefore define the dispersive properties of the prism in terms of energy and not velocity. Particles of the same charge but with different masses that have been accelerated through the same potential $\Phi$ will follow the same path through the prism: electrostatic prisms cannot be used to separate particles of different masses emitted by an ion source. They can, however, be used as velocity analysers when all the particles are of the same type. In this case, we have

$$\frac{\Delta v}{v_0} = \frac{1}{2} \frac{\Delta \Phi}{\Phi_0}.$$
The optical elements and the energy dispersion, \( D_e \), can be calculated from equations (10A.133) and (10A.134). Returning to real coordinates (see Fig. 115s), we obtain:

In the H plane

\[
f_H' = -f_H = \frac{r_0}{\sin (\omega_H \phi)}, \quad \omega_H = \sqrt{3 - n}; \quad (10A.136)
\]

\[
g_H' = -g_H = \frac{r_0}{\tan (\omega_H \phi)}.
\]

In the V plane

\[
f_V' = -f_V = \frac{r_0}{\sin (\omega_V \phi)}, \quad \omega_V = \sqrt{n - 1}. \quad (10A.137)
\]

\[
g_V' = -g_V = \frac{r_0}{\tan (\omega_V \phi)}.
\]

---

**Fig. 115s.** Positions of the foci and principal planes.

The electrostatic prism, like the magnetic prism, is equivalent to two centred systems of length \( L = r_0 \phi \), the cardinal elements of which are given by equations (10A.136) and (10A.137). If \( \omega_H \) and \( \omega_V \) are real, the two systems are convergent. If, on the other hand, \( \omega_H \) (or \( \omega_V \)) is imaginary, the circular functions must be replaced by hyperbolic functions, and the corresponding system is divergent. We recall that the term \( r_0/R_0 \) that appears in \( n \) can be negative, if \( R_0 \) is negative (the electrodes would then be convex in the \((Z, r)\) plane towards \( \overline{CZ} \); in Fig. 115q, \( R_0 > 0 \)). Summarizing the various possibilities, we have:

\[
n = 1 + \frac{r_0}{R_0} < 1 \quad (R_0 < 0) \quad \begin{cases} 
\text{H plane convergent} \\
\text{V plane divergent,}
\end{cases}
\]

\[
n = 1 \quad (\text{cylindrical condenser}) \quad \begin{cases} 
\text{H plane convergent} \\
\text{V plane equivalent to a drift space,}
\end{cases}
\]
$1 < n < 3$ both planes convergent,

$n = 3 \ (r_0/R_0 = 2) \ \begin{cases} H \text{ plane equivalent to a drift space} \\ V \text{ plane convergent,} \end{cases}$

$n > 3 \ \begin{cases} H \text{ plane divergent} \\ V \text{ plane convergent.} \end{cases}$

The properties of the cylindrical prism are thus similar to those of the magnetic prism with homogeneous field, since $f_V = g_V = \infty$. We have $g'_H = g_H = 0$, however, so that the foci are at $S$ and $S'$, for $\tan (\sqrt{2} \phi) = 0$ or $\phi_1 = \pi/2 \sqrt{2}$. By joining two prisms of angle $\phi_1$, therefore, we can focus particles from a point source at $S$ at the point $S'$. The image of $S$ will be a segment of a straight line (in first order) since there is no convergence in the $V$ plane, and the total deflection will then be (Fig. 115t)

$$\phi = 2\phi_1 = \pi/\sqrt{2} = 127^\circ 2'. \quad (10A.138)$$

This is true when the source at $S$, in the electric face of the prism, is in fact the crossover of a beam formed by an optical system placed in front of the prism. If, on the other hand, we attempt to define an “object” by means of a metal diaphragm placed at $S$, the electric face will move closer to the electrodes and we shall no longer have $\phi = 127^\circ 2'$; the exact position of the image must then be calculated with the aid of equation (10A.136).

For the cylindrical prism, equations (10A.136) become

$$f'_H = r_0/\sin (\sqrt{2} \phi), \quad (10A.139)$$
$$g'_H = r_0/\tan (\sqrt{2} \phi).$$

The spherical prism is equivalent to a rotationally symmetric optical system, since $\omega_H = \omega_V = 1$ (like the magnetic prism with index $n = \frac{1}{2}$); the system is stigmatic for an infinite number of pairs of conjugate planes.
We have

\[ f'_H = r_0 / \sin \phi, \]

\[ g'_H = r_0 / \tan \phi. \quad (10A.140) \]

Particles from \( S \) can be focused at \( S' \), if \( \phi = 180^\circ \), but we now have a point image (if aberrations are neglected).

**Toroidal Prism.** For \( 1 < n < 3 \), symmetric stigmatic operation can again be achieved in certain cases, but for only one pair of conjugate planes.

### 10A.2.5 Dispersion; Theoretical Resolving Power (First Order Approximation)

The dispersion, \( D_e \), is defined, in the Gaussian image plane, which lies at a reduced distance \( q/r_0 \) from \( S \), by

\[ D_e = \frac{\Delta x_i}{\Delta \Phi/\Phi_0} \quad (10A.141) \]

or

\[ D_e = \Delta x_2 + \frac{q}{r_0} \Delta x'_2, \]

so that

\[ D_e = \left\{ 1 - \cos (\omega_H \phi) + \frac{q}{r_0} \omega_H \sin (\omega_H \phi) \right\} / \omega_H^2. \quad (10A.142) \]

It can be shown (Bruck, 1958) that we again obtain the simple relation

\[ D_e = \frac{1 - G_H}{\omega_H^2}. \quad (10A.143) \]

From the relation between conjugates in the \( H \) plane, we can calculate \( q/r_0 \) when the object distance, \( p/r_0 \), is known; we have merely to write down the total transfer matrix \( \| T \| \) between the object \( P \) and its image \( Q \) and then set

\[ T_{12} = 0, \quad (10A.144) \]

which gives

\[ \tan (\omega_H \phi) = -\{ (p+q)/r_0 \} \omega_H / \left[ 1 - (pq/r_0^2) \omega_H^2 \right]. \quad (10A.145) \]

**Cylindrical Prism.** With symmetric operation, we have

\[ \frac{p}{r_0} = \frac{q}{r_0} = \frac{1}{\omega_H} \cot (\omega_H \phi/2) = \frac{1}{\sqrt{2}} \cot (\sqrt{2} \phi/2), \quad (10A.146) \]

\[ G_H = -1 \]
and
\[ D_e = 1. \]  
(10A.147)

If \( p \neq q, G_H \neq -1 \) and we find
\[ D_e = \frac{1}{2} (1 - G_H). \]  
(10A.148)

\( D_e \) is thus independent of \( \phi \) and of \( r_0 \). The *real separation*, \( d_c \), between particles of energies \( \Phi_0 \) and \( \Phi_0(1 + \Delta \Phi/\Phi_0) \) in the Gaussian image plane does depend on \( r_0 \) and is given by
\[ d_c = r_0 \Delta x_i = r_0 D_e \frac{\Delta \Phi}{\Phi_0} = \frac{1}{2} r_0 (1 - G_H) \frac{\Delta \Phi}{\Phi_0}. \]  
(10A.149)

**Spherical prism.** From equation (10A.143) we obtain
\[ \tan \phi = \frac{(p + q)/r_0}{1 - pq/r_0^2}. \]  
(10A.150)

The points \( P \) and \( Q \) (and \( G_H \) which is equal to \( G_V \)) can be determined graphically by Barber’s rule (see the section on the magnetic prism with homogeneous field). In the symmetric case,
\[ \frac{p}{r_0} = \frac{q}{r_0} = \cot (\phi/2), \]  
(10A.151)

\[ G_H = G_V, \]  
(10A.152)

\[ D_e = 2. \]  
(10A.152)

The dispersion is therefore twice as large as that of the cylindrical prism.

In the general case \( (p \neq q) \), we have
\[ D_e = (1 - G_H) \]  
(10A.153)

and
\[ d_c = r_0 (1 - G_H) \frac{\Delta \Phi}{\Phi_0}. \]  
(10A.154)

If we do not take the aberrations into account, we can define a theoretical first order resolving power, \( R_e = \Phi_0/\Delta \Phi \); the width of the image of an object \( a_0 \) is given by \( a_i = |G|a_0 \). If a slit of width \( s_0 \) is placed on the axis at \( Q \), we have
\[ R_e = \frac{r_0 D_e}{|G|a_0 + s_0}. \]  
(10A.155)

In reality, the practical resolving power will be poorer than the value predicted by equation (10A.155), since allowance must be made for the aberrations, especially the second order aperture aberration.
10A.2.6 Focusing in the Vertical Plane with a Cylindrical Prism

The cylindrical prism is the most common in practice, since it is easy to construct. The convergence is zero in the vertical plane, however, as we have already found, and this often limits the "clarity" of the system. It is in fact very easy to transform a cylindrical prism into a prism with non-zero convergence in the \( V \) plane: we have only to arrange that the equipotential surface \( \Phi(x, z) = \Phi_0 \) and the neighbouring surfaces have a positive vertical curvature, \( 1/R \). This is achieved by placing two supplementary circular plane electrodes, \( E_1 \) and \( E_2 \), above and below the electrodes of the condenser and holding them at the same potential, which is not equal to \( \Phi_0 \). If the field \( E_0 \) is positive, the outer electrode is held at \( \Phi_0 - V_1 \) and the inner electrode at \( \Phi_0 + V_1 \). In order to obtain \( R_0 > 0 \), we apply to \( E_1 \) and \( E_2 \) voltages that are slightly negative with respect to \( \Phi_0 \). The electrolytic tank yields \( R_0 \) as a function of the applied voltage for a given electrode geometry. In Fig. 115u(a), the form of the equipotential surfaces in a cylindrical condenser is shown. If we require \( R_0 = \infty \), we must have \( b \ll c \) (where \( 2b \) is the distance between the electrodes and \( c \) is their height). Figure 115u(b) shows the equipotentials when a potential \( -V_2 \) is applied to the supplementary electrodes. (In this figure, the potentials are measured with respect to earth—as they are in practice when we apply a voltage to a prism—(\( C_0 \) corresponds to \( \Phi = 0 \)).

![Diagram](image)

**Fig. 115u.** The conversion of a cylindrical prism into a toroidal prism. (a) Equipotential surfaces before modification. (b) Form of the equipotentials after application of a negative potential to the supplementary electrodes.

10A.2.7 Aberrations of Electrostatic Prisms

If we retain second and third order terms in the expansion for \( \Phi(x, z) \) and second order terms in \( v \), we can first of all calculate the second order aberrations in an ideal prism, bounded by plane virtual electric faces, using the rectangular model approximation (Ewald and Liebl, 1957;
10A.2.7.1 Aberrations of the ideal prism

The second order trajectory equations, in the form

\[ x'' + \omega_H^2 x - 2\delta = f(x, x', z, z', \delta) \]  \hspace{1cm} (10A.156)

are integrated; the right-hand side contains only terms of second order, and into it are substituted the (first order) solutions for \(x, x', z\) and \(z'\), obtained by integrating the corresponding homogeneous equation, for \(0 < \theta < \phi\). Thus in the exit face of the prism (suffix 2), for example, equation (10A.156) gives

\[ x_2 = (x_2) + \Delta x_2, \]  \hspace{1cm} (10A.157)

\[ x'_2 = (x'_2) + \Delta x'_2, \]

with

\[ (x_2) = H_{11} x_1 + H_{12} x'_1 + H_{13} \frac{\Delta \Phi}{\Phi_0}, \]  \hspace{1cm} (10A.158)

\[ (x'_2) = H_{21} x_1 + H_{22} x'_1 + H_{23} \frac{\Delta \Phi}{\Phi_0}. \]

(Suffix 1 characterizes quantities measured in the entry face.) Thirty coefficients are again necessary to describe the departures \(\Delta x_2, \Delta x'_2, \Delta z_2\) and \(\Delta z'_2\) completely (including the chromatic aberrations arising from an energy spread \(\Delta \Phi/\Phi_0 = \delta\)). These aberrations are given by expressions similar to those for magnetic prisms (see equation 10A.87). For the aperture aberration coefficient \(a_3\), for example (Wollnik, 1965), which is defined by

\[ (\Delta x_2)_{ap} = a_3 x'^2_1, \]

we find

\[ a_3 = \omega_H^{-1} \left[ \left\{ 4 \left( \frac{r_0}{R_0} - 1 \right) - 2b/3 + \omega_H^2 \right\} \{ 1 - \cos(\omega_H \Phi) \} \\ - \left\{ 2 \left( \frac{r_0}{R_0} - 1 \right) - b/3 \right\} \sin^2(\omega_H \Phi) \right], \]  \hspace{1cm} (10A.159)

where

\[ b = \frac{1}{2E_0 r_0} \left( \frac{\partial^2 \Phi(x, z)}{\partial x^2} \right)_{x=z=0}. \]
We observe that \( a_3 = 0 \) for \( \omega_H \phi = 2\pi \) (and the coefficients of the terms in \( x_1^2 \) and \( x_1 x_1' \) vanish at the same time), but this property cannot be exploited for in these conditions the dispersion \( \mathcal{D}_e \) vanishes.

**SPECIAL CASES:** *Spherical prism* \((r_0 = R_0, \omega_H^2 = 1, b = 3)\):

\[
(a_3)_{sph} = \sin^2 \phi - (1 - \cos \phi).
\]  

(10A.160)

If we wish to focus \( S \) at \( S' \) (\( \phi = 180^\circ \)), we have

\[
(a_3)_{sph, 180^\circ} = -2.
\]

*Cylindrical prism* \((R_0 = \infty, \omega_H^2 = 2, b = 1)\):

\[
(a_3)_{cyl} = \frac{1}{12} \{7 \sin^2 (\sqrt{2} \phi) - 8(1 - \cos (\sqrt{2} \phi))\} 
\]  

(10A.161)

For a prism for which \( \phi = 127^\circ 2', \) we have

\[
(a_3)_{cyl, 127^\circ} = -4/3.
\]

10A.2.7.2 *Passage through the fringing fields*

If we consider the passage through the real fringing fields, we see that, in the \( H \) plane, the trajectory \((C_0)\) corresponds to an incident trajectory \( T'_0 \), normal to the virtual entry face that we defined earlier and parallel to \( T_0 \) but shifted from it a distance \( \Delta x_E \) (Fig. 115v). Outside the ideal prism,

![Diagram](image)

**Fig. 115v.** Influence of the fringing field at the entry. \( E \): Electric face. (The \( H \) plane is represented.)

we use \( T'_0 \) as curved axis and inside the prism, the circle \((C_0)\). A trajectory from a point object \((x_0, x'_0, z_0 = z'_0 = 0)\) will strike the entry face \( E \) at \( x_E \) with slope \( x'_E \) and will then follow the trajectory \((C)\) inside the prism. \((C)\) appears to come from a point with coordinate \( x_1 \) (with respect to \((C_0)\)) and slope \( x'_1 \), lying in \( E \).

In order to obtain an actual deflection \( \phi \), the object must therefore be shifted a distance \( \Delta x_E \) (as in the case of magnetic prisms, see §10A.1.8).
A second order calculation shows (Wollnik and Ewald, 1965) that

\[ x_1 = x_E + \frac{1}{2} x_E^2 + \cdots, \]

\[ x'_1 = x'_E - \frac{1}{2R_1} (x_E^2 - z_E^2) - x_E x'_E + \frac{x_E}{3} l / 3r_0. \]

\(^{(10A.162)}\)

\( (l \text{ denotes the gap between the electrodes and the virtual electric face.}) \)

We here assume that the electric face \( E \) has a radius of curvature \( R_1 \) (Fig. 115w) in a vertical plane \((Z, s)\).

![Fig. 115w. Notation. \((V\text{ plane.})\)]

In the vertical plane, we have

\[ z_1 = z_E + \cdots, \]

\[ z'_1 = z'_E + \frac{1}{R_1} x_E z_E. \]

\(^{(10A.163)}\)

At the exit face \( F \), we can likewise calculate the corrected values \( x_F, x'_F, z_F \) and \( z'_F \) as a function of \( x_2, x'_2, z_2 \) and \( z'_2 \), the values corresponding to an ideal prism of angle \( \phi \). (For simplicity, we assume that the exit face has the same radius of curvature, \( R_1 \), as the entry face.) We have

\[ x_F = x_2 - \frac{1}{2} x_2^2, \]

\[ x'_F = x'_2 - \frac{1}{2R_1} (x_2^2 - z_2^2), \]

\[ z_F = z_2 + \cdots, \]

\[ z'_F = z'_2 + \frac{1}{R_1} x_2 z_2. \]

\(^{(10A.164)}\)

From the values at \( F \), we can calculate the corresponding values \((x_i, x'_i, z_i, z'_i)\) in the Gaussian image plane in the \( H \) plane, situated a reduced distance \( q/r_0 \) from \( S' \).
The departures $\Delta x_i$, $\Delta x'_i$, $\Delta z_i$ and $\Delta z'_i$ can therefore be calculated as a function of $x_0$, $x'_0$, $z_0$, $z'_0$ and $\delta$, and the corresponding aberration coefficients can be extracted. Of particular interest is the departure $\Delta x_i$ in the $H$ plane from which we can calculate the energy resolution. According to equations (10A.162) and (10A.164), $\Delta x_i$ will depend upon $R_1$, so that a partial correction at least of the spherical aberration will be possible. Up to the present time, however, no expression for the aberrations in the Gaussian image plane has been published; less work has been devoted to electrostatic prisms than to magnetic prisms.

10A.2.7.3 Resolving power

We can only give an approximate value for the resolution $R_e$ in a prism, in which the image of $S$ is formed at $S'$ (neglecting all the other aberration terms in comparison with the aperture aberration term $a_3 x_1'^2$), since to a first approximation $\Delta x_2$ is given by equation (10A.159). We add a term $r_0 \Delta x_2$ to the denominator of equation (10A.155).

**Spherical condenser** ($\phi = 180^\circ$, $|G_H| = 1$). In real coordinates,

\[
R_e = \frac{r_0 \phi_e}{|G|a_0 + s_0 + |a_3| x_1'^2 r_0} \tag{10A.165}
\]

or

\[
R_e = \frac{2 r_0}{a_0 + s_0 + 2 r_0 x_1'^2} \tag{10A.166}
\]

\[
= \frac{2}{(a_0 + s_0)/r_0 + 2 x_1'^2}.
\]

If, for example, we have

- $r_0 = 5$ cm,
- $a_0 = 0.25$ mm,
- $s_0 = 0.25$ mm (circular holes as object and image),

then

\[
R_e = \frac{2}{10^{-2} + 2 x_1'^2} = 200 (1 - 200 x_1'^2).
\]

The aberration has virtually no effect so long as $x_1' < 10^{-2}$ rad.

**Cylindrical condenser** ($\phi = 127^\circ 2'$, $|G_H| = 1$):

\[
R_e = \frac{r_0}{a_0 + s_0 + \frac{4}{3} r_0 x_1'^2} = \frac{1}{(a_0 + s_0)/r_0 + \frac{4}{3} x_1'^2}. \tag{10A.167}
\]
With the same values of \( r_0, a_0 \) and \( s_0 \) (slits as object and image), we find
\[
R_e = 100 \left( 1 - 133 \, x_1^2 \right) .
\]

10A.2.7.4 Direct calculation of the aberrations in the image plane

With the advent of fast computers, it has become possible to calculate the aberrations and the resolution directly, by integrating the general equations of motion from the object plane to the image plane; the values of the potential employed are either measured or calculated by solving Laplace's equation with the computer. The trajectories have recently been calculated in this way for beams of slow electrons, taking the space charge into account (Ballu, 1968; Barat, 1968). These calculations show that for beams of angular aperture \( \alpha > 10^{-2} \) in a cylindrical 127° prism, it is better to use entry and exit diaphragms in the form of inclined "horns" (Fig. 115x(a)) instead of plane diaphragms (Fig. 115x(b)). The potential distribution is disturbed less in the stray field zone with inclined planes.

![Diagram](image)

**Fig. 115x.** Examples of diaphragms (screens) and the appearance of the equipotentials in the \( H \) plane. (a) Plane diaphragm. (b) Inclined planes (thin walls).
PART 2

INSTRUMENTS
CHAPTER 11

ELECTRON OPTICAL INSTRUMENTS

11.1 LIGHT RADIATION
AND CORPUSCLULAR RADIATION;
THE SIMPLE CHARACTER OF PHOTONS

In the first part of this work, we have taken generous advantage of the ideas and methods of light optics to develop the optics of electron lenses. Despite the deep-rooted analogy between these two disciplines, however, there are features which are peculiar to electron optics, the original nature of which we have already indicated (Chapter 6); it is only rarely that we can transfer directly into electron optics results which have been obtained for glass lenses. This same characteristic reappears, in a still more pronounced form, when we begin to consider applications, and is apparent as soon as we try to classify the various instruments of electron optics. An illustration of this is the "mass spectrograph" and the "velocity spectrograph"; in light optics, we speak only of the "spectrograph" without qualification, since the optical spectroscope is unique like the property of the radiation which it analyses.

A light source emits photons which are characterized by the frequency $\nu$ of the electromagnetic wave which is associated with them. The energy of each photon is $h \nu \ (h = 6.6253 \times 10^{-34} \text{ joule sec}^{-1})$ and the velocity in vacuo is $c \ (c = 2.997 \times 10^8 \text{ m sec}^{-1})$ whatever the frequency. In a medium of refractive index $n(\nu)$, the phase velocity falls to $c/n$, and is thus determined as soon as the frequency is known. A single type of spectrograph is adequate to analyse the radiation at the various frequencies. Further, a primary luminous source—an incandescent metal or a luminescent gas, for example—does not reveal its structure, does not display fine geometrical variations which might have provided an incentive to build a microscope to study emission phenomena.

11.2 MASS AND VELOCITY SPECTROGRAPHS;
DIFFRACTION DEVICES

Corpuscular radiation, which we shall now consider, is not wholly defined when the nature of the particles of which it is composed is known; for the nature of the radiation to be completely defined, the energy of the particles
must also be known. In particle optics, therefore, two new types of instrument appear. The first of these, the mass spectrograph, sorts the particles according to their mass, or more precisely, according to the ratio of their charge to their mass, \( e/m \), irrespective of their velocity. Instruments of this kind contain a magnetic prism which is sensitive to the quantity \( e/m \), in conjunction with an electrostatic unit which is insensitive to variations in \( e/m \); the chromatic aberrations of the latter are useful, however, to compensate those of the magnetic prism and thus to provide a combination which is barely sensitive to the velocity dispersion of the particles which are being analysed. The mass spectrograph has no real analogue in glass optics.

There are "velocity filters", on the other hand, which are insensitive to the value of \( e/m \), and which are used to ensure that the velocity of a beam of particles of various masses and charges is homogeneous. The most important members of this second class of instruments are those which are used to analyse the velocities of the beams of positive or negative electrons which are emitted by radioactive bodies; since these are the particles which constitute the \( \beta \)-radiation (\( \beta^+ \) or \( \beta^- \)), the instruments are known as \( \beta \)-ray spectrographs, and can be constructed from any prism or lens which has a large chromatic aberration. \( \beta \)-ray spectrographs may have either a magnetic or an electrostatic lens, therefore, (cf. § 7.4.3) and velocity spectrographs either a magnetic or an electrostatic prism. These instruments, which are the electron counterparts of optical spectrometers, will be compared and described in detail in the final chapter.

Grating spectrometers have been employed in optics since the work of Fraunhofer. An analogue of the grating is to be found in electron optics also, but here the grating is never artificial; instead, it is the natural three-dimensional grating which is formed by the regular lattice of atoms within a crystal. This, we recognize, is an inevitable consequence of the shortness of the de Broglie wavelengths (§ 6.1.5), which are of the same order of magnitude as those of X-rays. For this reason, electron diffraction cameras perform the same function as X-ray spectrometers, analysing the structure of a crystal with the aid of a beam of monoenergetic electrons, of known velocity. The diffraction mechanism of de Broglie waves in a three-dimensional crystal lattice is the same as that which governs the diffraction of light at the lines of a plane Rowland grating, but the field of applications of the former—crystal structure analysis—is as original as it is extensive. This technique is still undergoing rapid development.

11.3 CATHODE RAY TUBES; OSCILLOGRAPHS

A property peculiar to electron lenses and prisms is the basis for another category of instruments, oscillographs and the devices which are derived from them. Once a glass lens has been made, it is inflexible, and to focus the image, either the lens, the object or the screen must be shifted. In elec-
tron optics, on the contrary, the optical properties of the lenses can be modified simply by varying the potentials of the electrodes (or the currents in the coils). Focusing, therefore, is extremely straightforward, and it is also possible to alter the direction of a beam very rapidly by varying the deflection of a prism. As a result, instruments which are most inconvenient and of limited use in glass optics become wholly practicable and are widely used when we consider their electron optical counterparts.

The most striking example of such an instrument, one of the commonest of electron optical devices, is the cathode ray tube, very often no more than the electron equivalent of the inconvenient "mirror oscillograph", namely, the cathode ray oscillograph. The extremely small mass of the electrons is of unequalled utility as regards the lateral deflection; whereas mechanical devices are restricted to acoustical frequencies (of kc/s, or less) the cathode ray oscillograph is capable of tracing out the sinusoidal curve \( x = x_0 \sin 2\pi vt \), for values of the frequency \( v \) as high 10,000 Mc/s. Such extreme speeds are as yet rarely exploited, and the applications which are of the greatest everyday importance—radar and television—are still far from requiring this extreme rapidity from oscillographs; the qualities of which they do take the fullest advantage can be summarized as convenience, simplicity and sensitivity, the precise meaning of which will appear in the next chapter.

A number of very specialized applications of the cathode ray tube have been developed, in which the properties of the tube which we have just discussed are linked to the phenomenon of secondary emission. The electron beam is no longer incident upon either a fluorescent screen or a photographic plate, but upon a "target" which has a large secondary emission coefficient. The target is said to be "explored" with an "electron probe", and the intensity of the stream of secondary electrons provides the "signal" from which we can extract information about the target. Similarly, in television, we reconstruct the image of the "target" at which the camera is aimed, on the screen of the receiver with the aid of a series of amplifiers, a radio transmitter, and finally a "television oscillograph". Here, however, the target is artificial—the mosaic image formed by an iconoscope, for example—since it is the optical image of the scene which is transmitted. But in these other devices which are used in physics, we obtain the image of the structure of a natural target on an analogous principle. These probe microscopes, or microanalysers, which are of relatively recent development, are extremely valuable in metallurgy and crystallography.

11.4 MICROSCOPES

This brings us to a class of instruments which are much nearer to their light optical originals, electron and ion microscopes. But here once again, the transmission electron microscope, which is the most typical member
of the class, makes it possible to observe new types of interaction within the object which create the contrast which is at the very foundation of the image. Further, it provides us with an original solution to the standard problem of the resolving power. No longer can we work with stigmatic lenses: the familiar Weierstrass conjugate points have no useful analogue in electron optics, and as a result, the role of the aberrations is most important. It is a triumph of electronics to have been able to show both in theory and in practice that despite this it is possible to form well-separated images of two points less than a millimicron (or 10 Å) apart, by a skillful balancing of the harmful effects of aberrations and diffraction, and by correcting defects due to mechanical shortcomings in the construction.

Electron (or ion) microscopy produces one further new field of applications—the microscopy of sources of electrons or ions, with which we can see all the physically interesting features of a natural structure. On the screen of an emission microscope, we are shown this new aspect of matter at a high magnification (which may reach 5000). This is microscopy in the proper sense of the word, as we can study such phenomena as allotropic transformations at a high temperature, or the rapid evolution of grain boundaries, in conditions which are poorly adapted to any other means of examination. Further, by employing the electron emission from a very fine point subjected to an electric field of several million volts per centimeter, we can examine the actual arrangement of the molecules or atoms which compose the surface of the crystal from which the point is formed at a magnification of more than a million times.

11.5 TELESCOPES AND IMAGE-CONVERTERS

Yet another instance of the diversity of electron optical methods is the electron telescope (or "image-intensifier"), not a particularly fortunate choice of name. In fact, it is principally of interest for the transformation of one luminous image into another, by means of a long chain of operations, the first of which employs the photoelectric effect. A faithful reproduction of the initial image eventually appears on the fluorescent screen of a cathode ray tube; the reproduction however, is an "improvement" upon its original, not only in that it may be magnified, but above all in that it may be 100 or 1000 times brighter. The function of the telescope, in short, is that of a "brightness amplifier". Further, the final image is visible whereas the initial image—which might well be produced by ultraviolet or infra-red light—need not necessarily have been; in this case, the telescope functions as an "image-converter".

In this descriptive summary of the instruments of electron optics, it is their distinguishing characteristics that we have stressed, mentioning no more than their fundamental properties and their essential parts. Naturally,
“electronics”—a word which nowadays embraces a whole section of physics—finds its place in all these instruments, which owe their efficiency and convenience—their quality, in brief—to the individual electronic apparatus, interesting but involved, which is associated with each of them. Within the limits of such a study as this, it does not seem possible to describe all this electronics, and we shall in fact only consider the basic structure of the cathode ray tube, ignoring completely the special electronic circuits which are associated with it.

Only the truly optical features of electron optical instruments will be considered in the following chapters of this book. The lavishness of possible subject matter has already led to the banishment of accelerators which produce beams of particles of very high energy from the present work. The reader is referred to the relevant volume of Flügge’s Encyclopaedia of Physics and to Linear Accelerators, edited by P. Lapostolle and A. Septier (North-Holland, 1970).

The order in which we have set out the various instruments in this introduction will not be followed in the rest of the book; we shall begin with those instruments which most resemble their optical counterparts, and continue in order of increasing complexity as far as the mass spectrograph, the study of which is an excellent preparation for the instruments of nuclear physics.
CHAPTER 12

THE CATHODE RAY TUBE

12.1 HISTORICAL INTRODUCTION;

THE IMPORTANCE OF THE CATHODE RAY TUBE

The cathode ray tube is the oldest of the instruments of electron optics. As early as 1896, Braun built the tube which is named after him, in which the electrons were obtained by secondary emission, and gave a green spot at the end of the receiver. There were neither stops nor a fluorescent screen, as the “spot” was produced by the natural fluorescence of the glass. It was, however, possible to deflect the spot by applying a magnetic or electric field perpendicular to the beam, and on the screen the spot traced out curves from which the temporal variation of the applied field could be deduced; the first cathode ray oscillograph had been created.

The most fruitful advance was made by Wehnelt, who replaced the capricious secondary source (the behaviour of which is so dependent upon the pressure which prevails in the tube) by a hot, oxide-coated cathode. In addition, he introduced the additional electrode which bears his name in the present-day Continental literature (the grid, in English). After this, the cathode ray tube could be said to involve genuine “electron optics”—fine, bright spots could be obtained, the brightness of which could be adjusted by varying the potential of the grid. It became possible to trace out well-contrasted pictures on the screen, by breaking them down into the curves which the spot followed under the action of two time-varying potentials applied to the deviating plates. The intensity of the spot was controlled independently by a third potential, which was applied between the cathode and the grid. The principle of the television receiver had been discovered. This is the principle, too, of the radar receiving unit, so rapidly developed as a result of the exigencies of the second world war. With three potentials, which can be varied at will in time, either by amplifiers (in the case of the potential which originates at an aerial and that which acts on the grid) or by scanning oscillators (which displace the spot), a whole map of the region which is being explored by the radio beam can be produced on the fluorescent screen.

The importance of radar and television to-day—in wartime as in peace—has made this one of the most highly perfected techniques of the age. Cathode ray tubes are mass-produced by machines just like radio valves.
A whole mass of research, often enough empirical, has led to a considerable improvement of the properties of Braun's and Wehnel's tubes—the fineness of the spot, its luminosity or the sensitivity—without introducing any fundamental modification of the basic design. A second lens has been introduced, which can be either electrostatic or magnetic, in such a way as to obtain a fine and brilliant spot, and the shape of the plates or coils which produce the deviation has been modified so that the spot remains as small as possible wherever it may be on the fluorescent screen. Again, great progress has been made in the chemistry and crystallography of fluorescent powders. Instead of 0·1 per cent, the efficiency with which the kinetic energy at the point at which the electrons hit the fluorescent screen is transformed into light can be of the order of 25 per cent, and a handsome range of colours is available, for colour television. In high tension oscillographs (functioning between 30 and 100 kV), the televised image is brilliant enough to be projected onto a cinema screen. Finally, technological advances (developments in vacuum techniques, and in sealing into glass, for example) have made possible a considerable reduction in the manufacturing cost of these instruments, which are virtually of everyday use in the sociological sense of the phrase.

So well stocked have the manufacturers' catalogues become that almost every conceivable requirement of the engineer, the physicist, or the television amateur is catered for. Even so, new types of tube appear on the market every year, for the inventive efforts of the physicists concerned keep the models available in a state of continual evolution. For this reason, a method of scaling, particularly well illustrated in the work of H. Moss (1950), has appeared which makes an ingenious use of dimensional relations to solve a number of the problems with which designers are faced. For example, suppose we have a direct vision cathode ray tube with an exceptional performance, which has a screen 38 cm in diameter; in what way must we modify the design to obtain the characteristics of a tube which is to be used in a projection television set, which has a screen with a diameter only one-third the size, but which is, of course, to provide no less detail in the image? The diameter of the spot will be reduced to one-third of its previous value, but the fact that the image is to be projected means that the luminous intensity must be greatly increased (perhaps some five times). The solution, however, has to be subject to certain restraints—there are quantities, like the specific emission of the cathode, which cannot, unfortunately, be varied between very wide limits. The dimensional relations which have been established by Moss provide an acceptable solution which is more than a simple geometrical similitude: the geometrical dimensions of the envelope will be one-third of those of the original tube, but the size of the electrodes in the gun is reduced only in the ratio 1 : 0·63; the direct vision tube is operated at 8 kV, while the projection tube requires 28 kV. The current at the spot of the latter reaches 60 μA, whereas it had previously
been 150 \( \mu A \). The laws by which we pass from one situation to another are thus rather complex, and for details of this ingenious method of "generalized similitude" the reader is referred to the original article by Moss.

This study of cathode ray tubes will be restricted to the basic faculties of focusing and deflecting the beam. Nor shall we consider construction techniques, again referring the reader to the article by Moss (1950) mentioned earlier, and also to the book by Say (1954). A much quicker and more effective way of learning, however, is to "perform an autopsy" on a discarded tube, attempting afterwards to reassemble it with the aid of a vacuum pump.† Similarly, we shall not attempt the description of the various auxiliary devices which have to be added to a cathode ray tube to transform it into a cathode ray oscillograph, a television receiver, or a radar receiver. For further information on such points the reader is referred:

(a) to the books and articles by Rider and Uslan (1950), Bonvalet (1952), Wilson (1953) and Say (1954), on cathode ray oscillographs;
(b) to Soller, Star and Valley (1948) for radar receivers; and
(c) to the publications of Maloff and Epstein (1938), Zworykin and Morton (1954), Delaby (1951), Delbord (1951), Puckle (1951) and Grivet and Herreng (1953) for information about television receivers.

In these are to be found details of the construction of scanning oscillators, amplifiers, and stabilized sources, in short, of accessories, which are easier to manipulate and construct than the tubes themselves and are of simpler structure.

It is convenient to conclude this section by pointing out that the same form of cathode ray tube is to be found only slightly modified in the television camera. In the image iconoscope are to be found all the elements which have been mentioned earlier—gun, focusing lens, deflection system. The difference between the two instruments resides not in the electron optics, but in the surface on which the beam impinges. The spot no longer sweeps across a fluorescent screen, producing light, but instead, across a mosaic of photoelectric cells; its purpose is to measure the charge which has been acquired by the cathode of each of these photoelectric microcells as a result of the light which has been shone upon it. For this new function, an extremely fine spot is necessary, but on the other hand, only a low beam current—a few microampères, ten or a hundred times less than in an oscillograph—is required. It is due to this that the gun of an iconoscope is of an essentially different design; again, further information is to be found in the relevant references which have already been quoted.

† Opening a television tube, it must be realized, is a delicate operation. The gun cannot be recovered by breaking the tube, which would implode, as this would be a dangerous procedure. The tube has first of all to be filled with air, by producing a leak near the point at which the tube had been sealed off.
12.2 FORMATION OF THE SPOT

12.2.1 The General Appearance of the Optical System

A cathode ray tube is an instrument which is sealed off at a static vacuum. Only in a few oscillographs, which are designed to be used at very high tension (60 kV) to study rapid phenomena (of frequencies above a few Mc/s), is the vacuum a dynamic vacuum, maintained by a diffusion pump as in the electron microscope. It is worth mentioning that Dufour used a dynamic vacuum in one of the earliest cathode ray oscillographs, while his “scanning system” consisted in moving the photographic plate. The most recent developments in this field are those of von Ardenne (1956). We shall limit ourselves to studying sealed-off systems, protected by a glass envelope the shape of which is well-known. All the electrical connexions are led out through the base, with the exception of the “post-acceleration” lead the purpose of which is to maintain the fluorescent screen and the surrounding part of the tube at a high potential (up to 40,000 V). As Fig. 116a shows, this connexion is led out of the tube through a glass-metal seal, for the obvious reason of insulation.

It is no more difficult to achieve the vacuum of \(10^{-6}\) mm Hg which prevails inside the tube than it is in the case of a radio valve, but it is less easy to maintain as the way in which the screen is made can lead to the production of volatile residues, which are not present in radio valves, and which have a most harmful effect upon the emissive properties of the cathode. To overcome this hazard, the tube is outgassed by a very careful baking, and an extremely active “getter” (made of one of the alkaline earths) is used, which absorbs the residual gas if it is heated while the tube is being evacuated. The envelope has to be able to withstand the stress due to atmospheric pressure over a considerable surface area, which creates difficult problems in glass or steel-and-glass technology (the base of a television tube 35 cm in diameter has to be capable of withstanding a force of the order of a ton weight).

The various parts of a typical tube (Fig. 116a) are, starting from the base:
—first an oxide-coated cathode \(K\) about 1 mm in diameter, in the form of a small disc of magnesium-nickel alloy (1 per cent Mg) coated with a mixture of barium oxide and strontium oxide. This is heated with a tungsten filament to the working temperature which is normally of the order of 1000°K, and is chosen in such a way that the mean emission at the cathode allows a long lifetime;
—a cylinder or ring, \(W\), which is known as the “grid” or “Wehnelt”; in normal operation, this electrode is held at a potential of the order of a few tens of volts, negative with respect to the cathode;
—an anode, \(A_1\), which accelerates the electrons to a few thousand volts and which in general consists of a metal cylinder together with a plane
plate in which a hole of a few millimetres diameter has been cut, through which the beam passes.

This combination represents the simplest type of gun, the triode gun, which makes the electrons converge to a very fine "crossover" in the neighbourhood of the hole in the anode. By varying the grid potential, the crossover can be moved away from the anode, but the linear magnification of this immersion lens rapidly becomes very large (Chapter 12) and the crossover of a prohibitively large diameter.

Triode guns are rarely used directly, but rather, in conjunction with a supplementary electron lens. A second anode $A_2$ is placed beyond $A_1$ and the two form a two-tube accelerating lens; the potential $V_2$, measured with respect to the cathode, is higher than $V_1$. By varying the ratio $V_2/V_1$, the convergence of the lens is varied, and the position of the focus can be varied between wide limits; the whole system constitutes a "tetrode gun".

The electron beam then passes through a series of electron prisms which attend to the deflexion, and finally falls on a fluorescent screen; the glass envelope is rendered conductive by depositing a layer of carbon (aquadag or lamp-black), so that the beam travels through a region which is genuinely field-free from the static point of view. The end of the tube which carries the screen is sometimes insulated from this layer and raised to a very high potential which accelerates the electrons during the later part of their trajectory (post-acceleration) to energies which may reach 30–50 kV.

The electrical circuit is closed by virtue of a thin metallic layer which covers the fluorescent screen and which is connected to the source of high tension. This layer is transparent to electrons and increases the brightness of the screen, by returning any light which could have been emitted into the interior of the tube back towards the observer (Epstein and Pensak, 1946). For further information on this point, the reader is referred to §12.2.5 on fluorescent screens; to §8.4 on guns, and to §8.1 on two-
cylinder lenses. Complementary information about the guns which can be used in cathode ray tubes is to be found in the articles of Morton (1946), Field (1946) and Moss (1946, 1946a, 1950).

The whole tube assembly is enclosed within a mumetal (or permalloy) shield, which eliminates the perturbations which exterior parasitic magnetic fields could produce; such fields are in particular produced by transformers in the supply unit. An alternating field with peak amplitude 0.5 gauss, acting over a distance of 4 cm in the neighbourhood of a gun operating at 1000 volts is enough to displace the spot 4 mm on the screen 20 cm away, and thus to mangle the image completely.

Nowadays, guns of a rather more complicated design are employed, with which the different adjustments to the gun can be clearly separated. "Focusing" consists in bringing the image of the crossover of the beam onto a fluorescent screen; this crossover lies at the exit of the triode system formed by K, W and A1, which here behaves as a strongly convergent lens. The focusing is carried out by varying the potential V1 of the anode, leaving the final accelerating potential of the electrons (V2) constant; V1 is also the potential which extracts the electrons, however, and there is in consequence a slight variation of the emitted intensity. The brightness on the other hand, is adjusted by varying w, the potential of the grid; the action resembles that of a triode valve, in which the emissive area of the cathode is varied. This variation of w is accompanied by a slight change in the convergence of the triode system, and hence by an axial displacement of the crossover which entails a slight defocusing. The coupling between the two adjustments is small enough in ordinary tubes, however, not to be troublesome; this explains the widespread use of the tetrode gun.

\[ g = 0 \quad w \quad v_1 \quad v_3 \quad v_2 = v_1 \]

\[ K \quad W \quad A_1 \quad A_3 \quad A_2 \]

Gun Unipotential lens

Fig. 116b. Diagram of a pentode gun.

The interaction of brightness and focusing can be suppressed by replacing A2 by a magnetic lens, situated outside the tube. Alternatively, the accelerating gap A1 A2 can be replaced by a genuine einzel lens by inserting a further electrode A3 between A1 and A2. These two electrodes are then connected electrically and held at an accelerating potential V2; the system can be focused by varying the potential V3 of the central electrode A3. The result is a pentode gun (see Fig.116b). The lateral extent being so great, the glass
envelope has to be large in diameter (5–10 cm), and this in turn allows us to use lenses with large openings and hence to reduce the spherical aberration. There are also advantages in using large diameter guns so far as the modulation of the brightness of the spot is concerned, as we shall see later. Figures 116c and 116d show examples of such constructions. In Table 1, we quote some characteristic parameters of cathode ray tubes (tubes constructed by RCA).

<table>
<thead>
<tr>
<th>Number</th>
<th>2 API-A</th>
<th>7 DP-4</th>
<th>16 AP-4</th>
<th>7 NP-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Oscillograph</td>
<td>Television set</td>
<td>Projection on a large screen</td>
<td></td>
</tr>
<tr>
<td>DIMENSIONS (in cm):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>18</td>
<td>35</td>
<td>55</td>
<td>45</td>
</tr>
<tr>
<td>Screen diameter</td>
<td>5.2</td>
<td>18</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>GUN:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut-off potential</td>
<td>- 60</td>
<td>- 27</td>
<td>- 40</td>
<td>- 125</td>
</tr>
<tr>
<td>Anode potential</td>
<td>1000</td>
<td>410</td>
<td>300</td>
<td>625</td>
</tr>
<tr>
<td>FOCUSING (the first electrode is the anode of the gun)</td>
<td>3-electrode lens</td>
<td>4-electrode lens</td>
<td>Exterior Magnetic lens</td>
<td>2-electrode lens</td>
</tr>
<tr>
<td>1000 V</td>
<td>410 V</td>
<td>6000 V</td>
<td>1200 V</td>
<td>625 V</td>
</tr>
<tr>
<td>250 V</td>
<td>6000 V</td>
<td></td>
<td></td>
<td>16000 V</td>
</tr>
<tr>
<td>1000 V</td>
<td>1200 V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCREEN POTENTIAL: (P.A.: post-acceleration)</td>
<td>1000</td>
<td>6000</td>
<td>12000 (P.A.)</td>
<td>80000 (P.A.)</td>
</tr>
<tr>
<td>DEFLECTION</td>
<td>Electric; sensitivity: 0.11 mm/V (vert.)</td>
<td>Magnetic; outside the tube (see the examples in § 12.3.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.13 mm/V (horiz.)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The reader will find a comparison of the deficiencies and advantages of tetrode and pentode guns in an article by Darbyshire (1955).

Oscilloscopes with a very high resolution (a spot of the order of 0.1 mm diameter) are provided with an improved optical system. First of all, the tetrode gun must be selected in such a way that for a given distance between gun and screen, the linear magnification is as small as possible; experience shows that it is best to use a two-tube lens (the two tubes formed by $A_1$ and $A_2$) in which the diameter of $A_2$ is considerably less than that of $A_1$ ($d_2/d_1 = 0.33$, for example). The size of the final spot is still further reduced by using an einzel objective lens $L$ in conjunction with $A_1A_2$; $L$ is placed
Fig. 116c. Cathode ray tube gun. (Photo: C. S. F., France - No. 504, 699, 63.)
Fig. 116d. The F8071 Thomson-CSF tube, with electrostatic focusing and deflexion, and with post-acceleration, for studying high frequency signals (up to 150 Mc/s). This tube is equipped with a deflexion amplifier using quadrupole lenses (which can be seen in the detailed view of the gun). Observations are made on a 120 mm × 80 mm flat screen. (Photo: C.S.F.)
at a suitable distance between $A_2$ and the deflexion plates. The latter, as we shall see later, behave like weak cylindrical lenses, which results in over-focusing of the deflected spot and elongation of the spot in the direction perpendicular to that of the deflexion. This can be remedied to some extent by introducing in front of the plates a focusing element which is itself astigmatic, the horizontal and vertical convergences of which can be adjusted independently. At best, the spot is then constituted by the circle of least confusion which lies mid-way between the two real line foci of the beam. An elegant way of achieving this is to replace the central electrode of the objective $L$ by two electrodes with elliptical openings, crossed, excited with two independently adjustable potentials (Cooper, 1961). In this way, we can obtain a spot, the diameter of which remains practically constant throughout all the useful part of the screen.

12.2.2 The Spot

(i) The dimensions of the spot

If we denote the ratio of the distances between the lens and the screen and between the lens and the crossover by $K$, and the focal length of the immersion lens which constitutes the gun by $f$, an application of the Lagrange–Helmholtz formula (§ 4.3.5) yields the diameter $d_s$ of the spot,

$$d_s = 2Kf\sqrt{\frac{T}{11600\Phi_0}}, \quad (12.1)$$

in which $\Phi_0$ is the potential of the fluorescent screen and $T$ is the temperature of the cathode. For the common kinds of television tube, the orders of magnitude are $T = 1160^\circ K$, $\Phi_0 = 10,000$ V, $K = 5$ and $f = 3$ mm, which leads to a value for the diameter in the region of 100 $\mu$. This has nothing to do with the spot into which the point is spread by the distribution of the thermal velocities of the electrons, and the effect of the aberrations of the lens has also to be superimposed. Nevertheless, the defects are not sufficiently serious to prevent the tube from being used in practice. In France, for example, the television image is broken up into 819 horizontal lines, which means that the spot on a screen 30 cm high must be 360 $\mu$ in diameter (that is $\frac{30 \times 10^4}{819}$). The margin for the other defects is quite adequate. The plates, on the other hand, deform the spot much more markedly, above all when the spot is close to the edge of the field, a point to which we shall return later. We can now understand one of the reasons why the tube is operated with such a high potential at the screen, some 40 or more kV. As the spot diameter varies inversely as $\sqrt{\Phi_0}$, this is a way of obtaining a finer spot; we shall reconsider this in § 12.2.4.
(ii) The intensity of the spot

If \( i_c \) represents the current density emitted at the cathode, and if we assume that \( i_c \) is uniform inside the region of diameter \( d_o \) over which electrons are emitted, then the current density in the spot, \( i_s \), is given by the formula developed by Langmuir (1937) which we have already obtained in § 4.3.5:

\[
i_s = i_c \left( \frac{11,600 \Phi_0}{T} \right) \alpha_s^2,
\]

where \( \alpha_s \) is the semi-aperture of the beam which produces the spot at the screen. In fact, of course, the current density is not uniform across the whole spot; Law (1937) suggested that the distribution law might have a Gaussian form, which was subsequently verified by Dosse (1940) and by Jacob (1939). More recently, Beam (1955) has invented a simple way of producing an image of the spot on the screen magnified forty times. It is thus possible to measure the intensity very precisely (Fig. 117). The spot is moved with the aid of magnetic deflexion over a metal plate in which an extremely small hole has been punched (with a diameter of about \( 2 \times 10^{-2} \) mm). This plate lies in the plane in which the spot would normally have been formed. For a given deflexion, part of the beam passes through the hole and impinges on the fluorescent screen which is placed a few centimetres beyond the plate. The deflexion device is still effective beyond the plate, however, so that the electrons move away from the axis as they travel between the plate and the screen, and the result, effectively, is a magnification. For example, consider the electrons which would have been 100 \( \mu \)

![Diagram](image-url)
from the middle of the spot; in order to pass through the hole, they have had to be deviated 100 μ across the plate, but as the electron beam continues to move away from the axis beyond the plate, the electrons strike the fluorescent screen some millimetres from the central point. We obtain, therefore, a magnified spot.

![Diagram](image)

**Fig. 118.** The apparent broadening of the spot produced by increasing the intensity (II) and the genuine broadening due to space charge (III).

The current density $i_c$ which is emitted cannot, however, exceed a certain value (0.25 A cm$^{-2}$) without the risk of exhausting the cathode and the danger of reducing its lifetime; (normally a cathode should last for at least 2000 hours). $\alpha_s$ can barely exceed 0.01, as otherwise the geometrical aberrations become too prominent, and naturally we cannot reduce $T$, as $i_c$ is dependent upon $T$, being a rapidly increasing function of temperature. In practice, $T$ has to remain in the neighbourhood of 1000°K, for the decrease of $i_c$ outweighs the increase of $\frac{1}{T}$ if $T$ is reduced. We can, therefore, only increase $\Phi_0$, which is a second reason for using high anode potentials.

### 12.2.3 The Effect of Space Charge on the Beam

The mutual repulsion between the electrons also prevents an intense electron beam from being perfectly focused. This real effect is not, however, to be confused with a deceptive effect which is always observed and which we shall first of all describe. It is easy to verify that if the brightness of a
television tube or an oscillograph is increased (by making the grid positive) the diameter of the spot appears larger. This impression can be explained in terms of the properties of the eye in the following way. The light distribution across the spot is represented by a bell-shaped curve, as shown in Fig. 118, and not by a uniform distribution; the eye, however, is sensitive to brightness only above a certain level, below which no impression is made. The apparent spot diameter, therefore, is obtained by drawing a line parallel to the abscissa axis at the height which corresponds to the threshold sensitivity of the eye; since the principal effect of increasing the brightness of the spot is to widen the distribution-curve for each value of the ordinate, it is plain that the diameter of the visible disc, obtained by the construction which has just been described, increases, quite irrespective of any broadening which may be produced by space charge effects.

Having disposed of this subjective illusion, we can pass on to an objective study of the broadening which space charge produces, which can be measured either with a Faraday cage, or by a number of other ingenious methods. We shall demonstrate that the space charge plays a role which is comparable with that of a weak divergent lens, supposed to be superimposed onto the lens of the cathode ray tube, but which cannot be adjusted. We shall examine the behaviour of the beam which leaves the lens with a radius \( r_0 \) and an angular aperture \( r'_0 \), negative since the beam is convergent (see Fig. 119). Normally, the beam would form a point spot after a distance

\[
z_0 = \left| \frac{r_0}{r'_0} \right|.
\]

We assume that the potential in that part of the tube which follows the lens is constant, so that the electrons travel at a constant velocity. If this potential is represented by \( \Phi_0 \), then the velocity, which is almost axial, is given by

\[
v = \sqrt{\frac{2e\Phi_0}{m}} \approx \frac{dz}{dt}.
\]

Consider now an electron on the edge of the beam—it experiences both the electric field \( E \) which is created by all the other electrons in the beam, together with the magnetic field which is associated with the current to which a beam of electrons in motion is equivalent. The effect of this magnetic field we can ignore, for being proportional to \( \left( \frac{v}{c} \right)^2 \) it is very small, of the same order of magnitude as a relativity correction. The electric field, which is radial, can be calculated by applying Gauss’ theorem to a small cylinder with axis \( Oz \), length \( dz \) and radius \( r \). The flux through the surface is given by

\[
\text{flux} = 2\pi r \cdot dz \cdot E_r
= \frac{1}{\varepsilon_0} q \pi r^2 \cdot dz.
\]
The charge density, however, is related to the current \( I \) and to the axial velocity, and hence to the potential, in the following way:

\[
\pi r^2 v_0 = I,
\]

\[
\varrho = \frac{I}{\pi r^2} \sqrt{\frac{m}{2e \Phi_0}},
\]

and consequently

\[
E_r = \frac{1}{2 \pi \varepsilon_0} \sqrt{\frac{m}{2e}} \frac{I}{\sqrt{\Phi_0}} \frac{1}{r}.
\]

The lateral equation of motion takes the form

\[
\frac{d^2 r}{dt^2} = \frac{e}{m} E_r = \frac{1}{2 \pi \varepsilon_0} \sqrt{\frac{e}{2m}} \frac{I}{\sqrt{\Phi_0}} \frac{1}{r}.
\]

On eliminating the time, remembering that the axial motion is uniform, we find

\[
\frac{d}{dt} = \sqrt{\frac{2e \Phi_0}{m}} \frac{d}{dz},
\]

and finally arrive at the equation for the trajectory

\[
\frac{d^2 r}{dz^2} = \left(\frac{1}{2} \beta^2\right) \frac{1}{r},
\]

with

\[
\beta = \frac{1}{2 \pi \varepsilon_0} \sqrt{\frac{m}{2e}} \frac{I}{\Phi_0^{3/2}} \approx 3 \cdot 10^4 \frac{I}{\Phi_0^{3/2}}.
\]  

(12.3)

Multiplying both sides by \( \frac{dr}{dz} \) and integrating, we find, on substituting the initial conditions, that

\[
r'^2 - r_0'^2 = \beta \log \left(\frac{r}{r_0}\right).
\]  

(12.4)

This equation has often been discussed in the literature—the reader is referred to the articles by Watson (1927), Zworykin (1933) and to a very detailed examination by Ívey (1954). We shall outline the method by which it may be integrated. We have

\[
\frac{dr}{dz} = \beta^{1/2} \left(\log \frac{r}{r_0} + \sigma^2\right)^{1/2},
\]  

(12.4(i))

with

\[
\sigma^2 = \frac{r_0'^2}{\beta}.
\]

We write

\[
\log \frac{r}{r_0} + \sigma^2 = \xi^2,
\]
so that
\[ r = r_0 e^{-\sigma^2} e^{e^z}, \]
\[ \frac{dr}{dz} = 2r_0 \xi e^{-\sigma^2} e^{e^z} \frac{d\xi}{dz}, \]  (12.4(ii))
and from equation (12.4 (i)):
\[ \sqrt{\beta} \, dz = 2r_0 e^{-\sigma^2} e^{e^z} \, d\xi. \]
The origin is selected at \( z = z_0 \), so that
\[ z - z_0 = \frac{2r_0}{\sqrt{\beta}} e^{-\sigma^2} \int_0^\eta e^{e^{e^z}} d\xi, \]  (12.4(iii))
and tables of the integral
\[ D(\eta) = \int_0^\eta e^{e^z} d\xi \]
are available (Dawson, 1898; Terrill and Sweeney, 1944a, 1944b).

If the beam is initially parallel to the axis \( (r_0' = 0, \sigma = 0) \), the relation takes the simple form
\[ z - z_0 = \frac{2r_0}{\sqrt{\beta}} D(\sqrt{\log \frac{r}{r_0}}), \]  (12.4(iv))
since
\[ D(0) = 0. \]

Given the radius \( r_0 \), the displacement and the factor \( \beta \) by which the beam is characterized, we can extract the value of \( D(\sqrt{\log \frac{r}{r_0}}) \) from (12.4(iv)). If we consult the curve plotted in Fig. 119a, we can obtain the value of \( \sqrt{\log \frac{r}{r_0}} \) on the abscissa axis, and hence \( \frac{r}{r_0} \) or the widening of the beam.

If the beam is initially convergent (see Fig. 119b), there will be a point with abscissa \( z_m \), where the radius of the cross-section is a minimum, \( r_m \), and the slope zero. From (12.4 (i)),
\[ r_m = r_0 e^{-\sigma^2} = r_0 \exp \left(-r_0'^2/\beta\right). \]  (12.5)
This would be the radius of the spot if only the space charge acted. In reality, the effect due to the aberrations of the lens and the gun must be added. The abscissa of the minimum is given by:
\[ z_m - z_0 = \frac{2r_0}{\sqrt{\beta}} e^{-\sigma^2} \int_0^\eta e^{e^z} d\xi, \]
or since $\sigma$ is negative,

$$z_m - z_0 = \frac{2r_m}{\sqrt{\beta}} \int_0^{|\sigma|} e^{\xi^2} d\xi = \frac{2r_m}{\sqrt{\beta}} D(|\sigma|). \quad (12.5(i))$$

**Fig. 119a.** The curve from which the broadening of the beam due to the effect of space charge can be calculated.

**Fig. 119b.** The appearance of the outermost ray of the beam: the zone of minimum cross-section (waist).
Equation (12.4 (iii)) takes the form
\[ z - z_0 = \frac{2r_0}{\sqrt{\beta}} \, e^{-\sigma} \left[ D \left( \left\lfloor \log \frac{r}{r_0} + \sigma^2 \right\rfloor \right) - D(\sigma) \right]. \]

But
\[ \log \frac{r}{r_0} + \sigma^2 = \log \frac{r}{r_m}, \]
and hence
\[ z - z_m = \frac{2r_m}{\sqrt{\beta}} \, D \left( \sqrt{\log \frac{r}{r_m}} \right). \tag{12.6} \]

We therefore first calculate \( \beta \), then \( \sigma \), and finally \( r_m \). Figure 119a gives \( D(\sigma) \) and equation (12.5(i)) then provides \( z_m \). For \( z > z_m \), we are brought back to the preceding case, as the initial conditions are now \( r = r_m \) and \( r' = 0 \) at \( z = z_m \). Equation (12.6) is in effect identical to (12.4(iv)).

If the space charge did not act, the beam would converge onto \( z = z_F \) (Fig. 119b). We have
\[ z_F - z_0 = \frac{r_0}{r_0}, \]
and
\[ r_m = r'_0 \, z_0 \, e^{-\sigma^2} = \sqrt{\beta \, \sigma} \, z_0 \, e^{-\sigma^2}. \]

From (12.5 (i))
\[ z_m - z_0 = 2 \sigma \, z_0 \, e^{-\sigma} \, D(\sigma), \]
and finally
\[ \frac{z_m - z_0}{z_F - z_0} = 2 \frac{r_0}{\sqrt{\beta}} \, \exp \left( -\frac{r'_0}{\beta} \right) \, D \left( \left| \frac{r'_0}{\beta} \right| \right). \tag{12.7} \]

In general, \( \frac{z_m - z_0}{z_F - z_0} \neq 1 \), except for one particular value of \( \sigma \); we consider a beam for which \( r_0 \) and \( r'_0 \) remain constant in \( z = 0 \), and we decrease \( \sigma \), increasing the factor \( \left( \frac{1}{\sqrt{\beta}^{1/2}} \right) \). The ratio \( \frac{z_m - z_0}{z_F - z_0} \) is at first equal to unity, increases to about 1.3, and subsequently decreases and becomes much smaller than unity for very small values of \( \sigma \) (see Fig. 120). The maximum occurs for \( \sigma^2 = 2.28 \). The influence of space charge will be lessened when \( \sigma \) is greater and hence \( \beta \) smaller.

Since \( I \) cannot be diminished too far without reducing the brightness of the spot, \( \Phi_0 \) must be increased. Yet again, we meet a reason for using a high screen potential. We shall see later that when it is not possible to work with a very high potential throughout the whole tube, we can nevertheless “post-accelerate” the electrons in the neighbourhood of the screen alone. The equations of motion above are no longer applicable in a region of accelerating electric field; Moss (1945), Sangster (1955) and Moak (1960)
have studied quantitatively the case of beam-widening under the action of space charge for the case of a constant electric field.

In conclusion, we should point out that the influence of space charge is progressively less as the beam diameter is larger. The space charge therefore has little effect during the passage through the lens which follows the gun.

![Diagram showing the position of the minimum cross-section with respect to the geometrical focus.](image)

**Fig. 120.** The position of the minimum cross-section with respect to the geometrical focus. (The influence of space charge increases when \( \sigma^2 \) decreases.)

### 12.2.4 Modulation of the Intensity of the Spot

Now that we have seen how a fine spot can be produced by the electrons, and reviewed the various causes due to which it may be broadened, we must analyse the mechanism by which we can vary the intensity of the current which falls on the spot. Modulating the spot is not in itself a difficult problem, but difficulties appear when we consider how to modulate the intensity while at the same time keeping the spot as fine as possible. In practice, we adjust the potential of the grid. The potential of this electrode is usually negative, a few tens of volts; if its potential is varied, the intensity varies correspondingly, and a characteristic curve can be plotted which relates the beam current to the modulation of the potential. This curve is similar to the characteristic of a triode valve; an example is shown in Fig. 121. The grid acts upon the brightness by modifying simultaneously both the area of the cathode from which the electrons are emitted and the emissive power. The experimental investigations of Moss (1946a) show that the diameter \( d_e \) of the part of the cathode which is actively emitting is proportional to the "conduction potential", that is, to the difference between the
Fig. 121. An example of the characteristic curve of a cathode ray tube (10BP4); this curve gives the intensity and brightness of the spot as a function of the Wehnelt (grid) potential, measured relative to the cut-off value.

actual grid voltage and the value at cut-off; we shall denote the potential at cut-off by $\Phi_c$. If $\Phi_w$ is the grid potential and $d_w$ its diameter, $d_c$ is given by

$$d_c = d_w \frac{\Phi_c - \Phi_w}{\Phi_c}.$$  

The cut-off potential is related both to the geometry of the gun and to the anode potential; experiment shows that it is proportional to the latter,

$$\Phi_c = K \Phi_a. \quad (12.8)$$

In modern tubes, the dimensions of the electrodes are chosen in such a way that the negative cut-off potential is of the order of a few tens of volts, while the maximum current is obtained when the grid potential is zero. The spot can thus be controlled with the aid of a simple voltage divider, so named because it takes no current and therefore provides negligible power. Also, we avoid the danger of drawing "grid current" from the beam. The current emitted at any point of the cathode depends both on the temperature at the point and on the space charge in the vicinity of the point, exactly as in a diode valve; the electron gun is of course no more than a variant of the latter. The characteristic curve of a diode valve, which relates the intensity $I$ to the potential of the plate, falls into two parts
THE CATHODE RAY TUBE

(Fig. 122). One section consists of the curve known as the Langmuir region which is described by

\[ I_c = p \Phi_0^{3/2}, \]

while the other is horizontal, and corresponds to saturation. The current at saturation is given by Richardson’s law, which states that

\[ I_s = I_0 T^2 \exp \left( -\frac{e \Phi_s}{kT} \right), \]

in which \( \Phi_s \) is the work function of the cathode.

![Diagram](image)

**Fig. 122.** The characteristic of a diode: I, saturation conditions; II, space charge limited conditions.

Returning to the electron gun, we find that here too the emission can become saturated if the temperature is low (900°K); this is apparent as the potential of the grid becomes ineffective as a regulator of the emission. This is, therefore, an exceptional situation which is only of interest if we are exploring the properties of the tube; the emission is constant, and is governed by Richardson’s law. In normal conditions, on the contrary, the characteristics are chosen in such a way that the intensity at saturation is far higher than the values which space charge allows when the grid potential is varied over the whole range of useful potential, between cut-off and full aperture. The mean emission is then given by a law similar to Langmuir’s; the coefficient \( p \) is known as the “perveance” of the gun. It is not a particularly useful measure of the quality of the guns which are used in cathode ray tubes, since in these it is the increase in the emitting area which plays the essential role, while \( p \) only varies over a rather narrow range. This parameter is less useful in this context than in connexion with the high-intensity guns which are used in klystrons.

Since the mechanism which governs \( I_s \) is so complex, it is difficult to obtain a theoretical estimate of the intensity of the spot. The intensity can,
on the other hand, be measured, and the measurements which Moss (1946a) and Jacob (1952) have made verify satisfactorily that \( I_s \) obeys a law of the form

\[
I_s = K \frac{(\Phi_e - \Phi_w)^{3/2}}{\Phi_e^2},
\]

(12.9)

in which \( I_s \) is the total intensity at the spot. \( K \) varies with the dimensions of the gun. The law which relates the potential by which the intensity is controlled to the intensity at the spot is not, therefore, linear. The linearity of this curve has been improved in certain modern guns without any serious loss in sensitivity by constructing the gun in a special form which has been studied in detail. It must be admitted, however, that the eye is extremely tolerant towards the defects which are present in a television picture, for example, as a result of the curvature of the modulation characteristic. The acceleration and the focusing of electrons are clearly separated functions in a pentode gun; with such a gun, therefore, we can obtain a final spot of which the position is almost independent of the grid potential. Throughout the process of modulation, the size of the spot remains constant; in the present state of the technique, it seems that it is by the distribution of the emission velocities of the electrons as they leave the cathode that the fineness of the spot is limited (Dufour and Wendt, 1954).

If a pentode gun is used, another way of modulating the spot is feasible. In the central zone of the unipotential lens which follows the triode gun, the electrons are considerably slowed down; if a supplementary negative potential is applied to the central electrode, the lens can be gradually transformed into a mirror (see Chapter 8), the outer part of the beam being halted first. Electrons which do not manage to pass the potential barrier are reflected back towards the cathode. It has been found necessary to design lenses which satisfy the following two requirements:

(i) The modulation characteristics should be as linear as possible, and very steep, so that a potential of about ten volts suffices to cut off a beam which has been accelerated through several hundred volts.

(ii) The spot must not be defocused during the modulation.

This can be achieved by employing a highly convergent lens. The divergent beam which emerges from the gun converges again into an extremely fine crossover at the point in the lens where the potential has a minimum, and then diverges again as it leaves the lens. A supplementary accelerating lens converges the beam onto the final screen. If the central electrode includes a stop of very small diameter, the ability of the lens to cut off the beam is considerably enhanced, and in addition, the optical properties of the two parts of the lens (before and after the diaphragm) are virtually independent. We can then hold the exit electrode at a potential appreciably higher than that of the anode of the gun. A particularly elegant solution of the problem is due to Schlesinger (1961); the cut-off potential of the lens
is about 12 volts for 500 volt incident electrons. The focusing does not vary during the modulation, and for the same current, the spot which is obtained is finer than that of an ordinary tube modulated by the grid.

12.2.5 The Fluorescent Screen

Finally, the electron spot has to be converted into a spot of light, the luminous intensity of which must vary as nearly as possible proportionally with the electron intensity. In §1.4, we have described the essential properties of the fluorescent materials. Here, we shall give the complementary results which concern cathode ray tubes in particular, and with the aid of which, the efficiency with which the kinetic energy of the electrons is transformed into light can be considerably increased.

![Graph](image)

**Fig. 123.** The yield of a fluorescent screen (in candles per watt) as a function of the surface density $c$ of the fluorescent substance (in mg cm$^{-2}$).

The fluorescent screens of cathode ray tubes are peculiar in that the image is observed by transmission through the layer and through the glass support, while in a microscope, for example, the observer usually looks at the face of the screen on which the electrons fall. Thin layers of fluorescent material have to be used, therefore, as a great deal of the light will otherwise be lost. Figure 123 shows the way in which the luminous yield at the screen depends upon the thickness of the fluorescent powder, and reveals that there is an optimum thickness, which varies with a whole set of parameters, but in particular with the size of the grains of powder.

For a given thickness of the fluorescent layer, the luminance of the screen is an increasing function of the accelerating potential and of the electron
intensity in the beam. This latter cannot be increased very markedly without the risk of shortening the lifetime of the tube considerably. Further, Fig. 124 shows that the fluorescent layer rapidly becomes saturated, and the efficiency with which the energy is transformed into light falls as the intensity rises. The potential of the screen, therefore, must be increased—this is one of the reasons why high accelerating potentials are used (Fig. 125). The

![Diagram](image)

**Fig. 124.** The reduction of the yield of a fluorescent screen with increasing current density (for a beam accelerated to 6000 volts).

![Diagram](image)

**Fig. 125.** The variation in the brilliance of a fluorescent screen (a television screen $P_4$ which fluoresces white) as a function of the accelerating potential of the beam.

The current densities corresponding to these curves are 1, 2, 3, 5 and 10 $\mu$A cm$^{-2}$.

yield at the screen diminishes though, when the potential is increased. The fluorescent products which are available commercially are mixtures; the composition of the ingredients is obtained empirically, and is denoted by various conventional notations ($P_1$, $P_2$ in America, $CR_1$, $CR_3$ in England). Table 2 lists the essential properties of some of these powders. The reader
will find a host of details in the book by Leverenz (1945) and in the articles by Hopkinson (1946), and Nguyen Tien Chi (1952). A fluorescent material consists of a base of some crystallized mineral material (a semi-conducting oxide or sulphide) which contains a small quantity of (metallic) activating material. The crystallized substance which forms the base has a considerable effect upon the efficiency of the transformation of energy, as Fig. 126 illustrates. The spectrum of colours which is obtained is usually described by a bell-shaped curve (Fig. 127). To obtain white, which is vital for tele-

![Graph](image1)

**Fig. 126.** The influence of the crystal symmetry of the base; I. zinc and cadmium sulphides activated with silver; II. zinc sulphide activated with silver.

![Graph](image2)

**Fig. 127.** The light spectra of fluorescent powders; $P_1$ gives a green spot, $P_4$ a white spot, and $P_5$ a blue spot.
vision tubes, two powders have to be mixed, and the spectrum has two maxima. The development of colour television has resulted in a great deal of research directed in particular towards the production of substances which fluoresce red.

![Diagram](image)

**Fig. 128.** A double fluorescent screen.

When the screen is struck by the electron beam, light is emitted, and a steady state is established as long as the bombardment continues; when the latter ceases, the luminous intensity decreases according to one or other of two laws:

\[ I = I_0 e^{-t/\tau} \]

or

\[ I = I_0 \frac{1}{1 + t/\tau} \]

The value of the time constant, \( \tau \), which measures the "remanence" of the screen can vary between wide limits. In television, this remanence plays an important part, for if it is too short, one frame fades away before the next one appears, and the screen flickers; if on the contrary it is too long, a number of images are mixed together.

Many screens to-day consist of a double layer, a design which was introduced in connexion with radar using the P.P.I.† system where each region of space is explored periodically. In Fig. 128, an example of a double layer is illustrated; the electrons impinge on a layer with a short persistence

† "Plan position indicator": the instrument traces out on the screen of the cathode ray tube what is virtually a map of the region which the beam explores.
### Table 2

<table>
<thead>
<tr>
<th>Base</th>
<th>activating substance</th>
<th>wavelength in Å at the intensity maximum</th>
<th>colour</th>
<th>persistence time/sec</th>
<th>candles per watt at 6000 V</th>
<th>candles per watt at 1000 V</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₀</td>
<td>Zinc silicate</td>
<td></td>
<td>green</td>
<td>5250</td>
<td>0.05</td>
<td>9</td>
</tr>
<tr>
<td>P₁</td>
<td>Zinc sulphide 48% and cadmium sulphate 52%</td>
<td></td>
<td>white</td>
<td>4440</td>
<td>5 x 10⁻⁹</td>
<td>10</td>
</tr>
<tr>
<td>P₂</td>
<td>Calcium tungstate</td>
<td></td>
<td>blue</td>
<td>4300</td>
<td>10⁻⁹</td>
<td>2</td>
</tr>
<tr>
<td>P₃</td>
<td>Zinc sulphide</td>
<td></td>
<td>blue</td>
<td>4580</td>
<td>0.05</td>
<td>2</td>
</tr>
<tr>
<td>P₄</td>
<td>Magnesium silicate</td>
<td></td>
<td>red</td>
<td>6740</td>
<td>0.1</td>
<td>2</td>
</tr>
</tbody>
</table>

which emits a blue light (zinc sulphide activated with silver). The light which is emitted then passes through the second layer and to the observer, this primary brilliance indicates the instantaneous position of the spot. Another part of this flux of primary light—the ultra-violet—is absorbed in the second layer, and produces a highly remanent fluorescence, yellow in colour, in this film of zinc and cadmium sulphide activated with copper. This provides a reminder of the appearance of the map between the times \( t \) and \( t + \tau \), which is useful in radar stations where the geography is fixed (at aerodromes) or slowly moving (at sea) while the positions of the aeroplanes or ships are changing rapidly. This double layer can also serve the purpose of increa-

![Diagram of an ion trap](image)

**Fig. 129.** An ion trap. A magnetic field perpendicular to the plane of the figure is maintained by a permanent magnet (see also Fig. 116c).

creasing the light yield at the screen; the electrons provoke ultra-violet radiation in the first layer, with a high efficiency both as regards the transformation of energy and the number of quanta. These ultra-violet rays then produce—again, quite efficiently—visible light in the second layer by photo-luminescence. In this way, as much as 10 per cent of the electron energy can be transformed into light, We should recall that the efficiency can be increased further by returning towards the front the fraction of the light which is emitted backwards, by the aid of a very thin metal layer which is spread across the bottom of the screen (Epstein and Pensak, 1946).

Finally, it is necessary to take precautions against the "ion spot" which is produced by the negative ions emitted by the cathode. These ions form a halo round the spot after the tube has been in use for a few hours, which diminishes the contrast considerably and is intolerable in television tubes (Schaeffer and Walcher, 1943; Broadway and Pierce, 1939). This halo is removed by building an "ion-trap" near the gun. This consists of two opposed deflexion systems, one electric, the other magnetic. The intensity is adjusted in such a way that the deflexions cancel one another for electrons; they will not, therefore, compensate one another for ions, which are far heavier and will be lost to the wall of the tube. Figure 129 is a diagrammatic sketch of a trap, in which an electrostatic two-cylinder lens is obtained by cutting a cylinder "obliquely". This lens will deflect the beam, and the compensation is provided by a small permanent magnet placed outside the tube.
THE CATHODE RAY TUBE

This trap also eliminates the positive ions which are created along the path of the beam of electrons by collisions with residual gas molecules; otherwise, there would be a risk that the latter might damage the emissive layer, as they are accelerated back towards the cathode.

12.3 DEFLEXION OF THE BEAM

12.3.1 Deflexion Plates

After leaving the second focusing lens, the beam passes through two pairs of plates which produce deflexions in the directions parallel to $Ox$ and $Oy$. In oscillographs which are to be used as measuring devices, the useful deflexion is only of the order of a few degrees, and can be obtained electrically or magnetically. In television tubes, on the contrary, which have to be as short as possible, the deviation is much larger ($\pm 45^\circ$ is usual, and recently, $\pm 55^\circ$ has been reached). We shall only consider the theory for more modest angles; to produce a tube capable of providing these high deflexions requires a laborious experimental step-by-step procedure.

In Fig. 130 an electrical system is shown. Two electrodes, symmetrical about the plane $yOz$, produce a field of which the principal component is directed along $Ox$; the equipotential plot is kept most nearly symmetrical if the plate potentials are disposed symmetrically, that is, if one is at a potential $+ \frac{\Delta \Phi}{2}$ with respect to the potential of the last electrode from which the beam emerged, and the other at $- \frac{\Delta \Phi}{2}$. In this situation, the axis $Oz$ remains at a constant potential, namely the potential of the layer of aquadag which is linked electrically to the final electrode, $\Phi_0$. This arrangement introduces complications into the supply circuits which we should often prefer to avoid, however, and generally speaking this would be done by holding one plate at the potential $\Delta \Phi$, and joining the other to the layer.

Fig. 130. The structure of electrostatic prisms which provide deflexion in the $Ox$ and $Oy$ directions.
of aquadag. This is the unsymmetrical arrangement which is frequently used, but which produces important defects, for the beam which is travelling in the neighbourhood of Oz has first of all to change its axial velocity, as within the plates the potential is approximately \( \Phi_0 + \frac{1}{2} \Delta \Phi \).

The appearance of the lines of force in such a system in these two cases is shown in Fig. 131. It is evident that even though the lines of force are similar inside the prism, their appearance is quite different outside and beyond the prism.

![Diagram](image)

**Fig. 131.** Two possible ways of supplying electrostatic prisms: (a) One plate is held at \( +\frac{1}{2} \Delta \Phi \), the other at \( -\frac{1}{2} \Delta \Phi \); this is the symmetrical arrangement. (b) One plate is tied electrically to the aquadag layer, the other is held at a potential \( \Delta \Phi \); this is the unsymmetrical arrangement.

The potential function can be expanded in powers of \( x \) and \( y \), the coefficients being functions of \( z \). Taking into account the symmetry of the system and the fact that the potential satisfies Laplace’s equation, the potential of the first prism in the symmetrically excited arrangement takes the form

\[
\Phi(x, y, z) = \Phi_0 + \Delta \Phi \left[ \varphi_1(z) x + \varphi_3(z) x^3 - \{3 \varphi_3(z) + \frac{1}{2} \varphi_1''(z)\} xy^2 + \cdots \right],
\]

(12.10)

while in the unsymmetrically excited arrangement, it takes the form

\[
\Phi(x, y, z) = \Phi_0 + \Delta \Phi \left[ \varphi_0(z) + \varphi_1(z) x + \varphi_2(z) x^2 - \{ \varphi_2(z) + \frac{1}{2} \varphi_0''(z)\} y^2 + \cdots \right].
\]

(12.10’)

The functions \( \varphi_1(z) \) and \( \varphi_2(z) \) play a role here equivalent to the role of the axial potential in electron lenses. They are characteristic of the plate system, and with their aid the trajectories can be calculated to first order, and the corrections to this first-order approximation—the aberrations—can be estimated. Magnetic deflectors consist of windings situated outside the tube—to modify and regulate them is therefore a straightforward affair. In tubes used for television, where the optimum conditions with a wide deflexion can be obtained only empirically, magnetic deflexion is preferable to electric deflexion as it is less easy to experiment with the latter. Further, magnetic coils have the advantage of occupying comparatively little space along the axis, which means that they are “superposable”; a beam which is deflected in the \( Ox \) and the \( Oy \) directions over
the same axial region will have less aberration. Figure 132 represents an example of a magnetic deflector. The system is characterized by the components of the magnetic induction, expanded as a power series in \( x \) and \( y \). For the case shown in Fig. 132, for example,

\[
B_x = \mu_0 n I \left[ k_1(z) + k_2(z)(x^2 - y^2) - \frac{k_1'(z)}{2} y^2 + \cdots \right],
\]

\[
B_y = \mu_0 n I \left[ -2 \left\{ k_0(z) - \frac{k_1'(z)}{2} \right\} x y + \cdots \right],
\]

\[
B_z = \mu_0 n I \left[ k_1'(z) x + \frac{k_2(z)}{3} x^3 - \left\{ \frac{k_2(z)}{3} - \frac{k_1'(z)}{2} \right\} x^2 y^2 + \cdots \right].
\]

(12.11)

The winding consists of \( n \) turns, carrying a current \( I \); the functions \( k_1(z) \) and \( k_2(z) \) are determined by the geometry of the system.

**Fig. 132.** Diagram of a magnetic prism. The electron beam is deflected towards the positive \( y \) direction.

### 12.3.2 The Trajectories in Electron Deflectors

The deflexion which is produced by a deflector system depends upon which ray is considered. The deflexion of a ray parallel to the axis will differ, therefore, from the deflexion of an inclined ray; and the deviation of a ray which enters the system at the axis will differ from the deviation of a ray at the edge of the beam. The electron beam will no longer be conical when it leaves the deflector, so that the spot which is obtained on the screen will be somewhat complicated in shape, instead of the point which is desirable. Deflexion systems introduce considerable aberration; we shall not examine the calculation of the trajectories in the general case, but simply give an example of a special case in the next paragraph. For further information, the reader is referred to the detailed publications by Wallraff (1935), Picht and Himpan (1941), Glaser (1949), Hutter (1947, 1948), Deserno...
(1953) and Wendt (1954). Here, we shall mention only a few details of the form of the traces which are found at the fluorescent screen, that is, of the aberration figures.

If we restrict the series expansions to terms of the first order, we can show that the prism produces a deflexion \( \theta_x \) which is the same for all the rays. The incident and the deflected rays, produced, intersect in a plane perpendi-

![Diagram](image)

**Fig. 133.** Deflexion of a trajectory in a cathode ray tube.

...cular to the axis which is the principal plane of the deflector (see Fig. 133). For a symmetricaly fed electrostatic system, for example, it is not difficult to show that

\[
\theta_x = \frac{\Delta \Phi}{\Phi_0} \int_{-\infty}^{\infty} \varphi_1(z) \, dz \tag{12.12}
\]

(cf. § 12.3.3 ii) in which \( \varphi_1(z) \) is the function defined in formula (12.10). If we denote the distance between the principal plane and the screen by \( L_x \), the spot is displaced through a distance

\[
\Delta X = \theta_x L_x, \tag{12.13}
\]

while the plates which produce a deflexion in the perpendicular direction displace the spot through

\[
\Delta Y = \theta_y L_y. \tag{12.13'}
\]

If we now take into account terms of higher orders in \( x \) and \( y \), the ray will be described more accurately—the point \( S' \) at which it intersects the fluorescent screen will have coordinates \( \Delta X + \varepsilon_x, \Delta Y + \varepsilon_y \). The corrections \( \varepsilon \), which are clearly small, will depend upon which trajectory we are considering, while the trajectory is determined by its slope \((x'_0, y'_0)\) where it enters the deflexion system. The slope \((x'_0, y'_0)\) is adequate, for the coordinates of the ray at the point at which it enters the prism are necessarily \( x'_0 L \).
and $y'_0 L$, as a result of the requirement that the spot shall be formed at the screen ($L$ is the distance to the screen). The magnitudes of the correction terms will also depend on the deflector, that is, upon $\theta_x$ and $\theta_y$. We can write, therefore,

$$\varepsilon_x = f(\theta_x, \theta_y, x'_0, y'_0),$$
$$\varepsilon_y = g(\theta_x, \theta_y, x'_0, y'_0).$$

To proceed further, we must make a calculation similar to the one in § 6.2, where the aberrations of electron lenses were analysed. $\varepsilon_x$ and $\varepsilon_y$ are expanded as a power series in the four variables, but fortunately there are a number of simplifying conditions which limit the number of terms of this expansion. First, if the deflectors are switched off, $\theta_x = \theta_y = 0$, the spot lies at the centre of the screen, and $\varepsilon_x = \varepsilon_y = 0$. The series, therefore, contains no terms independent of $\theta$. Further, terms of the first order in $\theta_x$ and $\theta_y$ cannot be present, as it is they which produce the original deviations $\Delta X$ and $\Delta Y$. The expansion must therefore begin with second order terms,

$$\varepsilon_x = A \theta_x^2 + B \theta_y^2 + C \theta_x \theta_y + D x'_0 \theta_x + E x'_0 \theta_y + F y'_0 \theta_x + G y'_0 \theta_y,$$
$$\varepsilon_y = A' \theta_x^2 + B' \theta_y^2 + C' \theta_x \theta_y + D' y'_0 \theta_x + E' y'_0 \theta_y + F' x'_0 \theta_x + G' x'_0 \theta_y, \quad (12.14)$$

and contains fourteen coefficients. The formulæ from which these coefficients can be calculated are to be found in the articles quoted earlier. Provided the deflectors are perfectly symmetrical, it is possible to cancel all the coefficients; since this perfection of the symmetry includes both the electrical and the geometrical properties, the potential must be fed symmetrically to the system.

The terms which depend upon $x'_0$ and $y'_0$ introduce astigmatism into the beam, but not field curvature. The circle of least confusion is at the image plane. There is no defect of the same nature as coma (which would be produced by terms containing $x'_0^2$) nor is there spherical aberration (which is a specifically third order aberration). On the other hand, there are terms in $\theta$ alone which displace each point of the spot and hence represent distortion. If $\theta_x$ were varied smoothly keeping $\theta_y$ constant, the spot would trace out a straight line so long as this distortion was negligible. By varying $\theta_y$ with $\theta_x$ constant, and conversely, it would be possible to trace out a pattern of squares on the screen. In reality, the effect of distortion is to deform this pattern of squares into a pattern of the form shown in Fig. 134. This defect would be particularly obvious in television tubes, for its effect upon the image is one to which the eye is especially sensitive, used as it is to judging proportions accurately (when we think of proportions in terms of the canons of aesthetics). In addition, it is not possible to base precise electrical measurements on the displacement of the spot when this aberration is present. Distortion, therefore, is a defect which we endeavour at all costs to eliminate.
These second order aberrations can, fortunately, be corrected by using two perfectly symmetrical electrical supply units with the same mid-point; this is the solution which is adopted in television sets. In this case, \( \varepsilon_x \) and \( \varepsilon_y \) contain third order terms only, and we find

\[
\begin{align*}
\varepsilon_x &= a \theta_x^2 + b \theta_x \theta_y + c x_0' \theta_x^2 + d x_0' \theta_y^2 + e \theta_x \theta_y y_0' \\
&\quad + f x_0' \theta_x + g y_0' \theta_x + h x_0' y_0' \theta_y, \\
\varepsilon_y &= a' \theta_y^2 + b' \theta_x \theta_y + c' y_0' \theta_y^2 + d' y_0' \theta_x^2 + e' \theta_x \theta_y x_0' \\
&\quad + f' y_0' \theta_y + g' x_0' \theta_y + h' x_0' y_0' \theta_x.
\end{align*}
\]

(12.15)

Fig. 134. Deformation of the scanning network by second order (trapezoidal) distortion.

Fig. 135. Deformation of the scanning network by third order distortion in a symmetrical system.

The terms which contain \( \theta \) alone again correspond to a distortion, but are less serious than their second order counterparts, as the effect is not trapezoidal; Fig. 135 gives a general idea of the deformation which third order distortion produces upon the basic square pattern.

The terms which contain \( x_0' \) and \( y_0' \) to first order give rise to astigmatism, which is accompanied by field curvature. It can be reduced either by using a curved screen, or by varying the convergence of the lens at the same time as the spot is deflected (Hoadley, 1951).
The terms which contain quadratic combinations of \( x_0' \) and \( y_0' \) have an effect upon the spot which is similar to that of coma in lenses. Figure 136 shows the appearance of the spot which is to be seen on the screen. There is, however, no aberration analogous to spherical aberration, as this would require a term in \( x_0'^3 \).

![Diagram](image)

**Fig. 136.** Some examples of the spots obtained in a cathode ray tube in which the spot (a) given by the gun has been deformed by astigmatism and coma.

Formulae from which the coefficients of third order aberration can be calculated are to be found in the articles which have been cited earlier. There are thirteen of these coefficients although there were sixteen second order terms, as certain natural equalities appear: \( h = g' \), \( h' = g \) and \( e = e' \). This theory has been verified in a remarkable way by Wendt (1954), who used an extremely fine pencil, which can be thought of effectively as a single ray, rather than a complete beam. A rotating magnetic field turned this beam about the \( Oz \) axis in such a way as to describe a conical surface, with its point at the screen. It is clear that if we introduce a deflecting system, it will act on this pencil just as it would act on the various rays which lay in each of the positions occupied by the pencil. The point of the pencil, therefore, traces out the outline of the spot on the screen which can be photographed easily.

### 12.3.3 An Example of an Electrical Prism

We shall give some details of the simplest of the electrical deflectors which consists of two plane parallel plates, a distance \( d \) apart. They are excited symmetrically (\( \Phi_0 + \frac{1}{2} \Delta \Phi \) on one plate, \( \Phi_0 - \frac{1}{2} \Delta \Phi \) on the other).

**i) The end correction**

We assume that a uniform electric field exists within the prism for a certain distance, beyond which it is zero. Recknagel (1938) has shown that this is a valid model of the real situation provided the length of the region of uniform field is regarded as being slightly longer than the length \( l_0 \) of the plates. (This is similar to the "equivalent length" which we have defined for quadrupole lenses.)

The behaviour of the lines of force in this model is shown in Fig. 137, together with the meaning of the correction \( \Delta l \). Recknagel showed that

\[
\Delta l = 0.318d \left( 1 - \log 0.636 \frac{d}{l_0} \right),
\]

(12.16)
which gives a correction of the order of 0.7d for \( d = 0.2\ l_0 \). We shall write

\[ l_0 + \Delta l = l, \]

and in the useful region, the potential is given by

\[ \Phi = \Phi_0 + \frac{\Delta \Phi}{d} x. \]

(ii) The equations of motion

The lateral motion is described by the simple equation

\[ m \frac{d^2 x}{dt^2} = e E = e \frac{\Delta \Phi}{d}, \]

which gives

\[ \frac{dx}{dt} = \left( \frac{dx}{dt} \right)_0 + \frac{e}{m} \frac{\Delta \Phi}{d} t. \]  

(12.17)

Combining this with the conservation of energy relation, we find

\[ \frac{1}{2} m v^2 = \frac{1}{2} m \left[ \left( \frac{dx}{dt} \right)^2 + \left( \frac{dy}{dt} \right)^2 \right] = e \left( \Phi_0 + \frac{\Delta \Phi}{d} x \right). \]

As the electron trajectory is always inclined only slightly with respect to the plates, we can assume that \( \frac{dx}{dz} \) is small compared with unity, and write

\[ \left( \frac{dz}{dt} \right)^2 \approx \frac{2 e \Phi_0}{m} \left[ 1 + \frac{\Delta \Phi}{\Phi_0} \frac{x}{d} \right] \left[ 1 - \left( \frac{dx}{dz} \right)^2 \right], \]  

(12.18)

\[ \left( \frac{dz}{dt} \right) \approx \sqrt{\frac{2 e \Phi_0}{m}} \left\{ 1 + \frac{1}{2} \frac{\Delta \Phi}{\Phi_0} \frac{x}{d} - \frac{1}{2} \left( \frac{dx}{dz} \right)^2 - \frac{1}{8} \left( \frac{\Delta \Phi}{\Phi_0} \right)^2 \frac{x^2}{d^2} + \cdots \right\}. \]

This equation displays the importance of the correction terms which even in this simple model are appearing little by little. If we neglect all the correction terms in the formula, we shall produce a "zero order" optics,
which will give the deflexion and the principal plane. It can be shown quite straightforwardly that the trajectory is a parabola, which implies that the principal plane lies at the centre of the useful part of the field. The magnitude of the deflexion is

$$\theta = \left( \frac{dx}{dz} \right) - \left( \frac{dx}{dz} \right) = \frac{1}{2} \frac{l}{d} \frac{\Delta \Phi}{\Phi_0}. \quad (12.19)$$

(iii) Sensitivity

By the sensitivity of a cathode ray tube is meant the potential difference which displaces the spot through its own width. Starting from formula (12.1) which gives us the width of the spot, we eliminate $\theta$, and find

$$\sigma_E = \frac{4 K f}{L} \frac{d}{l} \sqrt{\frac{T}{11,600}} \sqrt{\Phi_0}, \quad (12.20)$$

in which $L$ is the distance between the screen and the principal plane. The sensitivity of a cathode ray tube is in the region of one volt per diameter, and there would be an advantage in reducing $\Phi_0$, since $\sigma_E$ has to be as small as possible. This would reduce the brightness, however, and the influence of space charge would be accentuated.

(iv) The importance of the maximum deflexion

From the point of view of the electrical supply to the deflectors, it would be advantageous to use long plates, close together. This cannot be pursued to too extreme a limit, however, as the beam has to be able to reach the whole of the screen, and the angle $\theta$ has to be able to reach some maximum, $\theta_m$, which is fixed by the dimensions of the tube. Considering Fig. 138a, we see that $\theta_m \sim d/l$, so that if we increase $l/d$—and with it, the sensitivity—$\theta_m$ is diminished. For this reason, a great deal of research has gone into the study of more complicated types of system, with inclined or even curved plates. A full discussion of this point is to be found in the book by Klemperer (1953). With plane plates inclined at an angle $\alpha$ (see Fig. 138b), the lines of force are approximately circular, and we find

$$\theta \simeq \frac{1}{4 \tan \alpha} \log \left[ 1 + \frac{2l}{d} \tan \alpha \right] \frac{\Delta \Phi}{\Phi_0}, \quad (12.21)$$

and

$$\theta_m \simeq \frac{d}{l} + 2 \tan \alpha. \quad (12.21')$$

The full analysis shows that for a given value of $\theta_m$, the sensitivity is an increasing function of $\alpha$. In practice, the plates consist of a horizontal part and an inclined part, as illustrated in Fig. 138c. It would be still more
advantageous to use curved plates (Maloff and Epstein, 1938), parallel to the beam at the point at which its deflexion reaches a maximum, \( \theta_m \) (see Fig. 138d). The equation which the cross-section of the plates must satisfy is given by

\[
z - z_0 = 2y_0 \left( \frac{V_0}{V_d} \right)^{1/2} \int_0^{\log \left( \frac{z}{y_0} \right)} e^{au} \, du \tag{12.21''}
\]

in which \( 2y_0 \) represents the initial distance between the plates, \( \Phi_0 \) is the accelerating potential and \( \Phi_d \) the peak deflexion potential measured with respect to \( \Phi_0 \).

\[\text{Fig. 138. Different kinds of electrostatic prism. (a) A prism with parallel electrodes. (b) A prism with oblique electrodes. (c) The prism which is employed in the majority of tubes. (d) A prism with curved electrodes.}
\]

For the same deflexion \( \theta_m \), the sensitivity increases in the order (a) to (d). With the systems (c) and (d), very high values of \( \theta_m \) can be attained.

For \( z - z_0 = L \), we have \( y = y_s \).

Since the integral which appears in equation (12.21'') is the same as the integral which expresses the broadening of the beam which space charge produces (see § 12.2.3), we can calculate the form of the plates numerically with the aid of Fig. 119a.

(v) The aberrations

The results become more complicated when we take into account the aberrations. For example, suppose that the beam is of width \( x_0 \) when it enters the prism, the deviation \( \theta \) of the electrons varies from one ray to
another because the potential is not the same for all the rays. Calculation shows that the emergent beam is more convergent than the incident beam, but only in the plane $xOz$. The converging power of the equivalent lens would be given by

$$\frac{1}{f} = \frac{1}{2} \frac{l}{d^2} \left(\frac{\Delta \Phi}{\Phi_0}\right)^2.$$ \hspace{1cm} (12.22)

This is superimposed onto the convergence of the principal lens in the tube, but as it is produced by a cylindrical lens, the result is an astigmatism which can easily be seen in certain oscillographs when the spot is near the edge of the screen. This is the second order astigmatism which we mentioned in the general study.

![Diagram](image)

**Fig. 139.** The production of a field which is as uniform as possible over the whole depth of the prism. Prism $b$ has a larger useful zone than prism $a$.

If the second deflector is also perfectly symmetrical, we can see how it might be capable of producing a certain convergence in the other plane, which would partially compensate the astigmatism of the beam; the spot would still be broadened, however, as the crossover would be formed in front of the screen. As a result, if the trace of the beam is circular and as small as possible at the centre of the screen, elsewhere it will be elliptical in the directions $Ox$ and $Oy$ and practically circular along the directions at $45^\circ$ to the axes.

We shall not pursue the study of the aberrations of this model any further; were we to do so, we should find all the general terms of § 12.3.2. In conclusion, we should nevertheless mention two other possible causes of aberration.

If the beam is wide and the plates too narrow, the edges of the beam travel in a region in which the lateral overlap field disturbs the distribution of field in the deflexion system.

If the horizontal and vertical deflexion plates are too close together in the $Oz$ direction, the space between the two pairs of plates behaves like an astigmatic lens of very inferior quality (Cooper, 1961). This defect, which is apparent when the deflexion is strong, can be removed by separating the two pairs of plates by an earthed double shield.
12.3.4 Magnetic Deflectors

The simplest version, although it is rarely used, is the electromagnet, which produces a uniform magnetic induction $B$ over a distance $l$, and has no effect outside this region. In § 5.1.1 the reader will find a description of the principle behind the calculation of $\theta$

$$\theta = \sqrt{\frac{e}{2m}} \frac{B}{\sqrt{\Phi_0}}.$$  \hspace{1cm} (12.23)

The sensitivity here is characterized by the value of $B$ which displaces the spot through its own width on the fluorescent screen, and we find

$$\sigma_M = \frac{Kf}{l} \sqrt{\frac{2m}{e}} \sqrt{\frac{T}{11,600}}.$$  \hspace{1cm} (12.24)

This expression does not contain $\Phi_0$. We should remember, however, that the usual type of magnetic system is of the form illustrated in Fig. 132. The field inside is far from uniform, therefore, and calculation of the trajectories is complicated. Very fortunately, it is easy to make measurements, and excellent magnetic deflexion devices have been achieved by a series of successive adjustments (Woroncow, 1946; Wendt, 1954). As an example, we shall quote the characteristics of the system which is used in a radar tube. The number of turns is 2000, wound in the form shown in Fig. 132, the diameter is 7.5 cm and the length 5.6 cm; a current of 1.34 mA is necessary to deviate the beam through 1° when the electron beam is accelerated through 5 kV (Soller, Starr and Valley, 1948).

12.3.5 The Scanning Mechanism and the Writing Speed

Save in exceptional cases, the potentials which are applied to electrostatic plates do not remain constant, any more than do the currents which govern magnetic systems. The spot therefore will move about the screen, tracing out a curve which can be photographed. The speed with which the spot moves along this curve is known as the “writing speed”; and as soon as the sensitivity of the tube and the expression which describes the applied potential as a function of time are known, its calculation presents no problem. This speed is characterized by an “index” $\tau$, the time which the spot takes to move through its own width. In the case of an electrical system, fed with a potential $\Delta \Phi \sin \omega t$, we find

$$\tau = \frac{\sigma_E}{\omega \cdot \Delta \Phi}.$$  \hspace{1cm} (12.25)

The majority of tubes, provided they are supplied with a suitable frequency, are capable of providing a satisfactory record in an interval of the order of $10^{-8}$ sec. If the frequency of the supply is increased any further, it is found that the sensitivity begins to diminish, which is a result of the
transit time as we shall show in the next paragraph; it is this phenomenon which provides a natural limit to the writing speed.

It is to be noticed, too, that if the spot takes $10^{-8}$ sec to move through its own width, 300 $\mu$ say, its linear velocity is some 30 km s$^{-1}$. Few phenomena demand such extreme writing speeds, but they are useful nevertheless to display the form of the steep wavefronts on transmission lines, or of those provided by sources of decimetre waves, for direct observation.

![Diagram](image)

**Fig. 140.** The scanning system in a television tube. The return track, which is shown as a dotted line, is in fact rendered invisible by applying an appropriate polarization to the Wehnelt during the return.

At the present time, the electron intensity at the spot, the efficiency of the fluorescent screen and the sensitivity of the plate can achieve so high a quality that the writing speeds which can in consequence be attained are sufficiently rapid for the decrease in sensitivity resulting from the effect of transit time to become perceptible. We shall return to this point in § 12.3.6.

In a cathode ray tube which is to be used in a television receiver, a potential (or a current) which has a saw-tooth variation in time is applied to the system which provides the horizontal deflexion; the spot moves from one side of the screen to the other in a time $T$, then flies back to its starting point in a negligible time. Another saw-tooth potential with a period 819 times as long is applied to the other prism, so that, as Fig. 140 shows, the spot explores the whole screen by tracing out 819 practically horizontal lines in a time 819$T$. This time is chosen in such a way that 25 images are formed every second† (the frequencies of the two scanning oscillators,

† We neglect the small difference between two successive images which results when the return is interlinear.
therefore, are 25 c/s and 20,475 c/s). At each moment, only the potential of the grid need be varied to obtain a suitable brightness at the corresponding point on the screen, where we can watch a black and white image being traced out.†

In a cathode ray tube which is to be used as the panoramic receiver of a radar set, the potentials (or currents) which are applied to the two deflectors are of the form

$$\Delta \Phi_x = \varphi(t) \sin \frac{2\pi t}{T},$$

$$\Delta \Phi_y = \varphi(t) \cos \frac{2\pi t}{T},$$

in which $\varphi(t)$ is a saw-tooth function of period $T'$, small in comparison with $T$. The spot describes a star-shaped curve, as Fig. 141 shows. The signal which is received by the radar set is applied to the grid and hence governs the brightness of the spot at each point on the screen. For further details, the reader is referred to the book by Soller, Starr and Valley (1948).

![Fig. 141. The scanning in a panoramic receiver.](image)

If, finally, the cathode ray tube is to be used in an oscillograph, the signals which are being studied are applied to the deflectors after suitable amplification. The spot traces out a curve which describes the way in which one of the phenomena depends on the other. If the signals are periodic, and if the periods are simply related, the curve is closed and can be examined with ease thanks to the persistence of the screen. Otherwise, we have to resort to photography, and we should have to ensure that the trace was sufficiently brilliant for photography to be possible. From the article by

† The numbers in this paragraph relate to French television standards (Ed.).
Hopkinson mentioned earlier, we have extracted the following examples—
for a blue fluorescent screen (zinc sulphide), the writing speed of a spot
1 mm in diameter and 20 $\mu$A in intensity cannot exceed 10 km sec$^{-1}$ for
electrons at 3000 eV; this limiting velocity is practically independent of
the emulsion. With a green screen, the velocity cannot exceed 5 km sec$^{-1}$
with an orthochromatic emulsion, or even 2 km sec$^{-1}$ with a panchromatic
emulsion.

If the phenomenon is periodic, the problem is far less difficult. An ex-
posure of one second is sufficient with a blue screen, although it has to be
extended to 5 seconds with green screens. Further details are to be found
in the book by Puckle (1951). In modern cathode ray tubes with post-
acceleration, such as those with which high quality commercial oscilloscopes
are equipped, the exposure time with blue screens is very short, 1/50 or
1/25 of a second.

12.3.6 Transit Time as a Factor Limiting High Writing Speed

If we increase the frequency of the tension or current which is providing
the deflexion, the spot will clearly move more rapidly across the screen,
but a limit soon appears. To pass between the deflecting plates of length $l$,
an electron moving with velocity $v$ will take a time $l/v$ which has to be
negligible in comparison with the period $T$ of the applied potential. Only
in this case is it legitimate to regard the latter as constant during the time
of transit, and to apply the electrostatic arguments of the preceding para-
graph. Moreover, the "transit angle", $\varphi$, defined by

$$
\varphi = 2\pi \frac{l}{vT} = 2\pi \frac{l}{T} \sqrt{\frac{m}{2e\Phi_0}} = 1.05 \times 10^{-5} \frac{\varphi}{\Phi_0},
$$

becomes important only when the frequency is very high. For a potential
of 3000 V, and plates 3 cm long, the angle is $6 \times 10^{-3}$ for an applied fre-
quency of 100 kc/s. For 1 Mc/s, it reaches 0.060 and for 100 Mc/s, 0.60.
At this last frequency, it is obvious that the actual deflexion in the oscil-
lograph will bear little relation to the deflection which is calculated by elec-
tronic methods, and it is in fact considerably less.

We shall examine the simple case of an electrostatic system in a uniform
field, but the results can be extended to any deflexion whatsoever without
major modification. Once again, we consider the equation of motion in the
lateral direction, this time supplying the system with an alternating poten-
tial of period $T$. We find

$$
m \frac{d^2x}{dt^2} = e E_0 \sin \frac{2\pi}{T} (t + t_0).
$$

The time origin is selected to be the moment at which the electron enters
the condenser; at this instant, the field has the value $E_0 \sin \frac{2\pi}{T} t_0$. The longi-
From these equations, we deduce the trajectory equation and the deflexion angle as the beam leaves the plates; for the latter, we find

$$\theta = \frac{l E_0}{2 \Phi_0} \sin \left(2\pi \frac{t_0}{T} + \frac{\varphi}{2}\right) \frac{q/2}{\sin(q/2)},$$

(12.27)

whereas the value which we should have found if we had used a static field of magnitude equal to the mean value in traversing the plates, $E_0 \sin \left(2\pi \frac{t_0}{T} + \frac{\varphi}{2}\right)$, would have been

$$\theta_0 = \frac{l E_0}{2 \Phi_0} \sin \left(2\pi \frac{t_0}{T} + \frac{\varphi}{2}\right).$$

The actual sensitivity, taking the transit angle into account, is therefore

$$\sigma = \sigma_0 \frac{q/2}{\sin(q/2)},$$

(12.28)

in which $\sigma_0$ is the static sensitivity. $\sigma$, we recall, has to be kept as small as possible, since it represents the potential which is necessary to displace the spot through its own width. But $\sigma$ increases with $q$; so long as the transit time is not too long, we have effectively

$$\sigma = \sigma_0 (1 + q^2/24),$$

(12.28')

a decrease of 1 per cent in the sensitivity for $q = 0.48$. If $q$ becomes appreciable, the result is catastrophic. If, in particular, $q$ reaches $2\pi$, the instrument is useless, as its sensitivity is zero; an infinite potential would be necessary to move the spot through its own width. This extreme case can be observed with ordinary instruments by feeding the plates from a transmission line and a source with a frequency corresponding to decimetre waves.

Only one simple method of overcoming this problem exists, which consists of reducing the transit-time by increasing the potential at which the tube operates. As we have already seen, however, this has repercussions on the static sensitivity, which is reduced. By operating a cathode ray tube at 70,000 V, and moving the screen sufficiently far away not to lose too much sensitivity, potentials at frequencies up to 30 Mc/s can easily be examined. To go beyond this figure is, however, extremely difficult, and it is better to search for solutions of a new kind, by modifying the structure of the tube itself. Ingenious solutions have been suggested by Lee (1946), Owaki et al. (1950), Smith et al. (1952) and Bloom and von Forster (1954), which enable tubes to be built with a sensitivity which is still appreciable at 30,000 Mc/s—of the order of 0.1 mm V$^{-1}$ for a beam accelerated to a few keV. The writing
speed reaches some 15,000 km/s. Most of these tubes employ the properties of travelling waves to obtain deviation, so that the principle behind their mode of operation resembles that of the linear accelerator. The H.F. signal which is to be studied is led into the instrument along a Lecher line which passes into the interior of the envelope. The electric field associated with this

\[ \text{Fig. 142. The principle of "resonance deflexion" in cathode ray oscillographs for very high frequencies.} \]

wave provides the deviation. As Fig. 142 shows, however, the wave which has to follow the transmission line travels along a far longer path than the electrons which move in a straight line. At A, the beam will be very slightly deflected, after which it moves in a straight line until it encounters the wave again at B. The transit time of the wave between A and B is chosen in such a way that the electrons see an electric field at B which is in phase with the field they had seen at A. The beam is deflected again, therefore, in phase with the deflexion produced at A; by using a long enough line, we can obtain the required overall deflexion. Further details are to be found in a recent exposition of this topic by Hardy, Jackson and Feinberg (1956). At the present time, instruments are available with a writing speed of the order of 20,000 km/s.

12.4 THE POST-ACCELERATION

During this examination, we have come across numerous reasons for adopting as high a value of the anode potential as possible. The diameter of the spot will be less since the thermal velocity dispersion will have less effect (§ 12.2.2), the effect of space charge will be imperceptible (§ 12.2.3), and the brightness at the fluorescent screen will be intensified for a given value of the current (§ 12.2.5). Unfortunately, the accelerating potential also has an effect upon the sensitivity (§ 12.3.4) and as we have seen, an increase in \( \Phi _0 \) reduces the sensitivity of an electrostatic tube, but has no effect upon the sensitivity of a tube with magnetic deflexion. It is, therefore, of the greatest interest to operate a magnetic tube at a very high potential but, for an electrostatic tube—which would become less sensitive—this is not so.
Another inconvenience is that in a commercial instrument, it is dangerous to use high voltages (perhaps 50 kV) when the current is not negligible (reaching many milliamperes). To hold the cathode at such a potential is complicated—it is to be remembered that in the normal type of tube, the screen, the layer of aquadag and the anode are all earthed and it is the cathode alone which is connected to the high tension—since the cathode has also to be heated, and the beam modulation has to be introduced; both of these require either a highly insulated transformer, or H.T. condensers, and these are expensive pieces of apparatus.

To overcome these problems, Schwartz (1938) and later Rogowski and Thileen (1939) suggested that the original arrangement should be used, with an accelerating potential of a few thousand volts; only the screen and the region immediately adjacent to it should be held at a very high voltage with respect to earth, as shown in Fig. 116. This is known as "post-acceleration", and entails making a cut in the layer of aquadag near the screen, and supplying the latter with a potential of 25 or 30 kV. After deflexion, therefore, the electron beam is subjected to the influence of another electron lens, effectively a two-cylinder lens, across which the potential changes from the accelerating potential $\Phi_a$ to the post-acceleration potential $\Phi_d$, both measured relative to the cathode potential. The diameter of the spot at the screen, as a scrutiny of the analysis given in §4.3.5 shows, is now related to the potential $\Phi_a$; the formula for the spot diameter is still given by equation (12.1), but with the screen potential inserted, thus

$$d_s = \frac{A}{\sqrt{\Phi_a}}$$

(in which $A$ is a constant independent of the potentials).

The brightness of the screen is excellent, since it is governed by the potential of the screen itself. The space charge aberration too is diminished, as the electrons move more rapidly over the last part of their trajectory where the space charge is most intense. Finally, the technical problems which the construction of this system creates are lessened, since the steady voltage which has to be applied to the screen only requires a "post-acceleration terminal" (Fig. 125). What we have now to do is to confirm that this modification to the optical system has not reduced the deflexion sensitivity.

We shall demonstrate that the deflexion which the prism has to produce, in order to displace the spot through its own width when it is acting on a beam in a region in which the potential is $\Phi_d$, remains unaffected by the behaviour of the potential beyond the deflector. Whether the potential of the screen $\Phi_a$ is equal to $\Phi_d$, as in ordinary tubes, or whether it is considerably larger than $\Phi_s$, as in post-acceleration tubes, is irrelevant. In the latter case, the way in which the potential changes from $\Phi_d$ to $\Phi_a$ is also of no importance. The result is independent of the way in which the post-acceleration potential is applied, as it is of the type of focusing and deflexion
which can be either magnetic or electrostatic. It is not easy to demonstrate this property by extending the argument which was used to establish the value of the sensitivity if the structure of the post-acceleration field is unknown. An indirect line of reasoning due to Pierce (1941), on the other hand, is of general validity. We consider two diametrically opposed points on the spot, $A$ and $B$, shown on a vastly exaggerated scale in the transverse direction in Fig. 143. The rays $OA$ and $OB$ which have passed through the centre $O$ of the deflecting prism define the angular aperture $2\alpha$ of the pencil of rays passing through $O$ which would be produced by a diaphragm with a small hole placed at $O$. This is an obvious result if there is no post-acceleration, as in this case the rays $OA$ and $OB$ are rectilinear, but it is still true even if there is post-acceleration, despite the fact that the rays $OA$ and $OB$ are curved, for in this case $A$ and $B$ are the images of the points $A'$ and $B'$, which lie on the “virtual spot” which is formed by the part of the oscillograph which precedes the post-acceleration system so that $OA'$ and $OB'$ are straight lines. But if the deflector bends the rays through an angle $2\alpha$, it is clear that the rays which would have passed through $O$ and lain between $A$ and $B$ will now lie outside $AB$; in particular, the ray which would have passed through $A$ now intersects the screen at $B$. We can repeat the argument for every other position $O'$ of the point $O$ within the deflecting
prism, so that $2\alpha$, which is the natural aperture of pencils at the centre of the deflector, is equally the value of the deflexion to which the whole beam has to be subjected to displace it through its own diameter. The value of $\alpha$, however, depends only upon $\Phi_d$, the potential in the region of the deflector, and the cathode temperature, as we have already shown ($\S$ 4.3.5); if $C$ is a constant determined by the gun, therefore, we have

$$\alpha = \frac{C}{\sqrt{\Phi_d}}.$$  

If the deflector is electrostatic, the voltage $\sigma_E$ which is required to produce a deflexion $\alpha$ is [cf. $\S$ 12.3.3, formula (12.19)]:

$$\sigma_E = \alpha \frac{\Phi_d}{C_E} = \frac{C}{C_E} \sqrt{\Phi_d}$$  \hspace{1cm} (12.29)

(\text{where} \ C_E \text{ is a geometric constant, and depends only upon the deflector}) so that $\sigma_E$ depends not upon the screen potential $\Phi_d$ but upon the mean potential at the deflector, $\Phi_d$. We conclude, therefore, that the sensitivity of electrostatic tubes can be improved by introducing post-acceleration, as the mean potential at the deflector, $\Phi_d$, can be kept small for a given value of the brightness of the spot on the screen.

For a magnetic system, on the other hand, the current $\sigma_M$ which is required to produce a deflexion $\alpha$ is given by the formula

$$\sigma_M = \alpha \frac{\sqrt{\Phi_d}}{C_M} = \frac{C}{C_M}$$  \hspace{1cm} (12.30)

(cf. $\S$ 12.3.4), so that the sensitivity finally depends only on the cathode temperature (through the constant $C$) and the geometry of the system (through the constant $C_M$). It is completely unaffected by the presence or absence of post-acceleration; the technique is still useful, however, as a means of avoiding the risk of breakdown between the electrodes of the gun.

Finally, we should point out that we have not examined the defects which the post-acceleration lens produces in the marginal region; such defects do appear, and in order to reduce them, we try to use as uniform a field as possible, which means that the lens is extremely weak. For this reason, Lambert and Field (1946) apply their post-acceleration potential progressively along ten rings of aquadag in such a way that the equipotential surfaces are as nearly as possible regularly spaced concentric spheres. Sponsler (1955) has made a further improvement by replacing the rings by a helix of aquadag round the tube. The perturbation of the Gaussian rays which this produces is extremely small.
CHAPTER 13

THE EMISSION MICROSCOPE

13.1 INTRODUCTION

The image which is formed by an emission microscope is created by electrons emitted by the surface of the object itself. No illuminating system is required, therefore, and microscopes of this kind are simpler than the usual types. The electron optics possesses original features, and the sphere of application is specialized, for with an emission microscope, massive specimens can be examined directly, once they have been rendered emissive. This type of microscope is valuable, also, in that the structure of massive solids can be studied—instead of examining simply a rigid replica or an ultra-thin layer of the substance (see Chapter 20), the surface itself and its evolution can be examined.

Historically speaking, the first electron microscope was in fact an emission microscope; with the discovery of electron optics, the designers of cathode ray tubes destined for oscillographs were able to form an image of the emitting cathode on the fluorescent screen. The first photographs of this phenomenon were published in 1931–2. Subsequently, with the research that was directed towards improving the image, the oscillograph (which was ill adapted to play this new role) became clearly separated from the emission microscope, in which it became possible to obtain a resolving power at first equal to, and more recently much better than that which can be obtained with an optical microscope.

13.2 OPTICAL PROPERTIES

13.2.1 The Initial Velocities

In order to have a standard of comparison, we shall first mention the properties of a transmission microscope (see Chapter 15). Here, the electrons arrive at the objective with a high velocity and at a very small inclination, which can be adjusted at will with the condenser. The dispersion of the velocities about their mean value is very small, and the consequent chromatic aberration is negligible in the great majority of cases.

In an emission microscope, on the contrary, the electrons emitted at some point of the object have a very small velocity, and the most probable velocity may vary between a few tenths of a volt and a few volts according to
the particular case in question. The angles at which the electrons are emitted can lie anywhere between 0° and 90°, and no simple means of reducing this angular aperture exists. These features are unfavourable as far as the clarity of the image is concerned, and the resolving power of emission microscopes is undeniably inferior to that to be obtained with a transmission instrument.

13.2.2 The Resolving Power

By calculating the trajectories of the electrons which emerge from an object point, the minimum diameter of the beam in the neighbourhood of the Gaussian plane which corresponds to the object (§ 4.2.4) can be calculated; the resulting theoretical resolving power is (see § 8.3.2):

$$\delta_{th} = k \frac{\Phi_0}{E_0}$$

in which $\Phi_0$ is the most probable energy of the emitted electrons, measured in electron-volts, and $E_0$ is the electrostatic field at the point on the cathode which is being considered; $k$ is a numerical coefficient, close to unity. In the case of thermally emitted electrons with $\Phi_0 = 0.1$ V, $\delta_{th}$ is of the order of 400 Å if a modern objective for which $E_0 = 30$ kV cm$^{-1}$ is used. If we calculate the distribution of electron intensity across the beam in the vicinity of the image plane, either by taking the diffraction into account (Recknagel, 1943) or by using geometrical optics and investigating the shape of the caustic of the pencils of electrons (Septier, 1955) which is valid for small values of $\Phi_0$, we obtain a more favourable value of the theoretical limit. $k$ is found to be of the order of 0.1 or 0.2, so that in the example discussed earlier, $\delta_{th}$ is some 40 or 80 Å. In reality, this theoretical limit is unattainable without an objective stop because the spherical aberration of the objective itself intervenes.

We might consider improving this resolving power by stopping down the elementary pencils at the crossover of the beam with a hole a few microns in diameter (Möllenstedt and Düker, 1953; Fert and Simon, 1956, 1957; Simon, 1959). We shall summarize briefly the main results of the study of the role of this aperture which Fert and Simon have made. The objective is regarded as the combination of a purely accelerating section, length $\Delta$, followed by a convergent lens $L$ (see Fig. 144a). The results will still be approximately valid for normal immersion objectives.

We let $\alpha_0$ be the angle at which a ray leaves the axial object point $P$; at the entry to the lens $L$, this angle has become $\alpha_1$. If we place a diaphragm of radius $r$ at the exit of the lens, and if we suppose that the real image of $P$ is formed at infinity, the angle $\alpha_1$ cannot exceed an upper limit which is given by $\alpha_{1m} = r/a$ where $a$ is the distance between $L$ and the virtual image $P_1$ of $P$ which is formed by the accelerating system.
If we make allowances for the defocusing action of the hole in the accelerating anode, we find \( P_1 P = \frac{\Delta}{3} \) and the magnification of the virtual image at \( P_1 \) is \( G_1 = 2/3 \). The slope \( \alpha_1 \) is given by the relation

\[
\sin \alpha_0 \sqrt{\Phi_0} = \frac{2}{3} \alpha_1 \sqrt{\Phi},
\]

in which \( \Phi_0 \) represents the energy with which the electrons leave, and \( \Phi \) the accelerating potential; \( \Phi \gg \Phi_0 \).

---

**Fig. 144.** The resolution of an objective with a stop. (a) The path of the outermost ray which leaves \( P \) through the simple objective formed by an accelerating region \( KA \) followed by a thin lens \( L \) and a diaphragm of radius \( r \). The limiting angle is \( \alpha_0 < 90^\circ \). (b) Variation of the resolution \( \delta \) as a function of the energy at emergence, \( \Phi_0 \).

To the limiting angle \( \alpha_{1m} \) corresponds a critical energy \( \Phi_{0c} \) such that \( \Phi_{0c} = \frac{4}{9} \alpha_{1m}^2 \Phi \). If the electrons which emerge from \( P \) have an energy less than \( \Phi_{0c} \), none will be halted by the diaphragm; if \( \Phi_0 > \Phi_{0c} \) however, the diaphragm will stop all electrons which emerge at angles such that

\[
\sin \alpha_0 > \frac{2}{3} \alpha_{1m} \sqrt{\frac{\Phi}{\Phi_0}}.
\]
ELECTRON OPTICS

The diameter of the aberration disc in the plane of the circle of least confusion can be calculated from the formula:

\[
\delta = 1.2 \frac{\Phi_0}{\Phi} \Delta \sin \alpha_0(1 - \cos \alpha_0)
\]

(see, for example, Septier, 1954a). If \(\Phi_0 < \Phi_{0c}\), therefore, \(\alpha_0\) is not limited by the diaphragm, and we have

\[
\delta = 1.2 \frac{\Phi_0}{\Phi} \Delta = 1.2 \frac{\Phi_0}{E_c}.
\]

If \(\Phi_0 \geq \Phi_{0c}\), \(\alpha_0\) is restricted by the diaphragm to the value given above:

\[
\delta = 0.8 \sqrt[3]{\frac{\Phi_0}{\Phi} \Delta \alpha_{1m} \left(1 - \sqrt{1 - \frac{4}{9} \alpha_{1m}^2 \frac{\Phi}{\Phi_0}}\right)}.
\]

For \(\Phi_0 \gg \Phi_{0c}\), we should have

\[
\delta \approx 0.178 \Delta \sqrt[3]{\frac{\Phi}{\Phi_0} \alpha_{1m}^3}.
\]

Figure 144b shows the variation of \(\delta\) as a function of \(\Phi_0\), with the diaphragm present. The resolving power is clearly given by the maximum value, \(\delta_m\) of \(\delta\), which can be calculated with \(\Phi_0 = \Phi_{0c}\):

\[
\delta_m = 1.2 \frac{\Phi_{0c}}{\Phi} \Delta \approx 0.5 \Delta \alpha_{1m}^2,
\]

or

\[
\delta_m = 0.5 \Delta \frac{r^2}{a^2}.
\]

When a diaphragm is present, the resolution is independent of the velocity dispersion of the electrons emitted by the cathode.

With an objective for which \(\Delta = 2.6\) mm, \(a = 6\) mm, \(\Phi = 20\) kV, and \(2r = 30\) \(\mu\), we find \(\alpha_{1m} = 2.5 \times 10^{-3}\), \(\Phi_{0c} = 0.06\) V and \(\delta_m = 90\) Å.

The influence of the spherical aberration of \(L\) is of the order of 40 Å (for a focal length of 7 mm with \(C_s = 140\) mm)—the stop gives rise to a diffraction aberration which is of the same order of magnitude. With such a system, therefore, we can hope in theory to reach a resolution of the order of 100 Å. The experimental resolution obtained by Fert and Simon (1956) is of the order of 250 Å. This value is eight times better than that to be obtained with the best light microscopes, but forty times worse than that of the best transmission electron microscopes.

13.2.3 The Objective

The electrons leave the object with a very low energy, \(\Phi_0\), and arrive at the screen with an energy \(\Phi_i\) which corresponds to the accelerating potential. We return to the concept of refractive index \(n = \sqrt{\Phi}\); in object space,
\( n_0 \) is small—\( 0 \leq n_0 \leq 2 \) according to the particular case in question—while in image space, \( n_i \) is large—of the order of 100 or 200 in modern instruments. Here, therefore, there is a large disparity between the indices at the entry and the exit, unlike the situation in transmission optics where \( n_0 = n_i \). By analogy with the light optical situation in which this disparity between the indices is to be found (always, a far smaller disparity) the objective of emission instruments is known as an “immersion objective”.

The objective performs two distinct functions, as it serves both to accelerate the electrons which are emitted, and to provide a real magnified image of the object. From the very beginning, both of the two types of objective which are possible have been used—either the lens can perform both operations simultaneously, in which case it can only be electrostatic, or alternatively it can be “mixed”, when it consists of a part whose function is solely to accelerate, followed by an ordinary lens, which is more often magnetic, with which the beam is focused. It was with the aid of the former of these two types of objective that various research workers were able, between 1940 and 1941, to obtain a higher resolving power than the limit of the light microscope, of the order of tens of millimicrons (100 Å) (Mahl, 1936; Johannson 1933; Mecklenburg, 1942; Boersch, 1942). Kinder (1944), however, proved that a mixed objective could give equivalent results. At present, the best purely electrostatic solution at a given operating potential seems to be a mixed objective with an electrostatic lens (Septier, 1953 b).

(i) The electrostatic immersion objective (Fig. 144)

This is the name by which the earliest, and now the most common type of electrostatic objective is known. Originally, its construction was very crude—the electrodes simply thin sheets of molybdenum—but gradually, improvements inspired by the study of transmission microscope lenses have been made, with the result that the electrodes of modern objectives of this type, operating at high tension, are machined and centred to an extremely exacting degree of precision. There are only two electrodes, which lie in front of the plane cathode \( K \) which is in fact the surface of the object. The first electrode \( W \) (which is commonly known as the grid or Wehnelt, by analogy with the gun in a cathode ray tube) is mechanically rigidly linked to the anode \( A \), with the aid of insulating supports, but is held at a potential very close to that of the cathode. In general, the anode is earthed, and is joined to the positive terminal of the high tension; the insulated cathode is connected to the negative terminal. The object-cathode can be moved in any direction; by moving it perpendicular to the axis, the specimen can be explored, while a displacement along the axis provides a coarse focusing of the image at the screen \( E \). The fine focusing is achieved by varying the potential difference \( v \) between the grid and the cathode slightly; \( v \) reaches at most a few hundred volts when the accelerating
Fig. 145. (a) Diagram of the objective of Johansson's microscope (1933). (b) Diagram of the electrostatic objective. The appearance of the equipotential surfaces and of the electron trajectories which originate in a point $P$; $P'$ is the image of $P$ and $C$ is the crossover.

Fig. 146. The behaviour of the axial potential distribution in an immersion objective, in focus.

Potential of the anode is of the order of tens of kilovolts. The potential of the grid is usually negative with respect to that of the cathode.

The two functions of the lens, acceleration and focusing, are intimately bound together in this case. Figure 145 shows a map of the potential in such a system, through which the path of an elementary pencil is indicated. If the rays which emerge from a point object are to be reunited into a point image, the equipotentials which the rays cross must be curved convexly towards the cathode, and further, the axial potential distribution must be
of the form illustrated in Fig. 146, in that the second derivative in the vicinity of the object must be positive. Experiment shows that when the system is well focused, the electrostatic field \( E_0 \) at the cathode is weak. If we denote the accelerating potential by \( V \), and express \( E_0 \) in volt cm\(^{-1}\), the latter can never, it is found, exceed \( V \) volt cm\(^{-1}\); this imposes a limit on the theoretical resolving power (Septier, 1954). The field \( E_0 \) increases when the dimensions of the objective are scaled down, but this cannot be continued indefinitely because of the danger of continuous discharge, and because of the additional aberrations which appear when the radius of the holes is reduced too far. In good conditions, the minimum dimensions of an objective with a high quality factor operating at about \( V = 30 \text{ kV} \) are

\[
\begin{align*}
a & = 0.2 \text{ to } 0.4 \text{ mm} & \text{for the object-grid distance}, \\
b & = 0.2 \text{ to } 0.3 \text{ mm} & \text{for the grid}, \\
R_1 & = 0.65 \text{ to } 0.7 \text{ mm} & \text{for the anode.}
\end{align*}
\]

and

\[
\begin{align*}
c & = 2 \text{ mm} \\
R_2 & = 0.25 \text{ to } 0.5 \text{ mm}
\end{align*}
\]

In the optimum situation, the grid potential is close to zero. The focal length, in this case, is of the order of 2.5 to 3 mm, which allows direct magnification of 200 or 300 times in a long enough instrument; this is often adequate.

There are certain applications, however, in which we have to make allowance for the bulk of the parts which are merely accessories in light optics, but which are indispensable to the metallurgist; the interelectrode spacing has to be increased, \( E_0 \) decreases in consequence, and so does the resolving power. Hence the high tension \( V \) must be increased as far as the limits defined by the possibility of breakdown between the electrodes permit.

(ii) The objective with a magnetic lens, in which the two functions are separated

The plane cathode lies in front of an accelerating anode with a very small opening, behind which is placed a weakly convergent magnetic lens. We consider a typical point \( P \) (Fig. 147) of the object which is emitting electrons; as far as the lens \( L \) which follows the accelerating unit is concerned, the electrons originate in a virtual object \( P' \) which lies, to a first approximation, at a distance \( 2a \) from \( L \), as the trajectories are parabolae. For a system operating at 30 kV, we find \( a \geq 2 \text{ mm} \). As \( P' \) has to be effectively at the focus of \( L \), we shall have at the minimum \( f \sim 5 \) or 6 mm, making allowance for the thickness of the lens. To obtain a good image when the diameter of the opening in the anode is of the order of a millimetre, the plane cathode must be surrounded by a guard-ring held at the same potential, but placed slightly in front of it (this latter is to avoid field curvature). This guard-ring has a slightly unfavourable effect upon the axial potential, in that the field \( E_0 \) is slightly diminished. If, on the contrary, the opening in the anode
is a few tenths of a millimetre across, and if the latter is thin, a plane cathode can be used; the electric field $E_0$ is then virtually equal to the ratio $\Phi_i/a$.

Nowadays, a shielded objective of the standard type is used, feebly excited. The direct magnification provided by the objective alone is slight owing to its long focal length, and if a magnification of the order of 1000 is required, a second lens has inevitably to be introduced as projective. An excellent microscope can be constructed in this way, which possesses the further advantage of being capable of functioning at much higher accelerating potentials than the objective described earlier (Fert and Simon, 1956).

(iii) The wholly electrostatic objective, in which the functions are separated (§ 8.3.2)

This system works on the same principle as the lenses of the preceding paragraph. Once again, the accelerating system is separated from the lens itself which follows it. This lens is a weak three-electrode unipotential lens. For a given limit $V$ of the accelerating potential, this arrangement is an improvement on the standard type of objective, as the field $E_0$ is far greater (Septier, 1954; Düker and Illenberger, 1962).

13.3 THE VARIOUS TYPES OF MICROSCOPE

Modern microscopes are optically equal in quality, whether they be magnetic or electrostatic. A clearer classification is achieved by considering the way in which the electrons are emitted by the object, and we shall distinguish between:
(a) microscopes in which the emission is thermionic,
(b) microscopes which use secondary emission, and
(c) microscopes in which the electrons are emitted photoelectrically.
The difference between these instruments lies not in their optics but in the
arrangement of their cathodes. Finally, we shall also examine some very
different kinds of microscope, some of which have a specialized optical
system which can cope with the emanations from radioactive materials,
while others have no lenses at all, like the line or point microscope.

13.3.1 The Thermal Emission Microscope

This is the oldest type, but only a small number exists in the whole
world at the present time.

(i) The range of emitters

A wide variety of emissive objects can be examined:

*Pure metals and refractory metals* which emit spontaneously at high tem-
peratures: molybdenum, tantalum, niobium and tungsten. The minimum
emission for which an image directly magnified a few hundred times can
be obtained is reached when the metal is at a temperature of the order of
1200 to 1500°C. A powerful heating system is necessary therefore, and
it is useful to be able to reach a temperature as high as 2500°C if the recryst-
allization of these refractory materials is to be studied.

*Metals which contain impurities which diffuse to the surface* and which have
in addition an intense emissive power—thoriated tungsten and thoriated
molybdenum, for example, from which the cathodes of certain vacuum
tubes are constructed. The thorium is always present in only very small
quantities, but distributed over the surface of the material, it creates a
pronounced reduction of the potential barrier at the surface of the metal
on which it lies, so that an intense emission can be obtained at far lower
temperatures than with pure metal.

*Metals which are virtually non-emissive* (because they melt before they
begin to emit freely) *with activation of the surface*. The main agents which
are used for activation are the alkaline earth metals (barium and strontium)
and caesium; they form a monomolecular layer over the surface of the
object, uniformly distributed so that all the details of the underlying sur-
face are still present. The surface can be examined at about 800°C when it
is covered with a layer of barium or strontium, which is laid down by eva-
poration in vacuo within the body of the microscope itself (the carbonate,
or any other pure salt which on heating gives off the metallic constituent
without any harmful impurities, is used). The metal of the object must be
able to support a temperature of about 1200°C, however, which permits
"activation". To obtain an acceptable emission, and to render the under-
lying metal apparent, many cycles 800°C ⇔ 1200°C are sometimes necessary;
if the activation is to be accurately reproducible, and above all if the observations are to be of long duration, the microscope must be very thoroughly evacuated (to a pressure of less than $10^{-5}$ mm of mercury). In a new activation technique using barium, described and discussed in detail in an article by Heidenreich (1955), a solution of barium formate in a mixture of formic acid and isopropyl alcohol is used. Perfect polishing and cleaning of the surface which is to be activated are vital conditions for success, and the final result depends on the nature of the metal. Iron, nickel, copper, cobalt and chromium give intense and easily visible emission; the technique described above cannot be used with caesium, which is too easily oxidized, unless the whole operation is conducted inside a highly evacuated sealed-off glass envelope. Alternatively, the object can be plunged into caesium vapour while it is actually being examined; the vapour is either produced in an adjacent oven which is incorporated into the apparatus, or by heating a tungsten filament covered with caesium to about 700°C. The observations themselves are made at temperatures of the order of 500°C.

Complex emitters, such as cathodes formed from the oxides of the alkaline earth metals, together with a thick or thin oxidized layer; L-type cathodes, which consist of a piece of porous tungsten, through the pores of which the barium which is spread in the form of the oxide on the other face of the specimen can diffuse; thorium oxide cathodes, in which the metal support is covered with thorium oxide; cathodes consisting of the borides of the rare earths, obtained by sintering; metal cathodes directly alloyed to barium (Ni–Ba alloy, for example); these examples by no means exhaust the possibilities.

Composite objects, which contain numerous substances juxtaposed together, of which at least one is emissive; at the emission temperature, the ensemble must be a conductor, as the surface of the cathode has to be an equipotential if focusing is to be possible. In this way, it has been possible to examine such materials as the complex alumino-chrome agglomerates, and also many sorts of alumina which are conducting above 1000°C.

Finally, by simply changing the polarity, it is possible to transform a microscope which had been operating with thermionic emission into a microscope using thermally emitted ions. It is then possible to examine the surfaces of poor conductors which emit positive ions spontaneously (Couchet, Gauzit and Septier, 1951, 1952), for example the complex silicates which comprise the “steatites”.

(ii) The origin of the contrast

The contrast in the image is attributed to a number of different causes. (a) The irregularities of the surface can be seen, and it is often difficult to interpret the appearance of the image when the relief is very pronounced (of the order of a few hundredths of a millimetre, that is). The depth of
field of an immersion objective is, in fact, extremely small— theoretically it is zero—and in addition, even the smallest surface irregularity behaves like an elementary diverging or converging lens. The contrast can become inverted according to which plane in the neighbourhood of the surface corresponds to the focus—for example the edges of the irregularity shown in Fig. 148 appear dark or bright according as the focus corresponds to the plane \( P \) or the plane \( Q \). A sharp ridge appears as a dark line with two bright streaks parallel to it, and a similar phenomenon is seen if the irregularity is pointed.

![Diagram](image_url)

**Fig. 148.** The distortion of the image which is produced by surface relief.

(b) In the case of highly emissive cathodes formed from juxtaposed grains, an effect is produced by space charge—certain points of the image are dark not because they correspond to non-emissive zones, but because there is a slight depression on the emitter at which the emission is sufficiently intense to create enough space charge to repel the emitted electrons.

(c) On a specimen which has been polished and, after etching off the superficial amorphous layer produced by the polishing, annealed, the different crystals are visible in the form of grey regions, ranging between light grey and completely black. This can be explained quite simply, as the work function \( \Phi_s \) of the metal varies slightly according to the orientation of the crystal face. If the emission can be described by Richardson's law, 

\[
i = A T^2 \exp(-e \Phi_s/kT),
\]

it can be seen immediately that a change \( \Delta \Phi_s \) of a few tenths of a volt is sufficient to alter the intensity of emission by a factor 10. Exact measurements on single crystals have shown that \( \Phi_s \) can vary up to 20 per cent about its mean value, so that \( \Delta \Phi_s \) can exceed 1 volt for tungsten (Müller, 1955). Moreover, the emission microscope can be used to measure these variations, \( \Delta \Phi_s \), by keeping the temperature fixed, and bringing the image of successive crystals onto a Faraday cage. The problem is to identify the faces being studied. The contrast is better at lower temperatures near the point at which the emission begins to appear; when \( T \) is increased, however, all the crystals emit, and the completely black regions disappear. The different shades become uniform for reasons which are imperfectly understood at present—among these, the properties of the fluorescent screen play a part at high intensities.
Fig. 149. A niobium surface, displaying marked contrast between the different crystal faces, and gentle surface relief; $G = 600$ (Photo: Septier).

Fig. 150. Platinum crystals. The enlarged inset represents a part of the region enclosed by a black rectangle, and shows the slip lines which can be seen in the dark crystal; in the main figure, $G = 500$. in the inset $G = 6,000$ (Photo: Septier).
In Figs. 149 and 150, the appearance of various recrystallized metals is illustrated. The appearance of the surface of the activated metals is the same, and it seems likely that the activating material is selectively absorbed over the different facets of the crystals, and that the potential barrier $\Phi_s$, which is generally speaking reduced by the presence of this activating agent, varies appreciably according to the crystalline orientation of the face which is being examined.

In conclusion, we should mention that in the microscope using thermally emitted electrons, just as in the metallographical light microscope, the surface requires a special preparation—in order to avoid relief which would produce "ghosts", the surface has to be given an extended polishing (first mechanical, and then electrolytic), and after this, it must be cleaned chemically, by cathode sputtering, or more simply by raising it to a high temperature in vacuo (Gauzit and Septier, 1951).

(iii) Application to the study of electron emission processes

When it was first developed, this type of microscope was used to try to throw some light on the mechanism of electron emission itself—the structure of the emitter, the evolution of the latter with time, and the effect of aging. The work of Brüche and Johansson (1932, 1934) on oxide-coated cathodes should be mentioned particularly in this context. In these early experiments the image was well contrasted but only magnified about 65 times. In Fig. 151 (obtained by Septier), the same type of image is shown, but far more highly magnified. It was in the course of their research that Brüche and Johansson discovered (in 1933) that the crystals of the metal could be made visible by covering the surface with a very small quantity of barium, which was obtained by heating barium oxide. Schenk (1935) measured the relative emissivity of different crystals of nickel activated with barium; he found that the most extreme variation of potential was in the ratio 6:1, which corresponds to a variation in the height of the potential barrier of 0.2 V. Mahl (1936) examined the influence of absorbed oxygen on the thermal emission at thoriated tungsten and molybdenum cathodes.

The diffusion mechanism by which thorium travels from the interior of a specimen of tungsten or molybdenum to the surface has been directly observed by Brüche and Mahl (1935, 1936). The thorium travels rapidly from the interior to the surface, through pores which lie not only at the junctions of the grains but actually traverse them. The thorium distributes itself around the emitting "crater", and then evaporates. At the optimum temperature the surface is completely covered, the thorium which arrives more than compensating the thorium lost by evaporation. At low temperatures the emission is poor as the diffusion does not proceed smoothly; very high temperatures also have a deactivating effect, as the evaporation outweighs the diffusion. The beneficial effect on the emission of a preli-
Fig. 151. Armco iron daubed with barium carbonate. The emissive points are grains of carbonate. The crystal boundaries are just beginning to appear (Photo: Septier).

inary carburation of the support has been demonstrated by the same method. The migration of the thorium to the surface follows privileged directions before it disappears as a result of evaporation. All these emission mechanisms had been thought of many years before by Langmuir and his students, but their proofs were most indirect and were based upon a complicated theory, whereas in the microscope, the proof is direct and striking, and is perfectly credible without reservations.

More recently, Stranski and Suhrmann (1947) have studied directly the electron emission of the crystallized surface of tungsten and its relation to the structure and also the mechanism by which foreign atoms are adsorbed. The adsorption proves to be selective, and depends upon the atomic diameter of the impurity, the orientation, and the potential barrier at the crystalline face.

For crystallography and metallurgy, Brüche and Knecht (1934, 1935) have developed a more supple technique for barium activation, in which the barium is vaporized in the vicinity of the object so that the vapour condenses on the surface being observed. In this way, we can, for example, watch the allotropic $\alpha \rightleftharpoons \gamma$ transformation which occurs at about 900°C in iron. The transformation of an individual crystal of the specimen can be followed as the temperature is slowly varied, as Burgers and Ploos van Amstel (1937, a, b; 1938 a, b) for example, have shown; the same authors have also followed
the transformation of zirconium from the $\alpha$-phase to the $\beta$-phase at 600°C, and have recorded photographically the recrystallization of ferro-nickel. Recently, Rathenau and Baas (1951) have taken up the same research but with a high resolution microscope capable of high magnification, and studied in particular the growth of individual grains. They have attempted to elucidate the mechanism by which the junctions of the grains move during the recrystallization in a large number of alloys, amongst which the ferro-nickels are the simplest to observe.

Another possibility is to examine sections of wires, in order to determine the variations in the structure between the outer surface and the centre (Septier, Gauzit and Baruch, 1952)—in this way, it is possible to measure the diffusion of nitrogen in a niobium wire (Fig. 152). Using this same technique, it is likewise possible to study the diffusion of one metal through another, and precipitation and segregation in alloys.

![Fig. 152. Cross-section of a niobium wire, examined with a thermionic emission microscope. The wire is included in a small niobium plate. Right: pure niobium wire. Left: niobium wire nitridized in ammonia vapour. The layer of nitride (which is invisible optically) is clearly visible.]

13.3.2 Secondary Emission

(i) The basic principle

Every surface which is bombarded by particles with a certain energy emits electrons which are known as "secondary electrons". This emission is related to the energy of the incident primary particles, but for the majority of substances, the ratio $d$ between the secondary current and the primary current varies in the same way whether the primary particles be electrons
or ions. \( d \) reaches one for incident energies between 0.5 and 1 kV, then exceeds this value, but for high energies (in the region of tens of kilovolts) it generally falls to a value below unity. The multiplication factor \( d \) varies with the nature of the body which is being bombarded (varying between 1 and 15 for the whole range of polycrystalline materials) and with the nature of the crystalline face which is encountered.

The possibility of using secondary electron emission to form electron images was demonstrated by Zworykin in 1933. Mesccher (1938) made improvements to the earlier instrument after Behne (1936) had studied thin films by transmission; these attempts met with little success, however. The fact is that in all these attempts, some of the primary electrons reach the fluorescent screen, and since they cannot be focused at the same time as the secondary electrons, they confuse the image which is produced by the latter, by producing either a continuous parasitic illumination or intense parasitic spots. The resolving power which can be attained is very poor (a few tenths of a millimetre) as a result of the extreme velocity dispersion, and the project was therefore abandoned.

Recently, the method has been taken up once again by Bayh (1958) and with a highly developed experimental arrangement, he has been able to obtain extremely well-resolved images with 15 keV primary electrons (see § 13.4.2).

Mahl (1938) obtained an image of the cathode of a gas discharge tube with a pierced anode; the image which he obtained was of poor quality for a number of reasons but the experiment is of interest as it points to the possibility of obtaining emission of secondary electrons when the primary particles are ions. A similar, though improved process has recently been used successfully by a number of workers (Mönlenstedt and Düker, 1953; Septier, 1953a), who bombard the surface with positive ions. The ions which are reflected by the surface cannot get to the fluorescent screen, and the images prove to be excellent. The velocity dispersion of the electrons which are emitted is amazingly small (Mönlenstedt and Düker, 1953) in contradiction to the opinion which had been current before these experiments. The half-width of the spectrum is only 2 V, which means that we can hope for a resolving power comparable with that of the light microscope without having recourse to a small stop, which is so very difficult to place accurately at the crossover. Figure 153 gives some examples of the values of the coefficient of secondary emission for a few common metals.

The ions which impinge on a surface also give rise to positive and negative secondary ions. The latter are predominantly hydrogen ions, \( H^- \), and if the secondary electrons are eliminated by placing a weak magnetic field at the exit of the objective, it is possible to obtain an image of the surface which is being bombarded with the aid of these \( H^- \) ions (Bernard and Goutte, 1958). These ions originate in a layer adsorbed on the surface of the specimen; their emission is independent of the nature of the incident
ions, and occurs whatever the specimen may be when the vacuum is $10^{-4}$ or $10^{-5}$ mm Hg.

The positive ions which are emitted by a surface which has been perfectly cleaned and placed in a very good vacuum ($10^{-6}$ to $10^{-7}$ mm Hg) are mostly ions which are characteristic of the specimen, which have been torn away from the substance itself (Castaing and Slodzian, 1960). They too can be used to form an image of the surface; if we eliminate all the unwanted ions with a magnetic deflector, we can even form the image with one category of ions alone, which thus provides a map of the distribution of the corresponding element in the specimen (Castaing, Jouffrey and Slodzian, 1960; Castaing and Slodzian, 1962). The instrument with which images of this type can be obtained will be described later (§ 13.4.3).

![Graph](image)

**Fig. 153.** The secondary emission coefficient of different metals under the impact of ions of energy $E$. (a) Couchet (1954). (b) Hill, Buchner, Clark and Fisk (1939).
(ii) Contrast (secondary electrons)

Just as in thermal emission, the origin of the contrast can be attributed to a host of possible causes.

The relief. For normal bombardment, the only effect produced by the relief is that of a weak lens, and the appearance resembles that of the similar phenomenon in thermal emission. When the bombarding particles are incident obliquely, however, we benefit from a shadowing effect (the obliquity is usually of the order of 30° or 45°); the bombardment is less intense or non-existent on the slopes which are "invisible" from the source of the primary particles, and the more important relief throws a shadow across the neighbouring part of the surface. Nevertheless, we should repeat once again that this relief must remain very small if the resolving power is to be acceptable—as a general rule, we may say that the relief should be no higher than the value of the resolving power which we are trying to attain.

The different nature of the constituents of the object in the case of a composite object. In Fig. 154, the appearance of a chrome–alumina mixture is shown, in which the regions of alumina are lighter, but observations are still few in number; they have been obtained with specimens at ordinary temperatures. How to display the influence of the orientation of the crystal faces is not known except during the few seconds immediately after the
ion bombardment begins. The object has to be heated to a temperature of a few hundred degrees in order to eliminate the superficial parasitic layer; this layer, which is a consequence of the ion bombardment itself, consists principally of a deposit of carbon which is produced by the breaking down of organic molecules which are to be found inside the apparatus (in such substances as oils, greases, or solvents). A partial elimination can be achieved by heating the specimen to about 200° both before and during the experimental observation. Yet another difficulty appears, especially with those metals which are particularly susceptible to oxidization—all the contrast between the different grains is obliterated by a thin layer of

![Image](image_url)

**FIG. 155a.** The surface of perlite steel. The image obtained by secondary emission with ionic bombardment (Photo: Keller, see Möllenstedt and Hubig, 1954).

Direct $G = 580$; final $G = 2550$. 
oxide. Figure 155a is a photograph of the surface of a perlitic steel at 150°C, in which the different grains and the relief of the surface are perfectly visible (Möllenstedt and Hubig, 1954; Fert and Simon, 1956, 1957).

![Image](image_url)

**Fig. 155b.** The image of the surface of chrome steel, obtained by secondary emission; $G = 2400$ (Photo: Simon).

(iii) *Contrast (secondary ions)*

The experimental work of Bernard and Goutte (1958) has shown that the contrast which is obtained with $H^-$ ions is of twofold origin: the relief, and the nature of the metal—we could easily distinguish islands of gold evaporated onto aluminium, as the “emissive power” of the adsorbed layer (where the emission is of negative ions) would depend upon the nature of the metal.
The positive ion emission which is obtained with a purely electrostatic optical system (in which the convergence is independent of the mass) ought to be capable of forming very highly contrasted images of a heterogeneous specimen, as the emissive power for each of the characteristic ions is very different according to the various metals: the emission is particularly intense for silicon, aluminium and nickel, and weaker for beryllium and copper. Alkali compounds also give very intense emission. If we filter the beam behind the objective, and allow only one category of ions to pass, we obtain a very well-contrasted image.

(iv) Applications

The examination of surfaces with the aid of secondary emission widens the field of application of the emission microscope—it is possible nowadays to examine conducting surfaces at temperatures above a limit as low as 150 or 200°C, which is of great interest metallurgically where specimens of the light alloys are concerned. The resolving power, which is of the order of 2000 Å, is inferior to that which can be obtained with a reflexion instrument, but the data which are to be obtained are numerous and adequate for the study of recrystallization or of inclusions of foreign bodies.

If we form the image with each category of emitted positive ions successively, we can perform a kind of chemical analysis of the specimen.

13.3.3 Photoelectric Emission

(i) The basic principle

Photoelectric emission was first introduced in 1933 to obtain an image of the object without heating it. It was thought that by irradiating a clean and polished surface with intense ultra-violet light, it would be possible to produce a flux of electrons sufficiently dense to produce an appreciable magnification, and it was hoped that the contrast in the image of the crystals would appear just as in the case of thermal emission. Numerous experimental problems obstructed the development of this technique, which has in consequence been slow, and successive attempts have run up against very weak emission, and inhomogeneity of the velocities too marked for the resolving power to be other than poor, but above all, against the problem of rendering the crystals visible.

(ii) Applications

Zinc is a very good emitter of photoelectrons, and it was this element which was used in the earliest experiments, by Brüche (1933). The surface, which was freshly polished, was vividly illuminated by ultra-violet light,
and the electrons which were emitted were accelerated through 30 kV and then focused with a magnetic lens. On the image, which is only very slightly magnified, only the shadow relief produced by oblique illumination can be discerned. Later, Pohl (1934) and afterwards Mahl and Pohl (1935) showed how unfavourable is the influence of monomolecular surface layers of impurity and of the absorbed gas (or oxides) which are invisible in an ordinary microscope. Only the relief is visible when the specimen is cold; even after cleaning the specimen by prolonged red-heat, and subsequently cooling it, only the outlines of a few crystals can be made out on specimens of platinum and nickel, although the magnification is slightly higher (30 instead of 6). With the aid of a cathode made of many metals, the same workers were able to show that a high tension discharge in air at a very low pressure reduces the emissivity (except for the noble metals, gold, silver and platinum) whereas the same discharge in hydrogen increases it by a large factor. Finally, Gross and Seitz (1937) succeeded in making the crystals of a nickel specimen visible, in two different ways. In the first method, they varied the properties of the gas at the surface; at the beginning of the cooling which is carried out in vacuo, the gas is adsorbed to a different extent at the different crystalline faces which have been heated to red-heat. The contrast which is produced by this procedure rapidly vanishes, however, as the layer of adsorbed gas becomes too important at low temperatures. In the other method, a polished and outgassed specimen of nickel is covered with a thick layer of barium, held for a considerable time at a temperature of about 1000°C to obtain a well-contrasted image of the nickel with the thermal electrons, and only after this is it cooled again. On irradiating it with ultra-violet light, quite as good an image is obtained, with the same contrast between the different crystals. After the specimen has been cold for some time, the crystals fade away, and only the relief remains visible, but with fresh heating, which gets rid of the gas adsorbed by the barium, the crystals reappear. This is a difficult technique and is only suitable for the relatively refractory metals, as a temperature in the neighbourhood of 1000°C is necessary for outgassing and activation. It was abandoned after 1937.

Recently, however, Huguenin (1954, 1955, 1956) has developed a general procedure for examining cold specimens with a high magnification. He is able to destroy the amorphous surface layer without recourse either to thermal outgassing or to activation with barium, by scrutinizing carefully the ways of “cleaning” the surface for which the crystalline contrast is visible. The standard procedure of cathode sputtering in an inert gas like argon or a reducing agent such as hydrogen is inadequate; the grains are visible, but almost without contrast, as in fact the object is enveloped in a very fine layer of metal dust—the metal is that from which the other electrode of the discharge system is composed. By using a powerful ion gun, the contrast becomes visible as this drawback is avoided. Two examples of photo-
graphs obtained in this way are shown in Figs. 156 and 157. The resolving power is of the order of a micron on these pictures; this is a valuable method at low temperatures, as metals which are particularly susceptible to oxidation—uranium, for example—for which all the other techniques break down, can now be examined.

**Fig. 156.** The appearance of a copper surface after ionic scouring; photoelectric emission. $G = 100$ (Photo: Huguenin).

**Fig. 157.** The appearance of a silver surface after ionic scouring; photoelectric emission. $G = 100$ (Photo: Huguenin).
13.3.4 The \(\beta\)-ray Microscope and the Recoil Nucleus Microscope

A few attempts have been made to form an image of the surface of certain radioactive substances by using the natural \(\beta\)-emission. They have, however, run up against a number of difficult problems.

Firstly, the intensities are very small; exposures of many hours or even several days are necessary, although the magnifications are always small (between 1 and 10).

In addition, the particles emitted have a high velocity—\(\beta\)-rays have energies between 100 keV and 1 MeV; the magnetic lenses, therefore, must be able to accept rays emitted into a large solid angle but still be strong enough to focus electrons with an energy near 1 MeV within a reasonable distance.

Finally, the velocity spectrum is often very wide, so that the chromatic aberration is enormous.

In this context, we mention the articles by Marton and Abelson (1947) and by Barker, Richardson and Feather (1950); the latter obtained the image of a brass disc activated with radium B, magnified seven times, after an exposure of four and half hours. The electrons which are used have an initial velocity of 148 kV and a particularly narrow velocity spectrum, which allows the resolving power to be 6 \(\mu\).

In this rudimentary form, such microscopes as these present no practical interest, as modern direct auto-radiographs (in which a special photographic emulsion is placed in direct contact with the specimen) have a resolving power of a fraction of a micron, far better than that of a \(\beta\)-ray microscope.

Recently, attempts have been made to form the image of a radioactive layer emitting \(\alpha\) or \(\beta\)-rays by making use of the recoil atoms which leave the surface in the form of positive or negative ions with small energies as the elementary particles are emitted (Stühmer, 1960; Willax, 1960). The layer to be examined does not produce an intensity high enough to give an image which is visible to the eye on a fluorescent screen, and the optical adjustments are therefore made with a massive ion source situated in the plane of the layer (the first electrode of the immersion objective). As “objects”, the authors used metallic plates activated with the emanations from thorium; they obtained ion images of the surfaces activated in this way with magnifications between 10 and 40 times, with exposures of the order of two hours, and accelerating potentials between 10 and 30 kV.

13.3.5 The Importance of the Vacuum in an Emission Microscope

We have already seen how fundamental is the role which is played by the surface phenomena so far as the appearance of contrast in secondary emission microscopy or photoelectric emission microscopy is concerned, as the basic mechanism of these types of emission is controlled by these phenomena. Within instruments of this kind, therefore, it is most important
that the vacuum should be as good as possible and also as "clean" as possible, in the sense that it should be free of vapours which might con- dense or dissociate; this prevents the surfaces which are being examined from becoming contaminated, either by adsorption of gas or, above all, by the formation of a layer of carbon which will have originated in the molecules of grease or oil which always diffuse into the vacuum chamber, even when low temperature traps are employed.

At the present time, it seems highly desirable to employ the very latest developments of high vacuum technology when an emission microscope is being constructed.

The vacuum chamber and the optical system should be capable of being entirely outgassed by heating to some hundreds of degrees centigrade, which means that the whole assembly must be made of stainless steel, the joints must be metallic—made of aluminium or indium rather than of neoprene—and the insulators either quartz or a ceramic substance.

"Clean" pumps must be used, which will not produce troublesome vapours in the vacuum chamber. This requirement is best satisfied by the titanium pumps which appeared on the market a few years ago; the action of these pumps consists essentially in the adsorption of the residual gas by titanium vapour (which is similar to the effect produced by a "getter" in pumping down a radio tube). In certain types, ionic pumping is used as well—an intense electron beam falls onto a titanium plate (Huber, Schroff and Warnecke, 1959) which slowly evaporates and at the same time, ionizes the residual gas. The ions which are formed, together with the vapour which is loaded with adsorbed gas, are collected by a well-cooled outer electrode. If the pumping rate of the pump is to be increased, we can equally well use the evaporation of titanium droplets exposed to intense electron bombard- ment (Holland and Laurenson, 1959). Rates of the order of 2000 l sec$^{-1}$ are to be expected, and a vacuum limited to about $10^{-8}$ mm Hg.

In another type which is simpler to build and operate, the titanium is vaporized by cathode sputtering in a cold discharge; the discharge takes place within a tube which resembles that of a Penning gauge (a gauge which is used to measure low pressures and is among the more familiar vacuum measuring apparatus) provided with a titanium cathode. The gas is ionized by the discharge, and the ions which are collected at the cathode set off an intense cathode sputtering. By connecting a number of discharge units in parallel, it is possible to obtain high pumping rates (up to 4000 l sec$^{-1}$) which remain steady between $10^{-5}$ and $10^{-8}$ mm Hg (Jepsen, 1959). These pumps start at $10^{-2}$ mm Hg; here once again, we can obviate vapours from the oil by replacing the normal mechanical pump by an absorption pump in which the recipient is filled with activated carbon dipped in liquid nitrogen. When the pressure has fallen to $10^{-2}$ mm Hg, the connexion between this pump and the space to be evacuated is cut, and the titanium pump is set in operation. With such a pumping system as this, a vacuum
of $10^{-7}$ or $10^{-8}$ mm Hg can be obtained without any vibration, and this
latter property is another valuable asset in electron microscopy.

In conclusion, we should mention certain "molecular" rotary pumps
which have recently been developed, with which "clean" vacua of at best
about $10^{-6}$ mm Hg can be obtained, and which pump at several hundreds
of litres per second; unfortunately, the speed of the rotation entails a certain
danger of vibration, and if a very high resolution is desired, this is incon-
venient.

13.4 SOME ACTUAL INSTRUMENTS

13.4.1 The Thermal Emission Microscope

The model which we shall describe is a microscope which was designed
by J. Vastel and P. Grivet, and perfected by Septier and Gauzit (1950).
It was intended that it should be convenient to use, and that changing the
specimen should be rapid; further, that it should be easy to subject the latter
to various thermal treatments, as metallurgical research requires a variety
of temperatures. Figure 158 shows the appearance of the instrument.
The head, which contains the object support and the devices for heating
and shifting the object, can be separated from the body of the instrument
by means of a lever. The object, which has to be prepared in the form of
a thin plane disc 5 mm in diameter, is simply placed in a groove in the object
support. It is heated by electron bombardment of its upper face, by means
of an auxiliary electron gun which provides up to 50 watts, and with which
a maximum specimen temperature of 2500°C can be attained. The object
can be moved, in vacuo, as far as the point at which it would be withdrawn
from the case (Fig. 159), and in this position, it can be activated conveniently
in barium vapour which flows across its surface from a small lateral oven.
It is also possible to bake the metal for a protracted period without polluting
the electrodes of the objective. As an arrangement of this kind is sensitive
to vibrations, which would limit the resolving power to 2000 Å, it is neces-
sary to provide the upper part of the grid with an insulating refractory
ring, on which the base of the object-support rests during observations.

Figure 160 represents the second type of objective studied by Septier
(1953 b). The projector lens, which is one of the symmetrical unipotential
type with focal length $f = 16$ mm, has apertures of large diameter (17 mm),
which leave the electrons a wide channel when only the objective is being
used; the magnification of the latter is already between 100 and 250 (ac-
 accordance to the distance between the object and the grid) and its field is
very high. When the second lens is brought into action, the total magni-
 fication is between 1000 and 2500. The potential is normally between
20 and 30 kV, and the resolving power is in practice of the order of 1000 Å.
Fig. 158. The thermal emission microscope in use at the Laboratoire d'Électronique et de Radioélectricité (Paris). General appearance.

An instrument of the same kind but with various mechanical improvements has been described by Couchet (1954).

Heidenreich (1955) described a microscope similar to that of Mecklenburg (1942), using a double projector lens and an indirect cathode heating system, which was to be used for examining different metals at temperatures between 600 and 1000°C, after activation with barium. The resolving power was estimated to be 1000 Å.
An original type of thermal emission microscope has been described by Popp and Walcher (1957); the cathode consists of a thin biological section which is placed on a flat plate, heated by the Joule effect. In the immersion objective, we collect the alkali ions Na\(^+\) and K\(^+\) which are emitted by the heated tissue, and an image magnified 100 times is formed on a fluorescent screen. The resolution obtained is about 0.5 \(\mu\), and the images obtained are very rich in detail. To reduce the exposure time, these authors lay down a very thin layer of Apiezon oil on the surface of the photographic emulsion (using Ilford \(Q_1\) plates, for example); with a current density of the order of \(10^{-10}\) A cm\(^{-2}\), the exposure time is only between a half and three minutes.

**Fig. 159.** Detailed view of the "head" of the emission microscope.

- \(P_1'\) : the specimen-holder in the observation position.
- \(P_1\) : the specimen-holder in the activation position.
- \(F\) : oven.
- \(C\) : heating gun.
- \(W, A\) : objective.
- \(D_1, D_2\) : diaphragms.
Fig. 160. Purely electrostatic mixed objective (Septier, 1954).

- \( F \) : the electron beam which heats the cathode \( K \).
- \( K \) : the cathode (a metal disc) which is placed in the object-support \( P \).
- \( W \) : guard ring, held at the same potential as \( K \).
- \( A_1, A_2 \) : anodes (30 kV).
- \( L \) : electrostatic lens (at the potential \( V - v \)).
- \( I_1, I_2 \) : insulators.

13.4.2 The Secondary Emission Microscope

The microscope of Möllenstedt and Düker (1953) is an electrostatic instrument, with two stages, giving a direct magnification of 200. Figure 161a shows the arrangement of the immersion objective and of the gun with which the object is bombarded with ions. The whole "head"—the object-support, cathode and grid—is held at a potential of 40 kV, together with the accelerating electrode of the ion gun. The grid has a conical form so that the ions which bombard the object at an angle of 30° with the object plane can pass. The anode which is connected electrically to earth is mechanically an integral part of the head.

Beneath the anode is a diaphragm in which is cut a 25 \( \mu \) hole, together with its centring mechanism, the function of which is to reduce the aperture of the elementary pencils and to make the resolution better; despite 2 volts velocity dispersion, Möllenstedt has succeeded in obtaining extremely fine images by the introduction of this innovation and a resolving power better than 1000 Å. The ions are produced in a gas discharge gun, and are subjected to the accelerating potential of the full high tension; slower ions would be preferable, as the resulting electrons would have less velocity dispersion, but supplying the ion gun would be less convenient.

An improved version of this instrument has been described by Düker (1960). A microscope with which it is possible to obtain images by thermal emission, secondary emission (ions \( \rightarrow \) electrons) and photoelectric emission...
has been built by Bas (1960). Finally, Fert et al. (1958) have used a device with a magnetic objective to obtain resolutions of the order of 300 Å.

In the secondary emission microscope in which the primary bombardment consists of 15 kV electrons, which has recently been described by Bayh (1958), the arrangement resembles that used by Möllenstedt (Fig. 161a): the ion gun is simply replaced by an electron gun. By setting an aperture diaphragm of extremely small diameter (15 μ) just below the anode, a resolution of about 1000 Å can be obtained.

![Fig. 161a. The immersion objective and the ion gun, for secondary emission microscopy (Möllenstedt and Düker).](image)

- \( A \) : anode (electrode earthed).
- \( B, B' \) : apparatus for adjusting the contrast diaphragm.
- \( D \) : exit opening for the ions.
- \( E \) : ionization chamber.
- \( G \) : input for the gases.
- \( J \) : apparatus for controlling the ion beam.
- \( K \) : the contrast diaphragm.
- \( M, V \) : apparatus for varying the position of the object O.
- \( S \) : high tension lead.

### 13.4.3 The Positive Secondary Ion Microscope

A diagram of this instrument is given in Fig. 161b (Castaing and Slodzian, 1962). The primary ion beam (protons) bombards the plane surface of the specimen, and the positive ions which are emitted are then accelerated by a four-electrode immersion objective followed by a stop. They next cross the most delicate part of the whole apparatus, which is the homogeneous field magnetic prism analyser which filters out the various different
masses. The inclination of the entry and exit faces defines a stigmatic system for two particular points, $F$ and $F'$, which lie in symmetrical positions with respect to the prism. The point $F$ coincides with the crossover of the objective and a device to correct astigmatism is placed at $F'$. A projective lens magnifies the image which the objective has produced, and focuses it onto the perfectly polished cathode of a second immersion objective; this latter slows the ions down and behaves as an image converter (see Chapter 3 and § 14.4.2). The secondary electrons which are emitted are accelerated in this same objective, and form a magnified image on a fluorescent screen which is placed beneath the projective lens. In this way, the brightness of the image can be enhanced, a necessity for the feeble emissions of characteristic ions from the types of specimens which are used. Figs. 161c, d and e show the differences in the appearance of the surface of an Al–Mg–Si alloy when the silicon, the magnesium and the aluminium ions are successively selected.
13.4.4 A Photoelectric Emission Microscope

Huguenin (1955) constructed, during 1954 and 1955, an electrostatic two-stage instrument, which was wholly enclosed in a demountable metal casing. The direct magnification can vary between 100 and 200 (the objective magnification is 10, while that of the projective is between 10 and 20). The objective with plane electrodes (Fig. 162) has to operate with a long distance between the object and the grid (about 1 cm) to allow the illumination to enter through a lateral window at an angle of about 20°. The other dimensions have to be increased in the same ratio if the quality of the objective is not to suffer, so that the holes in the electrodes are one centimetre in diameter; under such conditions, the focal length is of the order of 15 mm. The ion gun can be moved about in vacuo, and can be brought under the anode; the object is lowered until it is in contact with the grid, and the surface cleaned for rather a long time, 5 or 10 min, according to the nature of the metal being examined, with an ion intensity between
100 and 200 \( \mu A \) (or alternatively, a current density between 5 and 10 mA cm\(^{-2}\)). By using hydrogen, a much finer etching can be achieved than with argon. When the specimen is ready for observation, the ion gun is retracted, the object raised 1 cm above the grid, and the ultra-violet light is switched on. The insulation of the grid must be carefully protected from the metal particles which are released at the object while it is being etched by the ions. The body of the microscope is blackened inside, bent, and provided with a series of blackened diaphragms, all designed to prevent the ultra-violet light which is reflected by the object and the metal parts of the objective from reaching the fluorescent screen and masking the electron image.

![Image of a microscope](image)

**Fig. 161 d. Magnesium ion image.**

Perfected versions of the electrostatic photoemission microscope with which a high resolution (1000 Å) can be obtained, in which an aperture diaphragm is placed beneath the anode of the immersion objective, have recently been described by Koch (1958) and Bethge et al. (1958). The
source is once again a high pressure mercury vapour lamp, and the operating potential is in the range 20—40 kV. The objective and the general arrangement recall those of Möllenstedt and Düker (Fig. 161a).

All these emission microscopes are described in a recent article by Möllenstedt (1963).

---

**Fig. 161c.** Aluminium ion image.

---

**13.5 THE FILAMENT MICROSCOPE AND THE SPHERICAL (OR POINT-PROJECTION) MICROSCOPE**

Besides the emission microscopes in the proper sense, much simpler instruments have been conceived in which a highly magnified image is produced without using any electron lenses at all. The accelerating potential
of several kilovolts is simply applied between the cathode (which is either a point or a wire) and a curved conducting fluorescent screen spherical or cylindrical in shape; the cathode constitutes the centre or the axis of the screen.

Fig. 162. Diagram of a photoemission microscope (Huguenin, 1955).

- **S**: ultra-violet source.
- **L₁, L₂**: quartz optical system.
- **P**: object support.
- **K**: object; **W**: Wehnelt; **A**: anode.
- **C**: (retractable) ion gun for scouring the surface.

This type of instrument owes its inception to Johnson and Shockley (1936) and to Johnson (1938) who were studying the electron emission of polycrystalline filaments and the adsorption of activating atoms at crystals of known orientation. The filament which is to be studied, of radius **a**, is stretched along the axis of a cylindrical glass tube of radius **b**. The inside wall of the tube is covered with a layer of some fluorescent substance, in which a wire bent into the form a helix of large pitch is embedded (Fig. 163); it serves the purpose of an anode. A potential difference of a few thousand volts is applied between the wire, which is maintained at a high temperature, and the surrounding helix. Electrons which are emitted by the wire travel radially in straight lines towards the wall of the tube, where they form a pseudo-image of the wire. In a cross-section, the magnification **G** is given by the ratio of the two radii **b/a**, which can be considerable, but in a meridian plane the magnification always remains unity. This produces a specific
deformation of the image, as all the details appear to be stretched out in the transverse direction. The extraction field \( E_0 \) at the surface of the wire increases rapidly as the radius \( a \) is reduced (\( E_0 \approx K/a \)); this explains why the electrons which are emitted in all directions at any given point are very quickly forced back towards the mean trajectory of the pencil, and hence, despite their initial divergence, produce a very small spot at the image. The transverse resolving power can be better than a micron, although

![Diagram of a "Wire" microscope.](image)

Fig. 163. "Wire" microscope.

the longitudinal resolution is only one or two tenths of a millimetre. Nevertheless, so exciting and so valuable was the information which this first instrument provided that the technique has since been considerably developed. The lack of symmetry in the magnification was first eliminated by Martin (1939) who replaced the wire and cylinder by two concentric spheres, as shown in Fig. 164. The object may be a single crystal of tungsten, for example, polished into a sphere and then etched away to display the various faces. This hollow sphere is heated from within by electron bombardment, and is placed at the centre of an evacuated spherical glass envelope of which the inner surface is coated with fluorescent material. A ring placed between the sphere and the wall, and held at a potential of a few kilovolts, accelerates the electrons which are emitted. As before, the magnification is given by
the ratio of the radii of the envelope and the object. The instrument has been used (Benjamin and Jenkins, 1940) to measure the variations in the emissivity of the different crystalline faces, and to follow the modification to the emission which is produced by selective adsorption of various substances, some activating agents, others contamination.

![Diagram](image)

*Fig. 164. "Sphere" microscope.*

Finally, Müller has raised the magnification to an extraordinarily high value, after protracted study and improvement of the instrument (since 1937). The radius of the sphere he has reduced in the extreme, so that the microscope becomes the point-projection instrument which is widely used at the present time, and which we shall describe in detail in the next section.

### 13.6 THE POINT-PROJECTION MICROSCOPE

#### 13.6.1 Description

A fine tungsten point, which is obtained by chemical action on a wire \( \frac{5}{100} \) mm in diameter (the metal is etched away by liquid sodium nitrite) is mounted in a glass tube the other end of which is almost spherical and is coated with the fluorescent screen as in a cathode ray tube (Fig. 165). A second wire is welded to the first close to the point \( P \), and the two wires are led out of the tube through seals; the point, therefore, can be held at a high temperature by passing a current of a few ampères. A large potential difference can be applied between the point and the fluorescent screen,
and for this, it is profitable to cover the fluorescent material $E$—after it has been spread over the surface of the tube—with a thin layer of collodion (a few hundredths of a micron thick) over which a metallic layer $C$ about a hundred Ångströms thick is spread by evaporation in vacuo. The sealed-off high tension lead is also coated during the evaporation process, so that the potential at the screen is well defined. In addition, the metallic layer behaves like a mirror, and returns in the forward direction part of the light which is emitted by the screen towards the interior of the tube; this produces a marked increase in the brightness (§ 1.4.2).

The tube is evacuated with the greatest possible care, using a mercury vapour diffusion pump with a liquid air trap. The whole operation must be carried out extremely carefully, and in particular, the whole assembly of tube and trap must be dried for at least 24 hours, heating the metal wires to red heat at the same time, if possible. After this, the tube must be baked again for one or two days with the trap immersed in liquid air. The tube is now ready for use, as the vacuum will be of the order of $10^{-10}$ mm Hg. Alternatively, we can work with a sealed-off tube by vaporizing a "getter" in the tube before isolating it from the pump. We shall see below how the quality of the vacuum in the tube can be estimated in a simple way. By applying a potential difference of a few kilovolts between the point (which is the negative electrode) and the fluorescent screen, we obtain an extremely large electric field at the surface of the point of the order of $V/r$, $V$ being the total potential drop, and $r$ the radius of curvature of the point. The radius of curvature of the tip of the point can easily be reduced to less than a micron, so that if $V$ is 10 kV, the field at the point exceeds $10^8$ V cm$^{-1}$. 

![Fig. 165. The principle of the point-projection microscope (shown schematically).](image)
This is a strong enough field to tear electrons out of the tungsten point; these electrons follow effectively radial trajectories, and form a kind of image of the surface of the point on the fluorescent screen. The magnification of this image is huge, as it is clearly given by the ratio of the radii of curvature of the fluorescent screen and the point which is emitting the electrons; for \( r = 1 \mu \) and \( R = 10 \text{ cm} \), the magnification is \( 10^6 \).

### 13.6.2 Emission from a Clean Metal Point

If we heat the point to a high temperature (about 2500°, say) for a few seconds, and then apply a high tension of a few kilovolts, we see a regular pattern of bright patches and dark regions on the fluorescent screen.

![Bare tungsten point, radius 3500 Å; the equilibrium array obtained after annealing at 1300°K (Photo: E. W. Müller).](image)

(Fig. 166a). The emissions of electrons, therefore, is not isotropic—rather, its spatial distribution is dictated by the orientation and the symmetry.
of the crystal lattice at each point which is emitting. In fact, the point consists of a single tungsten crystal, as a result of the smallness of its diameter. Certain crystallographic directions correspond to maximum emission, others to no emission at all. An attempt to explain the presence of these directions in which there is no emission has been made by Müller, by relating them to the direction in which propagation of the conduction electrons within the body of the lattice is forbidden; these are the directions in which there is Bragg reflection of the electron waves at the lattice planes. Unfortunately, such an interpretation would require the conduction electrons to have most unreasonable energies—to explain the lack of emission in the [310] direction, for example, the electrons would have to have an energy of 37.4 eV, whereas calculation of the Fermi energy gives 5.8 eV if there is one conduction electron per atom, 9.2 eV if there are two.

Subsequently, Müller (1939, 1943), Benjamin and Jenkins (1940), Smoluchowski (1941) and Ashworth (1951) have suggested a variety of explanations. As we are at present primarily concerned with the use of field emission in microscopy, we shall not consider the emission from a bare metal surface further.

**Fig. 166b.** A thin layer of barium, absorbed onto a hemispherical tungsten crystal, 2000 Å in radius. The mean coating is 1/20 of a monatomic layer thick (Photo: E. W. Müller).
13.6.3 The Adsorption of Foreign Substances at the Surface of the Point

If, after outgassing the point by heating it to a high temperature, we lay down a very thin layer of some foreign substance on the surface of the point by evacuation under a vacuum—a metallic film of caesium or copper for example—we can study the distribution of the foreign atoms over the surface of the point by applying the high tension and examining the modifications to the cold emission which ensue. This distribution can be modified by heating the point to a moderate temperature, but even at the normal temperature numerous scintillations are to be seen which are produced by migration of the foreign atoms across the surface in an attempt to group themselves into a more stable configuration (Fig. 166b). It is possible, therefore, to study the adsorption of the substance which is added to the tungsten surface; this adsorption is particularly favoured at those points of the surface at which the local orientation of the lattice is such that the mutual arrangement of the tungsten atoms allows the foreign layer to be in epitaxy. Increasing the temperature of the point accelerates the processes of migration and rearrangement, but at too high a temperature the layer begins to evaporate and the image produced by bare tungsten reappears. In a similar way, more complex processes, such as the oxidization of a fine metallic deposit, can be studied (Ashworth, 1948).

13.6.4 Estimation of the Quality of the Vacuum in the Point-projection Microscope

It would be extremely difficult to determine the pressure of the residual gas within the microscope by any of the usual methods, without the aid of a Bayard–Alpert gauge, as this pressure has to be far below $10^{-8} \text{mm Hg}$ if a serious study of some phenomenon is to be undertaken. Fortunately, however, the behaviour of the microscope itself enables us to make a rapid estimate of the residual pressure; we might say that the instrument plays the role of an almost unlimitedly sensitive gauge for its own benefit.

If we examine the pattern which is produced by the electrons which are emitted when the temperature of the point is high enough to denude its surface, the diagram changes progressively; it becomes hazy, bright points appear in the field of view, the overall brightness falls, and the image settles down into an equilibrium configuration perturbed by rapid local intensity fluctuations.

This phenomenon must be interpreted as the result of the adsorption of a continuous film of impurities—oxygen for example—at the tungsten surface. If a known pressure of oxygen has been admitted beforehand into the tube (although never more than $10^{-6} \text{mm Hg}$) it is found that the duration of the period during which the pattern is settling down into
its equilibrium position is inversely proportional to the pressure. Conversely, by measuring this "contamination time", we obtain a rapid absolute—if approximate—measure of the residual pressure in the chamber, even though it would have been too small to be measured by the usual methods. If it is only an order of magnitude that is required, we can consider that a monomolecular layer is sufficient to bring the pattern into equilibrium, and that this layer is formed atom by atom by chance molecular impacts at the point; each molecule which strikes the point is assumed to adhere to it without rebounding. Reasoning in this way, Ashworth (1951) found that a contamination time of 1 minute corresponds to a residual pressure of $10^{-8}$ mm Hg, 100 min to $10^{-10}$ mm Hg, and so on.

13.6.5 High Resolution Microscopy with Field Emission

(i) Examination of the individual adsorbed particles

We examine the dark parts of the pattern which correspond to the non-emissive regions of the point, immediately after the point has been outgassed at high temperature. We can then watch the brilliant points appear and after a very short interval (less than a second), separate out into two adjacent spots; these turn around one another at first, and then either vanish or separate completely into two brilliant spots which stray independently about the field of observation. This behaviour can be explained by ascribing it to the adsorption of a single oxygen molecule, which then breaks up into two atoms attached to adjacent sites in the tungsten lattice; subsequently, these two atoms can separate completely and either disappear as a result of evaporation, or migrate towards an emissive region where their image is swamped by the general brightness. A confirmation of this interpretation is provided by the experimental result that the spot does not divide into two if the residual gas is a monatomic gas such as argon.

This remarkable result suggests that the resolving power of the image which is obtained by field emission is sufficiently high for two emissive points to be distinguished when they are only separated by the distance between two adjacent tungsten atoms, about 2 or 3 Å. As we shall see, the structure of the field emission microscope is such that a resolving power of this order is indeed attainable, at least in favourable cases.

(ii) The resolving power of the point-projection microscope

In order to establish the limit of resolution, either we can consider electron trajectories and the geometrical aberrations only, or we can take into account the diffraction of the electron waves. We shall perform the geometrical analysis first.
We assume that the electrons are emitted from some given point of the surface—which we suppose to be completely spherical—of the tungsten crystal, with an initial velocity \( u_t \) corresponding to an energy \( eV_t \), and that their initial direction is arbitrary. If \( u_t \) were zero, the electron would follow a radial path and arrive at some point \( P \) of the spherical screen (which is held at potential \( V \)). The non-zero initial velocity smears the impacts out over a small circle, centre \( P \), diameter \( D \); we can calculate the latter by taking into account the fact that the electrons which arrive at the rim of the circle must have left the sphere of radius \( r \) tangentially, and subsequently travelled along a hyperbolic path (as the repulsive force is of the form of a Coulomb force); the angle, \( 2\theta \), between the asymptotes of this hyperbola is very small, and one of its foci lies at the centre \( F \) of the tungsten crystal. When the electron hits the screen, it is virtually travelling along the asymptote \( OH \), and is moving with velocity \( u_R \) (Fig. 167a).

![Diagram](image)

**Fig. 167a.** The appearance of the trajectories in the vicinity of the point.
(Note: \( OH = OA \).)

Conservation of momentum gives us the relation \( FH \cdot u_R = FA \cdot u_t \), and the geometrical properties of the hyperbola give \( OF \cos \theta = OA \) or \( FA = OF(1 + \cos \theta) \) and \( FH = OF \sin \theta \). From these relations, we obtain

\[
\frac{FH}{FA} = \frac{1 + \cos \theta}{\sin \theta} = \tan \frac{\theta}{2} = \frac{u_t}{u_R} = \sqrt{\frac{V_t}{V}},
\]

\[
D = 2R \theta \approx 4R \tan \frac{\theta}{2} = 4R \sqrt{\frac{V_t}{V}}.
\]

We have still to estimate \( V_t \), and here the calculation becomes much less precise. The velocity of the thermal motion of the electrons (which corre-
sponds to an energy of about 0.02 eV at normal temperatures) is not the only velocity to contribute to the tangential motion of the electrons as they leave the cathode. The major part of this tangential velocity is produced by distortion of the electrostatic field in the vicinity of the emissive point (a single adsorbed atom, for example). The energy $eV$, produced in this way has been estimated to be 10 eV by Ashworth (1951). For $V = 20$ kV and $R = 5$ cm, we obtain $D = 4 \times 5 \sqrt{\frac{1}{2000}} = 0.44$ cm. If we regard the spots on the screen as being just separated when the distance between their centres is just greater than their radius, and if we denote the magnification by $G = R/r$, the resolving power $\varepsilon$ is given by $\varepsilon = \frac{0.22}{G}$ cm; for $r = 50$ Å (which is an extreme case), with $G = 10^5$, the resolving power is $\varepsilon = 2.2$ Å.

Benjamin and Franklin (1940), starting from a completely different point of view, have obtained an estimate of the resolving power taking the dif-

---

**Fig. 167b.** Phthalocyanine molecules absorbed on a clean tungsten surface between the (011) plane (at the bottom of the photograph) and the (112) plane (at the top). All the molecules lie parallel to the [011] direction (Photo: E. W. Müller).
fraction of the electron waves into consideration. They regarded the electrons which arrive at the screen as having originated in a virtual aperture traced on the spherical surface of the cathode. If we denote the diameter of this aperture by $d$, but otherwise use the same notation as before, the diameter of the spot at the screen will be $dR/r$ in the absence of diffraction. We can, however, regard the angular aperture of the beam as seen from the screen, as being $d/R$, so that the effect of diffraction will be to enlarge the diameter of the spot by an amount $\lambda R/d$, in which $\lambda$ is the wavelength associated with the electrons as they arrive at the screen. The total diameter of the spot, therefore, is $D = R\left(\frac{d}{r} + \frac{\lambda}{d}\right)$, which is a minimum for $d = \sqrt{\frac{150}{V}}$.

We can consider this value, then, as defining the best resolving power. With $V = 20$ kV, $\lambda = \sqrt{\frac{150}{V}} = 0.1$ Å and $r = 50$ Å, the resolving power is $2.2$ Å and the minimum diameter of the spot is $0.44$. This fits in well, besides, with the fact that the electrons which form the image have an energy of about $10$ eV when they leave the object or equivalently, a wavelength of $4$ Å, and are therefore incapable of distinguishing between details less than half a wavelength apart, that is, about $2$ Å.

We might try to combine these two ways of examining the resolving power, by regarding the emitting atom as a small aperture of diameter $d$, emitting electrons in all directions. This localization into a small region about the point on the cathode which is emitting electrons results in an uncertainty $h/d$ in their initial tangential momentum. The total momentum of the emitted electrons, therefore, must be equal to at least $h/2d$; in consequence, the spot which would have a diameter $dR/r$ if tangential emission at zero velocity were possible is enlarged by $\frac{4Ru}{u} = \frac{4R}{d} \frac{h}{2mu} = \frac{2R\lambda}{d}$.

The total diameter of the spot is $R\left(\frac{d}{r} + \frac{2\lambda}{d}\right)$ which is analogous to the result obtained earlier; it has a minimum when $d^2 = 2\lambda r$, with the intermediate result that the electrons which are emitted by a detail which lies at the limit of resolution must have an energy of about $10$ eV. This estimate of the resolving power shows definitely that it is possible to resolve a pair of adjacent atoms, but only in exceptional cases, and it is quite remarkable that this limiting resolving power seems to have been achieved in the examination of a pair of oxygen atoms. The explanation is to be sought for in the extreme simplicity of the point-projection microscope, as a result of which it is insensitive to the numerous causes of instability such as mechanical vibrations or electromagnetic imperfections which are of primary importance in restricting the quality of the image in a microscope with lenses. To obtain a resolving power of $2$ Å is, nevertheless, exceptional; $5$ Å, however, which is common, is adequate to provide a spectacular
analysis of the structure of molecules such as phthalocyanine (Müller, 1950, see Fig. 167b).

In conclusion, we mention that Müller (1951) has succeeded in obtaining images with ions extracted by field emission. By applying a positive high tension (20 kV) to the point, which is surrounded by hydrogen at a pressure of $5 \times 10^{-3}$ mm Hg, the molecules which are attached to the tungsten surface become detached in the form of ions, and travel towards the screen. The image is less bright but finer (see Fig. 167c) and according to Müller, this process renders the individual steps of atoms visible. Too high an accelerating potential, however, produces confusion of the image. Müller

Fig. 167c. The ion image, obtained by field emission, of the central part of a tungsten point; the dark plane (011) is at the centre, and is surrounded by the four dark planes (112) of about 700 Å radius, annealed at 2000°. Helium pressure: 0.2 μ Hg; potential: 19,000 V; magnification: 1,300,000.

The “peary” rings which surround the main planes represent the edges of lattice planes with small indices. Certain planes with high indices are fully resolved.

(Photo: E. W. Müller.)
explains this as the result of ionization of the hydrogen molecules before they arrive at the point.

A general review of field emission phenomena, in particular in connection with microscopy, has been written for the new *Encyclopaedia of Physics*, 21, (1956), 176–231, by R. H. Good and E. W. Müller. A detailed examination has also been made by Dyke and Dolan (1956), and more recently, a very complete review article has been published by Müller (1960).
CHAPTER 14

THE IMAGE-CONVERTER AND THE IMAGE-INTENSIFIER

14.1 THE BASIC PRINCIPLE

An image-converter transforms an invisible image which has been obtained by irradiating the object with infra-red or ultra-violet light into an image which is visible to the eye. The instrument which effects the same transformation of the image which has been formed with an objective which provides visible images is known as an image-intensifier, and in this case, it is the brilliance of the original image which has to be reinforced. Both types of instrument, however, work on the same principle: a suitable objective throws the optical image onto a light-sensitive layer, which emits electrons. The latter are powerfully accelerated, and focused, and produce the final image on a fluorescent screen or a photographic plate.

It was due initially to military requirements that these instruments have been developed to such a high degree; they are used to examine areas which are illuminated with infra-red transmitters at night or in fog, and to detect distant machinery which gives off infra-red radiation (artillery batteries and aircraft, for example). Physics too has benefitted from this progress, however; astronomers, for example, in their attempts to expand the limits of the visible universe, have been using “amplifiers” for fifteen or twenty years now, under the name of “electron telescopes”, and it seems that the recent attempts to use these techniques in the microscope are likely to be successful. Finally, the need for image intensifiers with a very high gain has been felt since the invention of plastic fibre scintillators which emit photons when penetrated by fast ionizing particles; the search has been tenaciously pursued, therefore, both for new optical systems and for ways of perfecting existing systems, especially the photoelectric layers.

14.2 THE MAIN TYPES OF INSTRUMENT

14.2.1 The Image-converter with Uniform Field

The prototype of all these instruments is that of Holst, de Boer and Veenemans (1934)—shown in Fig. 168. It consists of an evacuated glass envelope, two of the faces of which are semi-transparent parallel metallic
planes, a few millimetres apart, between which a potential difference of a few kilovolts is maintained. One face is covered with a photosensitive layer while the other carries the fluorescent screen.

The very intense electrostatic field enables a bright spot of small radius to be formed by each point which is emitting, so that the resolving power is acceptable. The radius of the spot can easily be calculated, and proves to be given by

\[ r = 2d \sqrt{\frac{\Phi_0}{\Phi}} \]

in which \( \Phi \) is the accelerating potential, \( \Phi_0 \) is the mean energy of the electrons which are emitted making a grazing angle with the cathode, and \( d \) is the distance between the cathode and the screen. This tube, which was widely used during the last war because of the simplicity of its construction, is, however, only suitable for examining brightly illuminated objects. If the object is dim, the brilliance at the screen must be augmented by increasing the potential difference which is applied between the two planes; this produces parasitic emission, however, which results in an unwanted continuous background, as a high potential drop corresponds inevitably to a high field at the surface of the cathode, and hence to parasitic field emission.

### 14.2.2 Image-conversion with a Magnetic Field

To overcome this drawback, the anode and the cathode must be placed further apart, and this also reduces the risk of poisoning the photoelectric layer with the fluorescent sulphide. The image, however, must now be focused on the screen. In the first attempts, a combination of a homogeneous electrostatic field \( E \) which accelerated the particles and the homogeneous magnetic field, \( B \), of a long solenoid which focused them was used (Fig. 169). The distance between the cathode and the screen can now be several centimetres. The electrons spiral round the lines of force of the magnetic field, intersecting them from time to time, and by varying \( B \) we can arrange that

![Fig. 168. The Holst converter.](image-url)
one of the foci lies on the fluorescent screen. This system produces a clear image of an extended plane surface with unit magnification. In order to make the electrostatic field more uniform, and to avoid the perturbations which parasitic charges on the glass walls would produce, a conducting layer which has a high ohmic resistance is evaporated onto the latter. This system is still used in some image-forming tubes in television for tele-cinema use; the elementary pencils of electrons pass through a double scanning system, and fall on the cathode of a photo-multiplier (Farnsworth, 1934). As in Holst's system, the magnification is unity and the image is upright.

Figure 170 shows a variant of the same idea, in which a short shielded lens forms the image of a cathode $C$ on a screen $S$; $C$ is encircled by a guardring $A_1$, and the electrons are accelerated between $C$ and $S$. This arrangement is used in super-emitron television tubes (McGee and Lubzinsky,
1939; Iams, Morton and Zworykin, 1939), in which the electron image is formed directly on the silver oxide mosaic. The anode in this case consists of a metal layer (M) on the outer walls of the tube, the accelerating potential is a few hundred volts, and the magnification is about 5.

In the devices described above, the magnetic field can be produced by permanent magnets, as it does not exceed a hundred gauss, although it does fill a volume of about 500 cm$^3$—even so, the instrument is still heavy and cumbersome, and military requirements have led to the development of a light electrostatic instrument.

14.2.3 The Use of an Electrostatic System

(i) Evolution and technology

The earliest image-converters using an electrostatic electron optical system (see Zworykin and Morton, 1936) consisted of two tubes of the same diameter, each closed at one end by a plane disc—one of the latter supports the cathode C, the other the screen S (see Fig. 171). The slit between the two tubes acts as an accelerating lens, the magnification of which varies between

0.2 and 2 according to the relative lengths of the tubes. It is, in short, a diode system, and lacks flexibility since the focus is fixed once and for all by the geometry; further, in common with all systems with a plane cathode, this system suffers from too much field curvature and distortion. The quality of the image has been considerably improved by Morton and Ramberg (1936) by using a curved cathode; the curvature cannot, however, be increased as far as the electron optics requires, as this would distort the image formed by the preceding glass optical system too acutely at the edges. During the last war, the result of extensive research into this problem was the development of a high quality electron optical system which was purely electrostatic (Morton and Flory, 1946; Schaffernicht, 1948; Eckart, 1954).

![Fig. 171. An electrostatic two-cylinder converter.](image-url)
These devices are all sealed off. Special studies have been devoted to producing fluorescent screens which are capable of very high brightness—zinc sulphide activated with silver has received particular attention—and photosensitive layers with a high quantum efficiency (10 to 20 per cent) which can be used in the same envelope without either lowering the photoelectric sensitivity or reducing the secondary emission coefficient of the screen below unity, which would make its potential unstable. This inconvenience can be avoided by adding a layer of aluminium to the screen (§1.4). A special method for evacuating tubes of this kind is always used. The screen is formed first of all and baked, and the tube is pumped down before the somewhat complex photoelectric layer is introduced; then, without destroying the vacuum, two small evaporators are introduced through two lateral openings in the tube, and with these, first a lining layer of silver or antimony vapour is inserted, and then a photoelectric layer of caesium. These evaporators, which have flexible connexions, are afterwards retracted magnetically into lateral tubes, which are cut off with a blow-lamp. A different procedure has been suggested by Zacharov and Dowden (1960), which prevents, as far as possible, the electrodes and insulators in the tube from becoming polluted by caesium vapour. The cylindrical tube of the converter is divided into two parts by a flat glass disc; after forming the screen in one of the compartments, Sb and Cs are evaporated on the face of the disc which lies in the other compartment. The disc is then turned round in vacuo, so that the photocathode faces the optical system and the screen.

This technique has been highly developed in France, for another purpose, by the astronomers Lallemand and Duchesne, who have built an image-intensifier which is extremely sensitive to visible radiation: this is the electron telescope. In fact, it is an ideal photon receiver, in which the photons fall on a photoelectric layer which, unlike a photographic plate, has no threshold of sensitivity. Each electron which is emitted is accelerated to an energy at which it produces an impression on the photographic plate whereas in the case of very feeble illumination there would have been none.

(ii) An infra-red converter with high-quality optical properties

A modern infra-red converter is illustrated in Fig. 172 (Schaffernicht, 1948). It is a triode system in which the curved cathode $C$ (of radius about 12 cm) and the metallized glass wall $M$ which surrounds it constitute the first electrode, which is held at a potential $V_2$. The first anode $A_1$, which is about 6 cm from $C$ and is held at a potential $V_1$, has an opening of 1 cm; the other anode $A_2$ which lies about 2 cm from $A_1$ and is connected to earth (together with the metallic layer of the fluorescent screen) has a 7 mm opening. The screen is 162 mm from $C$. The electrodes are made of very thin sheets of aluminium. This device is used with a potential $V_1$ of 18 to 20 kV and a potential difference ($V_1 - V_2$) of 7 kV. In all instruments
of this kind, the magnification $G$ (which is always less than unity) varies with different designs between 0.7 and 0.1; this increases the image brightness which is multiplied by $1/G^2$ in any particular case and hence by a factor between 2 and 100. The image is focused by varying $V_1$.

![Diagram](image.png)

Fig. 172. An electrostatic three-electrode infra-red converter.

By using electrodes of more complicated shape, or by increasing their number to six, it has been found possible to make these converters work at potentials higher than 50 kV without increasing the field $E$ at the sensitive layer. Post-acceleration is used in other types.

(iii) An image-intensifier with a very high yield: the "electron telescope"

Image-intensifiers very similar in design to those described above have been built commercially (Linden and Snell, 1957). A fine-mesh grid which is placed parallel and very close to the cathode plays at the same time the role of intermediate electrode and control grid; in its latter capacity, this grid can be used either to cut off the beam at will by applying a negative potential, or to provide pulsed operation. The image is again formed on a fluorescent screen.

Lallemand and Duchesne’s "electron telescope" is quite a different affair. Its structure originally resembled that of a converter very closely, but successive improvements have gradually modified it (Duchesne, 1953a, 1961). Figure 173a represents a cross-section of the optical system of one of the latest models, which has a magnification 1/7. A supplementary electrode has been added, between the cathode and the first anode, and is held at 200 volts; the potential of $A_1$, however is 2000 volts, and the final potential lies between 25 kV and 50 kV. Figure 173b shows the outside appearance of the optical assembly, with the plate-container attached to the base. The whole instrument is surrounded by a glass envelope with high tension leads
and numerous accessories. The cathode, which is all-enveloping, consists of a piece of copper, in the centre of which the metallized sensitive layer, 25 mm in diameter, is to be placed. Each caesium–antimony layer is kept in an extremely thoroughly evacuated thin glass envelope in a side arm until it is required (Lallemand and Duchesne, 1954). When a good vacuum has been attained in the glass envelope of the instrument, and the electrodes and photographic plates have been outgassed, the small glass phial is broken and the layer spread in place, by moving magnetized parts from outside. The plates are cooled in liquid nitrogen to prevent the photocathode from being poisoned by the water vapour which they contain. The vacuum can be improved by using getters, and the most recent sensitive layers, which are themselves cooled to about $-60^\circ$C, have a lifetime which is sufficiently long for the whole instrument to be detached from the pumping unit for several hours (Lallemand, Duchesne and Wlerick, 1960) and

Fig. 173a. Section of the optical system of a recent electron telescope. 1: glass insulator; 2: copper cathode; 3: first anode ($V = 200$ V); 4: second anode ($2000$ V); 5: third anode (20 to 50 kV); 6: high tension leads; 7: cassette of photographic plates.
Fig. 173b. The optical unit of the electron telescope of Lallemand and Duchesne; the cassette of plates is visible in the lower part.
to be placed immediately behind the eyepiece of an astronomical telescope. With an overall operating potential of some 30 or 40 kV, the plates can be used in conditions where their sensitivity is maximum. Figure 173c shows the whole “telescope” which is mounted at the eyepiece of an astronomical telescope.

Fig. 173c. General diagram of the electron telescope installation. 1: (glass) entry window; 2, 3, 4, 5: the optical system; 6: plate-carrying drum (for 6 plates); 7: the magnetic system which controls the rotation of the plate-carrier; 8: photoelectric layer, placed inside its glass envelope 9; 10: joint; 11: cooling system for the plates (immersed in liquid nitrogen); 12: hammer for breaking the envelope 9, moved by means of the piece of steel 13, which is in turn attracted by the mobile magnet 15; 14: electro-magnet to set 15 in action; 16: radioactive carbon (now replaced by a titanium pump). The cathode is connected to a block of copper which is placed in a third lateral arm, and immersed in liquid nitrogen.

(iv) Image-intensifiers with a thin window

The solution to the problem of making the best possible use of the energy of the photoelectrons suggested by Lallemand and Duchesne is evidently by far the most satisfactory, the more so in that the resolving power of the instrument is a maximum in these conditions, as we shall see later. The lifetime of the cathode is limited, however, and an exceptional degree of experimental skill is needed to set up and operate the instrument.
Attempts have been made to produce sealed-off tubes commercially, in which the yield remains very good but the photographic emulsion is outside the evacuated chamber in which the cathode is situated. The electrons fall on a fluorescent screen and the resulting photons make a direct impression on the photographic emulsion without any intermediate glass optical system which would inevitably collect only a small proportion of the light emitted by the screen. If an extreme loss of resolution is to be avoided, however, the fluorescent material and the emulsion must be practically in contact; this explains the need for extremely thin windows, the whole area of which must be of uniform thickness, and which must be vacuum-tight and capable of supporting atmospheric pressure. Mica walls of large diameter (2.5 to 5 cm) have been produced by Zacharov and Dowden (1960), which are capable of withstanding pressures of several atmospheres without fracture. A special arrangement enables us to apply the flexible photographic emulsion to the outer face of the window. According to Zacharov, the final yield of the system is not far from that of an image-intensifier, but the resolution is clearly worse.

Tubes of a slightly different design are used in astronomy (Hall, Ford and Baum, 1960), which contain an even thinner membrane (500 to 1000 Å), whose function is to separate the conversion region from a second less highly evacuated compartment in which the emulsion is situated. When the tube is in use, the emulsion and the membrane are in contact. The electrons are accelerated and pass through the membrane to make an impression on the photographic plate; the resolution is much better than in the preceding case, as the scattering produced by the layer of aluminium is very small.

Tubes of the same type have been built by Hiltner and Pesch (1960), who use thin foils of aluminium or silicon monoxide evaporated on a layer of plastic. This plastic is destroyed by heating to 350°C in vacuo, and the life of the cathode is on the average several weeks. A considerable improvement is obtained if aluminium oxide windows, which are robust and leak-tight, are employed.

14.3 PRINCIPAL CHARACTERISTICS

14.3.1 Sensitivity

The sensitivity of the whole instrument can be defined in terms of the minimum power emitted by a test object \( P_m \) which produces a perceptible image on the screen (or the photographic plate). A visible image is produced as soon as the eye can discern a difference of brightness or darkness between the object and the continuous background illumination, \( B_o \), which is produced by parasitic currents. A photometric measurement of the luminous flux which arrives at the layer, and in terms of which \( P_m \) is defined, is made;
the sensitivity of the system is dependent, above all, upon the sensitivity of the photoelectric layer, and the efficiency of the fluorescent screen or photographic plate.

(i) The sensitive layer

The most sensitive photoelectric layers for making observations in the infra-red consist of caesium on oxidized silver; the maximum sensitivity is somewhere about 0.8 $\mu$A, but acceptable images have been obtained with radiation of a few microns. The sensitivity can reach 20 or 40 $\mu$A lumen$^{-1}$ (or about $10^{-8}$ A watt$^{-1}$ in the near infra-red), with a quantum efficiency of about 3 per cent; in the visible region, the most common choices are layers of caesium–antimony, which are easy to prepare. Their sensitivity can reach 50 $\mu$A lumen$^{-1}$, with a quantum efficiency of 10 per cent while their maximum sensitivity is around $\lambda = 4,200$ Å. For special applications, however, far more sensitive layers are employed.

A layer of SbCsO gives 60 $\mu$A lumen$^{-1}$ towards 4,500 Å, with a quantum efficiency of 45 per cent; in the ultraviolet, the efficiency remains very good, even in the far ultra-violet.

Recently, a layer consisting of antimony and the three alkali metals, sodium, potassium and caesium, has been developed (Sommer, 1955). This layer is capable of giving 150 $\mu$A lumen$^{-1}$, and the efficiency may reach 20 per cent. Its main use is in certain types of television camera (Gibbons, 1960). Once again, the maximum sensitivity is around 4,200 Å.

The sensitivity of the instrument is limited by the parasitic electrons emitted by the layer (the "dark current") which produce an unwanted illumination at the screen and reduce the contrast. One part of this current is due to thermal emission, and obeys the law $I_1 = A T^2 e^{-h/\lambda}$ while the other is due to field emission, $I_2 = C E^2 e^{-d/E}$ (in which $E$ is the electrostatic field). $I_1$ can reach $10^{-12}$ A cm$^{-2}$ at normal temperatures. Because of this possibility, the layer is cooled in liquid nitrogen and the field $E$ is kept very small (at most 1000 V cm$^{-1}$) in order to keep the parasitic currents as low as possible. Another source of parasitic electrons is field emission at the electrodes or the insulating walls. A very thorough examination of this phenomenon, which also produces a continuous background on the fluorescent screen even when no illumination at all is present, has been made by Essig (1960). Several precautions may be taken to reduce it to the maximum:

Strong fields must never be produced at any point of the structure, and bombardment of the insulators by the parasitic electrons produced at the electrodes is to be avoided as far as possible, as this produces very high fields at certain points on the insulators, at which the emission is in consequence reinforced;

Contamination by caesium vapour is to be avoided, as we have already mentioned.
Finally, the tube should, if possible, be cooled in liquid nitrogen, a technique similar to the one used to reduce the background noise of a photomultiplier.

In a commercial tube working at about 10 kV, the darkness current at the screen can be kept down to $10^{-15}$ A cm$^{-2}$, which corresponds to the threshold sensitivity of the eye, but this current increases very rapidly to $5 \times 10^{-13}$ A cm$^{-2}$ at 20 kV (Essig, 1960).

By taking this phenomenon into account, we can explain the way in which the sensitivity of an image-converter varies with the accelerating potential; the minimum luminous flux at which the image is just detectable falls at first as the potential is increased, passes through a minimum and then begins to increase. This increase is due to the reinforcement of the continuous background by the increasing darkness current.

(ii) *The sensitivity of the receiver*

The brightness, $A$, at a screen is proportional to the current density, $j$, which is incident upon it; if $V_0$ is the threshold potential, $B = B_0 j (V - V_0)^\alpha$ with $\alpha \approx 2$ for aluminized screens near $V = 20$ kV. Precise photometric measurements have shown that the brightness is strictly proportional to the illumination of the cathode between 5 and 50 kV for reasonable values of the illumination. Equally, it depends upon the thickness of the screen, and passes through a sharp maximum when the thickness is equal to the depth to which the electrons penetrate (cf. §12.2.6). It seems advantageous, therefore, to increase $V$ to the maximum, but this, as we shall see, reduces the resolving power; an acceptable compromise must be made, taking into account the increasing dark current which we have already mentioned.

When the instrument records directly on a photographic emulsion, we work in the zone of maximum sensitivity of the plate, which corresponds to a potential $V$ of 30 or 40 kV for the best emulsions. The sensitivity of modern emulsions has been examined both at normal temperatures and at the temperature of liquid nitrogen by Duchesne (1953b). We are faced with two very different problems; either we try to obtain the image of a star with a considerably shorter exposure than would be needed for direct photography, or else we should like to record on the plate as much information as possible about those regions of the sky of which the brightness is very feeble, and of which we are largely ignorant. In the first case, the research is directed towards finding very fast emulsions, which have at the same time a sufficiently fine grain for the resolving power not to be diminished—at present, Crumièr Spécial D or Ilford G 5 give the best results. The exposure can be reduced by 50 or 100 times. Figure 174:1 is a photograph of Saturn, taken with an exposure time of half a second.

In the second case, we try to reduce to the minimum the continuous background of the plate so that every possible useful detail appears, even if it is very poorly contrasted. If very fine grain nuclear plates are used
Fig. 174:1. Photograph of Saturn, taken with an "electron telescope" on an Ilford K2 plate with an exposure of 1/2 sec (Photo: Wlérick, Rösch and Dupré).

Fig. 174:2. The recording of a very low luminous flux by counting the traces of the photoelectrons. (a) Microphotometry of a grating, obtained by ordinary photography on a Kodak Rochester 103—0 plate; the period of the grating is 200 μ, the exposure time, 100 sec. (b) Microphotometry of the same grating, obtained by electron photography on an Ilford G5 plate; electron magnification: 98; exposure: 1 sec. (c) Electron photography, as in (b), but with an exposure of 1/10 sec; the grating is still distinguishable against the continuous background. (d) Detection by counting the traces corresponding to (c). The ordinate of each horizontal line is equal to the number of traces within a rectangle of 26 × 170 μ² (its longer sides parallel to the lines of the grating).
(Ilford G5 or Kodak NT2A, for example), each electron gives a perfectly recognizable trace provided it has been sufficiently accelerated. If the number of photo-electrons is not too high, it is then possible to detect the image which is formed by an extremely small light flux, by counting the number of electron traces. This method is limited by the “veiling” (produced by silver grains which are developed even outside the traces) and by the electron “background noise” (which is produced by parasitic electrons emitted by the photocathode). By using recently prepared emulsions and short exposure-times to limit these disturbing effects, the method proves to be most effective: Lallemand, Duchesne et al. (1959) have found a gain of $10^4$ over the direct recording of the photon flux on the plate, as Fig. 174:2 displays. (In each case, the most sensitive type of plate has been used, either to photons or to electrons as the case may be.)

14.3.2 Resolving Power

The resolving power is restricted by a number of factors:

- the defects of the optical system;
- the grain size of the screen (or plate);
- the reduction of contrast which can occur when the phosphor is saturated by high local brilliance; with mobile objects, the contrast may be reduced by the “remanence” of the screen.

We shall examine only the first two of these possible factors.

(i) The resolving power of the optical system

The resolving power in the neighbourhood of the axis is calculated from the formula $d = k \frac{\Phi_e}{E_0}$ where $k \approx 1$, and $d$ is the diameter of the aberration disc, just as in all emission systems (cf. § 13.2.2). For a caesium cathode, used in the infra-red, we can take $\Phi_e$ to be 0.5 V; for $E_0 = 500$ V cm$^{-1}$, we should have $d = 10 \mu$. Off the axis, $d$ can be as large as 100 or 150 microns even when a curved cathode is used, as there is an effect due to residual field curvature and to the aberrations of the rays which are far from the axis. The average quality of the image can be improved by choosing the focus to be a certain distance from the centre (Fig. 175a). $d$ can be reduced by increasing $E$, but we have to make a compromise between the mean resolving power which is wanted and the “darkness current” of the layer.

In converters with a homogeneous magnetic field, the resolving power is given by a formula of the type

$$\delta = k \frac{\Phi_0}{E} = k \frac{\Phi_0}{\Phi} L,$$

in which $L$ is the length of the tube and $\Phi$ is the accelerating potential; the values of $k$ quoted by different authors lie between 1.6 and 2, if we select for $\Phi_0$ the potential which corresponds to photoelectrons of the maximum
energy (for an Sb–Cs cathode, this is 2.3 V). The measured resolving power is some 10 μ for a potential of 5 kV (Wendt, 1955), over the whole of the screen.

![Graph](image)

**Fig. 175.** (a) Resolving power δ, as a function of the radial distance R from the axis. (1) image focused on the axis; (2) focus at 7 mm from the centre.

(b) Semi-conductive mirror converter (Schaffernicht).

- C: semi-conductive layer.
- S: transparent fluorescent screen.
- K: electron source.
- O: anode.
- A: mirror.
- M: magnifying glass.

(ii) The resolving power of the receiver

Even in the most favourable cases, the grain of a screen obtained by sedimentation is still about one or two microns in diameter, so that the resolving power of screens of this kind cannot be less than 5 or 10 μ; the thickness, however, is equally relevant—it has been estimated that the resolving power is effectively equal to the depth to which the electrons penetrate into the layer, and hence to the thickness of the screen under the conditions of maximum brightness. At 20 kV, this depth is some 10 μ, while at 80 kV it is 100 μ. The resolving powers which are found at the screens of the infrared converters described earlier are about 50 μ at the centre. This result can be improved by a factor of 3 or 4 by using screens of evaporated zinc oxide (Studer, Cusano and Young, 1951) and with these the resolution can reach 15 μ (the distances are to be understood as having been projected back into the object plane of the layer, taking into account the fact that the magnification is less than unity). It is clear, then, that at the centre it is the grain size of the screen which governs the resolving power of the instrument, while at the edges the latter is limited by the aberrations. In laboratory instruments, however, a curved screen can be provided, so that
the resolving power is virtually constant over a region of about 30 mm diameter.

In recent instruments, where the electrons record directly on a photographic plate after passing through a thin window, the resolution is worse than that of the optical system as we must take the scattering produced by the window into account; a value of 30 Å can be reached, however, which is adequate in many applications.

If the image which is recorded by the emulsion is provided by the photons emitted by a screen placed against the thin window, the resolution is of the same order of magnitude (30 μ on the axis).

In the "electron telescope", where all these perturbing influences are removed, the resolving power is the same as that of the optical system itself, about 10 μ at the centre.

14.4 THE DEVELOPMENT AND THE FUTURE OF THESE INSTRUMENTS

14.4.1 Improvement of the Sensitivity

(i) Converters using secondary emission

A new type of image-intensifier tube was proposed by Sternglass in 1953, in which the secondary emission phenomenon was employed in cascade. The tube is an image-converter tube with homogeneous electric and magnetic fields (see Fig. 169), in which thin, planar, regularly spaced dynodes are placed perpendicular to the axis. Each dynode consists of a very thin metal foil (facing the incident electrons) on which a layer of an insulator with a very high secondary emission coefficient is deposited. The photoelectrons are accelerated towards the first dynode, and on passing through the metal and the insulator, produce secondary electrons which are in turn accelerated towards the second dynode; the same cycle recurs in cascade, with an amplification of the current at each stage. Finally, the electrons which emerge from the last dynode fall onto a fluorescent screen.

Considerably improved versions of this type of instrument have been described by Wachtel, Doughty and Anderson (1960), who have developed a technique for the very difficult task of making the dynodes; it has been possible to obtain secondary emission coefficients of 5 with films of potassium chloride, 500 Å thick, deposited on aluminium, and an accelerating potential of 4 or 5 kV. Layers of barium fluoride deposited on aluminium give the same yield, and have a longer life. With tubes of this kind, we can hope to obtain a resolution of 40 or 50 μ, with six dynodes 15 mm apart,
a magnetic field of 400 gauss, inter-dynode potentials of 2·5 or 3 kV, and a potential of 10 kV between the last dynode and the screen. In this way, electron gains of 1500 have been obtained; gains in brilliance of $10^4$ have been measured (Anderson, 1960), using a layer with a sensitivity of $15 \mu A \text{lumen}^{-1}$, an inter-dynode potential of 4 kV, and a potential drop of 15 kV between the last dynode and the fluorescent screen (zinc sulphide activated with silver). The limiting resolution is about $40 \mu$, inside a circle 25 mm in diameter.

Gains of about $10^5$ have been measured in blue light (with a photocathode and a screen for which the spectral response is virtually identical) by Wilcock, Emberson and Weekley (1960), using a tube containing five dynodes, supporting a total potential drop of 35 kV between the cathode and the screen. Once again, the resolution is about $40 \mu$.

(ii) Converters in cascade

An alternative suggestion for increasing the gain is to couple several converters together in cascade. The best solution, due to Schaffernicht (1948) consists of setting the screen of the first converter close to the photosensitive layer of the second, to reduce the loss of resolution as much as possible. Practical versions of two- and three-stage tubes have been described by Stoudenheimer (1960). Each stage is a converter with a purely electrostatic electron optical system (see Fig. 175b) and unit magnification.

In the two-stage tubes made by RCA, the first photocathode is a layer of Sb–Na–K–Cs and the second, of Sb–Cs. The maximum sensitivity of both of these two layers is in the blue region, around 5000 Å and their characteristics satisfactorily match those of a screen of zinc sulphide activated with silver. The two stages are separated by a “sandwich”, which consists of a thin glass sheet (about 0·1 mm thick) on one side of which is the screen and on the other, the Sb–Cs cathode. Although less sensitive, this type of cathode is preferred to the “three-alkali” cathode, being easier to prepare. More recently, three-stage tubes of a similar design have been produced (Stoudenheimer, Moor and Palmer, 1960) with a potential of 10 kV at the first stage and 20 kV on each of the two succeeding stages. At the maximum sensitivity, which occurs at about 4200 Å, the radiant-flux conversion gain—the ratio of the number of photons which emerge to the number of photons which are incident upon the photocathode—is given by

$$G_{th} = S_1 S_2 S_3 V_1 V_2 V_3 E_1 E_2 E_3 K^2$$

in which $S_1$, $S_2$ and $S_3$ denote the sensitivities of the three layers respectively, in amp watt$^{-1}$, $V_1$, $V_2$ and $V_3$ are the accelerating potentials, $E_1$, $E_2$ and $E_3$ are the efficiencies of the three screens and $K$ is an adaptation factor, with the aid of which the relative sensitivities of the adjacent photocathodes and screens can be taken into account.
With
\[
\begin{align*}
S_1 &= S_2 = S_3 = 0.06 \text{ A watt}^{-1}, \\
V_1 &= 10,000 \text{ V}, \quad V_2 = V_3 = 20,000 \text{ V}, \\
E_1 &= E_2 = E_3 = 0.1, \\
K &= 0.9,
\end{align*}
\]
we find \( G_{th} = 7 \times 10^5 \). In practice, gains of \( 10^8 \) have been obtained with some experimental tubes.

The resolution of the image-converter is of the order of 40 \( \mu \) in the neighbourhood of the axis, and falls to 60 \( \mu \), 10 mm away from the axis. This decrease is due to the field curvature of the imaging field in the second stage. To counteract this, a photocathode curved in the opposite sense to the image which is formed upon it by the first optical system would have to be used, but this would worsen the aberrations of the first stage: we are forced to compromise; a further loss of resolution arises from the passage of the photons through the support of the intermediate screen.

(iii) The channelled image intensifier

For applications in which an enormous gain in brilliance is required rather than a very high resolution, special amplifiers have been developed in several laboratories, in which the optical system is wholly different from those described above. The cathode is divided into small units, and the electrons emitted by each unit are accelerated towards a fluorescent screen along channels formed by the successive dynodes which amplify the current by secondary emission exactly as in the photomultiplier. The resolution is clearly fixed by the transverse dimensions of the individual channels, and depends neither upon the number of stages nor upon the velocity distribution of the electrons which emerge from the cathode. The dynodes, which are held at progressively higher potentials, are placed one after the other in such a way that all the secondary electrons which are emitted by

Fig. 176 a. Channelled image-intensifier (McGee). (I) The "venetian blind" type. (II) The model with connecting cylindrical tubes.
any one dynode subsequently fall onto its successor. One possible structure, which has been proposed by McGee and his collaborators (1960), is shown in Fig. 176a, and another due to Burns and Neumann (1960a, b) in Fig. 176b; some typical trajectories have been included.

![Diagram of electron optics](image)

Fig. 176 b. Another type of channelled image-intensifier with successive dynodes (Wachtel, Doughty and Anderson). Some typical electron trajectories are shown. The period of the structure is $\frac{19}{1000}$ of an inch, which explains the considerable technological difficulties encountered in the construction.

The dynode surfaces are composed of evaporated layers of an Ag–Mg alloy, which has a secondary emission coefficient of 5 or 6 for an accelerating potential of 200 V. Work is at present in progress on large scale models, with the object of measuring the electron gain of the different channels. The hardest problem which has still to be solved seems to be the insulation between the stages; poor insulation produces micro-discharges, which result in an intense continuous background at the final screen.

(iv) The use of feedback in the image-intensifier

A great deal of nuclear research—in the domain of plastic fibre scintillators, in particular—requires high gain amplifiers. We should like to be able to detect the passage of ionizing particles which produce only a very few photons in their passage through a unit of the counter. If we define the quantum efficiency, $R$, of a converter as the ratio of the number of photons emitted to the number of photons incident, present-day single-stage tubes can provide

$$20 < R < 50,$$

whereas the amplification which is required here is of the order of

$$10^5 < R < 10^8.$$
Some authors have suggested that the image formed on the screen should be fed back to the cathode by an appropriate glass optical system in such a way that this image coincides with the original image. The system must use pulsed operation, so that the final gain in the brilliance can be controlled. A discussion of this problem and of the various different designs which have been proposed is to be found in an article by Perl and Jones (1960).

(v) The use of "memory" tubes

The optical image may be formed upon a photosensitive cathode, which produces an electron image on a layer of insulator with a high secondary emission coefficient; alternatively, it may be formed, either directly or after amplification by an image-converter provided with a fluorescent screen, on a photoconducting layer. In the first case, the electrons, and in the second, the photons, lead to the appearance of a "potential image" at the surface of the receptive layers. This image is then scanned by a beam of slow electrons, which erases the differences of potential across the layer; an electrical signal which is proportional to the potential difference at any point is obtained on a special electrode which is in contact with one face of the layer; this signal is subsequently transmitted to an ordinary television tube, whose scanning is synchronized with that of the beam which "reads" the image. The beam in the television tube is modulated by the signal, and on the screen of this tube we obtain the final image, improved, however, by a marked increase in brilliance.

This class of image-converters closely resembles tubes of the type which are used as television cameras, and further, allows us to benefit from the long integration time of the sensitive receptive layer, before actually reading the potential image (Beurle and Slark, 1960; Heimann, 1960).

14.4.2 Application to Microscopy

In ordinary light microscopy, converters can be used to examine the infrared and ultraviolet. In electron microscopy, the resolution in the image plane is often no better than 50 μ, for a magnification of 50,000 (which corresponds to 100 Å at the object) and amplifiers can be useful to reinforce the image so that beams of very weak intensity can be used. It is in ion microscopy, where the lack of brightness at the screen is the main difficulty, that an image amplifier would be most valuable.

One solution would be to replace the screen of the microscope with a semi-transparent fluorescent screen painted onto the glass wall of the converter, on the other side of which would be the light-sensitive layer. This layer must be kept sealed off in its own envelope, as the vacuum in the microscope is inadequate to ensure that it has a reasonable lifetime. With such an arrangement, however, the resolution of the whole device will be
considerably reduced by the diffusion of light in the glass wall of the converter. The thickness of the glass between the layers cannot be reduced below a few tenths of a millimetre, as it has to be capable of supporting atmospheric pressure; on the light-sensitive layer, it will be an emissive spot, a few tenths of a millimetre in diameter, which corresponds to a point on the electron or ion image.

Three ways of reinforcing the image without loss of resolution have been tried.

The electron (or ion) image which is obtained on the screen of the microscope is recorded with a very sensitive television camera, which transmits it, after suitable amplification, to the screen of a television receiver. With the aid of special circuitry, the contrast can be considerably reinforced by this procedure, and any troublesome continuous background can be suppressed.

An image-converter using secondary emission and deprived of its photosensitive layer can be employed. This procedure was suggested by Möllenstein (1953) and Le Poole (1954), and has recently been tried out experimentally by Castaing, Jouffrey and Slodzian (1960), as a means of reinforcing ion images.

The principle is as follows—after passing through the objective \( L_1 \) at 40 kV, the ion beam is slowed down to 10 kV, and the first low-magnification image \( (G_1 \approx 100) \) is formed on a perfectly polished metal plate (stainless steel, for example), which acts as the cathode of an immersion objective. The secondary electrons which are emitted are accelerated in the opposite direction and are focused by the immersion objective on a fluorescent screen (see Fig. 177), through which the primary ions have been able to pass by means of a small hole pierced close to the focus of \( L_1 \); the final magnification can reach \( G_2 \approx 10,000 \). If we suppose the secondary emission coefficient equal to unity, the gain in brightness between a direct ion image

---

**Fig. 177.** Diagram of an image-intensifier for ion microscopy.  
\( L_1 \): electrostatic objective.  
\( L_2 \): immersion objective.  
\( AB \): ion image.  
\( A'B' \): secondary electron image.  
\( C \): polished metal plate.  
\( S \): screen.
with $G = 10,000$, and the image obtained in this way with the same magnification is very much greater than unity, as the plates are considerably more sensitive to electrons than they are to ions. The first attempts in this direction have led to gains of the order of 100, with an ion microscope working with Li$_7^+$. Further, the cathode of the immersion objective can be made from a metal which has a very high secondary emission coefficient when bombarded with ions.

Finally, a new device for intensifying the image has recently been built by Haine, Ennos and Einstein (1960). This device makes use of the conductivity which is induced in a thin film of selenium by fast electrons (50 kV).

![Electronics diagram](image)

**Fig. 178. Image-converter with a semi-conductor element, for electron microscopy (Haine).**

Figure 178 shows the principle: an amorphous layer of selenium is deposited onto a metallized plastic layer which acts as signal electrode; the latter is connected both to a source of polarization and to an electronic amplifier which is in turn connected to the modulation electrode of an ordinary television tube. The local conductivity of the selenium layer varies with the density of the incident beam, although only in the direction normal to the layer. These variations are "read" by a second electron beam, this time of very slow electrons, which scans the lower surface of the selenium layer. In the absence of any "illumination", the potential of the lower surface is uniform and equal to the potential of the cathode of the "reading" gun, by virtue of the phenomenon of secondary emission with a coefficient $d$ less than unity. Between two sweeps of the beam, when the layer is "lit" by the 50 kV electrons, the charges on the lower face flow towards the polarized electrode, and create local variations of the surface potential. The "reading" beam provides charges which restore a uniform potential, which results in a variation of the equivalent potential at the signal electrode, an effect due to the capacitance of the selenium layer. This variation, after suitable amplification, is used to modulate the intensity of the electron gun of a television tube.

The resolution in the plane of the electron image is about 80 $\mu$, but this can be improved to 30 $\mu$, which is in practice adequate for high magnification (for $G = 10^8$, we should have a resolution of 3 Å at the object).
Electron densities of $10^{-12}$ A cm$^{-2}$ at 50 kV, for which the image would be barely perceptible even on a very good fluorescent screen, produce acceptable images on a television tube with this procedure.

14.5 AMPLIFICATION OF THE BRILLIANCE OF RADIOSCOPIC IMAGES

In one recent design, the usual arrangement is used, but the screen which is fluorescent to X-rays is placed within an evacuated envelope on a very thin wafer of glass on the opposite side of which is the light sensitive layer. In another instrument (Eckart, 1954), an insulating layer only a few tenths of a micron thick separates the screen from the layer.

![Diagram of an image-intensifier for radioscopic images.](image)

**Fig. 179.** Diagram of an image-intensifier for radioscopic images.

- $A_1, A_2$: electrodes.
- $S$: fluorescent screen.
- $C_1$: fluorescent layer, sensitive to X-rays.
- $T$: electron trajectory.
- $M$: microscope for examining the image.

In connection with this, we must also mention the design of Teves and Tol (1952; Fig. 179), which has the following characteristics—the operating potential is 25 kV, the electron magnification is 1/9, and the gain in brilliance is about 1000. A small microscope is attached to the screen and provides a final magnification of unity. The surface examined, however, is some 25 cm$^2$ in area, and only technological problems at present prevent this area from being doubled or tripled.

The compactness of this instrument allows us to use it in conjunction with a television camera to retransmit radioscopic images which are formed in a region which is dangerous for an observer (physico-chemical treatment
of radio-active bodies, for example, can be examined from a distance). Tubes of this type are now produced commercially by several firms.

A slightly different model has been constructed by Timofeev and Sorokina (1960). Inside a vacuum chamber of the type illustrated in Fig. 179, the convex portion is covered with a thin film of aluminium, transparent to X-rays, on which the fluorescent material is deposited (a few tenths of a millimetre to obtain the maximum brightness).

The caesium–antimony photosensitive layer is evaporated directly on this material, facing the anode. The grid is connected to the cathode; the anode is a tube with an entry stop, behind which the fluorescent screen is placed. The shape of the electrodes is such that the potential distribution \( \Phi(r, z) \) is of the form

\[
\Phi(r, z) = \Phi_0(2z^2 - r^2 + 2pz).
\]

In the useful region, the equipotential surfaces are hyperboloids of revolution; the trajectories perform one semi-oscillation in a radial plane between the cathode and the final screen. The magnification depends upon the model, and is either 1/6 or 1/9. The resolution is limited above all by the thickness of the fluorescent material at which the primary image produced by the X-rays is formed; in the cathode plane, the resolution is of the order of 0.1 or 0.2 mm. The increase of brightness between the primary image and the final image produced by the electrons is of the order of 1000 for \( \Phi = 25 \text{ kV} \) and \( G = 1/6 \). In the model for which \( G = 1/9 \), gains of 3000 have been attained. In some models, the useful area of this cathode is 10 cm in diameter.
CHAPTER 15

THE ELECTRON MICROSCOPE; STRUCTURE AND SPECIAL FEATURES

15.1 THE NOTIONS OF RESOLVING POWER AND USEFUL MAGNIFICATION

The electron microscope provides a highly magnified image of a small object which is being irradiated with electrons. The importance of this type of microscope lies in the fact that by using electrons, the effective magnification \( G \) can be 100,000 times or more, whereas with light, it is only with difficulty that \( G \) can be increased beyond 1500 times. Simple though this statement is, it is easy neither to analyse this fundamental property in detail, nor to give a precise definition of it, as we shall see in Chapter 16.

![Diagram](image)

Fig. 180. (a) The form of the curve which gives the intensity at the image spot as a function of the distance from the centre. (b) The limiting position for the images of two points which can be "separated". The resultant intensity is represented by the dotted curve.

In preparation for this final goal, we shall tabulate the various factors which affect the magnification. In this way, we shall obtain criteria from which we can judge the essential quality of the instrument, understand the stages in the historical evolution of its construction, and compare the different variants which exist at present.
Like all optical instruments, the microscope has intrinsic imperfections. The image of a point is not a point but a small spot in which the brightness decreases outwards from the centre (Fig. 180a) according to a law which depends on the operating conditions. When diffraction phenomena preponderate, a set of small concentric rings is produced about the central maximum, with progressively decreasing intensities. As it is only the central peak which is of interest, however, the distribution of (luminous or electron) intensity can in general be represented to a reasonable approximation by a bell-shaped curve; often, the Gaussian error function is selected to describe this curve analytically, when a definite model is required for a calculation.

The limit of resolution, or simply "resolution", corresponds to the minimum distance between two maxima of the light distribution (or image points), at which the two can still just be separated. This distance is always referred back to object space, and is, therefore, the quotient of the real distance observed at the image and the magnification of the latter. The expression "resolving power" (pouvoir séparateur) which is frequently used for this minimum distance is not a particularly fortunate choice, as the resolving power increases when this minimum distance decreases. Let us consider the images of two bright points very close together (Fig. 180b) for which there is no possibility of separating the respective luminous intensity curve. The curve which represents the resultant intensity at each point has a slight depression which becomes less and less perceptible as the two points are brought closer and closer together. We decide arbitrarily to call two images "resolved" (or separately discernible) if the distance between the two maxima is equal to the width at the point at which the two bell-shaped curves have fallen to half their maximum height (the width, that is, above which the intensity is more than half the maximum intensity). This clearly does not exclude the possibility that the minimum might still not be obscured when the two curves are slightly closer together. The relative depth of the minimum does, however, decrease very rapidly as the two images approach each other, and in practice it is very difficult to resolve two images which are separated by a distance which is less than their half-width. This, therefore, is an excellent convention—it is almost equivalent to the more well-known definition which is in general the one which is adopted in optics when only diffraction is present (cf. Bruhat and Kastler, Optique, Masson, Paris, pp. 233 and 239).

The image formed by an (optical or electron) microscope has a certain direct magnification. Whether the image can or cannot be further magnified depends upon its quality; this additional magnification would be effected with a photographic enlarger, and would give the image a useful magnification which is determined by the resolving powers both of the instrument and of the human eye—this magnification has to be adequate to magnify the smallest details in the object to a size at which the eye can discern them.
If $\theta_{\text{eye}}$ and $\theta$ are the resolutions of the eye and the microscope respectively, we have

$$G_u = \frac{\theta_{\text{eye}}}{\theta}$$

($\theta_{\text{eye}}$ corresponds to an object placed at the least distance for distinct vision, about 20 cm from the eye). There are many cases, however, in which it is convenient to work with double the magnification which we have just defined. This magnification is adopted to avoid straining the eye by using it continually at the limit of its capability—in this way, the final image can be scrutinized without fatigue.

In the majority of research with the microscope, the image is first of all directly recorded in the apparatus, and this first negative is then magnified by the usual photographic procedure which produces the final positive print; it is this overall final magnification which has to be at least equal to $G_u$. The limit of resolution with the naked eye is $\theta_{\text{eye}} \approx 0.1$ mm, which gives the following useful magnifications for various types of microscope:

<table>
<thead>
<tr>
<th>Type of microscope</th>
<th>$\theta$</th>
<th>$G_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical</td>
<td>0.2 $\mu$</td>
<td>500</td>
</tr>
<tr>
<td>Ultra-violet</td>
<td>0.12 $\mu$</td>
<td>800</td>
</tr>
<tr>
<td>Electrostatic</td>
<td>1 m$\mu = 10$ Å</td>
<td>100,000</td>
</tr>
<tr>
<td>Magnetic</td>
<td>0.5 m$\mu = 5$ Å</td>
<td>200,000</td>
</tr>
</tbody>
</table>

Using a magnification higher than $G_u$ would not reveal any further detail; two points which are just resolved with a magnification of 100,000 will simply be further apart if the magnification is increased, but no fresh details will appear between them. The image will be larger but more blurred.

15.2 A COMPARISON OF THE ELECTRON AND OPTICAL MICROSCOPES

The use of glass objectives with wider and wider apertures, which are continually being more satisfactorily corrected, might suggest that the increase of the magnification would lead to an unlimited increase in the resolving power of instruments of this type. It is soon clear, however, that however perfect the instrument may be as far as geometrical optics is concerned, it is never capable of forming distinct images of two points on the object which are very close together; biological specimens are rich enough in details for this to be verified by simple observation.

It was Abbe who, more than eighty years ago, gave the first explanation of the origin of this limit, and provided a quantitative theory which was
soon to be confirmed independently by the original work of Lord Rayleigh. He showed that the limit of resolution of the microscopes in use at the time was restricted not by the degree of optical or mechanical perfection of the instrument but by the phenomenon of diffraction which is inherent in the nature of light, and which is due to its wave-like character; the wavelength of the light determines the order of magnitude of the smallest detail which will be visible. No objective, however perfect it may be from the point of view of geometrical optics, can resolve details which are separated by less than about half the wavelength of the radiation in the medium surrounding the object. The results, therefore, ought to be improved in proportion as the frequency of the radiation is higher and the refractive index of the medium surrounding the object is larger.

For visible light, the wavelength $\lambda$ of the colour to which the eye is most sensitive, yellow, is about $0.5\mu$. The liquids in which the object is immersed have a refractive index which does not exceed 1.6. This shows that the limiting resolving power of the optical microscope is in the region of $0.2\mu$, as we have already stated. This value had been attained before the end of the nineteenth century, and the only improvement which has been made since, before the advent of electron microscopy, is the use of ultraviolet light and the optical properties of quartz, which reduces the wavelength $\lambda$ to about $0.25\mu$ and results in a resolving power of about $0.12\mu$.

It is not possible to use electromagnetic waves of shorter wavelength, as no refringent media exist with which the rays could be suitably refracted, and the latter would moreover be difficult to produce. Kirkpatrick and Baez (1950) have tried to use X-rays, and the resulting technique is interesting, but no real microscopes have as yet been built successfully, and magnifications exceeding 100 have not been attained.

The work of L. de Broglie has established that a wave is associated with every moving particle, and that both the frequency and the wavelength which characterize this wave depend upon the energy of the particle. The wavelength $\lambda$ which he attributed to the waves associated with the particle is given by

$$\lambda = \frac{h}{p}, \quad p = \frac{mv}{\sqrt{1 - (\frac{v}{c})^2}}.$$ 

If the velocity is low, $(v/c)^2$ can be neglected in comparison with unity, and we obtain

$$\lambda = \frac{h}{mv}.$$ 

If the particles are electrons with charge $e$, accelerated through a potential drop $V$, we have $eV = \frac{1}{2}mv^2$, so that

$$\lambda (\text{Ångströms}) = \frac{12.25}{\sqrt{V (\text{Volts})}}, \quad \text{(and } 10\text{ Å} = 1\text{ mμ).}$$
In Fig. 181, the variation of the wavelength of the waves associated with electrons and with various ions is shown as a function of the accelerating potential. Electrons accelerated through a potential of 60,000 V, for example, have a wavelength of 0.05 Å, which is 100,000 times shorter than the wavelength of blue light. Electron radiation is immediately advantageous in microscopy, therefore, as its wavelength is very short and the diffraction phenomena which are proportional to $\lambda$ are extremely small. A very high resolving power is to be expected with an electron instrument.

![Graph](image)

**Fig. 181.** The wavelength of the associated wave, plotted as a function of the accelerating potential for various ions and for electrons.

At first sight, an electron microscope ought to be much more "magnifying" than a light microscope. If this advantage is to be achieved, we must first find a way of building electron lenses as perfect as the corresponding glass lenses as far as the correction of aberrations is concerned; this, unfortunately, is an extremely difficult problem to solve (cf. Chapter 7), and at present, the geometrical qualities of glass lenses have not been equalled with electron optical lenses. We are reduced, therefore, to using beams
with very small numerical apertures in the electron microscope, of the order of $10^{-3}$ (instead of unity, which is often attained in optics). This small aperture produces, amongst other defects, diffraction effects which vary inversely with the aperture. The possibilities of improving the resolving power which are implicit in the shortness of the wavelength of the waves which are associated with electrons are far from being fully realized. The quality of electron optical images is only 100 times better than that of the best light image, as a result of the defects of electron lenses.

In the laboratory, under ideal conditions, a resolving power of 6 or 8 Å has recently been obtained, but only by choosing a particularly favourable type of object. This is a record, and normally, 15 or 20 Å is reckoned to be highly satisfactory. In the large majority of applications, however, a resolution of 50 or 100 Å is sufficient, which corresponds to a useful magnification between 20,000 and 40,000. Most of the microscopes available commercially are of high enough quality to enable the average experimental scientist to obtain photographs of this quality habitually.

15.3 THE GENERAL PROPERTIES OF THE INSTRUMENT

To each of the two types of electron lens—the electrostatic lens and the magnetic lens—corresponds a different type of microscope; these two types have retained, despite the similarity between their structures and their quality, a pronounced individuality. In France, the microscopes which have been constructed have, until recently when a magnetic microscope has appeared, all been electrostatic; in England and the United States, however, magnetic microscopes have been built almost exclusively. In Germany, on the other hand, the two versions have never ceased to be developed side by side since the inception of the subject, and their progress has been parallel. In conclusion, we should mention a Swiss microscope in which a combination of both kinds of lens is used, and point out the fact that Russia and Japan have mass-produced both types since the war.

15.3.1 The Analogy with the Structure of the Light Microscope

Although the details of the construction differ considerably from one instrument to another, it is quite clear that in all the variants the structure of the electron microscope is analogous to that of the light microscope; this is illustrated in Fig. 182, in which the paths of the rays and the positions of the images are shown in both types of instrument.

The “projection” light microscope comprises a condenser which concentrates the light from a powerful lamp onto the object to illuminate it, an objective which forms a highly magnified image of the object, and
a projective lens which magnifies this first image a second time and projects it onto either a photographic plate or a screen.

The same arrangement is to be found in the electron microscope. The object is no longer illuminated by light, but by a beam of electrons which emerges from a “gun”, similar to the gun of a cathode ray oscillograph. This electron beam is directed onto the preparation, either directly or after passing through a magnetic or electrostatic condenser which plays the same role as the “second lens” in a cathode ray oscillograph. The electrons then pass through an objective, a projective lens, and finally form a magnified image either upon a screen for direct observation or upon a photographic plate or film. The only complication, which will be new to the person used to an optical microscope, is that the whole length of the instrument has to be evacuated (to about $10^{-4}$ to $10^{-5}$ mm Hg).

The focusing, which was for a long time obtained by moving the object about within the vacuum, is nowadays produced by modifying the focal length of the objective electrically.

**Fig. 182.** Diagrams showing the points of comparison between an electron microscope and a projection glass-optical microscope.
15.3.2 Depth of Field

This similarity of structure does not prevent wide disparities from appearing between the characteristic magnitudes in the optical microscope and its electron optical junior; these stem from the fundamental difference between the numerical apertures of the two objectives, unity for a glass lens, $10^{-8}$ for an electron lens.

![Diagram](image)

**Fig. 183.** The depth of field (to simplify the drawing, $d$ and $D$ have been drawn on a larger scale than the rest of the diagram).

We consider a lens (see Fig. 183) which forms the image of a point $B$ on a screen at the point $B'$; another point $A$, distance $D$ from $B$, will be imaged at a point $A'$, and on the screen we shall see a blurred circle of diameter $d'$. We refer the length of this diameter back to the original object plane by considering the circle of diameter $d'$ as the image of a circle of diameter $d$ lying in the plane of $B$. We shall say that the image of $A$ is "well focused" on the screen provided $d'$ is less than the "width" of the diffraction disc which surrounds the image point $B'$, or referring back to the object plane, $d \leq \rho$ (in which $\rho$ is the resolution). As the distance $D$ is small, together with the angle $\alpha$ in the case of the electron microscope, we find $\frac{d}{2} = \alpha D$. A glance at Fig. 183 shows that there is another point $C$, virtually symmetrical to $A$ with respect to $B$, which is also "well focused" onto the screen; the depth of field, therefore, is given by $p = 2D$, and so

$$p = \frac{d}{\alpha} \approx \frac{\rho}{\alpha}.$$ 

In normal working conditions, $\alpha \approx 10^{-8}$ and $\rho = 30$ Å, we find that the depth of field is close to $3 \mu$. If a poorer resolution is acceptable, however, the depth of field is considerably greater, and can be as high as a fraction of a millimetre when the resolution is the same as that of an optical microscope. This is an extremely large value, as can be seen by comparing it with the depth of field of an optical microscope which is in the neighbourhood of $0.2 \mu$, and hence of the same order of magnitude as the limit of reso-
lution;† in the electron microscope, the depth of field is 1000 times higher than the limit of resolution. This depth of field is enormous, also, if we compare it with the "field of view", the portion of the object which is seen at the image—for a direct magnification of 20,000, the latter is in the region of 5 μ.

It should be realized, however, that to be able to increase the depth of field is not always a particularly important advantage in electron microscopy. Indeed, the restricted depth of field of the optical microscope is often valuable, since for thick objects with very tiny details (which are often encountered in biology), different cross-sections of the object can be examined by varying the focal length; the image which is obtained, therefore, represents a section which is thinner than the original specimen, while the details of the other parallel sections only appear diffusely at the image. We can, therefore, build up the structure of the image in the third dimension by examining a number of different cross-sections. In electron microscopy, such a procedure is impossible; the whole object—which has of course to be very thin in order to be transparent to electrons—comes into focus simultaneously as soon as any particular detail is in focus. It is most difficult, especially with thick biological sections, to localize the relative positions of the various details in space.

The large depth of field of the electron microscope is, from another point of view, an advantage, since focusing, which would be an extremely critical operation if the aperture of the instrument were not so small, is made considerably easier; electron microscopists are already faced with enough difficulties in focusing correctly, which are due to numerous secondary causes which will be discussed later.

15.3.3 Stereographic Reconstruction

Another consequence of the large depth of field is the possibility of obtaining a "stereographic reconstruction" of the object—this entails taking two photographs of the same object after tilting the object plane 5 or 10 degrees each way from the axis. If these two images are examined through an ordinary stereoscope, the effect of stereoscopic relief is striking. Alternatively, the twin photographs can be projected onto a single screen using two projecting beams polarized at right angles, and the composite projected image examined with polarizing spectacles which filter off the component which is needed to reconstitute the relief for each eye.

† In a glass optical microscope, the depth of field can of course be increased by reducing the aperture α, but only at the expense of the resolution. Good resolution and outstanding depth of field are, therefore, incompatible properties. Generally speaking, a compromise must be made; an optical microscope is difficult to use at its optimum resolution, as various details of the object which lie in different planes are seen simultaneously, and they cannot all be clear at the same time. In practice, the best results are often obtained when the depth of field is between twice and five times the resolution.
Fig. 184. The method of obtaining stereoscopic images by rotating the object plane between two successive photographs.

Fig. 185. A simple means of obtaining stereoscopic images. The illuminating beam is slanted between two successive photographs.
Two different methods are in use. In one of these, which provides excellent images, the object plane is differently orientated for each of the two photographs. For this, we can turn the object about an axis perpendicular to the beam (see Fig. 184) and it is the object-support of this kind, which has been described by von Ardenne (1940c) and Marton (1945), which is usually employed. The other method, which is simpler, was first recommended by Marton and Schiff (1941); the object is fixed, but in the neighbourhood of the object, the direction of the optic axis is varied. This can be done by deflecting the beam in such a way that it passes obliquely through the object (Fig. 185), with the aid of a magnetic or electrostatic field; the resolution which is obtained with this method is less high as the beam remains inclined with respect to the axis of the objective. As a result of the simplicity of its construction, however, Kinder (1946) has been able to obtain the two stereoscopic images simultaneously on the two halves of a fluorescent screen—the effect of relief can then be reconstructed in the microscope itself where it will be directly visible. So far, however, the total magnification is only 100 in this procedure.

Once the two stereoscopic photographs have been taken, the form of the object in space can be reconstituted with the aid of instruments similar to those which are used to determine altitude from stereoscopic aerial photographs (Eitel and Gotthardt, 1940, 1942; Heidenreich and Matheson, 1944; and Helmcke, 1954). The separation into layers can be determined to within about ±50 Å; so far, the applications of this method have been rather limited as it can usually be replaced by the shadow method which is considerably easier to use.

15.3.4 Depth of Focus at the Image Plane

Suppose that the geometrical image A' of a point A (Fig. 186) is produced at a fluorescent screen or a photographic plate, the structure of which permits the actual resolution to be d (and in a good instrument, d is clearly smaller than the diameter of the diffuse spot which surrounds the image point A'). Displacing the screen along the axis through a distance x spreads the image out by a further amount, 2xα'. Provided that the spreading out which is produced by this adjustment is definitely less than d, a dis-

![Fig. 186. The depth of focus in an electron microscope.](image-url)
placement of this kind will have no effect on the resolution which is observed. The screen, therefore, can be moved between the two positions I and II without deterioration of the image clarity; the "depth of focus" \( I \), the distance between I and II that is, is given by:

\[
2 \frac{l}{2} \alpha' = \frac{l\alpha}{G} = d,
\]

or

\[
l = \frac{dG}{\alpha},
\]

in which \( G \) is the magnification of the image. For \( d \sim 10^{-3} \text{ cm} \) (photographic plate), \( \alpha \sim 10^{-4} \) and \( G = 10^4 \), we find that \( l = 10^4 \text{ cm} \); the depth of focus, therefore, is effectively infinite.

This explains why it is not at all important that the photographic plate should be situated in exactly the same plane as the screen, and why, on the contrary, it is often placed some distance from the screen upon which the image is focused. This fact can be used to simplify considerably the way in which one is replaced by the other, and in the old commercial Philips instrument, great advantage was taken of this possibility by placing the camera quite a long way in front of the screen, where the image which is photographed is one-quarter of the size of the image at the focal point (Le Poole, 1947; van Dorsten, 1948).

Another consequence of this great depth of focus is that the convergence of the projective lens can be varied over a wide range to alter the magnification, without the final image ceasing to be clear.

### 15.4 Examining and Recording the Image

#### 15.4.1 The Screen

(i) Characteristics

Electron images are transformed into light images by a screen covered with some fluorescent substance. In the early stages of this subject, the photographic image was obtained by photographing the screen, but this method is no longer used except in very exceptional cases such as the recording of cinematographic images (Gauzit, 1954b; and Arnal, 1956). In all modern instruments, the electrons act directly on the photographic layer—this means that the process can be much faster, the resolution higher, and the gradations of contrast can be far wider.

A fluorescent screen, whether it be examined by reflexion or by transmission (Deubner and Hieber, 1954) is characterized by its luminous yield and by its own resolving power. The substances which are habitually em-
ployed are zinc sulphide and cadmium–zinc sulphide, activated by traces of various kinds of impurity in such a way that they fluoresce greenish-yellow, the colour to which the eye is most sensitive. More details about fluorescent screens are to be found in § 1.4, and in Chapter 12 (§ 12.2.5); here we shall simply give a few characteristics which are peculiar to microscope screens. The optimum thickness of the fluorescent layer has been studied systematically by von Borries (1948). Between 50 and 100 kV, he found that the best results were obtained with 44 mg cm\(^{-2}\) of fluorescent material (or a layer of about 100 \(\mu\) thick). In general, the minimum intensity which must be incident on a good screen, observed directly, is about \(5 \times 10^{-11}\) A cm\(^{-2}\) between 40 and 80 kV. In these conditions, the eye perceives a sufficiently bright image for the observer to adjust the focusing precisely. In this operation, the resolution of the screen is an important quality. The resolution of the image which is seen on the screen depends both upon the size of the grains of the fluorescent material and upon the scattering of the beam within the layer; in particular, as the penetration increases, the resolution decreases as the potential is raised. The normal resolution for a screen is 70 or 100 \(\mu\) (Hinderer, 1942). The grain size in the finished screen, which is composed of a fine powder, is very variable and depends upon the way in which the fluorescent layer is prepared; the method of doing this must be chosen rather carefully so that the finest particles lie on the surface.

(ii) Construction

One method of preparing the screen is by sedimentation (in alcohol, with silicate of soda as binding material; or in acetone, with 0.1 per cent of collodion); alternatively, a solution of cellulose acetate thoroughly mixed with the fluorescent material can be poured over a glass plate, and after the solvent has evaporated, a thin cuticle remains which is stripped from the glass surface and stuck on the metal support of the screen. The surface which had been in contact with the glass is the surface which is observed. Screens which are very brilliant because they contain almost no binding material (this is essential when the electrons are replaced by ions—see Chapter 18), and which have a high resolution, can be obtained by sprinkling the metal support of the screen with the fluorescent material, after first spreading a thin film of oil over the support. A glass plate is then pressed against the outer surface of the fluorescent layer, thus making it perfectly smooth.

When the electron optical system is to be used at low magnification, the image has to be examined with the aid of a high-magnification light microscope in order to focus the system correctly, but we are then faced with two difficulties. The grain-size of the screen limits the resolution of the images, and the depth to which the electrons penetrate the phosphor
becomes important. The numerical aperture of the light microscope has to be reduced until its depth of focus is of the same order of magnitude as this penetration depth, and screens free of any structure must be employed. Attempts to produce the latter have been made in which either certain monocryals are used (von Ardenne, 1939), or certain kinds of uranium glass, or special zinc oxide screens (Kopp and Möllenstedt, 1946, 1947); crystals, which are perfect as far as the absence of granular structure is concerned, have the disadvantage of being rather insensitive (a hundred or a thousand times less sensitive than ordinary screens).

A technique is available with which screens free of granular structure can be obtained in the form of a thin layer, by evaporation of zinc sulphide followed by a suitable annealing (Feldman and O'Hara, 1957). These thin screens, which are grain-free and are about as thick as the penetration depth of electrons accelerated to a few keV, can thus be examined at their brightest with an objective of large angular aperture; this has recently enabled microscope images to be examined at very low accelerating potentials (Nixon, 1958).

15.4.2 The Photographic Plate (see also § 1.5)

(i) Plates and films

The camera is to be found beneath the screen and is situated like the latter in the vacuum; the shutter is often simply the screen itself. Non-sensitized emulsions are used, mounted either on glass or on film, although the latter has the drawback of containing certain volatile substances which means that the pumping-time is rather long; emulsions spread on film, however, seem to be faster to electrons than the corresponding plates. Certain films have the additional disadvantage of being too good insulators—parasitic deformations are then seen in the image which is recorded, caused by charges which have been unable to flow away (Couchet, Gauzit, and Septier, 1951a and 1951b).

(ii) Choice of exposure time

The exposure time is usually estimated by the operator on the basis of his visual impression of the brilliance of the screen, but just as in ordinary photography, a photometer is a most useful accessory. An equal intensity photometer can be constructed with the aid of an auxiliary electric light bulb of which we can vary the brilliance; the light which is emitted is filtered by a coloured plate and produces a bright spot of the same intensity beside the image (van Dorsten, 1950). The current which falls on the screen can also be measured, in terms of the potential difference which it produces across a large resistance of the order of tens of megohms which is
connected between the screen and earth (Hamm, 1951). A more accurate and robust instrument, which is, however, more expensive, consists of an exterior photomultiplier which is placed against the glass window (Cuckow, 1946; Bishop, 1950; Frei and Hirsfeld, 1951; Gauzit, 1954b; and Tristram, 1956).

Long exposures are to be avoided, as the probability that the image will be displaced during the exposure (by an accidental vibration, or by mechanical or electrical drift) is too high; very short exposures, too, are not advisable with the microscopes commercially available as the shutters are too coarse and produce vibrations which continue during part of the exposure. The best exposure time is about three or four seconds. Extremely good results can, however, be obtained with an exposure of a second or less provided that an "electrostatic shutter" is installed, which diverts the beam except during exposures, or alternatively, an electromagnet which acts on a very light shutter under the projective.

(iii) Sensitometry

Photographic emulsions are characterized by their sensitivity, their contrast and their resolution. Their behaviour in the presence of electrons is, generally speaking, analogous to their behaviour in the presence of light; in the electron case, the properties of emulsions vary in the same sense but over a considerably more reduced domain than in the optical situation. The photographic plate is extremely sensitive to the arrival of electrons (von Borries, 1942, 1943, 1944); whereas between 10 and 100 incident photons are necessary before a grain of bromide can be developed, a single electron can render one or even several grains ready to be developed. The energy yield (which is defined as the number of watt cm$^{-2}$ which provide a given intensity) is, however, about twenty times weaker for electrons than for light. This explains the failure of the reinforcing fluorescent screen technique, which had been intended to increase the sensitivity of the photographic emulsion to electrons (von Borries, 1942; Baker, Ramberg and Hillier, 1942, 1943). For the same reason, neither a colouring material to increase the sensitivity nor a material which reduces the sensitivity has any effect on the speed of the emulsion (Nissen, 1944).

When the high tension is increased, the sensitivity of a solution is reduced; the optimum operating potential is in the neighbourhood of 50 kV. Recent attempts at very low tension microscopy (see § 19.2.8) have re-awakened research into this question, however, with the object of finding an emulsion on the commercial market which still has a good sensitivity at 10 kV. Van Dorsten and Premseila (1960) recommend the Gevaert 9 E 56 films which only require an increase in current density of 1.6 times to obtain the same degree of blackening when we pass from 52 to 13 kV, whereas the films which are in everyday use in microscopy would require the intensity to be increased by ten to a hundred times.
The forms of the curves which describe the photographic density of photographic emulsions in the presence of electrons resemble those which are produced by light; the base, however, is relatively longer, and has a weaker curvature, so that the straight line portion is only reached at the density 1·5 (whereas for light the value is 0·5). From Fig. 187, we can see that the rectilinear part of the photographic density curve corresponds to charge densities between $10^{-11}$ and $5 \times 10^{-9}$ C cm$^{-2}$.

![Graph showing the relationship between incident charge density and photographic density](image)

**Fig. 187.** Curves showing the way in which the blackening of the photographic plate varies with the incident charge density (in C cm$^{-2}$).

Just as in the case of light, slower plates have a better resolution and a higher contrast, or what is equivalent, a more vertical rectilinear portion. Since the "shadow technique" has been introduced, however, the contrast seems to be adequate, and the present tendency is to protect the object from being heated by reducing the illumination and using faster and faster emulsions which are, of course, less contrasty.

The resolving power, which varies widely between one type of emulsion and another, is difficult to measure. Von Ardenne (1939) found that the resolution of the best ordinary emulsion was 10 $\mu$; von Borries (1944) after a precise and extensive examination, concluded that the result should be about 30 $\mu$, and this estimate is generally accepted to-day. Certain special emulsions have, of course, a much finer grain; for example, Kopp and Möllenedstedt (1946) considered certain plates of the Schumann type, and found a resolving power of one or two microns, but the sensitivity is far below that of normal emulsions. The price of this type of plate is high, and manipulation is difficult, so that despite the better contrast, it is only rarely employed.
A very thorough survey of the response of photographic emulsions to electrons has just been prepared by Valentine (1966). The reader will find much helpful information about the most recent emulsions in this article, as well as practical advice about developing films and plates.

15.4.3 The Choice of Direct Magnification

(i) The influence of the grain of the plate

The grain of the plate affects the magnification of the image at the fluorescent screen or on the photographic negative; as the microphotographs are always enlarged photographically, but not beyond about twenty times because of the grain size of the emulsion, the total magnification which can be attained is fixed at about twenty times the direct magnification. In order to exploit the qualities of the objective as fully as possible, an electron microscope should have a direct magnification of the order of one twentieth of its useful magnification.

The lower limit is given by

\[ G_{\text{min}} = d/q, \]

in which \( d \) and \( q \) are the resolutions of the plate and the instrument respectively (von Ardenne, 1939). The upper limit cannot be so simply defined, as many different emulsions with grains of widely different sizes are available, which can support enlargement between about five and twenty times (or even 100 times for certain emulsions which have been developed recently). As fine-grain emulsions are slower than ones with coarse grains, we must resign ourselves to a compromise, as a short exposure avoids many of the sources of blurring which are due to slow drifts or instabilities of the potentials, or the lens currents.

To take a photograph of which the resolution is to be 20 Å, for example, with an emulsion with a resolving power of 50 μ, the electron magnification has to be at least 25,000. Experience shows that the best results are obtained by taking the photograph at the smallest electron magnification which is compatible with the resolution of the emulsion, and then exploiting to the maximum the optical properties of photographic enlargement.

(ii) Heating the object

Among other advantages, this rule means that the object is treated as gently as possible—given the minimum current density which is admissible at the image \( I_{\text{min}} \) (which is of the order of \( 10^{-11} \) A cm\(^{-2} \) for an exposure
of 10 sec) we can calculate the minimum current $I_{\text{min}}$ which passes through the object; we find

$$I_{\text{min}} = I_{\text{min}}' \left( \frac{d}{\varnothing} \right)^2.$$ 

In the example above, $I_{\text{min}}$ is of the order of $6 \times 10^{-5}$ A cm$^{-2}$.† We see, therefore, that the value of using a plate with a high resolution resides in the fact that for a given exposure the intensity to which the object is subjected will vary with the square of the resolution associated with the grain, provided we can assume that the sensitivity remains constant.

(iii) *A measure of the quality of emulsions*

Since fine-grain plates have a lower sensitivity, the advantage which we have just mentioned is considerably diminished by the fact that $I_{\text{min}}$ is in-

![Graph](image)

**Fig. 188.** The dependence of the *Bildpunktarbeite* upon the accelerating potential, for various emulsions (B. von Borries). (a) Agfa Normal. (b) Agfa Mikro. (c) Agfa Kontrast. (Developer—Agfa Medialon 85 g $l^{-1}$, 18°, 8 min.)

† At this intensity, a considerable power is being concentrated on the object. At a potential of 50 kV, the incident power is 300 watts cm$^{-2}$, whereas the earth only receives 0.1 watts cm$^{-2}$ from the sun. Von Ardenne (1939) compares the current $I_{\text{min}}$ with the critical electron density, $3 \times 10^{-6}$ C cm$^{-2}$ at which living matter is just killed by the effect of ionization; his conclusion is that at present it is not possible to observe living matter in an electron microscope working at a very high magnification, in the potential range 50–100 kV.
increased. We ought, therefore, to characterize the quality of an emulsion by some index in which speed and fineness are weighed against each other. Baker, Ramberg and Hillier (1942, 1943) and von Borries (1944) have taken as a measure of the quality the coefficient $A$ which is defined by

$$A = U q_{0.5} \cdot d^2,$$

and which they call the Bildpunktarbeit (image-point energy); $d$ is the least distance resoluble at the plate, $U$ is the accelerating potential, and $q_{0.5}$ is the charge density per unit of surface which darkens the plate to an optical density of $d = 0.5$. The product of $U$ and $q_{0.5}$ is a measure of the energy which must be incident upon unit area of the surface of the plate in the form of flux of electrons to produce the stated photographic density, so that $A$ is an energy which explains the name which von Borries, to whom this definition is due, gave it.

This measure does indeed take the speed of the emulsion into account, since $q_{0.5}$ is proportional to the exposure for a given current density at the image plane. If we introduce this final current density $i_{0.5}$ at the plane in which the plate lies, $A$ takes the following, more suggestive form which emphasizes the role of the exposure time

$$A = U i_{0.5} td^2.$$

The smaller the coefficient $A$, the better is the emulsion. Figure 188 shows von Borries' results, which describe the behaviour of $A$ as the potential is varied for various emulsions.

(iv) The role of the eye—the use of a magnifying-glass for focusing

The choice of the direct magnification is also affected by the fact that the image has to be inspected visually in order to focus it onto the fluorescent screen, and the smallest details have to be capable of being separated by the eye. The resolution of the eye in normal conditions is about 0.1 mm, so that if two details 20 Å apart are to be distinguished, the electron magnification must be at least 50,000 which is far too high for the photographic conditions to be good, since the intensity at the screen falls as the square of the magnification. For this reason, a compromise between the exigencies of eye and plate is often adopted, in which the adjustment is made and the photograph taken at some intermediate magnification between 25,000 and 30,000, say.

There is, however, an optical way out of this dilemma. At a direct magnification of about 10,000, the brilliance at the fluorescent screen is quite adequate to prevent the eye from being too fatigued, and to allow good photographs to be taken, but this is not a high enough magnification for details near the limit of resolution of the instrument to be distinguished—
the contrast is too slight and the brightness is still low enough to diminish the resolving power of the naked eye.† We can, therefore, simply use a small microscope which magnifies the fluorescent image five or ten times; little brilliance is lost, and a great deal of resolving power is gained. If the objective has a very wide aperture, the image intensity and the intensity at the object are equal—for this, the pupil of the eye must be wholly covered by the exit pupil of the eyepiece. The only disadvantage of this artifice is that the field of view at the fluorescent screen is reduced to less than two centimetres, but the overall magnification can easily be of the order of 100,000 times and in research in which a high resolving power is necessary, this method is frequently employed. Finally, we should point out that the method can only be used if the grain-size of the fluorescent screen is less than $0.1/g$ in which $g$ is the magnification of the eyepiece. To satisfy this condition, the fluorescent powder must be specially prepared, or else the smaller grains of an ordinary powder must be selected by sedimentation; pounding a coarse powder does not produce a good screen as the sensitivity of the grains is considerably reduced by so rough a fracture, and no subsequent reheating can restore it to its previous value. The reader interested in this point can usefully consult a detailed study by von Borries (1948).

15.5 THE IMPORTANCE OF THE PROJECTIVE LENS

15.5.1 The Role of the Projective

The purpose of this lens is to provide a direct magnification large enough for the resolution of the objective to be exploited to the full in an instrument which is short and hence mechanically very stable; the overall resolution, therefore, is determined by neither the photographic plate nor the fluorescent screen but by the objective. For this reason, it was possible to dispense with a projective lens in the early electron microscopes the resolution of which was not very high. Attempts were made to retain this simple structure even when the optical properties became more refined, and screens and photographic emulsions with a very fine grain were tried (with a resolution of between 4 and 12 μ); in this manner, Kopp and Möllenstedt (1946, 1947) were able to obtain images with a useful magnification of 30,000 by invoking all the resources of modern screen and plate techniques. This is a fine achievement, but difficult and costly, and recent work by Möllenstedt has shown that fine-grain plates do not give as good a performance with electrons as their resolving power with light would lead us to hope. Parasitic effects due to the imperfect conduction of the electrons are the cause.

In the electron microscopes which are available commercially, the inter-

† The reader is referred to § 14.4.2 for details of the use of image-intensifiers.
mediate image which is formed by the objective is in general magnified by one or even two successive projective lenses. The advantage of introducing two successive stages is partly that the length of the structure is reduced to the minimum (for rigidity) but also because various auxiliary functions of the instrument become simpler, as we shall see later. The “object” for the projection system is the image formed by the objective, and we choose the position of this image to be near the focus of the projector—that this is the “classical focus” (§ 4.3.4), we must state explicitly, as this objective image is a virtual object so far as the projective lens is concerned.

The path of a pencil of electrons through the projective is quite different from the trajectory in the objective. In the former, the image of an object with a relatively large surface area has to be formed with pencils of electrons with an extremely slender angular aperture—that of the objective divided by its magnification—in the neighbourhood of $10^{-8}$. The object, which is virtual, can, however, be placed well within the electric or magnetic field of the lens so that with the same design, the projective can be given a considerably shorter focal length than the objective. The projective, therefore, has to give a good image of a considerable area (from a fraction of a mm$^2$ up to 1 mm$^2$) in such a way as to fill the screen on which it is observed for an extended range of magnifications. The spherical aberration no longer plays a direct role, as the pencils are very fine, and the aberration which is important is the distortion. It has the effect of magnifying the peripheral region of the intermediate image more than the central part, so that the final image displays “cushion” distortion in these conditions (§ 7.2.2) which is particularly pronounced as the magnification is low. It is this defect which provides a lower limit to the possible magnification, and which decides the minimum value which is acceptable for a given projective. To work at a smaller magnification, a projective with, for example, electrodes or pole-pieces of a larger diameter would have to be used, but this would increase the minimum focal length and reduce the maximum magnification. Only by building a more complicated system can we avoid the necessity of making a compromise between these two contradictory conditions; this entails placing a second projective lens after the first (§ 15.5.2). With a single projective, the magnification can be varied in the ratio 1 : 10. The magnification of the projective might be adjustable between about 25 and 250, for example, which would give a total magnification between 2000 and 20,000. As we have explained earlier (§ 15.4.2), the convergence of the projective can be varied over wide limits without affecting the resolution of the final image in consequence of the extreme depth of field of the objective.

We may equally well use a double projector (Rang and Weitsch, 1956), the electrostatic version of which contains four electrodes. This is in fact simply an electrostatic lens with two outer electrodes at earth potential, and two independently excited central electrodes. It is less bulky than two
separate lenses placed side by side, but it is nevertheless still possible to obtain a very wide range of magnifications; if the two central electrodes are at the same potential, the focal length is of the order of 20 mm when the potential is three-quarters of the cathode potential, whereas if one electrode alone is excited and the other earthed, the focal length is \( f_1 = 5 \) mm in one case, \( f_2 = 2.2 \) mm in the other. For the same excitation potential, the magnifications of each of these three possible arrangements are in the ratio \( 1 : 3 : 10 \), provided the positions of the principal planes are taken into account.

15.5.2 The Use of Two Projective Lenses

(i) The advantages of this combination

Many modern instruments have two projective lenses (see Fig. 189), an arrangement which was first proposed by Marton (1944, 1945). This combination, which is more compact and considerably more flexible, is advantageous both mechanically and optically. In particular, the direct magnification can be varied continuously over an extremely wide range (from less than 1000 to more than 50,000) without either the object having to be shifted or the lens replaced. In addition, the diffraction pattern which corresponds to a zone which has been examined microscopically can be obtained without altering the position of the object; a simple variation of the excitation of a lens is all that is necessary to produce on the screen the diagram which corresponds to the part of the object which had been examined previously (see Chapter 21). In this way we can select some detail of the crystalline form and analyse it without error. The mechanical advantage of using two projectives should be underlined as well. For the same direct magnification, the three-stage instrument is far less cumbersome, and can therefore be made far more rigid and free from the aberrations which mechanical vibrations would produce.

(ii) The path of the rays and the production of the magnification

In practice, we should like above all to reduce the distortion to the minimum, and to make focusing as easy as possible. Experiment shows that the best results are obtained when it is the potential of the first projective that is varied, while the second has a constant convergence which corresponds to a magnification of about 100; in these conditions, the field in the second lens, and therefore the distortion which it produces, are small. The intermediate projective has large openings to reduce distortion and quite a large minimum focal length, in general about two centimetres. It may produce a radial diminution of the magnification, which leads to "barrel" distortion—Hillier (1946) has considered this case in detail.
(iii) Some practical examples

In the first case which we shall consider (see Fig. 189a), the two projective lenses $P_1$ and $P_2$ function identically, progressively magnifying the image, and providing an overall magnification which is the product of the magnifications of each of the two separate stages. When the convergence of $P_1$ is reduced to diminish the magnification, the image plane $I_2$ remains stationary while the object plane $I_1$ moves progressively towards the objective $O$; the convergence of the latter will have to be slightly increased to

Fig. 189. The various ways of using a microscope with two projective lenses (from top to bottom): (a) Normal use, at high magnification. (b) Use as a diffraction device, the two projectives giving a magnified image of the diffraction pattern which is formed in the focal plane of the objective. (c) Use at low magnification. (d) Extreme case—the instrument behaves like an instrument with a single projective.
keep the image in focus when we move towards the smaller values of the magnification, although there is a certain latitude in the focusing. The minimum value of the magnification (which is of the order of 3000 in the normal types of instrument) is fixed by the fact that the convergence of the objective cannot be increased beyond a certain point.

It is, however, no problem to continue decreasing the convergence of the first projective $P_1$, and there comes a moment (see Fig. 189b) at which the focal plane on the object side of this lens coincides with the focal plane on the image side of the objective, the plane in which the objective focuses the "diffraction pattern" of the object at infinity. $P_1$ in this case works at a magnification of about unity, so that on the final screen, we find an image of the diffraction pattern magnified about a hundred times by the second projective. This diffraction arrangement was first employed by Boersch (1936); in this situation, as van Dorsten et al. (1950) have pointed out, an image is formed in the focal plane on the image side of the first projective $P_1$ at magnification $f_1/f$ ($f_1$, which is of the order of a few centimetres, is the focal length of $P_1$ while $f$, which is of the order of a few millimetres, is the focal length of the objective). By placing a diaphragm in this plane, of which both the size of the opening and the position can be varied, we can isolate a small region of the object (a crystallite, for example) the diffraction pattern of which we should like to examine; the area of this zone can be less than a square micron if $f_1/f$ is large enough (50, for example) and the diaphragm is small ($50 \mu$).

If the convergence of $P_1$ is reduced still further, the focal plane on the object side of this lens passes beyond the objective. It is then possible to obtain an image with a very small magnification (about 500 times), if the power of the objective is reduced simultaneously, as Fig. 189c shows. The reason for this is that as far as $P_1$ is concerned, the electrons appear to have originated in the (virtual) image of the original object which is formed by the objective and lies "upstream" from the objective; the distance of this image increases in proportion as the convergences of $P_1$ and of the objective decrease. We can go on reducing the convergence of $P_1$, and if the latter is to produce an image at $L_5$, the objective must produce an image beyond $L_5$ (see Fig. 189d). In the limit, $P_1$ will cease to have any effect, its convergence will be zero, and the microscope will function as a two-stage instrument with a magnification of about 5000.

In the course of the series of adjustments described above, in which the adjustment of the second projective is left untouched, we have seen that the magnification begins at a very high value, of about 50,000, falls at first to nearly zero (this is the region which is useful for diffraction) and increases again to a mean value of about 5000 for example. Only two regions, on each side of the diffraction zone, cannot be used as a result of distortion or the value of the field. In case (d), the beams which form the intermediate image do not intersect in a crossover, so that the barrel dis-
ortion which results can compensate the cushion distortion of the second projective. Detailed analyses of all these focusing procedures have been published by Le Poole (1947) and by Challice (1950), who illustrate them with numerous numerical examples. Several modern instruments work in the way just described, for example the Metropolitan Vickers EM 3 (Haine, Page and Garfitt, 1950), the RCA type EMU (Simard, Burton and Barnes, 1945) and the CSF type M 9 (Grivet and Regenstreif, 1950).

Sometimes, still more complex systems are employed; certain instruments have three projective lenses, although only two are ever simultaneously in use (Le Poole, 1947; van Dorsten et al., 1950) while in other designs, certain lenses can be retracted mechanically while the instrument is in use so that their electrodes do not restrict the field (the Trub-Tauber Microscope, type KM 4 is an example).

15.5.3 The Projective Lens Stop

A field stop is usually placed within or beneath the projective; its purpose is to collect any electrons which have been reflected by the walls of the tube which supports the lenses, or scattered by the gas in the instrument (Hillier, 1950). In the electrostatic microscope, the stop has a supplementary function; it is placed just above the projective, in the focal plane on the object side, and it helps to eliminate rays which are following transgaussian trajectories far from the axis and which would converge onto the centre of the image where they would form a transgaussian star. The diameter of this diaphragm, which is about 1 mm, must be determined with care so that the field is not purposelessly reduced at small values of the magnification, but the transgaussian rays are interrupted.

15.6 MEASUREMENT OF THE CHARACTERISTICS OF A MICROSCOPE

15.6.1 Measurement of the Magnification

This is a fundamental operation with an electron microscope, though it is difficult and often neglected. In fact, various factors introduce a degree of uncertainty into the process of calibration, which is rather a long operation, so that the exact value of the magnification is often not known to within better than 30 per cent. First of all, the distortion, which is equivalent to an increase of magnification as we move from the centre towards the edge of the image, must be taken into account. Further, the position of the object with respect to the focal plane of the objective is not absolutely fixed, so that the focal length of the latter must vary to keep the image in focus. In a magnetic microscope, both the magnetic hysteresis of the lenses and variations in the high tension can be sources of error. Generally speak-
ing, and contrary to the opinion of certain authors, it seems that the value of the magnification can only be determined to within $\pm 10$ per cent (or $\pm 5$ per cent if the microscope is electrostatic).

We now discuss the various methods which have been used to measure the magnification.

The magnifications of the objective and the projective (this method is suitable only for instruments with a single projective) are measured separately in two stages, and the total magnification is given by their product. Von Borries and Ruska (1940) examined holes with diameters sufficiently large to be measured accurately with an ordinary microscope.

The object can be laterally displaced through a known distance (measured with the aid of a micrometer, or by interferometry) and the magnification calculated from the corresponding shift of the image.

The electron image can be compared with an optical image; the useful magnification of an ordinary microscope is restricted to about 1000 times, however, so that a particle must be at least $25\mu$ in diameter if its dimensions are to be determined accurately. This is a prohibitively large diameter for electron microscopy, as the field is too small and there will be distortion.

In the most accurate method, the one which is in general use, an object of which the dimensions can be measured by some other method and related to the usual units of length is examined through the electron microscope. Spheres of various materials can be used (glass, silica, latex) which are uniform in shape and calibrated by centrifuging, by interferometry or by the scattering of X-rays. Backus and William (1949) and Gerould (1950) have found that the polystyrene latex marketed by the Dow Chemical

---

Fig. 190. The image of a replica of a diffracting lattice, with which the magnification in an electron microscope can be measured.
Company consists of highly regular spherical particles (2500 ± 25 Å in diameter). This method has, however, been the subject of a certain amount of dispute; Yudowitz (1951) and Danielson et al. (1952) have measured the diameter of the same particles with the aid of X-rays, and found that the diameters were 2740 Å ± 1 per cent and 2687 ± 0.5 per cent while Kern (1950) measured the diameter—electron microscopically—to be more than 3000 Å, when the preparation is “shadowed”.

Another solution is to examine the replica of a grating (with at least 5000 lines cm⁻¹; see Fig. 190) the mesh length of which can be determined optically with a precision which is for the most part adequate. This replica must not be in an organic glaze, which is deformed when it is stripped off; instead, a layer of silica or alumina is necessary. The accuracy of the method (Rouzé and Watson, 1953) is limited at once by the fact that the spectroscopic measurement gives only an average value of the mesh-length, and not the local value which is observed in the microscope, and by the difficulty of measuring the exact distance between two corresponding points on two lines of the grating (the grooves may not be identical, for example, or the lines too wide). The most accurate method at present seems to be the one due to Farrant and Hodge (1948), who examine a glass fibre about 5 μ in diameter. This latter is measured interferometrically to a very high accuracy.

15.6.2 Measurement of the Resolving Power

(i) The practical determination of the resolving power

The theoretical definition of resolving power is based on the curve which describes the luminous intensity across the image of a bright point. Such images are only very rarely encountered in electron microscopy, however, except perhaps in “dark field” observation (see Chapter 19). No object is known with holes sufficiently close together to play a role analogous to that of diatoms in light optics. The finest details which are to be seen in electron optics are the images of small particles deposited on a transparent film (Fig. 191). The criterion which is used to define resolving power can be extended to this case, but the possibility of “resolving” the images of two small particles depends not only upon the distance between them but upon their individual opacity as well, that is, upon their scattering power. The empirical resolution and the scattering power are intimately related therefore, so that this is a most uncertain method of measuring the actual resolving power.

Great patience is needed in practice, to choose the zone on the image in which the particles with the smallest dimensions have collected together. Metal particles are usually employed, and obtained from a colloidal sus-
pension of gold, silver or platinum (in an aqueous solution of Argyrol at\textsuperscript{1}1/1000 for example). Alternatively, a metal can be evaporated on the support in vacuo to produce, for example, a film of chrome, gold or platinum 5 or 10 Å thick. Von Ardenne (1941) suggested that the arrangement shown in Fig. 192 might well be used; in this, a razor-blade is placed 0.1 mm from the object-support, so that when the metal is evaporated onto the support, a transition zone 0.5 μ thick is formed. The supporting film must be as thin as possible and of the lowest practicable atomic weight. The electron intensity too must be as low as possible so that both recrystallization and the growth of the particles by contamination are prevented.

Figure 191 is a photograph of an aggregate of fine particles of an alloy of gold with palladium. The smallest particles which are distinguishable on this photograph are slightly less than 10 Å in size; the photograph was taken by Haine in his laboratory at Aldermaston with a Metrovick microscope, the resolving power of which he has coaxed down to 5 Å at the optimum operating conditions. With a similar type of object (an evaporated platinum-iridium film), Weichan (1966) has been able to discern separated grains, 4 Å in size, using a microscope with accelerating voltage 80 kV, an electron beam incident with angular aperture $10^{-2}$ and a contrast aperture of $10^{-2}$.

Haine (1954) has studied the possibilities of using objects of this kind in practice to determine resolving power. The difficulties are considerable, for if the resolving power to be measured is of the order of 5 Å, the object which we have to use must be of atomic dimensions. Haine suggested that this problem could be avoided by using the Fresnel diffraction fringes which are also used to detect astigmatism (§ 16.4). This was not, fundamentally, a new suggestion, but it was based on original reasoning and a detailed study and suggests convincingly that this technique leads effectively to a correct estimate of these extreme resolving powers. The measurement is indirect, and the calculation of the resolving power entails introducing the spherical aberration constant, which has to be measured separately; nowadays, however, the values of the latter are well known.

A certain amount of research has gone into finding a different definition of resolving power which could be measured practically and which would be independent of the nature of the object. So far, however, no suitable test object has been discovered, and the image-quality which results almost always depends as much upon the nature of the specimen and the competency with which it has been prepared, as the quality of the instrument itself. A particular scrutiny of this point has been published by the American Electron Microscopical Society (1946), in which the various possible definitions are examined, but no method could be recommended in particular for its precision as a result of this analysis.

As we shall see later, a pseudo-image of the lattice planes can be obtained with a periodic object (a monocrystalline layer, for example) which
Fig. 191. The image of the grains of a gold-palladium alloy deposited onto a collodion film. \( G = 1,000,000 \) (Photo: Haine and Mulvey).

Fig. 192. The preparation of a test object by von Ardenne’s method for measuring resolving power.
may, under certain conditions, have a higher resolution than we could anticipate from two neighbouring points or grains; the nature of the test object must therefore always be specified. Using an instrument with a practical resolution between neighbouring points of about 5 Å, it has proved possible to form images of the (200) lattice planes of gold, for which \( d = 2.04 \) Å (Komoda and Otsuki, 1964), of the (220) planes of gold, \( d = 1.44 \) Å (Komoda, 1966) and of the (220) planes of copper, \( d = 1.27 \) Å (Watanabe et al., 1966).

(ii) Estimation of the quality

This empirical method was proposed by Hillier (1945). A negative is enlarged to the maximum when a photograph is being printed, up to the point at which the details become blurred. This gives an experimental measure of the useful magnification and hence of the resolution, using the relation

\[
\theta = \frac{\theta_{\text{eye}}}{G_u}.
\]

This method, useful though it is for rapidly comparing several negatives produced with the same instrument or with different instruments, is not very accurate when it is an absolute value that is required; it depends on such subjective effects as the keenness of the eye and above all on the photographic contrast. The contrast plays a most important role in this way of estimating the image "quality", as the image of a "shaded" object always seems considerably better than the image of the same object in its normal state, even though the latter may often contain far more detail. If the estimate is to lean on the "clarity" of the detail (a more or less round particle, or a crystal boundary for example), a careful printing of the photograph onto "hard" paper can also improve the resolution artificially (by a factor of up to two), while in reality much of the detail of the negative is being lost. Nevertheless, a trained observer can very easily judge the degree to which his instrument has been well adjusted by examining a few photographs of a "test-object" to which he is used and the magnification of which is known to him.

(iii) Typical experimental results

Figure 193 shows the various values of the resolution which have been obtained since the creation of the electron microscope; a factor of about three separates the values which correspond to the electrostatic microscope from those which correspond to the magnetic microscope, a difference which we pointed out at the beginning of this chapter. It does seem, however, that the quality is tending asymptotically towards a limit, despite the increasing numbers of studies which are being devoted to the instrument.
Fig. 193. Chronological chart, giving a general idea of the progress made in electron microscopy over the last thirty years. $E$: electrostatic microscope; $M_1$: magnetic microscope, non-periodic specimens; $M_2$ magnetic microscope, periodic specimens.

The best results which have been published at the present time are the following:

With the electrostatic microscope:
- Boersch (1944) 28 Å
- Grivet and Regenstreif (1950); with a Bertein stigmator 20 Å

and with the magnetic microscope (with a stigmator):
- Hillier and Ramberg (1947) 10 Å
- Leisegang (1954) 6–8 Å
- Haine and Mulvey (1954) 6–8 Å
- Engel, Koppen and Wolff (1962) 4 Å
  (non-periodic specimen)
- Komoda and Otsuki (1964) 2 Å
  (periodic specimen)
- Watanabe, Shinagawa and Shirota (1966) 1.3 Å
  (periodic specimen)
CHAPTER 16

THE FACTORS WHICH GOVERN THE RESOLVING POWER IN ELECTRON MICROSCOPY

In particle microscopy as in optical microscopy, the defects which limit the fineness of the image can be classified into three categories, according to the degree to which their presence is unavoidable. The first of these categories which we shall examine is diffraction, the influence of which is, even to-day, so very difficult to eliminate (Gabor, 1949 and 1951) that we must consider it to be inevitable. The final resolving power is determined by the diffraction, which is a characteristic of the radiation that is used in the microscope.

The theoretical limit of resolution of a microscope is determined by the combined effect of diffraction and spherical aberration. Finally, the practical resolution of any instrument also depends on the supplementary aberrations that arise from constructional defects, or the observer himself, or from the mechanism of interaction between the beam and the object to be observed. We shall consider each of these points in succession. Much complementary material will be found in a very searching recent article by Ruska (1966).

16.1 THE FINAL RESOLVING POWER;
DIFFRACTION ALONE

16.1.1 The Abbe Formula

If we suppose that the lenses are perfect in the sense that they have no aberrations, which is quite possible in light optics, the resolving power is fixed by the wavelength of the radiation employed and by the aperture of the instrument. Only diffraction of the beam at the openings in the stops provides a limit to the quality of the image. This ideal resolving power is given by the general formula due to Abbe, a derivation of which is to be found in the book by Louis de Broglie (1950), which states that

\[ q = \frac{0.61 \lambda}{n \sin \alpha}, \]

529
in which \( \lambda \) is the wavelength of the light which is being used, \( n \) is the refractive index of the medium which surrounds the object, and \( \alpha \) is the semi-angle of the cone defined by the useful rays which emerge from some point on the object. If the optical system which is employed allows the angular aperture to be high (\( \alpha \approx 70^\circ \)), the formula becomes \( g \approx 0.61 \frac{\lambda}{n} \), or roughly \( \lambda/2 \).

16.1.2 The Optical Microscope

In an optical microscope, therefore, when the objective is corrected and of wide aperture, and the rays are immersed in a liquid the refractive index of which is close to that of the face of the objective at which they enter (\( n \approx 1.5 \)), the final resolving power which is obtained with \( \lambda = 0.5 \mu \) is 2000 Å which corresponds to a useful magnification of about 500. If the ultra-violet radiation provided by a mercury arc is used, the ideal resolving power is about 1000 Å, and the useful magnification about 1000. The term \( \sin \alpha \) can never be larger than unity, and it is virtually impossible to find a liquid with a refractive index larger than that of cedar oil, for which \( n = 1.6 \). There is, therefore, an upper limit to the useful magnification attainable.

16.1.3 Attempts to Use X-rays

It is tempting to try to go down to the wavelengths of X-rays. Two different ways of exploiting this idea have been followed, and interesting results have already been obtained; we shall outline the essentials of these.

The first idea was to try and construct a very fine X-ray source, and then to project a highly magnified "X-ray shadow" of the object which was being studied onto a fine-grain photographic plate. The finer the source and hence the smaller the penumbra, the sharper is the image. The whole problem, therefore, is to construct a suitable source; recently, the extremely tiny spot which is obtained when the gun replaces the photographic plate in an electron microscope has been used (Cosslett and Nixon, 1952, 1953; Cosslett, 1954). It is thus the smallest possible image of the tip of the filament which is employed. This is not an easy exercise, and the operation of the electron microscope in reverse is described by exactly the same theory as the normal electron microscope. The anti-cathode, which is a very thin leaf of tungsten, serves at the same time as an exit window for the X-rays; the object can thus be placed very close to the X-ray source, and the resolving power which is bounded above all by the diameter of this source can equal or even exceed the resolution of an optical microscope. It is also very easy to record stereoscopic images (see § 22.4).
Alternatively, we can take advantage of the fact that a truly “X-ray” microscope can be built by using the focusing properties of a system of crossed cylindrical mirrors which totally reflect the radiation. Even with soft X-rays and a reflecting surface with a high atomic number, however, there will be total reflexion only for extremely small angles of incidence, of the order of half a degree. The numerical aperture of such an objective is, therefore, very limited, and the resolving power has barely reached a micron at the present time.

16.1.4 Diffraction in the Electron Microscope

As we saw in the last chapter (§15.2.3), the wavelength which is associated with an electron beam accelerated to 40 kV (a normal value of the potential in electron microscopy) is of the order of 0.06 Å, or 100,000 times shorter than the mean wavelength of visible light (0.6 μ). All other things being equal, therefore, it seems that the final resolving power should be 100,000 times better (or of the order of 0.002 Å); and the useful magnification 100,000 times larger (and hence about 50,000,000).

Unfortunately, no lens is stigmatic† in electron optics, and so we cannot work with the large apertures of glass optics; aberrations will play a considerable role, for it is they which limit the aperture which can be employed. The notion of a final resolving power determined solely by diffraction is meaningless in electron microscopy, and we must now introduce a second factor into the resolving power, the spherical aberration coefficient.

16.2 THE OPTIMUM RESOLVING POWER; DIFFRACTION AND SPHERICAL ABERRATION

16.2.1 The Geometrical Aberrations

We have already studied these aberrations (Chapter 7); they are of varying degrees of importance, and in microscopy, only the spherical aberration has any effect upon the optimum resolving power. Moreover, only the spherical aberration of the objective is relevant as the beams of electrons at the projective are of very small aperture (§ 15.4). The spherical aberration of the objective is, however, very important, especially as no way is known of correcting it (§ 7.7.2) at least so long as we keep to axially symmetric lenses (§ 7.7.6). The only way of correcting the objective would be to use a mirror as projective (§ 7.7.3), and this has never been tried; Mahl and Pendzich (1942) and Le Rutte (1952) have certainly tried—with success—to replace the

† In this context, “stigmatic” is to be understood as meaning “capable of forming a point image of a point object”, and not simply “devoid of astigmatism” (Translator’s note).
projective lens by a mirror, but this was only to simplify the construction and to make the instrument less cumbersome by folding the electron trajectories back upon themselves.

### 16.2.2 Spherical Aberration

An electron lens converges marginal rays more rapidly than rays close to the axis. The image in the Gaussian image plane which corresponds to an axial point object is spread out over a spot of radius $\theta_s$ which is proportional to $\alpha^3$ (§ 7.2.2) thus:

$$\theta_s = C_s \alpha^3.$$  

With the aid of geometrical optics, we can determine the distribution of intensity across the emergent beam (see Fig. 194a) and the minimum beam cross-section is to be found in a plane situated at a distance $\Delta f = -\frac{3}{4} C_s \alpha^2$

---

**Fig. 194a.** The structure of an electron beam, symmetrical about the horizontal axis.

**Fig. 194b.** The wave surface of an astigmatic axially symmetric pencil.
from the Gaussian image plane where the radius of the beam is $\frac{1}{4} C_s \alpha^3$. The beam is described by a caustic curve and would become infinitely intense if diffraction were ignored. The intensity falls as soon as we move away from the caustic.

The constant $C_s$, which has the dimensions of length, depends both upon the form of the lens and upon its convergence. It satisfies a relation of the form

$$C_s = K_s f.$$ 

which shows why lenses with short focal lengths are of interest: the spherical aberration will be correspondingly small as well. Numerous values of this constant $K_s$ are quoted in §7.6, and in Chapter 17. Here, we simply mention that the best values at present are $K_s \approx 1$ for magnetic objectives (Liebmann, 1952) and $K_s \approx 50$ for electrostatic objectives (Bruck and Grivet, 1947). $C_s$ cannot be reduced to zero, so that to reduce spherical aberration, we can only reduce the aperture; in practice, apertures of 0·01 radians or less are used, but this means that the diffraction phenomena, which we had hoped would be negligible at such short wavelengths, reappear.

16.2.3 The Diffraction Aberration

Diffraction, the effect and the importance of which in a perfect lens we have already discussed, is of course also present in an uncorrected lens; to a first approximation, we assume that the effect is the same as the effect in a perfect lens of which the aperture is $\alpha$. The waves which are associated with the particles are diffracted by the diaphragms and spread the image of a point object out over a disc of radius $\varrho_d$ given by

$$\varrho_d = \frac{0.61 \lambda}{\sin \alpha} \simeq \frac{0.6 \lambda}{\alpha},$$

for small values of $\alpha$.

The diffraction aberration therefore decreases as $\lambda$ which, we recall, is related to the accelerating voltage $\Phi_o^*$ by the formula

$$\lambda = \frac{h}{p} = \frac{h}{(2m_0 e \Phi_o^*)^{1/4}},$$

and for electrons,

$$\Phi_o^* = \Phi_0(1 + 0.978 \times 10^{-6} \Phi_0).$$

It is therefore advantageous to work at very high voltages.

16.2.4 Calculation of the Theoretical Resolving Power

All we have to do now is to combine the effect of diffraction with the effect of spherical aberration. As Fig. 195a shows, the spherical aberration
coefficient $q_s$ and the diffraction aberration coefficient $q_d$ vary in opposite senses with the aperture of the beam; there must, therefore, be an optimum aperture, $\alpha_{opt}$, at which the resultant aberration is least. A number of different methods of making an approximate calculation of this optimum condition all give comparable results only slightly different from the rigorous theory which was first established by Glaser (1943) and later by Bremmer (1950).

Von Ardenne (1940) regarded the total aberration as the quadratic mean of the two aberrations; he assumed that the defects would be combined like the luminous intensities of two beams with independent phases, so that

$$ q = \sqrt{\left(\frac{0.6\lambda}{\alpha} \right)^2 + (C_s \alpha^3)^2}, $$

which is minimized when

$$ \alpha_m = C_1 \sqrt{\frac{\lambda}{C_s}} \quad \text{with} \quad C_1 = 0.77, $$

where it takes the value

$$ q_{opt} = C_2 \sqrt{C_s \lambda^3} \quad \text{with} \quad C_2 = 0.91. $$

In a note by Bruck (1947a) a comparison of the various conventions which have been suggested by the different authors is to be found—a straight-
<table>
<thead>
<tr>
<th>Distribution of particles in the beam</th>
<th>( C_1 )</th>
<th>( C_2 )</th>
<th>References</th>
<th>Position of the screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambert’s cosine law</td>
<td>1.4</td>
<td>0.43</td>
<td>Conrady (1919) Haine and Mulvey (1954)</td>
<td>Plane of least confusion</td>
</tr>
<tr>
<td></td>
<td>1.13</td>
<td>0.56</td>
<td>Glaser (1943)</td>
<td>Gaussian image plane</td>
</tr>
<tr>
<td>Gaussian law</td>
<td>0.92</td>
<td>0.783</td>
<td>Glaser (1950)</td>
<td></td>
</tr>
</tbody>
</table>

forward linear superposition would give \( C_1 = 0.47 \) and \( C_2 = 1.2 \) (Grivet, 1948).

The full calculation of the size of the diffraction disc either in the Gaussian image plane or in the plane of least confusion of a lens with third order spherical aberration has been made by Conrady (1919), by Glaser (1943, 1950), by Bremmer (1950) and by Haine and Mulvey (1954). The Kirchhoff formula has to be used, and some expression for the angular distribution of the electrons which emerge from the object point must be assumed. We quote below some of the values which have been obtained for the optimum aperture and the optimum resolving power; an example of the calculation involved in obtaining these is given in the book by L. de Broglie (1950).

Both the empirical methods and the complete calculations, therefore, lead to values of \( C_1 \) and \( C_2 \) close to unity. It is interesting to notice that the parameter \( C_1 \) has a physical significance; if we consider the spherical (Gaussian) wave surface and the wave surface perturbed by spherical aberration (Fig. 194b), it is easy to show that the extreme disparity between these two surfaces at the edge of the diaphragm is \( \frac{1}{4} C_1 \lambda^4 \), or in the optimum situation, \( C_1 = \frac{\lambda}{4} \).

This brings us back to the well-known rule formulated by Lord Rayleigh, which states that the spherical aberration should be corrected just up to the point at which the two wave surfaces mentioned above are a quarter of a wavelength apart. To correct further is no advantage, as it is by the diffraction that the resolution will be limited. Lord Rayleigh’s calculations, therefore, give \( C_1 = 1 \).

The parameter which characterizes the quality of an objective is therefore \( (C_0)^{1/4} \lambda^{3/4} \); with a magnetic objective, this quantity is about 5 Å (Liebmann, 1952), while it is about twice this value for an electrostatic objective. For the best magnetic and electrostatic objectives which are at present available, the best theoretical resolution which can be obtained varies between the following limits, according to the value of \( C_2 \) which is chosen:

- \( \rho_{\text{opt}} \) (magnetic): between 3 and 10 Å,
- \( \rho_{\text{opt}} \) (electrostatic): between 7 and 20 Å.
In Fig. 195b, the curves which describe the behaviour of the resolving power and the optimum aperture as a function of the spherical aberration constant are depicted. Detailed discussions of these problems have recently been given by Liebmann (1955) and by Haine (1954).

The calculations that led to the preceding values were based upon the best objectives used in practice, with which the object was not (or hardly) immersed in the field. Recent experimental work by Riecke (1962a, b) and Ruska (1962, 1964) has, however, shown that by using a special type of condenser, it is possible to employ an objective (condenser-objective) in which the object is situated at the centre of the field; we can thus take advantage of the fact that the spherical aberration is much lower than in the standard kinds of objective. Even with high fields (and saturated pole-pieces), the focal length of this condenser-objective remains short (1·64 to 2·1 mm) and the spherical aberration coefficient $C_s$ varies from 0·5 to 0·6 mm for electrons between 50 and 200 kV. At 1000 kV, we should still have $f = 3·6$ mm and $C_s = 1·1$ mm. The table below (Ruska, 1966) shows the theoretical resolving power of such an objective as a function of the accelerating voltage $\Phi_0$ when the objective is operating at an excitation
corresponding to \( k^2 = 3 \) (the distribution \( B(z) \) is treated as a Glaser bell-shaped curve).

<table>
<thead>
<tr>
<th>( \Phi_0 (\text{kV}) )</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f (\text{mm}) )</td>
<td>1.64</td>
<td>1.82</td>
<td>2.08</td>
<td>2.71</td>
<td>3.61</td>
</tr>
<tr>
<td>( C_s (\text{mm}) )</td>
<td>0.49</td>
<td>0.55</td>
<td>0.62</td>
<td>0.81</td>
<td>1.08</td>
</tr>
<tr>
<td>( \alpha_{\text{opt}} (10^{-2}) )</td>
<td>1.16</td>
<td>1.03</td>
<td>0.90</td>
<td>0.73</td>
<td>0.62</td>
</tr>
<tr>
<td>( \rho_{\text{opt}} (\text{Å}) )</td>
<td>2.94</td>
<td>2.29</td>
<td>1.77</td>
<td>1.23</td>
<td>0.98</td>
</tr>
</tbody>
</table>

The values of \( \alpha_{\text{opt}} \) and \( \rho_{\text{opt}} \) have been calculated for \( C_1 = 1.13 \) and \( C_2 = 0.56 \) (Glaser, 1945).

Despite the increase of \( C_s \) with \( \Phi_0 \), the immense value of raising the accelerating voltage of electron microscopes is obvious.

16.2.5 The Resolving Power for a Crystalline Lattice

The angular aperture of the beam after it has passed through the object is restricted by the objective aperture (usually to a value \( \alpha_0 = \alpha_{\text{opt}} \)). Electrons that are scattered through an angle \( \alpha_1 > \alpha_0 \) by a point in the object are halted, and this mechanism provides the main source of the contrast seen in the image; two neighbouring grains of a body, \( A \), will be distinguishable only if the scattering of electrons by the support, \( B \), has a different amplitude and provides enough contrast between the image of \( A \) and that of the background \( B \) (see Chapter 19).

In this case, the angular aperture of the beam illuminating the object is equal to \( \alpha_0 \)—this situation has been investigated in the preceding section: the image-forming electron beam is incoherent (there is no phase relationship between two electrons emerging from an object point).

If the object is a thin monocrystal, a new mechanism comes into play: diffraction of a proportion of the electrons at the lattice planes. A point \( A \) on a lattice plane is the source of two classes of electrons, “direct” electrons that are scattered by varying amounts and electrons “diffracted” through an angle \( \theta_1 \); between the two families, there is now a certain degree of coherence. If \( \theta_1 < \alpha_0 \), the objective will give a normal image of the point \( A \), formed by the cone of scattered electrons, together with a pseudo-image where the direct rays and the diffracted rays intersect. Because of the spherical aberration of the objective (the longitudinal aberration is \( \Delta z = C_s \theta_1^2 \)), the images will not be in exactly the same plane; on defocusing slightly, we can thus bring an “image” of the lattice planes into view. If the angular aperture \( \alpha_2 \) of the illuminating beam is considerably diminished (\( \alpha_2 \ll \alpha_0 \)), without altering the objective aperture, the diameter of the aberration patch corresponding to the pseudo-image \( A' \) will be reduced; furthermore, the slight defocusing makes the normal image of \( A \) less distinct, which enhances the contrast, and we can thus observe an “image” of the lattice planes on the screen with a much better resolu-
tion than that suggested by the formulae of the preceding section. The resolution can be increased still further by tilting the illuminating beam through an angle $\beta = \theta_1/2$ (Fig. 196). This "image" may be regarded as the interference pattern formed by the direct beam and the diffracted beam (see §19.2.6 and Dowell, 1963). The contrast and hence, the resolution, can be yet further increased by illuminating the object with as coherent a beam as possible; in particular, a special gun with very high brightness using a pointed cathode is employed.

16.3 THE RESOLVING POWER IN PRACTICE; THE INFLUENCE OF THE CHROMATIC ABERRATION

16.3.1 "Practical" Defects

The theoretical resolving power which we have just defined represents a limit which is attainable only if the instrument is constructed with extreme mechanical precision and is perfectly adjusted. It is an indication of the kind of resolving power we should hope for at the final image under the best conditions of observation—a very thin object, with very strong contrast. These optimum conditions are, however, extremely difficult to realize, and normally, the resolving power of every actual instrument is limited by the various mechanical or electrical defects, and by the difficulty of focusing the image exactly on the screen. In everyday working conditions, the practical resolution lies between 10 and 25 Å. The disparity between
this and the theoretical resolving power depends on a wide variety of factors, of which it is not easy to give a complete list.

The practical resolving power is a function of certain defects which we can place at first in the same category, the category which contains all defects which can always be reduced below some limit chosen beforehand by a suitable design and by taking special care. These are the defects which can be corrected; they are very numerous, and can originate in several different ways. Among the most important are the chromatic aberration, the ellipticity aberration, misalignment of the lenses, the influence of parasitic magnetic fields, and mechanical vibrations. These practical defects can be rendered negligible in the exceptional conditions which we have been calling “optimum”, by ingenious construction or correction procedures and by an astute choice of the conditions of observation; in this way, the practical and the theoretical resolving powers become almost identical in the best instruments: a factor of perhaps 2 or 3 separates them.

With the best modern magnetic microscopes, working at 80 or 100 kV, a practical resolution of 4 Å has been attained (Engel et al., 1962) at the image of a non-periodic object.

In the category which we have just defined, three classes can clearly be separated:

(a) Chromatic aberration, which may depend either upon the object or upon the electrical supply.

(b) Instabilities in the convergence of the objective, arising from fluctuations in the magnetizing current.

(c) Imperfect focusing (which we can regard as an error in the magnetizing current of the objective of a magnetic microscope, or as a voltage shift in an electrostatic microscope).

(d) Ellipticity aberrations (astigmatism).

(e) Residual aberrations (vibrations, parasitic fields, movement of the object heated by the beam, and other causes).

The last aberrations, (e), are negligible in good modern microscopes. We have therefore to arrange that the total aberration, (a)–(d), is as small as possible by careful construction of the lenses and their supplies.

We may consider these aberrations as the result of a number of statistically independent causes, as von Ardenne does, so that the practical resolving power is given by

\[ \rho_{\text{pract}} = \sqrt{\rho_{\text{opt}}^2 + \rho_{\text{chrom}}^2 + \rho_{\text{ellipt}}^2} \]

\[ = \rho_{\text{opt}} \left\{ 1 + \left( \frac{\rho_{\text{chrom}}}{\rho_{\text{opt}}} \right)^2 + \left( \frac{\rho_{\text{ellipt}}}{\rho_{\text{opt}}} \right)^2 \right\}^{1/2}. \]

Following Ruska (1966), we can gather all the aberrations (a), (b), (c) into the term \( \rho_{\text{chrom}} \).
The form of this expression reflects the method which is used to improve the practical resolving power by making adaptations to the objective; first the optimum resolving power is made as good as possible, then, once the instrument has been built, we endeavour to check this by direct observation. To do this, the aberrations of the other kind must be reduced empirically until they no longer enlarge the “image spot” perceptibly.

16.3.2 Chromatic Aberration

This aberration has been examined in detail in § 7.4. It is a consequence of the fact that the trajectories of the electrons depend upon their energy. As the electrons which leave some point on the object do not all have the same velocity, they will not converge into a point image but onto a small spot. The radius of this chromatic aberration disc (referred back to the object plane) is given by

$$\varphi_c = K_c \cdot f \cdot \alpha \frac{\epsilon}{\Phi},$$

where $K_c$ is a dimensionless constant which characterizes the shape and the disposition of the electrodes of the lens, and which we can call their “form factor”, $\alpha$ is the angular aperture of the beam, $f$ is the focal length, and $\epsilon$ is the variation in the energy of the electrons about the mean value $\Phi$. Any particular lens is characterized, therefore, by the value of the product $C_c = K_c \cdot f$ which has the dimensions of a length, and which can either be calculated by numerical integration if the form of the field is known (Glaser, 1940; Dosse, 1941; and Liebmann, 1953) or measured experimentally. We quote a few of the results:

<table>
<thead>
<tr>
<th></th>
<th>$K_c$</th>
<th>$f$ (mm)</th>
<th>$C_c$ (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetic objective</td>
<td>0.75</td>
<td>2.4</td>
<td>1.8   [Dosse (1941)]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[Liebmann (1953)]</td>
</tr>
<tr>
<td>Condenser-objective</td>
<td>0.6</td>
<td>1.82</td>
<td>1.1   Ruska (1966) (100 kV)</td>
</tr>
<tr>
<td>100 kV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrostatic objective</td>
<td>3.5</td>
<td>5.6</td>
<td>20  Boersch (1954)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>25  Bruck and Grivet (1950)</td>
</tr>
</tbody>
</table>

The coefficient $K_c$ cannot be reduced to zero by a judicious choice of the field distribution within the lens. It is $\epsilon$, therefore, that we must modify. We can, on the other hand, reduce the chromatic aberration considerably by working at very high voltage (Dupouy and Perrier, 1966).
16.3.3 The Causes of Velocity Dispersion

(i) The spectrum of the initial velocities

The electrons which leave the cathode do not all have the same velocity, and each source is characterized by a velocity spectrum which represents the "density" of electrons which have an initial energy $eV$; the most usual example is provided by the thermal electrons which have a Maxwellian velocity distribution which is described by

$$n(V) \, dV = \left( \frac{e}{kT} \right)^2 \exp \left( -\frac{eV}{kT} \right) V \, dV,$$

in which $n(V) \, dV$ is the fraction of the electrons emitted with an energy between $eV$ and $e(V + dV)$, $k$ is Boltzmann's constant, and $T$ is the absolute temperature. The maximum of the curve, the abscissa of which represents the most probable energy, occurs at

$$V_m = \frac{kT}{e} = \frac{T}{11,600},$$

or about 0.25 volt for normal working conditions ($T \approx 2700^\circ$K) with a tungsten filament. The majority of the electrons are emitted with velocities between zero and one volt, and about one half of these have a velocity below about 0.45 V, so that $\varepsilon \approx 0.45$. Calculating numerically with these data, we find that for a magnetic objective, the resultant aberration can be reduced to 0.03 m$\mu$, but that even when it is ten times larger (as it is in an electrostatic objective), its effect is still negligible in comparison with the other aberrations. Finally, we should point out that in principle we can reduce this aberration by using a cathode which works at a lower temperature (a matrix or sintered oxide cathode).

We can equally well produce the electrons in a gas discharge at reduced pressure. In this way, Induni (1947) and Wegmann (1953) have obtained images with a resolution of about 20 Å, which is an indirect proof that the dispersion of the initial velocities remains below $\pm 0.5$ V. Möllenstedt and Düker (1953) have recently been able to measure the velocity distribution curve of this source directly, and of a variant upon it suggested by Hailer (1938). The two curves have the same half-width, of about 2 eV, irrespective of the nature of the gas and of the material of which the cathode is composed. Figure 197, which was obtained by Möllenstedt and Düker, shows the distribution curves (A) for a discharge tube at 40 kV and (B) for the thermal electrons which are emitted by a tungsten wire at 2900$^\circ$K. The spectrum for a cold cathode is two or three times more widely spread out than the curve for thermal electrons. At the larger velocities, it has a long tail but fortunately the corresponding intensities are small enough to be negligible.
(ii) The spectrum which describes the interaction with the object

Transmission through the object results in energy losses of the order of a few electron volts (up to 10 or 30 V), and we shall examine them in some detail. Since the losses are discrete, we can filter out the electrons which have been slowed down, and on this point, the reader is referred to Chapter 19. At present, we shall simply emphasize that these discrete losses represent the ultimate source of chromatic aberration in very thin objects (20 mμ). When the thickness is made greater, these losses become increasingly more important, and affect more and more of the electrons, but the width of the energy spectrum of the electrons that have been slowed down decreases very rapidly when the operating voltage is increased. Dupouy and Perrier (1966) have calculated that the chromatic aberration arising from these losses is twenty-five times smaller at 1000 kV than it is at 100 kV.

(iii) Fluctuations of the potential and the current

A variation in the energy of the electrons can also be a result of a fluctuation of the accelerating potential. The problem is very different according as the lenses are electrostatic or magnetic. In the magnetic case, a fluctuation of the lens current is effectively equivalent to a fluctuation of the voltage.
(a) Magnetic objective. The trajectory equation in a magnetic objective
(§ 5.2.2) depends on the function \( \frac{\Phi}{B^2(z)} \), in which \( \Phi \) is the accelerating potential
and \( B(z) \) is the magnetic induction on the axis, and is a function of the
current, \( i \), in the coils. If \( i \) is small, the iron is not saturated and \( B(z) \) varies
in proportion to \( i \). If we are trying to work under the optimum conditions,
(small spherical aberration) however, we are forced to use strong lenses and
hence large values of \( i \). The iron, in this situation, is saturated and the
function \( B(z) \) is deformed so that when \( i \) is increased, the ordinate increases
too but less at the centre of the bell-shaped curve by which \( B(z) \) is
represented than at the tails. The curve flattens out and its half-width
increases, for the reason that the permeability of the iron falls while \( B \)
increases.

In general, the convention suggested by Haine (1954) is adopted, to fix
the degree of stabilization. This requires that the radius of the chromatic
aberration disc which is produced by fluctuations of the high tension shall
always be less than one tenth of the optimum resolving power. This con-
dition is stringent but feasible.

If we are far from saturation, \( \phi/i^2 \) must remain constant since the focal
length, for any given geometrical arrangement, depends upon this parameter
(§ 9.4.3). Experiment shows that the stabilization must be to at least a fraction
of one part in ten thousand, both for the high tension and for the current.
This is a technical problem which can now be solved commercially, while in
the laboratory the limit can even be one part in \( 10^5 \) (Haine, 1950; Bruck, 1953).

With saturation, the phenomena become more complex; Liebmann (1953)
and Lenz (1953) have thoroughly explored this situation to discover whether
it is in fact true, as many designers have believed, that the increase in the
half-width of the field can compensate the increase in the total excitation.
If this were so, we could relax the conditions on the stability of the supply
to the lenses appreciably, and then operate them in these conditions at
the maximum magnification. This research shows that it is possible to
construct lenses with very little aberration, but not, however, without any.
The anticipated advantage does indeed exist, but is not as pronounced
as had been hoped. Even in this case, extreme stability is necessary, but
\( 10^{-4} \) or even a little less can be tolerated. One further point is that the chro-
matic aberrations of the magnification and the rotation must be considered;
fortunately, these latter can be completely corrected (Ito, 1955). This
problem can be solved most elegantly by coupling the fluctuations of \( \Phi \n\) and those of \( i \) electronically, when both are caused by fluctuations of the
mains supply to the whole installation, in such a way that the condition

\[
\frac{2\Delta B(z)}{B(z)} = \frac{\Delta \Phi}{\Phi}
\]

remains satisfied as exactly as possible (Dupouy et al., 1964).
We should mention that there are no current fluctuations at all when superconducting lenses running in the persistent-current mode are employed (see Chapter 9).

(b) Electrostatic objective. In the electrostatic version, stabilization of the potential is far less important an exercise. The differential equation of the trajectories is, in fact, homogeneous so far as the potential is concerned (§ 4.2.2) so that if all the potentials are multiplied by the same factor, the focal length does not vary. So that this condition shall be satisfied, all the potentials are obtained from a single source by a system of potentiometers. For tensions of the order of 60 kV, this precaution becomes inadequate, as there is a slight aberration which is relativistic in origin (§ 4.4) and the differential equation of the trajectories is no longer homogeneous with respect to the potentials; a stabilized high tension supply is again necessary, but stabilized to a lesser degree than in the case of magnetic lenses since only a correction term has to be eliminated: a stabilization of about one part in a thousand is adequate. This simplification is one of the main advantages of using an electrostatic system.

16.4 THE PRACTICAL RESOLVING POWER;
THE INFLUENCE OF DEFECTS IN THE MECHANICAL CONSTRUCTION

16.4.1 The Various Causes of Asymmetry

So far, we have been assuming that the fields within the lenses have perfect axial symmetry. In practice, however, this ideal state can be achieved neither for electrostatic nor for magnetic lenses, and at present it is not possible to reach resolving powers of the order of 10 Å even when the lenses are very carefully constructed. Even though the most expensive machine-tools are available, it is not possible to achieve a mechanical precision sufficient to avoid an ellipticity of the field large enough to have a detectable effect. Further, in the magnetic case, there will be local variations in the permeability of the iron which have a more pronounced effect the nearer the iron is to saturation; at present, this heterogeneity cannot be eliminated by metallurgical treatment.

The magnitude of the mechanical astigmatism of a magnetic lens is also related to the size of the metal grains. An examination by Leisegang (1956) showed that for an identical treatment, a nickel steel (with 50 per cent Ni) with grains of 30 μ produces an astigmatism which is thirty times smaller than that of carbon steel with three millimetre grains.

In electrostatic objectives, the leads also produce a lack of symmetry in the field (Bertein, 1949).
A supplementary source of astigmatism appears when the instrument has been in use for some time, as the field inside the lens becomes modified by the small quantities of insulating deposit which are produced by the interaction between the electrons and the residual gas in the evacuated tube and also, quite often, by fragments of the object which break off and become fixed to the electrodes of the objective. Also, a brown substance is deposited continuously on the stop of the objective in the region which is struck by the electron beam; Ellis (1951) and Ennos (1953, 1954) have made a particular study of this contamination, as this insulating layer, which sometimes increases at 60 Å per second, conducts away the incident charge only with difficulty, and is not necessarily axially symmetrical. The parasitic electric field which is produced can result in an important astigmatism, and the influence of this defect is increased as the dimensions of the diaphragm are reduced. This contamination of the objective diaphragm seems to be less rapid when we use a mercury diffusion pump but in fact oil pumps are usually employed because of the cumbersomeness of the liquid-air supply which is required for trapping mercury vapour in the pump. We should, however, point out that although the use of a mercury diffusion pump does not eradicate all the sources of organic vapour (elastic joints, spots and residual deposits of cleaning fluids, for example, always exist within the system) it does reduce them in number.

The real solution of this contamination problem is, however, to surround the zone containing the object and aperture as completely as possible by a metal screen; the latter is vigorously cooled with the aid of a reservoir of liquid nitrogen (Heide, 1964; Ruska, 1966), although the temperature of the specimen-holder and the aperture are kept close to the ambient temperature. The same zone of the specimen can then be studied for several hours at very high magnification without any noticeable decrease in the resolving power (and without any carbon contamination appearing at the object). Reduction of the size of the electron beam is highly advisable; the quality of the vacuum in the microscope column should also be improved, by lowering the pressure but above all, by achieving a “cleaner” vacuum, free of organic molecules. The use of an ultra-high vacuum \((10^{-8} - 10^{-9} \text{ mm Hg})\) with ion pumps, in conjunction with cooled zones, seems to be inescapable in future.

16.4.2 Ellipticity Astigmatism (§ 7.5.2)

Ellipticity of the diaphragms of the lenses produces an astigmatism of the beam which we have considered in detail in § 7.5.2. The best “focus” is obtained when the circle of least confusion lies at the screen—the radius of this circle \(r_e\), is given by

\[
r_e = C_e \alpha,
\]
in which \( C_e \) is the constant of ellipticity aberration which has been calculated by several authors (Bertein, 1947a, 1947b and 1948; Bruck and Grivet, 1947; Sturrock, 1951; and Hillier and Ramberg, 1947). We mention only that the order of magnitude of the coefficient is given by

\[
C_e = \delta a,
\]

in which \( \delta a \) is the difference between the axes of the ellipse which forms the outline of the diaphragm. Present-day methods of machining are not capable of reducing the difference between the axes to below a micron, so that if the optical aberration is to be reduced, \( \alpha \) must be diminished; we distinguish, therefore, between two cases.

In the first of these, the value of \( \varrho_e \) which corresponds to the optimum aperture of §16.2.4 is considerably smaller than \( \varrho \) (more than ten times, for example). This implies that

\[
\varrho_e = \delta a \cdot C_1 \sqrt{\frac{\lambda}{K_s f}} < 0.1 C_2 \frac{\lambda}{K_s f},
\]

or, since \( C_1 \simeq C_2 \simeq 1 \),

\[
C_e = \delta a < 0.1 \sqrt{\lambda K_s f}.
\]

In this case, we use the optimum aperture, but this does require an objective which has been carefully machined and corrected.

In the other case, where the objective is only passably good, we must not work at the optimum aperture, as \( \varrho_e \) would then be much larger than \( \varrho_{\text{opt}} \) which would be absurd. The aperture must be reduced, but this means that the diffraction becomes all-important while the spherical aberration which behaves like \( \alpha^3 \) has virtually no further effect. The roles have changed, for the problem now is to determine the optimum aperture and resolving power by combining the astigmatism and the diffraction which are now the two defects which are important. Using a convention similar to the one described in §16.4.2, Bruck and Grivet (1947) found that

\[
\varrho_{\text{opt}}' = 0.85 \lambda \sqrt{\frac{C_e}{\lambda}}, \quad \alpha_{\text{opt}}' = 0.71 \lambda \sqrt{\frac{\lambda}{C_e}}.
\]

We must, however, verify that the spherical aberration is indeed considerably less than \( \varrho_{\text{opt}}' \) (more than ten times smaller, say). This requires that

\[
\varrho_s = K_s f (0.71)^3 \left( \frac{\lambda}{C_e} \right)^{3/2} < 0.1 \times 0.85 \lambda \sqrt{\frac{C_e}{\lambda}},
\]

or in practice, replacing \( C_e \) by \( \delta a \),

\[
\delta a > 2 \lambda \sqrt{K_s f}.
\]

When this condition is satisfied, we should work at the aperture \( \alpha_{\text{opt}}' \).

If, in conclusion, the constants of the objective are such that \( \delta a \) lies in the transition region between the above limits, the three defects have to be combined to give the optimum operating conditions.
Fig. 198. The practical resolving power. The dotted curve represents operation with the optimum aperture calculated as in § 16.3.1; the full curve represents the case when the optimum aperture is calculated as in § 16.4.2.

To obtain the practical resolution, the various aberrations are combined in the way described in § 16.3.1. Figure 198 shows the way in which the resolution varies as a function of $C_e$ in two cases, in one of which the aperture is $\alpha_{\text{opt}}$, and in the other, $\alpha_{\text{opt}}'$. An elliptical objective will be operated at $\alpha_{\text{opt}}'$, but it will never be very bright since $\alpha_{\text{opt}}'$ is always markedly less than $\alpha_{\text{opt}}$, and the difference between these has a very pronounced effect upon the brightness, which is proportional to $\alpha^2$.

16.4.3 Observation of the Ellipticity Astigmatism

Measurement of the ellipticity astigmatism of an objective is extremely valuable, as it gives us a direct overall measure of the quality of its construction, and suggests the characteristics of the “stigmator” which will be necessary to correct the aberration. Various methods of detecting and measuring the astigmatism in the image have been suggested.

One of the simplest methods (Hillier and Ramberg, 1947) is based on an examination of the Fresnel fringes which are formed at the edge of an opaque object, the objective being slightly defocused in such a way as to produce dark fringes outside the shadow (the “real diagram”; see Fig. 199a).

We suppose first of all that the objective is free of astigmatism. The image of any object (specks, of smoke-black, for example) is surrounded by a diffraction fringe which is particularly clearly visible if the object is illuminated at a small angular aperture. The appearance of this fringe is markedly different according to whether the convergence of the objective is slightly higher (overfocused or positive fringe) or slightly lower (under-
focused or negative fringe) than the value which corresponds to an exact focus.

The negative fringe appears as a brilliant line, attached to the edge of a dark image. The edge of this image is, therefore, extremely clear, as the contrast is far sharper than it is when the focusing is exact, the situation which in fact corresponds to minimum contrast. This explains why one tends always to underfocus the image slightly when focusing by eye. The positive fringe, on the other hand, is a dark fringe, again running around the image, and separated from it by an interval which becomes wider as the overfocusing is made more pronounced. When the image is exactly in focus, the fringes disappear, and the edge of the image seems slightly blurred; the contrast is clearly a minimum, and a trained operator can use this as a criterion.

We consider now an objective which has a certain ellipticity astigmatism; if the image is observed between the two line foci, the edges of the image which are parallel to one direction will be accompanied by a positive fringe, while the edges parallel to the perpendicular direction will have a negative fringe. Detecting a slight astigmatism, therefore, involves being able to distinguish positive fringes which are only very slightly separated from the image and which correspond to a very slight overfocusing (since the image which is being examined lies between two focal lines which are themselves very close together). The possibility of doing this is itself related to the clarity of the fringes which certainly depends upon the aperture of the illuminating beam but also on the stability of the image. It is clear, therefore, that for the astigmatism to be visible, the apparatus must be absolutely clean, and free of mechanical or electrical instabilities. When this has been achieved, however, the fringes can be made clearly visible simply by reducing the aperture of the illumination, even though the overfocusing be very small indeed, and it is easy to make them symmetrical with the aid of a correction system. Towards the end of the correction process, however, it will be necessary to work very close indeed to the true focus, as any but a small overfocusing produces a fringe far from the object and symmetrical in appearance even when the objective has a pronounced astigmatism (on this point, examine the top part of Fig. 199b).

The correction procedure consists of adjusting the stigmator (§ 7.7.5) until the Fresnel fringes are identical in all directions. The best test object is provided by the holes in a film of collodion covered with a thin layer of some heavy metal or of carbon (Koenig, 1951) to increase the contrast of the fringes. The sensitivity of the method is limited by the problem of seeing the fringes on the screen, as the distance between the fringes and the image is of the same order of magnitude as the resolving power. According to Haine and Mulvey (1954), if a sufficiently intense illumination were used (between 2 and 5 A cm\(^{-2}\)), with a fine grain screen, and a high direct magnification of the order of 50,000 (200,000 is possible in the Siemens UM 100)
Fig. 199a. Microphotographs of grains of smoke-black (Photos: Baillie and Castaing).

The objective is free of astigmatism. On the left, overfocusing with a positive fringe; in the centre, in focus, minimum contrast; on the right, underfocusing, with a negative fringe.
Fig. 199b. Microphotographs of grains of smoke-black (Photos: Baillie and Castaing). The grain in question is indicated with an arrow. The objective is slightly astigmatic, and the arrow points in the direction of one of the line foci: top overfocusing—positive fringe; centre, focusing onto the circle of least confusion—the fringe is partly negative and partly positive; bottom, underfocusing—negative fringe.

together with an additional microscope with a magnification of ten times, it would be possible to correct an astigmatism visually until the focal lines became less than 0.05 μ apart. Leisegang (1954), who followed an analogous procedure, only achieved Δf = 0.16 μ. A better correction can be obtained, although more laboriously, by directing the correction procedure step by step photographically.

Other methods of observing the astigmatism have been suggested, and for further information the reader is referred to Part I and to the
original publications (Castaing, 1950a and 1950b; and Bertein and Regenstein, 1949). Van Dorsten (1948, 1950) has shown that the astigmatism can be observed directly by magnifying it artificially; to do this, the beam is deviated to and fro by an alternating magnetic field in the vicinity of the gun. The effect is of increasing the aperture, and a very faint astigmatism becomes visible.

16.4.4 Correction Systems (Stigmators)

To correct the ellipticity astigmatism, a correcting device has to be inserted within the field of the lens; this corrector has to be free to rotate about the axis in such a way as to restore perfect axial symmetry to the field, and it was this type of system which was used by Hillier and Ramberg (1947). At first, they tried to obtain a sufficiently symmetrical system by constructing the pole-pieces with the utmost care, but they found that the field asymmetries could not be reduced below 0.5 μ (which corresponds to 10 Å aberration) as the inhomogeneities in the iron and the parasitic fields of the small cracks which remained were still uncorrected. To improve this arrangement, they then inserted eight iron screws, each inclined at 45° to the next, in the outer part of the space between the pole-pieces. These screws modified the field distribution, and were adjusted in such a way as to eliminate all perceptible asymmetry from the image. RCA microscopes have been fitted with a device of this type for a long time now. The correction process is long, as the operation of the microscope has to be interrupted a great many times to withdraw the pole-piece, and it may take a whole day to achieve complete correction. A considerable improvement is to use a magnetic objective with double windings (Le Poole and Le Rutte, 1949), which leaves a space free for the correction screws which can then be adjusted from outside the instrument while it is still working. Leisegang (1954) has described a mechanical system of this kind, in which two small blocks of soft iron (0.1 mm³ in volume) are moved about in the neighbourhood of the lower face of the objective, while remaining diametrically opposed; by adjusting their height and their azimuth, a good degree of correction can be obtained in a few minutes. This method cannot, however, be transferred very simply to the case of electrostatic lenses, as the region near the central electrode becomes highly susceptible to electrical breakdown if electrodes with small radii are introduced there.

The other method of obtaining correction, which is far more flexible, is copied from an optical method. The basis of this method, which was proposed by Bertein (1948), is the creation of a supplementary potential along some particular stretch of the beam, the effect of which is to

† This method may well have been thought of earlier, and independently, by Mahl, who mentioned it in 1944 in a request for a patent, but never published it.
correct the harmful effects of ellipticity in the neighbourhood of the optic axis by producing an equal asymmetry of the field of the opposite sense. We have explained the principle of this in §7.7.5.

A very simple corrector can be constructed by producing a transverse electric field with a combination of four plates (Fig. 200) each at right angles to the next. These constitute a negligibly convergent cylindrical lens which is markedly astigmatic, in that it is convergent in the plane \(xOz\) and divergent in a plane normal to this; it is, in fact, a very weak quadrupole lens (cf. Chapter 10). As the line foci of this lens are not necessarily parallel to those which are produced by the astigmatism of the lens, provision has to be made for some kind of adjustment with which either the correcting device or the field alone can be rotated about the axis. This can be arranged either mechanically (Haine, 1954), or by rotating the field by adding a second system of electrodes, rotated through 45° with respect to the first set, which are of a new shape (Rang, 1949; see Fig. 200). With the two electrical controls both the direction and the intensity of the astigmatism can be varied. The earlier design has been simplified by earthing the negative electrodes which reduce to a single pair (Fig. 200).

In another of the possible methods of obtaining correction, a field with elliptical equipotentials with variable eccentricity is created in the neighbourhood of the optic axis by means of six electrodes in the form of a regular hexagon: a “sextupole” lens. To these is applied a potential of the form

\[ V_k = v \cos \left( \theta - 2 \frac{K \pi}{3} \right), \]

in which \(K\) represents the number ascribed to one of the electrodes which are numbered from 0 to 5. This produces an astigmatism which is proportional to \(v\) in magnitude, and of which one of the line foci is inclined at an angle \(\frac{1}{2} \theta\) to the zero-th electrode. Systems of this kind have been built by Rang (1949) and by Castaing (1950a).

Each of these corrective systems is usually placed as close to the exit face of the objective lens as possible, as this is the place at which the diameter of the beam is smallest. The corrector can therefore be smaller, and as the

---

**Fig. 200.** Diagrams showing some stigmators.
distances both between the electrodes themselves and between the electrodes and the beam are slight, only small potentials are required, which is very convenient. Bertein (1948) calculated that for an ordinary electrostatic objective, with electrodes at a radial distance of 5 mm, potentials of the order of a hundred volts would be adequate for correcting an aberration as high as $\Delta f = 10 \mu$. Only the form of the field in the neighbourhood of the axis matters; the electrodes can be bars either parallel or perpendicular to the axis, or planar or curved plates, and generally speaking a well-designed stigmator requires a supply of only a few hundred volts.

The optic axis of the stigmator must be aligned to coincide with that of the objective, as otherwise the consequent lack of symmetry would prevent a complete correction from being possible. To satisfy this condition, either the stigmator is provided with a mechanical adjustment or alternatively, deflecting plates with which it can be displaced are set round the beam at some point before it reaches the stigmator; and there is yet another possibility which is to supply the stigmator itself unsymmetrically, which shifts its optic axis; in any case, it is very easy to check whether the alignment has been successfully achieved since if the centring is poor the image moves when the intensity of the field in the stigmator is varied. The optimum position can be found very rapidly by temporarily applying a small alternating potential to the corrector.

16.4.5 Defects Produced by Misalignment of the Lenses

These are less serious defects than the ones discussed above, as they can be made negligible by building the system carefully. The electrostatic lens case has been studied theoretically (Bruck, 1947a; Bruck and Grivet, 1947). We differentiate between defects in the centring, and defects in the inclination. Let us suppose that all the diaphragms are perfectly round and parallel to one another. If the central plate of one of the lenses is eccentric by a distance $\Delta e$ with respect to the common axis of the two outer plates, an astigmatism is produced with constant

$$C_e = C_s \left( \frac{\Delta e}{f} \right)^2.$$

If a perfectly round lens is tilted through an angle $\beta$ with respect to the optic axis of the system, an astigmatism is again produced,

$$C_e = C_s \beta^2.$$

The calculation for a normal electrostatic lens (which has $\alpha = 3 \times 10^{-3}$, $C_s = 10 \text{ cm}$, and $f = 0.5 \text{ cm}$) shows that a $10 \AA$ aberration corresponds to $\Delta e \simeq 10^{-3} \text{ cm}$ or $\beta \simeq 2 \times 10^{-3}$. It is easy to centre the three electrodes to within a micron, by using a reference spindle which is provided with
an optical microscope for sighting. Any mechanical tilt of the electrodes can be compensated by suitably inclining the beam, by adjusting the gun in fact. The beam has to be made to pass through the middle of the central electrode and the adjustment is checked by varying the potential of the objective by a few hundred volts (for a total potential of 50 kV); if the system is well adjusted, the image on the screen should not move, as the effect of a variation $\Delta V$ in the potential is a displacement given by

$$\Delta y = G \cdot K_c \frac{\Delta V}{V} \cdot \Delta e,$$

in which $G$ is the total magnification and $K_c$ is the chromatic aberration constant of the objective. With the aid of this relation, we find that if the aberration is to be less than 10 Å, the image must not move more than 1 or 2 mm for $\Delta V = 250$ V. In conclusion, we would mention that it is difficult to ensure directly when the instrument is being built that the inclination of the objective with respect to the axis of the projective lens or of the beam with respect to the axis of the objective is smaller than $2 \times 10^{-3}$ radians.

16.4.6 Parasitic Fields

(i) Electric fields

The metallic case of the instrument eliminates the influence of any outside electric field, but very often, charges which originate in the electron beam itself (direct, scattered or secondary electrons) accumulate on any insulating materials which are directly in the path of the beam and produce parasitic electric fields which usually vary in an irregular way. The result is an irregular displacement of the image which depends on the intensity of the beam. These insulating materials can be fragments which have broken away from the object and fallen into the lens, or insulating deposits which have formed on the metal surfaces. Further, any metallic part which is struck by an intense beam of electrons rapidly becomes covered with a brownish layer of some insulating substance, or contamination, as we have already mentioned in § 16.4.1.

Generally speaking, a microscope cannot be kept in proper working order without a thorough cleaning at least every six months, and a weekly cleaning of the parts which are close to the beam. Wherever possible, steel wool is used for the cleaning while the more delicate parts are rubbed with a tuft of cotton-wool impregnated with alumina. The grids upon which the object is supported are cleaned in 20 per cent formic acid, or heated to red-heat (Eaves, 1952).

The contamination proves to be particularly troublesome in that it forms both on the object under observation and upon the contrast dia-
phragm of the objective, and this prevents the observation of any one region of the specimen from being very prolonged. Indeed, in certain unfavorable cases such as thin metal layers, a few seconds bombardment contaminates the object sufficiently badly for observation at high resolution not to be possible; focusing under these circumstances is most difficult, and various ways of reducing the contamination or even suppressing it entirely have been proposed. Heating the surface which is being bombarded to 200° is most effective (Ennos, 1953), but the object which is being examined is not always left unscathed by such treatment. Leisegang (1954) places the object in a region which is chilled, so that any organic vapours will be condensed. Sweeping the instrument with a gaseous current (which often happens involuntarily, as a result of small leaks) seems to reduce the contamination within the system considerably as well; directing a jet of air at low pressure onto the surface of the specimen reduces the contamination appreciably and can even get rid of it completely in the regions which are bombarded heavily (Castaing and Descamps, 1954). At present, it is contamination which represents the main obstacle to obtaining high resolutions conveniently. If the specimen can support this, the contamination can be remedied by bombarding it with ions during the examination itself—thin metallurgical sections (Castaing, 1954) and massive specimens observed by reflexion (Fert, 1954) are examined in this way.

Another cause of instability can originate in the electrostatic charge of the specimen itself. If an intense illumination is used, the electrons which are stopped by the object may not be able to leak away sufficiently rapidly if the object is a poor conductor, and the object charges up. Elsewhere, where the secondary emission coefficient is high, there may be areas which are positively charged. It is possible, therefore, for different zones to become charged with opposite signs, and hence to repel or attract one another. The object, as a result, becomes deformed as various parts creep slowly about and these electrostatic forces are even sufficient, quite often, to tear the thin film which supports the object. A metal layer deposited under vacuum first of all strengthens the plastic film mechanically, but above all it has the valuable effect of making it conductive, which gets rid of the harmful charges by conducting them away to earth, or at least reduces their influence considerably.

(ii) Magnetic fields

The electron beam is deflected at points along its trajectory by parasitic magnetic fields. If this were absolutely necessary, the influence of constant fields could be eliminated by adjusting the centring of the beam, but this would introduce an astigmatism which would in turn have to be corrected with a stigmator, and it is better to eliminate any appreciable field by adequate screening. Variable fields produce a lateral vibration of the beam,
which blurs the image in a particular direction; these magnetic fields are usually the result of transformers, motors or leads through which intense currents are passing, which have been placed too close to the instrument. The need to choose a good single earth connection for the whole instrument has been demonstrated by Grube (1949), as the use of "multiple earths" can lead to intense currents, and hence to harmful magnetic fields.

Von Ardenne (1940) finds that if the resolution is to be 1 μm, any residual alternating field perpendicular to the axis must be reduced to about 10^{-4} gauss, which is a difficult condition to achieve. The most sensitive parts of the beam are those at which the velocity of the electrons is reduced, and in particular, the cathode, the centres of electrostatic lenses (Mahl, 1947), and the region immediately beneath the objective. In an electrostatic microscope, the beam is screened by surrounding the whole of its trajectory with one or several concentric mumetal (or permalloy) cylinders between 1 and 5 mm thick; at the objective, the beam is much more carefully screened, and certain parts of the lens are solid mumetal. In the magnetic microscope, the same precautions are taken for the regions between the objective and the projective lenses, and between the latter and the screen; nevertheless, there are other critical zones, at the lenses themselves and above all at the objective—the metal from which the pole-pieces are made is too close to saturation to be much help, and it is the yoke which is primarily effective which explains why it is of generous dimensions. Tables of the "coefficient of reduction" which describe the effect of cylindrical screens are to be found in the books by von Ardenne (1940) and by Kaden (1950) and in the article by Charles (1947) for various geometrical arrangements; Charles also describes a method of verifying the efficiency of the screening by a measurement of the induction.

16.4.7 Indistinctness and Disturbances Produced by Mechanical Movements

It is difficult to distinguish these defects from the one described above simply by examining their influence upon the image, but their effect is often most pronounced.

A common occurrence is thermal migration at the object; every part of the object which is bombarded by the electron beam has a rather high temperature, which produces local expansion (cf. § 20.7). The thermal equilibrium (at which this displacement ceases) is only reached quickly if the object-support is a good enough conductor of heat; for this reason, and also because they are more heat-resistant, electrolytic grids are preferable to woven grids. Conversely, it is harder to reduce this migration if a transparent object-support is used, as this is subject to a number of stresses, and is very flexible. An effective solution is to work at as low an intensity as possible, and the heating of the object is certainly reduced if the illumination is reduced to a very narrow region (a few microns, say) around
the point being examined. A condenser lens and suitable diaphragms seem therefore, to be an advantage, since the brightness of the beam is increased, and hence the total intensity at the object may be diminished. The mechanical vibrations which are transmitted to the system from the ground, from the ducts of the pump (as a result of the motor of the pump or of boiling oil or mercury) or which are created by certain of the mechanical accessories of the instrument (the shutter of the camera, for example) have the particularly troublesome effect of displacing the object-support relative to the objective; in some instruments, they can also produce a slight oscillation of the lens system and hence of the optic axis.

16.5 "REVOLUTIONARY" ATTEMPTS TO IMPROVE THE RESOLVING POWER

We shall conclude this chapter with a few details of the attempts which have been made to improve the resolving power by modifying the very principle upon which the instrument operates. Theoretically, spectacular progress is to be anticipated, but in practice, the same is not true as the host of technical problems are as yet only poorly solved. The instruments which we are about to describe cannot compete with the ordinary electron microscope.

16.5.1 The Ion Microscope

(i) The reason why ions are of interest

The expression for the optimum resolution

$$\theta_{\text{opt}} = C_2 \lambda^{3/4} C_{s}^{1/4},$$

shows that we could make a great deal of progress if we were to use ions instead of electrons, as the wavelength of the waves which are associated with ions is about a hundred times shorter than the value which corresponds to electrons accelerated through the same potential. If the ion which is used is of mass $M$, and the electron mass is $m$, the gain in resolving power is described by the factor $(M/m)^{3/8}$, or 17 for $H^+$, 43 for $\text{Li}^+$ and 105 for $\text{Cs}^+$. The optimum aperture

$$\alpha_{\text{opt}} = C_1 \left(\frac{\lambda}{C_s}\right)^{1/4},$$

falls in the ratio $1: \left(\frac{m}{M}\right)^{1/8}$; for example, with $\text{Li}^+$ ions, the aperture at which the best resolution would be obtained would be 3.2 times smaller than the optimum aperture of the objective in the case of electrons. In so small an
aperture, it is difficult to obtain a sufficiently intense ion beam, but as the potential gain in resolving power is so huge, it is still appreciable if a larger aperture than $\alpha_{\text{opt}}$ is used. Unfortunately, the chromatic aberrations and the mechanical aberrations are as important in the ion microscope as in the electron microscope and although modern technical advances have enabled these defects to be reduced to less than the optimum resolving power for electrons, the relative importance is reversed in the ion situation. At present, there is always a mechanical or chromatic aberration which is the essential factor in limiting the practical resolving power.

(ii) Experimental arrangements

The suggestion to use ions instead of electrons in the microscope was made originally by Knoll and Ruska as early as 1932. Koch (1936) and Walcher (1943) obtained the image of a grating with a resolution of a millimetre.

The problem was later taken up again more successfully by Boersch (1942–7) who had at his disposal an instrument which performed excellently with electrons (with a resolution of 28 Å). With ions, the photographs which he took of zinc oxide corresponded to a resolution of only about 500–800 Å, and after various attempts, Boersch abandoned his work, faced as he was with such extreme experimental difficulties; in his conclusion, he expressed the opinion that it would be better to find ways of perfecting the electron microscope than to try to use ions. The source which he was using was of the Kunsmann (1926) type, modified in the ways suggested by Koch (1936) and Walcher (1943), and produced Li$^+$ ions of too feeble an intensity to produce a visible image on the screen. Boersch was obliged, therefore, to make the preliminary adjustments with electrons. The direct magnification of his photographs was between 1000 and 2000, but even with five or ten minute exposures, the pictures were very under-exposed.

A more systematic attack upon the problem was made by Magnan and Chanson (1945, 1949), who demonstrated the value of using protons, and a high tension of 300 kV. They designed a microscope which was adapted to operate in this new way, but which was extremely difficult to build and adjust. Photographs which correspond to a useful magnification of 3000 have been published (Chanson and Magnan, 1951) with an operating potential of 50 kV. More recently (Magnan and Chanson, 1955), the practical resolution has been forced down to 10 or 20 Å in visual examination, but so far no photograph at a high magnification has been published. The protons were provided by a high-frequency source of the Thonemann (1948) type, with the modifications of Moak et al. (1951). Certain improvements enabled Magnan and Chanson (1955) to obtain a very stable source, with a high electrical efficiency, providing 90 per cent
of atomic ions, but the velocity spectrum of the ions (which is wider than a volt) produces a large chromatic aberration which limits the resolution. A much simpler arrangement has been built by Gauzit (1951, 1953 and 1954a) who converted an ordinary commercial microscope by replacing the electron source by a source of alkali ions, the latter due to Couchet (1951, 1954). At 60 kV, the practical resolving power which he was able to obtain was 80 Å, effectively half as good as the value obtained with electrons. The photograph in Fig. 3 was taken under these conditions.

The experimental difficulties are very numerous; photographic plates and fluorescent products which give good results with electrons are no longer so good with ions. Plates with a high concentration of AgBr, which had been intended for nuclear physics, can be employed. The vacuum must be excellent, as otherwise the ions may be neutralized by collision with the residual gas.

16.5.2 The Shadow Microscope

In the shadow microscope, the image of a thin object is obtained by conical projection from an electron source of very small dimensions. The source may be real, as it is in the cold emission microscope (§ 13.6) but in the shadow microscope in the proper sense (Boersch, 1939, 1940) the source is virtual and consists of the highly reduced image of a real source. The aberrations of the system of reducing lenses have no effect on the fineness of the image, as the clarity of the latter depends only upon the Gaussian diameter of the image-source (and not upon its real diameter, which may well be made considerably larger by the aberrations of the reducing system) and upon the Fresnel diffraction at the specimen. The aberrations of the optical system by which the image-source is formed only have the effect of producing a distortion in the image of the specimen which increases indefinitely when we try to reduce the importance of the Fresnel diffraction by increasing the nominal magnification (cf. § 7.6.2 and Castaing, 1950a, 1950b). The result is not a limitation on the resolving power, but upon the possibility of using a theoretically infinite resolving power, and this limitation can only be overcome if we have some means of reconstructing the real image from the actual image obtained. The diffraction microscopy due to Gabor (1949) is based on just such a procedure.

16.5.3 Gabor’s Microscope

In the method which Gabor suggested, the object is illuminated by a virtual electron source extremely small in (Gaussian) diameter, that is, by a highly coherent radiation. A photographic plate is placed beyond the object, just as in the ordinary shadow microscope, and records the diffraction pattern which is formed by the interference between the electron waves
which arrive directly from the source and the waves which are scattered by the object. This pattern (which Gabor calls the "hologram") is transformed into a transparent positive, and reprojected with the aid of a light source, $S'$, which has the same aberrations as the original electron source, $S$. By sighting an eyepiece through the hologram onto the plane originally occupied by the object, a highly magnified "reconstructed" image is seen, in theory free of aberrations. The electron wave which is incident on the plate contains all possible information about the structure of the object, and with a knowledge of the form of this wave, it is, in principle, possible to reconstruct an image, the resolving power of which is the same as that of an electron lens with all the geometrical aberrations corrected, in which the illuminating beam has the same semi-aperture as the projecting beam. Unfortunately, the photographic plate only records the amplitude distribution in the wave, irrespective of the phase relations—an analogous limitation is present in the Bragg X-ray microscope. The ensuing loss of information is revealed physically in a most singular fashion by the simultaneous appearance of a second reconstructed image placed symmetrically behind the source. This "false image" becomes more troublesome the nearer it lies to the real image; this, most inconveniently, is the case when the projection source is placed very close to the object in order to illuminate only a very small area of the latter, and to reduce the exposure to an acceptable value. For further information about this brilliant method, the reader is referred to Dyson (1950), Haine and Mulvey (1952), Haine and Dyson (1950) and Gabor (1951).

16.6 NEW TENDENCIES IN ELECTRON MICROSCOPY

We have just reviewed the extremely numerous factors which broaden the spot which we call the "image" of a point. The principal part of this chapter has been given to defects which in theory can be corrected, but which are less easy to eliminate in practice. This explains the asymptotic form of the curve which illustrates the progress which has been made in improving resolving power. Technical advances allow us to approach the limit which is fixed by the optimum resolving power, but every fresh step is harder than the last. We might well imagine that the resolution would be improved by an increase of the operating potential. We have already seen that the theoretical resolution limit is lowered when the associated wavelength $\lambda$ is reduced (§ 16.2). The theoretical chromatic aberration too is diminished when $\Phi_0$ is increased, since the energy spectrum of the electrons emitted by the filament and the discrete amounts of energy which the electrons lose in passing through the object (see § 19.1) remain constant.
in absolute value. Furthermore, when the object is thick, the velocity dispersion introduced by the passage of the electrons through matter decreases very rapidly as $\Phi_0$ increases. When $\Phi_0$ exceeds 1 MV, however, it becomes ever more difficult to keep the fluctuations of the accelerating voltage within the range $10^{-5} - 10^{-6}$, as it is not easy to obtain stabilities comparable with those of standard microscope generators (which operate in the range $50 < \Phi_0 < 100$ kV). At present, therefore, the aggravated experimental problems have prevented the resolving powers attained in the various attempts to use very high tensions in microscopy (Müller and Ruska, 1941; von Ardenne, 1941; Zworykin, Hillier et al., 1941; van Dorsten, Le Poole et al., 1947; Tadano, Sakaki et al., 1956; Tadano et al., 1958; Kobayashi et al., 1958; Dupouy and Perrier, 1960, 1964) from reaching the level of those obtained with ordinary electron microscopes. The improvements that have been achieved during the years 1962–6, and which are the fruit of unremitting labour, have been such that the practical resolution has come considerably closer to the theoretical limit and we can reasonably hope that the same will sooner or later be true of very high voltage microscopes. In addition, the new perspectives that have been opened up before the experimentalists by the possibility of studying thick metallurgical or biological specimens together with this real hope that a practical resolution of about an Ångström will be reached, explain the spectacular increase in the number of very high voltage microscopes (500 kV microscopes: Shimadzu et al., 1966; Tadano et al., 1966; 750 kV: Smith et al., 1966; 1 MV: Watanabe et al., 1966, Nishigaki et al., 1966). The absolute record is still held by Dupouy and his colleagues (Dupouy and Perrier, 1960, 1960a, 1964a, 1964b; Dupouy et al., 1960b, 1961) with $\Phi_0 = 1.2$ to 1.5 MV, but even more ambitious designs are being studied (3 to 5 MV, Gilroy et al., 1966). To go beyond 1.5 MV while keeping the focal lengths of the lenses short and their aberration coefficients small, new lines of approach will be necessary, however: superconducting lenses with very high fields and extreme temporal stability (see §17.4.5) or quadrupole lens systems equivalent to round lenses. The familiar types of high voltage generators and electrostatic accelerating columns will certainly give way to linear accelerators and, at the other end of the instrument, new methods of recording the image that are satisfactory at very high energies will have to be found.

A further point is that microscopes are being operated more and more with a "clean" vacuum, in the vicinity of the object at least, so as to avoid contamination of the specimen by organic vapours and molecular fragments of carbon compounds, broken by the electron beam; if deposits of these are allowed to form, image contrast rapidly deteriorates and hence the resolution becomes poorer.

The results which are being attained allow us to hope nevertheless that
resolving powers of the order of an Ångström unit will be reached in a few years, resolving powers which are, that is to say, of the same order of magnitude as the radius of an atom. The conditions for the microscopic examination of an atom have been studied by Gabor (1948) and Louis de Broglie (1950), who showed that as the beam electrons set the atoms in vibration, they will disturb the image. In the light of the recent experimental results described above which show that crystal lattices can be observed, we may conclude that these predictions were too pessimistic. Since the resolving power (for periodic specimens) has been reduced below 10 Å, a considerable amount of work has been devoted to this problem, and especially to the determination of the contrast in the image plane. If an atom is to be "seen", the contrast between the image point and the continuous background must be at least 5 per cent. Heidenreich and Hamming (1965) have shown that the classical amplitude contrast arising from the inelastic scattering of electrons outside the aperture would be of the order of 1 per cent for isolated atoms. Phase contrast, caused by interference between the diffracted waves and the direct wave near the image plane, could on the other hand be higher. An analytic investigation of this problem must, however, take the spherical aberration of the objective into account, for this deforms the wave-surfaces, as well as possible defocusing at the objective. A first calculation of the interference in the vicinity of the image plane was performed by Glaser (1949); this was reconsidered by Heidenreich and Hamming (1965) and later by Eisenhandler and Siegel (1966a).

It is not unreasonable to think that it will be possible to observe atoms in the near future, when objectives with little spherical aberration become available, and we operate at very high voltage (since the contrast is proportional to the relativistic increase in electron mass) and very low temperature (to reduce the thermal motion of the atoms in the crystal lattice to the utmost). The phase contrast will be enhanced by placing an aperture resembling a “Fresnel zone plate” at the exit plane of the lenses (Hoppe, 1963; Lenz, 1963; Hanszen et al., 1964, 1965; and Eisenhandler and Siegel, 1966b).
CHAPTER 17

THE OPTICAL ELEMENTS OF A MICROSCOPE

17.1 THE OBJECT ILLUMINATION

17.1.1 The Structure and the Function of an Electron Gun

Formally, the gun resembles the electron gun of an oscillograph, but the properties are rather different (§ 8.3). The source of electrons is in general a filament of tungsten or tungsten–molybdenum, about 0.1 mm in diameter, bent into the form of a V (Dlugosz, 1950), and raised to a temperature between 2600 and 2800°K by Joule heating† (which entails a current of a few amperes). The grid, which surrounds the filament, is an electrode in which a circular hole between 1 and 3 mm in diameter has been bored and which is centred on the point of the filament. The accelerating potential is applied between the filament and the anode—the latter is normally earthed. A small potential difference (one-hundredth of the total potential drop, for example, or about 500 V) is applied between the filament and the grid. The diameter of the opening in the anode is also a few millimetres, but its size is not critically important.

Microscope guns are of two types. The first type is wholly analogous to the gun of an oscillograph, and provides a divergent beam of aperture $\lambda_0$ (which is of the order of a few thousandths of a radian) which appears to have originated in a virtual source of extremely small diameter ($2r_0 \approx 10$ to 40 $\mu$, for example). This beam could be used directly to illuminate the object, but in fact a condenser—usually magnetic—is always added, so that the way in which the object is illuminated can be modified.

To avoid using a condenser, Bruck and Bricka (1948) and Steigerwald (1949) have modified the form of the grid so that the beam converges slightly as it leaves the gun, and forms a real crossover a distance $L$ from the accelerating anode. The angular apertures are of the same order of magnitude as those quoted for the preceding case if $L$ is of the order of 10 or 15 cm; the density of the current which passes through the object

† A thorium oxide-coated tungsten filament is sometimes used—this increases the emissive power considerably.
is then high, and may reach one or two A cm\(^{-2}\). In Fig. 201a, Bruck and Bricka’s gun is shown schematically, and in Fig. 201b that of Steigerwald, which is called *Fernfokus*. It is not possible to obtain a crossover as small as the virtual source of a gun provided with a condenser lens with guns of this type; the diameter \(2r_0\) is reduced only with difficulty below

\[ 0.3 \text{ to } 0.5 \text{ mm, for } L \approx 10 \text{ to } 15 \text{ cm}. \]

It is, however, possible to retain only the central part of the beam, in which the current density is greatest, by placing a very small diaphragm (20 to 50 \(\mu\) in diameter) beneath the anode.

The fundamental quality of a microscope gun is the electron brightness \(B = \frac{I}{\pi \Delta_0^2 \cdot \pi r_0^2}\), which is measured in A cm\(^{-2}\) sterad\(^{-1}\). This expression

---

**Fig. 201.** Two types of gun which work without a condenser.
(a) Gun with spherical electrodes (Bruck and Bricka, 1948); \(h = 17\) mm.
(b) *Fernfokus* (Steigerwald, 1949). The essential figures for a possible version (Braucks, 1958) are as follows:
\(a \approx 45\) mm, \(d_w = 50\) mm, \(d_A = 20\) mm,
\(d \approx 60\) mm, \(b = 0\), \(\tan \varphi = 0.55\),
\(h = 10\) mm, \(a_k = 1\) mm (\(a_k\) = diameter of the bore in the conical inset of the Wehnelt).

c) Gun employing a point cathode (Sakaki and Möllenstedt); the various parameters have the following values: \(d_1 = d_2 = 0.2\) mm, \(h = 8\) mm, \(d = 1\) mm,
\(l = 0.7\) mm, \(t = 0.5\) mm, \(V_g = 550\) V (or \(d = 2\) mm, \(l = 1.4\) mm, \(t = 1\) mm, \(V_g = 200\) V).
defines a mean brightness. Placing a stop close to the anode is a very useful artifice in electron probe devices, as a source of reduced area is obtained, with the same brightness as the central zone (see Chapter 21).

The maximum brightness, \( B_m \), can be roughly estimated with the aid of Langmuir's formula:

\[
B_m = \frac{\eta e \Phi}{\pi k T} \approx 3.7 \times 10^3 \frac{\eta e \Phi}{T},
\]

in which \( \eta e \) is the maximum current density at the centre of the cathode measured in A cm\(^{-2} \), \( \Phi_0 \) is the accelerating potentials in volts, \( k \) is Boltzmann's constant \( \frac{1}{11,600} \) eV °K\(^{-1} \) and \( T \) is the absolute temperature of the cathode (°K).

For a tungsten cathode held at temperature \( T = 2500°K \), with \( \Phi = 50 \) kV, we find that the brilliance is \( 2.2 \times 10^5 \) A cm\(^{-2} \) sterad\(^{-1} \); the order of magnitude accuracy of this result has been confirmed by experiment (Haine, 1947, 1950, 1957; Haine and Einstein, 1952; Boersch, 1954). If \( T \) is increased to \( 2800°K \), a brightness of the order of \( 3 \times 10^6 \) A cm\(^{-2} \) sterad\(^{-1} \) can be obtained but the emission is then restricted by space charge and the lifetime of the filament is also considerably shortened. Haine, Einstein and Borcherds (1958) have measured the brightness at this temperature, and found values \( B_m \approx 1.6 \times 10^6 \) A cm\(^{-3} \) sterad\(^{-1} \) in a divergent gun with a condenser.

The whole illuminating system forms the "head" of the microscope—it is articulated in various ways; with the knobs provided, the filament can be moved laterally and in height with respect to the grid, and the beam can be "aligned", by moving the whole system bodily transversally, or by rotating it around the object (for the case of the gun without a condenser).

17.1.2 Electron Sources

(i) The heated cathode

A tungsten filament is currently used. The electrical power which has to be dissipated to obtain a suitable emission is of little importance, as in microscopy it is the lifetime of the filament which is all-important; a value of 100 hours is readily acceptable, and von Ardenne (1940) finds that the temperature dependence of lifetime is given by the values in the following table:

<table>
<thead>
<tr>
<th>( T ) (°K)</th>
<th>2300</th>
<th>2420</th>
<th>2525</th>
<th>2640</th>
<th>2750</th>
<th>2825</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime (hours)</td>
<td>287000</td>
<td>30700</td>
<td>4360</td>
<td>770</td>
<td>150</td>
<td>53.5</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Haine et al. (1958) give the following empirical formula for the lifetime $t$:

$$t = \frac{50}{j_c} \text{ hours}$$

($j_c$, measured in $\text{A cm}^{-2}$, is the maximum current density at the cathode).

Again, it is of no particular interest to increase the specific intensity incident upon the object beyond the values which are ordinarily obtained with a tungsten wire, as the power which is dissipated as heat in any object which is not exceptionally thin is already close to the maximum which can be tolerated. For this reason, the improvements which are being made at present consist in giving the gun a more complicated structure, so that only the minimum region of the object which lies within the field of the microscope is illuminated (an area of some 20 $\mu$ by 20 $\mu$).

It is only with this kind of gun, and in exceptional circumstances (a very short exposure, or a very thin object, for example) that it can be advantageous to increase the specific emissivity of the source.

In general, the tungsten filament which is employed is bent into the form of a hair-pin (see Fig. 201), and it is the point of the filament, situated on the axis, which effectively constitutes the electron source. To obtain beams of very small aperture, and higher brightness, Hibi (1954) and later Sakaki and Möllenstedt (1956) and Swift (1960) have studied cathodes in the form of a very fine point, of which the useful radius of curvature is a few microns—this point is formed at the tip of a wire by electrolytic polishing. The wire is heated from behind the point by leading in a lateral current. This wire may either be of tungsten or of platinum, coated with barium and strontium oxide, which require only very slight heating. The aperture of the resulting beam is of the order of 1 to $3 \times 10^{-5}$ radians, and this type of source is particularly well adapted for electron interference experiments, as the beam is highly coherent. A gun using a point cathode is shown in Fig. 201c.

(ii) Induni's cold cathode

In the first electron microscope which Knoll and Ruska (1932) constructed, the illuminating electron beam originated in a gas discharge, at low pressure (about $10^{-8} \text{ mm Hg}$), between an anode and a cold cathode; the discharge, however, was unstable and difficult to control, and variations of potential of more than one per cent were found between the terminals, which produced a variation in the energy of the electrons of the same order. For this reason, a hot cathode was used instead in subsequent instruments, and it was generally admitted that a cold cathode is not suitable for a microscope (von Ardenne, 1940; Gabor, 1945; and Zworykin et al., 1949).

Induni (1947, 1955), however, showed that this conclusion must be revised, by designing a source which resembled that of Knoll and Ruska in appearance
but which he could control easily and with which he could obtain images with a resolution of about twenty Ångströms. Figure 202 indicates the structure of this type of source—a potential of 40 kV is applied to the central rod (which is made of aluminium or 18/8 stainless steel); the anode,

![Diagram](image1.png)

**Fig. 202.** Induni's cold cathode: cross-section of the electrodes.

![Diagram](image2.png)

**Fig. 203.** Diagram illustrating the principle behind the cold cathode; the separation of the ionization zone from the acceleration zone is shown.

which is connected to earth, is a concentric cylinder about 5 mm away, with a large opening cut in it. A very fine diaphragm limits the flow of gas. The lifetime of this gun is virtually unlimited, and the geometrical features and the specific intensity of the beam which it provides are satisfactory and comparable with those which are obtained with a hot cathode. The electrons which are provided by this type of gun are virtually monoenergetic; the reason for this can be understood by examining Fig. 203 which illustrates the principle of the gun. The chamber contains air at a low pressure; any ion in this chamber (which is virtually equipotential) is slowly drained off towards the diaphragm \( D \) by the weak field which overlaps through the aperture in this diaphragm. It then flies towards the high tension plate \( P \), which in turn emits secondary electrons with a very small initial energy. These electrons cannot create ions before they reach the chamber, as the distance \( l \) between \( P \) and the diaphragm \( D \) is less than their mean free path.
Some of them do produce ions inside the chamber, and thus maintain the discharge, but these ions can only subsequently produce electrons at the plate $P$, and neither in the chamber (where their energy is too small) nor in the accelerating region between $P$ and $D$ (where the trajectory is too short). The electrons all originate in $P$ and will therefore be mono-energetic when they leave the diaphragm $d$; this is not the situation in an ordinary discharge tube, where the electrons may be produced anywhere within the field and emerge (through a Lenard window, for example) with widely varying velocities. The basic advantage of the Induni cathode is that the ionization region and the acceleration region are separate.

17.1.3 The Electrical Supply Circuitry—Autobias

Our aim is to achieve stable behaviour at the gun, in the sense that we strive to obtain an electron current $I_e$ of constant intensity. Variations in $I_e$ would alter the brightness, affect the specimen heating (and hence give rise to transverse migrations of the object) and above all, create fluctuations in the high tension itself since high tension sources have a relatively high impedance. The variations in $I_e$ are essentially due to fluctuations of the filament temperature $T$, as the emissivity varies exponentially with $T$. Present-day high tension sources are stable enough (one part in $10^8$ or $10^9$) for their fluctuations to have virtually no effect upon the emission.

Figure 204 shows two ways in which the potential of the gun can be supplied. Every effort must be made to make the supply as stable as possible. In the arrangement (a), the filament is heated with the aid of a high frequency transformer (100 kc/s) which is easily made as both the primary and the secondary windings have only a few turns; any thermal fluctuation at the filament is eliminated by using a high frequency. The high tension is applied through a potential divider in such a way that the point of the

![Fig. 204. The two possible ways of supplying an electron gun.](image)

(a) The Wehnelt, or grid, is polarized with batteries. (b) The Wehnelt is polarized automatically by the potential drop created by the gun current as it passes through the polarization resistance $R_s$. 
filament is held at a constant potential. The grid is supplied with the aid of a potentiometer connected to batteries.

The system (b) which is biased automatically is, however, more usual. The filament may still be heated at a high frequency, but often (as illustrated in Fig. 204b), it is fed by an accumulator, while the current is varied with a well-insulated rheostat. The bias is produced by the potential difference across a high resistance (between 0.5 MΩ and a few dozen MΩ). A condenser is often added to eliminate a.c. ripple. The negative bias \( V_g \) of the grid is proportional to the current, \( I_e \). When \( I_e \) increases, the bias also increases and hence \( I_e \) tends to diminish: we therefore have a self-regulating mechanism for stabilizing the current.

![Diagram](image)

**Fig. 205.** The difference between the behaviour of the gun supplied in these two ways (see Fig. 204).

(a) The anode potential is increased by an amount \( \Delta V \). The yield of the gun with independent polarization is increased by \( \Delta I_t \), whereas that of the automatically polarized gun is increased only by \( \Delta I_s \). (b) The filament temperature increases by an amount \( \Delta T \). The increase in the output of the independently polarized gun is considerably larger than that of the automatically polarized gun. (c) The influence of the resistance \( R_e \) on the characteristic \( I_e = f(T) \) of a triode gun; the appearance of a kind of saturation.
Figure 205a shows $I_e$ as a function of the bias, $V_g$. When $V_g$ is constant (batteries), a variation $\Delta V$ of the potential of the high tension supply results in the variation $\Delta I_1$, shown in Fig. 205a; when, on the contrary, the system is autobias, the working-point moves from $A$ to $B$, and the change $\Delta I_2$ is far smaller. In Fig. 205b, two curves representing the emission at slightly different temperatures $T$ and $T + \Delta T$ are plotted; the current is stabilized by an analogous mechanism, using autobias. In this latter case, the stabilization is better to the extent that the two curves are close together. If, however, we examine the curves which describe the variation of $I_e$, as a function of the temperature, $T$, for a given voltage $V$ (Fig. 205c) it is apparent that if $V_g = \text{constant}$, the curve is of the exponential form which we should expect, whereas if the system is autobias, the variation in $I_e$ tends to become stabilized, even though at first it is more acute as $V_g$ is initially very small. A kind of "saturation" appears, which has its origin in the autobias mechanism itself; the saturation is more noticeable the larger is $R_s$ or, what is equivalent, the closer $V_g$ lies to the cut-off potential $V_{oc}$ of the gun. The total current falls rapidly as $R_s$ increases.

For each type of gun, therefore, we must bear two factors in mind: (i) we should select a working point in the region where the saturation curve is as flat as possible, after ensuring that the filament temperature is relatively low in order that the lifetime shall be reasonable; (ii) at the same time, we should, by modifying one of the dimensions, for example (the filament-grid distance, or the diameter of the opening in the grid), make this working point coincide with the region of maximum brightness of the gun, which is a function of $V_g$ (see the section on triode guns).

The focusing of a microscope gun therefore involves a meticulous experimental study. An example of such a study is to be found in an article by Haine, Einstein and Borchers (1958).

Further improvement on the above arrangements is still possible, with the result that we can obtain a saturated region in which $I_e$ remains virtually constant when the heating current varies by as much as 30 per cent. A biasing system composed of a potential source $V_g$ in series with a resistance $R_s$ is used; this system may be formed very straightforwardly with a triode, the grid of which is held at a positive potential, $v$, which can be adjusted, and the cathode of which is placed in series with a resistance $r$ (Dolby and Swift, 1960). The valve then behaves like a voltage generator, with e.m.f. $V = -\mu v$ and internal resistance $R_s = (\mu + 1) r$; $\mu$ is the amplification factor.

The current $I_e$ can be described by a law of the form

$$I_e = K \left( \frac{V_{oc} - V}{V_{oc}} \right)^n \quad n \approx 2,$$

in which $V_g$ is the bias, $V_{oc}$ the cut-off potential of the gun, and $K$ a coefficient which depends only upon the emission of the cathode (which is a
function of $T$). The accelerating potential $V$ and the geometry of the gun only affect $V_{gc}$, since for all guns, $n \simeq$ constant.

If we define a stabilization index, $\varepsilon$, for fluctuations of $T$, in the following way:

$$\varepsilon = \frac{d I/I}{d K/K},$$

it is clear that $\varepsilon = 1$ when $V_g = \text{constant}$ (battery-operated polarization) which corresponds to a complete absence of regulation. When $V_g$ tends towards $V_{gc}$, $\varepsilon$ decreases, which clearly displays the fact that the stabilization is much better in an autobiased gun the nearer it is to cut-off. For $V_g = 0.8V_{gc}$ for example, we should find that $\varepsilon \sim 0.1$, a tenfold gain with respect to fixed polarization guns.

With the final arrangement, we can always hope to gain yet another factor of ten when $V_g = 0.8V_{gc}$.

### 17.1.4 The Role of the Condenser

The condenser is a magnetic lens, with which we can adjust simultaneously the angular aperture of the beam which falls on a point of the object, and the current density which passes through the object.

Consider the situation shown in Fig. 206, in which the crossover, which, as we explained in § 17.1.1 plays the part of the electron source, illuminates the object directly. The pencil of electrons which passes through a point on the object has an angular aperture which is given by

$$\alpha_0 = \frac{r}{p_1},$$

![Fig. 206](image)

Fig. 206. In the absence of a condenser lens, the object is illuminated at a fixed angle $2\alpha_0$. The condenser produces an image of the crossover which acts as the source and with which the angular aperture of the illumination can be varied.
in which \( r \) is the radius of the crossover. We now insert a lens of convergence \( C \) at a distance \( k p_1 \) from the crossover. This lens will produce an image of the crossover which will now play the part of the source of the illumination. A simple calculation then shows that

\[
\alpha = \alpha_0 \frac{1}{k (1 - k) p_1 C - 1}.
\]

It must not be forgotten, however, that the condenser contains a stop of radius \( R \), so that the aperture of the beam can certainly not exceed

\[
\alpha_m = \frac{R}{(1 - k) p_1}.
\]

Finally, therefore, we shall obtain the curve shown in Fig. 207, which displays the dependence of \( \alpha \) upon the convergence of the condenser. There are, therefore, two possibilities: (1) We can obtain a large aperture, accompanied by a high electron density at the object, which we shall use for focusing visually; in this case, the image of the crossover is formed in the immediate vicinity of the object. (2) A small aperture, on the other hand, has the advantage of increasing the resolution of the instrument, but compels us either to go over immediately to photographic recording because of the low intensity or to take special precautions and accept the resulting fatigue to the observer. The latter has to remain for at least a

---

**Fig. 207.** The variation of the angular aperture of the illuminating beam as a function of the convergence of the condenser.
quarter of an hour in darkness, before attempting the adjustment, while
his eyes become adjusted into a state of maximum sensitivity; the optimum
physiological condition characterizes the properties of what is known as
"scotopic" sight.

17.2 GENERAL CHARACTERISTICS OF
MICROSCOPE LENSES

It may perhaps surprise the reader to meet a chapter on electron lenses
at this point in the volume, but recently, important work has been performed
on these lenses and it is useful to have a rapid summary of the results as
the lenses are the vital organs of the microscope; the construction of the
objective represents the summit of perfection of experimental technique.
All the lenses with high magnification which are used in microscopes possess
certain characteristics as a result of which they form a well-defined class
within the plethora of lenses which are used in television tubes, image
converters, and mass spectrographs. It is, therefore, a fit point at which
to bring together the new results.

17.3 ELECTROSTATIC LENSES

17.3.1 Significant Geometrical Parameters

The electrostatic einzel lens, with its three electrodes, has so far always
been used in the symmetrical form. To avoid intense fields (the effect of a
point), the rims of the holes must be rounded, and the lens is defined geo-
metrically by a rather large number of parameters (Fig. 208). These are:
- the interelectrode spacing \( \Delta = z_2 - z_1 \);
- the thickness of the central electrode \( e = 2z_1 \);
- the diameter of the diaphragm of the central electrode \( D = 2R_1 \);
- the diameter of the diaphragms of the outer electrodes \( d = 2R_2 \);
- the radius of curvature \( \varphi \) of the edges of the electrodes.

The parameter \( \Delta \) cannot be less than a certain limit, as it represents the
distance between two metal parts at different potentials, and the minimum
is determined by the breakdown condition.

Experiment shows that the influence of the thickness of the outer elec-
- trodes on the distribution of potentials is negligible. The curve shown in
Fig. 209 (Regenstreif, 1951) shows that the axial potential directly below
an external electrode is virtually \( \Phi_2 \). This result shows that even a marked
lack of symmetry between the entry aperture and the exit aperture only
modifies the focal length very slightly.
Fig. 208. The geometrical parameters of an electrostatic microscope lens.

Fig. 209. The value of the axial potential directly below the outer electrode (the abscissa denoted by \( z_2 \) in the text is here rechristened \( z_0 \)). The potentials of the central and outer electrodes are \( V_1 \) and \( V_2 \) respectively. It can be seen that \( y \) is always close to unity (Regenstreif, 1951).

So far, the influence of \( \varrho \) has not been studied. Certain workers have chosen \( \varrho \) to be equal to \( e/2 \) (Bruck and Romani, 1944) while others select fixed values for \( \varrho/L \) (Lippert and Pohlit, 1952). We write

\[
L = 2\Delta + e.
\]

Finally, the diameter of the outer electrodes only plays a minor role, at least provided \( d \) is small, as the electrons are moving at high velocities when they are level with these electrodes.
We conclude, therefore, that the parameters which are optically important are $\Delta$, $e$ and $D$. Regenstreif's results have been given in the first part. We shall complete this description of lenses with a résumé of the results of a series of precise measurements which have been made recently by Lippert and Pohlit (1952, 1953 and 1954) which have yielded valuable data.

17.3.2 The Results Obtained by Lippert and Pohlit

These results relate to lenses of which the central electrode is at cathode potential. Figure 210 gives values of $L/f$ as a function of the parameters $D/L$ and $e/L$.† The lenses which are defined by $\partial f/\partial v = 0$, where $v$ is the initial velocity of the electrons ("apparent" achromatism), and which are

Fig. 210. Curves showing the convergence of a series of einzel lenses (Lippert and Pohlit, 1952).

† These curves are slightly different from the ones which Lippert and Pohlit (1952) have published, as they are reproduced from unpublished papers which have been kindly put at our disposal by these authors. A few points which represent results obtained in our own laboratory have been added.
free of distortion lie on the dotted line near \( e/L = D/L \). For \( e/L < D/L \), the lens has cushion distortion and \( \partial f/\partial v \) is positive. Between the achromatic lenses and the curve \( L/f = 0 \) (at which the lens is afocal) lies a region in which the lenses suffer from barrel distortion, for which \( \partial f/\partial v < 0 \). Beyond

![Diagram](image)

**Fig. 211.** Curves giving the convergence of a series of einzel lenses (after Regenstreif).

is the transgaussian region (where the focal length is negative), at the beginning of which the distortion is again cushion, but \( \partial f/\partial v \) is still negative.

Figure 211 shows the same set of curves, as calculated theoretically by Regenstreif. The correspondence between theory and practice is satisfactory, as we can see, making allowance for the imperfection of the theoretical model adopted which had to be simplified before the calculations could be concluded.

Figure 212a represents the behaviour of the abscissa of the focus in the same coordinate system. Any of the lenses which correspond to \( p/L > 0.5 \) can be used as objective. The first curve, \( p/L = 0.5 \) corresponds effectively to the parallel curve \( L/f = 1.6 \) of the preceding diagram; the focal length of such an objective is in practice determined only by \( L \), therefore—the interelectrode spacing fixes the lower limit of the focal length; the focal
Fig. 212a. Curves giving the abscissae of the foci of a series of einzel lenses (Lippert and Pohlit, 1952).

Fig. 212b. Curves giving the spherical aberration and the convergence of a series of electrostatic einzel lenses, plotted from the results obtained by Lippert and Pohlit (1953) and Bruck and Romani (1944).
length of an "optimum objective", working in the unipotential fashion, is therefore about 60 per cent of the total thickness of the lens.

Figure 212b gives the values of the reduced spherical aberration coefficient $C_s/L$ as a function of the parameters $e$ and $D$. This set of curves is based upon the results of Lippert and Pohlit, together with those of Bruck and Romani (1944); only the region which corresponds to conditions in the neighbourhood of an objective for which $L/f \geq 1.6$ is represented. The spherical aberration constant has a minimum in a region which, on the "convergence diagram" lies between distortion-free lenses and objectives. It can be seen that the spherical aberration constant for objectives is reduced according to the extent to which the focus lies within the electrode.

17.3.3 Einzel Lenses with Unsymmetrical Electrodes

At present, lenses of this kind have rarely been studied, still less used in practice, despite the fact that in an experimental investigation of the spherical aberration constant of various einzel lenses, Seeliger (1948) found that one very unsymmetrical lens, with a central electrode, had an extremely small spherical aberration constant ($K_s \approx 3$).

To determine experimentally the optical elements of these lenses is a particularly delicate task. Heise's method (1949) which is suitable for symmetrical lenses has been extended by Everitt and Hanssen (1956) to the case of lenses with any kind of asymmetry. Hanssen (1958) has also shown that the optical properties of any unsymmetrical lens are well represented by those of a doublet of two thin lenses, the more convergent of which lies on the side in which the field $E_z = -q'(z)$ is stronger. It has long been known in ordinary optics, however, that the spherical aberration of an objective is least when the surface of the lens with the greatest curvature—where the convergence changes the most rapidly—faces the object. In the same way, an unsymmetrical einzel lens should produce less spherical aberration than a symmetrical lens with the same convergence. Hanssen (1958) has been able to confirm this hypothesis by making such a comparison. Septier

![Fig. 213. Asymmetrical lens with small spherical aberration for an object placed in the neighbourhood of $F (C_s = 2.5)$; all quantities in mm; $f_{\text{min}} = 10$ mm.](image-url)
(1960) has recently determined the optical properties of a number of electrostatic lenses used with very wide aperture incident beams, and gives values of $K_s$ for several unsymmetrical lenses together with an analysis of the influence of the various geometrical parameters on $K_s$. It is in fact possible to obtain values of $K_s$ of the order of 2·5, whereas even the best symmetrical lenses only give

$$12·5 < K_{s(min)} < 20.$$  

These high quality lenses (Fig. 213) do, however, have the disadvantage of being relatively weak ($f_{\text{min}} \approx 8$ to 10 mm) and can thus only be used as low magnification objectives either in the composite immersion objectives described in §8.3.2, or finally in electron probes where we need the focus to be outside the electrodes (see §21.5).

17.4 MAGNETIC LENSES

17.4.1 Construction of the Electromagnet

(i) The form of the magnetic circuit

A magnetic lens is an axially symmetric electromagnet or permanent magnet, through the pole-pieces of which is bored a channel along which the beam of electrons passes.

The field in the gap of a magnetic lens must be intense, and localized into a short axial distance; a field which overlaps the edges of the gap too far is not only useless but harmful, as it produces centring problems. There is one factor, however, which limits the extent to which the field can be localized and intensified, which is the saturation of the pole-pieces; this, therefore, it is convenient to reduce to the maximum if not to eliminate

![Fig. 214. The possible forms of the core of a microscope lens. (The central channel is not shown.)](image-url)
altogether by a judicious choice of lens design. The designer is likewise guided in his choice of the shape of the pole-pieces by considerations of a practical order; for example, the objective has to be designed in such a way that the object can be easily displaced radially.

The great majority of magnetic lenses lies somewhere between two extreme arrangements, the asymmetrical version shown in Fig. 214a, and the symmetrical version shown in Fig. 214b. For an objective, the principal advantage of the asymmetrical arrangement is that the focus can be completely detached and the mechanism for displacing the object can, in consequence, be simplified considerably. More extreme arrangements have been used in certain exceptional cases (Fig. 214c); one of these is the electron probe microanalyser, where the beams of X-rays have to leave the object at an appreciable emergent angle, but this is a very special case, in which we are forced to use the lens in unfavourable conditions (Castaing, 1954; Mulvey, 1958).

(ii) *The distribution of magnetic field; the optimum form for the pole-pieces*

We now consider the symmetrical arrangement. The lines of force are distributed as if the symmetry plane $P$ of the gap had been replaced by a plane of some infinitely permeable substance. A symmetrical lens, therefore, can be regarded as two asymmetrical lenses in juxtaposition and it was this

![Diagram](image)

*Fig. 215. The significant parameters in a magnetic lens.*

property which led Mulvey (1953) to study the distribution of magnetic flux in a single asymmetric lens; he began with a lens with unbored pole-pieces, away from the saturation conditions, and performed the experiments on the lens itself, using a fluxmeter. The meaning of the various parameters is shown in Fig. 215, and to begin with we shall suppose that $\theta = 0$, so that $D_1 = D_2$. 
A short coil is moved along the core, starting from the gap, towards the base of the core which is a distance $L_{\text{max}}$ from the gap. The field distribution clearly shows that the flux which passes through this coil increases steadily with the distance between the coil and the gap; the induction $B$ in the core at a distance $L$ from the gap increases from the value $B_c$ (which is obtained in the gap) for $L = 0$ to a considerably larger value which is obtained for large values of $L$. The graph of Fig. 216a, which Mulvey obtained, shows the way in which $B/B_p$ varies with $L/D_1$; this variation is more rapid when the separation $S$ between the pole-pieces is wider, and consequently, the field within the gap itself is less localized.

This is a particularly disadvantageous form to give to the pole-pieces—if the lens excitation is increased, the value of the induction rapidly reaches the saturation induction $B_s$ of the mild steel which is used at the junction between the pole-piece and the end-face. The useful induction $B_p$ in the gap is always less than $B_s$; if we require $B_p$ to be equal to $B_s$ (16 to 20 kG for ordinary mild steel; 22 kG for Armeo iron; 25 kG for special cobalt steel), the form of the pole-pieces must be chosen in such a way that the induction decreases steadily within the iron as we move away from the gap, despite a magnetic flux which, on the contrary, increases steadily since the lines of force of the overlap field are to be added to those of the useful field. The solution is to construct pole-pieces in the form of a truncated cone ($\theta = 0^\circ$); the cross-section then increases with $L$, and for any given value of $S/D_1$, we can find a value of $\theta$ such that $B/B_p$ is a decreasing function of $L$ (see Fig. 216b). For too wide an angle, however ($\theta \geq 70^\circ$), the total flux becomes too large and the dimensions of the magnetic circuit too great, to no purpose; $\theta = 50^\circ$ is a suitable value.

There is no point in flaring the core indefinitely; as soon as the induction reaches a low enough value, the diameter of the pole can remain constant (Mulvey, 1953), equal to $D_2$. The induction begins to increase again with $L$, and $D_2$ must simply be chosen sufficiently large for $B$ to reach its maximum at $L_{\text{max}}$, and to remain of the order of $2B_s/3$, for example. $D_2$ must be kept as small as possible to avoid giving the windings too large a mean radius, as too much Joule heat would be dissipated in supplying the lens normally, and to prevent the whole assembly from becoming prohibitively cumbersome. In Fig. 216c, curves describing the variation of the induction with the distance $L$ are shown, in the two cases $S/D_1 = 0.2$ and $S/D_1 = 0.3$. The full curves correspond to a core of total length $L_{\text{max}} = 3D_1$, and the dotted curves to $L_{\text{max}} = 6D_1$. From Mulvey's study, we can extract the optimum characteristics of the magnetic circuit: $\theta = 50^\circ$, $S/D_1 \approx 0.2$ and $D_b/D_1 \approx 3$. The presence of an axial hole of diameter $D$ has no appreciable effect upon these conclusions—the total flux is practically the same after the channel has been bored. If $D$ is increased, the induction within the pole-pieces increases; to diminish this effect, we have only to choose a value of the diameter $D_1$ such that $D_1 \geq 3D$. 


Fig. 216. The value of the magnetic induction on the axis of the magnetic core. (a) For a core of constant cross-section ($\theta = 0$). (b) For a core which ends in a truncated cone (Mulvey, 1953). (c) The value of the magnetic induction on the axis of a magnetic core for various values of the parameters (Mulvey, 1953).
A very complete study has been made by Durandeau (1955, 1956), Durandeau and Fert (1957), and Dugas, Durandeau and Fert (1961), which allows us to build lenses with the minimum weight in which saturation appears at the position of the gap so that the induction is highest in the useful zone. The form of the gap is shown in Fig. 217a, and the shape of

![Diagram](image)

(a)

![Diagram](image)

(b)

Fig. 217a. Section of the faces of the poles. The optimum values of the parameters are: \( e \approx s/3 \); \( (\pi/2 - \gamma) \approx 25 \) to \( 30^\circ \); \( S_1 \geq N I/500 \); \( 2l \geq S + \gamma(S^2 + D^2) \) (Durandeau and Fert, 1957).

Fig. 217b. The magnetic circuit of a high powered lens and the variation of the mean density, \( \Phi/\Sigma \) of the induced flux in the polar core. The broken line corresponds to the pole-faces in the form of truncated cones with cylindrical cores (Mulvey, 1953); the full line corresponds to the case where both pole-faces and cores are in the form of truncated cones (Durandeau, 1956; Dugas, Durandeau and Fert, 1961).

the pole-pieces in Fig. 217b. The truncated conical section, of angle \( \gamma \) is followed by a second truncated cone of angle \( \theta_2 \approx 12^\circ \), rather than by a cylinder of diameter \( D_2 \); the magnetic flux then remains constant over the whole length of the second cone (Fig. 217b). The thickness of the cheek is such that \( e \approx \frac{r_{\text{max}}}{2} \), so that at the junction the cross-section is effectively constant.

For small excitations, the field distribution along the axis is the same whatever the form of the poles in any lens in which the ratio \( S/D \) remains unaltered (\( S \) now represents the total width of the gap). Replacing the
actual distribution of induction by the distribution which would be produced by a solenoid, we arrive at the following formulae for the maximum induction $B_m$ on the axis (of a symmetrical lens) and the width, $2a$, of the curve representing the distribution $B(z)$ at the point at which it has fallen to half the maximum height (Fig. 217c):

$$B_m = \mu_0 \frac{NI}{S^2 + 0.45D^2} = B_1 \frac{S}{L},$$

in which

$$L = \sqrt{S^2 + 0.45D^2},$$

and

$$2a = 0.96L.$$

![Diagram](image)

**Fig. 217c.** The field distribution on the axis in the region of the gap (symmetrical lens).

These formulae are valid for low excitations ($B_1 < 12,500$ gauss for ordinary mild steel; $B_1$ represents the field in the gap $S$ when $D = 0$). $B_m$ is always smaller than $B_1$. The length $L$ which we have just defined plays an extremely important role in these lenses, and as we shall see later, it allows us to argue in reduced coordinates.

Durandau has pointed out that the central hole of diameter $D$ must be cylindrical for a certain distance $2l$, as otherwise $B(z)$ is deformed; it is convenient, therefore, to increase the diameter, both to facilitate cleaning and to avoid the possibility of electrons being reflected from the walls—in general, we choose $2l \geq S + \sqrt{S^2 + D^2}$. 
(iii) The influence of saturation

According to Durandeau, we can establish a threshold of the excitation \( NI \), in ordinary mild steel, beyond which the first saturation symptoms will appear in the vicinity of the gap. If \( S \) is expressed in millimetres, we find that

\[
NI = 1000S
\]

![Graph](image)

**Fig. 217d.** The variations of the maximum induction \( B_m \) and of the width \( a \) as functions of the excitation, expressed in \( NI/S \) as unit; the pole-pieces are of soft iron and the lens is symmetrical. Curve I: \( B_m \) (in teslas, or webers m\(^{-2}\)). Curve II: \( a/a_0 \) (Dugas, Durandeau and Fert, 1961).

is virtually true. For \( NI > 1000S \), the width \( 2a \) of the distribution \( B(z) \) increases slightly, and the maximum induction, \( B_m \), increases more and more slowly when \( NI \) increases (Fig. 217d). The lens properties are then well represented by a fictitious increase of \( S', D' \) and \( a' \); for these new values \( S', D' \) and \( a' \) of \( S, D \) and \( a \), we have

\[
\frac{S' - S}{S} = \frac{D' - D}{D} = \frac{a' - a}{a} \approx 0.14 \frac{NI - 1000S}{1000S}
\]

When the induction is very large, the difference can reach 10 per cent. In every case, we can write \( \frac{aB_m}{nI} = \text{constant} \).
Further, if the magnetic circuit is poorly chosen, a parasitic induction $B$ appears in the central channel near the zones where saturation occurs. This induction is extremely harmful, as it behaves like a weak lens placed before or behind the main lens, and the optical centring of the lens becomes very difficult.

If we are to be confident that the saturation will set in at the gap, the distance between the bases of the two truncated cones which form the first part of the poles (see Fig. 217a) must be greater than $S_1$ (mm), given by:

$$S_1 = 500NI.$$  

This condition determines the initial diameter $D_1$ of the core of the pole. When the excitation is small ($NI < 1000S$), the circuit can be considerably simplified.

(iv) The power which is necessary for excitation

The minimum internal radius $r$ of the winding is dictated by the diameter $D_2$ of the core. Let $r'$ be the external radius of the coil of length $h$, $s$ the cross-sectional area of the copper wire which is used, and $N$ the number of turns. The total cross-section $S = h(r' - r)$ of the winding will be most efficiently filled by contiguous turns of carefully wound copper wire coated with the minimum thickness of enamel which will provide the necessary insulation. We have $Ns = kS$, in which $k$ is a factor smaller than unity, and equal to at most 0.7, say, to fix our ideas.

If $NI$ is the excitation, and $\rho$ the resistivity of the copper, the resistance of the whole winding is

$$R = N \pi \rho \frac{(r + r')}{S},$$

and the power expended in the winding by virtue of its excitation is

$$W = RI^2 = \frac{R}{N^2} (NI)^2 = \frac{\pi \rho (r' + r)}{kS} (NI)^2 = \frac{\pi \rho (r + r')}{kh(r' - r)} (NI)^2.$$

The power which is required for any given excitation is therefore independent of the cross-sectional area of the wire, and hence of the driving potential also. Accumulators (6 V) are therefore equivalent to the modern regulators (250 V). There is an advantage in reducing the power, not so much to ease the load on the supply as to reduce the heating of the lens; this can be done by increasing $h$, which means increasing the size of the lens generally and produces an undesirable elongation of the core, but above all, by reducing $r$ as far as possible; it is necessary, therefore, to prevent $D_2$ from being increased unnecessarily. To give an idea of the magnitudes involved, consider a copper winding with internal radius $r = 2$ cm, external radius
r' = 6 cm, and length h = 10 cm; for k = 0.7, we find
\[ W = 1.5 \left( \frac{NI}{1000} \right)^2, \]
or about 6 watts for 2000 A turns, and 24 watts for 4000 A turns (a very high excitation). In the hypothetical case—clearly most unfavourable—in which the lens is thermally insulated, the heating is about 4° hour⁻¹ for 2000 A turns and 16° hour⁻¹ for 4000 A turns (the total mass of the lens is assumed to be 15 kg and its mean thermal capacity 0.1). In reality, the thermal contact between the windings, the yoke and the mounting provides an efficient and adequate cooling.

17.4.2 The Optical Properties of the Lenses

(i) The normal elements

In the projective lens or "projector" which is used to magnify the image formed by the lenses which precede it, it is the normal focal length \( f_1 \) which interests us, and above all, the very strong convergence. The spherical and chromatic aberration coefficients are of little interest, as the aperture of the beam is very small, but we must try to work in conditions of minimum distortion.

A detailed study of the variation of \( f_1 \) with the excitation for a number of different geometries has been carried out by Liebmam and Grad (1951) and Liebmam (1952). The excitation can be characterized by the quantity
\[ k = \frac{NI}{\sqrt{V^*}}, \]
in which \( V^* \) represents the potential \( V \), after correction for relativistic effects, at which the microscope is operating:
\[ V^* = V(1 + 0.98 \times 10^{-6} V). \]

The coefficient of \( r \) in the trajectory equation in a rotating coordinate system:
\[ r'' + \frac{eB^2}{8m_0 V^*} r = 0, \]
is thus proportional to \( (NI/\sqrt{V^*})^2 \).

Further, the image rotation \( \theta \) is also proportional to \( k \). The focal length \( f_1 \) at first decreases as \( k \) increases, passes through a minimum \( f_{1m} \) for \( (NI) = (NI)_0 \), and subsequently increases again. The abscissa, \( z_1 \), of the image focus, measured with respect to the centre of the lens (which is assumed to be symmetrical) decreases steadily, passes through zero and becomes negative for very high values of \( k \).
Liebmman, and later Durandeau, have pointed out that the curves giving $f_t$ as a function of $k$ for different geometries can, to all intents and purposes, be reduced to a single curve which describes the variation of $f_t/f_{1m}$ as a function of the ratio of the excitation actually applied to the excitation

![Graph showing optical elements of a magnetic lens](image)

**Fig. 218a.** The optical elements of a magnetic lens, shown as a function of the excitation; the reduced curves relate to symmetrical lenses. $f_t$: asymptotic focal length; $f_0$: immersion focal length; $z_f$: abscissa of the asymptotic focus; $z_0$: abscissa of the immersion focus; $C_e$: chromatic aberration constant; $C_s$: spherical aberration constant; $f_{1m}$: minimum asymptotic focal length, for $(NI) = (NI)_0$ (Durandeau and Fert, 1957).

$(NI)_0$ (Fig. 218a). The same is true of the curve $z_f/f_{1m}$, to an approximation which is in practice adequate, within the useful ranges of $D/S$ (from 0.5 to 5) and of $(NI)/(NI)_0$ (0.25 to 1.6); the differences remain below 4 or 5 per cent. If, further, we plot the curves giving $(NI)_0 \sqrt{V^*}$ and $f_{1m}/L$ as functions of $D/S$, we see (Fig. 218b) that for $0.5 < D/S < 5$,

$$(NI)_0 \approx 13.5 \sqrt{V^*},$$

and

$$f_{1m} = 0.5L = 0.5 \sqrt{S^2 + 0.45D^2}.$$
From Fig. 218b we see that $L$ is a better choice of reference length than $(S + D)$, the quantity selected by Liebmann (1955). Given $(NI)_0$ and $f_{1m}$, we can obtain the optical properties of any symmetrical lens from the universal curves. For asymmetrical lenses, $L$ has simply to be replaced by

$$L_A = \sqrt{S^2 + 0.45 \left( \frac{D_1 + D_2}{2} \right)^2},$$

in which $D_1$ and $D_2$ are the diameters of the channels bored through the two pole-pieces (Durandveau, Fert and Tardieu, 1958). Here we can again calculate $f_{1m}$, and obtain a single curve for $f_1/f_{1m}$, which coincides with the symmetrical lens curve. The curve which gives the abscissa of the focus is slightly separated from the curve which corresponds to $D_1 = D_2$. If $D_1/D_2 < 1$, the new curve lies below the curve for a symmetrical lens; if $D_1/D_2 > 1$, the converse is true. The distance between the curves always remains slight, however.

![Diagram](image)

**Fig. 218b.** The minimum focal length $f_{1m}$ and the corresponding excitation as functions of $D/S$ (Durandveau and Fert, 1957).

The possibility of obtaining extremely short focal lengths has been studied by Liebmann together with van Ments and Le Poole (1947). Were it not for the fact that the iron becomes saturated, the minimum focal length could be made arbitrarily short simply by reducing the width $S$ of the gap. The field in the gap is limited by saturation to about 26 kG, however, with the best steel known (Permendur or AFK 2 produced by the French Imphy company). In this way, Liebmann (1952) calculated that
for \( D = 0.6 \text{ mm} \), the following absolute minima of \( f_1 \) are reached: \( f_1 = 0.63 \text{ mm} \) for \( S/D = 2 \) and \( V = 40 \text{ kV} \); \( f_1 = 0.93 \text{ mm} \) for \( S/D = 3 \) and \( V = 90 \text{ kV} \); and \( f_1 = 1.43 \text{ mm} \) for \( S/D = 4 \) and \( V = 210 \text{ kV} \)—these calculated results agree well with the measurements made by Ruska (1944). The value \( f_{im} \) is reached for relatively wide gaps.

If the distribution \( B(z) \) is represented not by a bell-shaped curve, which is the appearance which it usually has, but by a rectangular model of height \( B_0 \) and length \( L_1 \), the cardinal elements can be calculated easily; we have simply to write \( B = \text{constant} \) in the differential equation above. The trajectory then has the form:

\[
r = r_0 \cos \omega L_1 \quad \text{with} \quad \omega = \frac{B_0}{V^*} \sqrt{\frac{e}{8m_0}},
\]

for a ray incident parallel to the axis. The value of \( B_0 \) is defined by

\[
B_1 S = B_0 L.
\]

We then find:

\[
\frac{f_1}{L_1} = \frac{1}{\omega L \sin \omega L_1} \quad \text{and} \quad \frac{z_1}{L_1} = \frac{1}{2} + \frac{1}{\omega L \tan \omega L_1},
\]

or in the reduced coordinates

\[
\frac{f_1}{f_{im}} \simeq \frac{\sin 2.03 \sin (NI)}{(NI)_{0} \sin \left( 2.03 \frac{(NI)}{(NI)_{0}} \right)},
\]

when the excitation is weak \([(NI) < 1000S \text{ or } (NI)/(NI)_{0} < 0.774] \).

(ii) The immersion elements

It is most important to know the immersion focal length \( f_0 \) and the position of the immersion foci when an electron microscope objective is being designed, since the object is placed at the immersion objective focus.

The immersion focal length \( f_0 \), expressed as a function of \( k = NI/\sqrt{V^*} \), does not pass through a minimum; instead, it decreases steadily as the excitation is increased, for the values of \( k \) used in practice at least. If, moreover, it were to reach a minimum for a certain value of the excitation, with pronounced saturation present, the objective would not be achromatic. Only \( f_0 \) would be stationary—the magnitude of the objective would be achromatic, but since the abscissa of the immersion objective focus continues to decrease when \( k \) is increased, the object would not be corrected for axial chromatic aberration; the practical resolution is thus limited (see § 16.3). The only effect of increasing \( NI \) or diminishing \( V \) is to increase the curvature of every point on the trajectory, and to bring the point at which the latter first intersects the axis steadily closer to the object.
Just as for \( f_1 \), universal curves for \( f_0 \) and the abscissa \( z_0 \) of the immersion focus can be constructed; these quantities need simply to be divided by the length \( f_{1m} \). These curves are shown in Fig. 218a. When the lenses are asymmetrical, the curve \( f_0/f_{1m} \) remains very close to the corresponding symmetrical lens curve for \( 0.5 < D_1/D_2 < 2 \); the same is true of the curve \( z_0/f_{1m} \), always provided that \( L = \sqrt{S^2 + 0.45D^2} \) be replaced by

\[
L = \sqrt{S^2 + 0.45\left(\frac{D_1 + D_2}{2}\right)^2}
\]

in the formula which gives \( f_{1m} \).

(iii) The aberration coefficients

A knowledge of the spherical aberration coefficients of lenses which are to be used as objectives is most important, as the beams which pass through them are of considerably wider aperture than the beams which pass through projective lenses. The spherical aberration coefficient, \( C_s \), and the chromatic aberration coefficient, \( C_c \), are thus determined for an object lying at the immersion focus \( F_0 \) with image at infinity, which is the condition in which the lens is normally employed.

For slight excitations, \( F_0 \), moreover, coincides with the asymptotic focus \( F_1 \) (according to Durandneau, this is true when \( (NI)/(NI)_0 < 0.774 \)).

For any given lens, irrespective of the excitation it may have, the coefficients \( C_s \) and \( C_c \) decrease when \( NI \) increases. In Fig. 218a, the curves which show the variations of \( C_s/f_{1m} \) and \( C_c/f_{1m} \) as a function of \( (NI)/(NI)_0 \) are plotted. These are virtually universal curves, and here again, the difference between these curves and the curves obtained for asymmetrical lenses is very slight. The coefficient \( C_c \) is always smaller than \( f_0 \), and for very high excitations, \( C_c/f_m \approx 1 \). \( C_s \), on the other hand, is much larger than \( f_0 \) for weak lenses, while for strong lenses, \( C_c/f_{1m} \leq 0.5 \).

A detailed study of the problem of determining the objective lens with the smallest aberration coefficients, \( C_c \) and \( C_s \), has been undertaken by Liebmann (1951, 1952). If we consider a weak lens in which \( D \) is varied for a given value of \( S/D \), the axial induction \( B_m \) and the working potential \( V* \) being held constant, we find that at first \( C_s \) decreases rapidly as \( D \) is increased, then passes through a minimum, and subsequently increases slowly. Liebmann uses the parameter

\[
k^2 = 0.022 \frac{R^2 B_m^2}{V*},
\]

in which \( R = D/2 \), and whatever may have been the value of \( B_m \) initially selected, the minimum corresponds to \( k^2 \approx 5 \). We thus obtain an optimum value of \( R \), which is given by

\[
R_0 \approx 15 \frac{\sqrt{V*}}{B_m} \text{ cm},
\]
and so

\[(C_s)_{\text{min}} = 0.267 R_0 = 4.05 \frac{\sqrt{V^*}}{B_m} \text{ cm.}\]

For other values of \(S/D\), slightly different values of \((C_s)_{\text{min}}\) are obtained. The search for the absolute minimum of \((C_s)_{\text{min}}\) is bounded by saturation phenomena, which are in turn linked to the value of the induction \(B_p\) which prevails between the pole-pieces off the axis, and which differs the more widely from \(B_m\) as \(D\) is made larger. Liebmann (1951) finds that this absolute minimum corresponds to the zone \(1 < S/D < 2\); with a carefully constructed objective, we can hope to obtain \(C_s \approx 1\) mm. An analogous study, with the object of reducing \(C_s\), reveals that again an optimum lens exists which resembles closely the minimum spherical aberration lens.

The defect which it is essential to eliminate from a projective lens is the distortion. In another article by Liebmann (1952), the ways of determining the parameters which correspond to a projective with minimum distortion are to be found. The distortion of the magnification always vanishes for a particular value of \((N I)/(NI)_0\). For every value of \(D/S\), Liebmann recommends that the lens should always be operated near \((NI) = (NI)_0\).

(iv) **The resolving power of a magnetic objective**

Using the results which had been obtained by Liebmann and Grad (1951), Mulvey (1953) plotted the curves which represent \(C_s^{\frac{1}{3}} \varpi^{\frac{1}{2}}\)—for various values of the ratio \(S/D\) and of the field in the gap—as a function of the diameter \(D\) of the opening in the pole-pieces. To each value of \(S/D\) and of the field, there corresponds a minimum of the parameter which describes the resolution (the optimum resolution); this minimum is reached for an excitation of about 6000 A turns when \(V = 50\) kV, and about 12,000 A turns when \(V = 500\) kV.

Liebmann (1955) plotted analogous curves to show the behaviour of \(C_s^{\frac{1}{3}} \varpi^{\frac{1}{2}}\) as a function of \(S\) for various values of the field \(B_p\) in the gap and the ratio \(S/D\). The curves vary very little with \(S/D\) provided this ratio remains between 0.5 and 2 (which is usually the situation). These curves have very flat minima, and by sacrificing a completely negligible degree of resolution, a great many ampère-turns can be saved. For example, \((C_s^{\frac{1}{3}} \varpi^{\frac{1}{2}})_{\text{opt}} = 5\) Å for \(B_p = 24,000\) gauss (2.4 T), \(V = 50\) kV and \(S = 0.36\) cm; without noticeably reducing the resolution, however, \(S\) can be made 0.2 cm, which reduces the excitation from 6700 A turns to 3700 A turns.

The chromatic aberration always remains close to the focal length (see Liebmann, 1952).

**17.4.3 Some Practical Examples**

From the article by Durandeau Fert and (1957), we have extracted the numerical data and diagrams necessary to construct a magnetic microscope with high resolution.
All the lenses have the same magnetic circuit, but the extremities of the pole-pieces are of different shapes.

For \( V = 75 \text{ kV} \), which corresponds to \( V^* = 80 \text{ kV} \), and values of \( D/S \) between 0.5 and 2, we have

\[
(N I)_0 \approx 13.5 \sqrt{V^*} = 3900 \text{ A turns}.
\]

(i) The objective

For an immersed object lying in the plane of the upper pole-piece (see Fig. 219a), we should use \( S = 3 \text{ mm} \), \( D = 4 \text{ mm} \); \( f_{1m} \) is then approximately 2 mm, but \( z_0 = 1.5 \text{ mm} \), or \( z_0/f_{1m} = 0.75 \). The curves of Figs. 218a and 218b

---

**Fig. 219a.** The pole-pieces of an objective (with an immersed object). \( O \): object; \( P \): object-support; \( A \): access to the gap (diaphragm); \( S = 3 \text{ mm} \); \( D = 4 \text{ mm} \); \( f_0 = 2.1 \text{ mm} \) (Durandau and Fert, 1957).

**Fig. 219b.** A magnetic objective. (\( V = 50-75 \text{ kV} \); \( z'z \): axis of the lens; \( P \): pole-pieces; \( C \): body of the lens; \( B \): windings; \( c \): channel to give access to the gap; \( g \): socket for the joint (to prevent leaks) (Durandau and Fert, 1957).

**Fig. 219c.** Section of a projective lens \( S = D = 1.7 \text{ mm} \); \( f_{1m} = 1.20 \text{ mm} \) for \( N I = 3900 \text{ A.t. and } V = 75 \text{ kV} \) (Durandau and Fert, 1957).
give the following data:

\[
\frac{NI}{(NI)_0} = 0.85 \quad \text{or} \quad NI = 3300 \ \text{A turns} \quad \text{for} \quad V = 75 \ \text{kV},
\]

\[
\frac{f_0}{f_{im}} = 1.04 \quad \text{or} \quad f_0 = 2.1 \ \text{mm},
\]

\[
\frac{C_c}{f_{im}} = 0.8 \quad \text{or} \quad C_c = 1.6 \ \text{mm},
\]

\[
\frac{C_s}{f_{im}} = 0.9 \quad \text{or} \quad C_s = 1.8 \ \text{mm}.
\]

Since the specimen-holder must be convenient to move, the arrangement shown in Fig. 219b is adopted.

(ii) The projective lens

The angular aperture of the beam is small, but the emergent beam may well be inclined at a large angle to the optic axis, and hence the form shown in Fig. 219c is to be recommended.

We can select

\[ S = D = 1.7 \ \text{mm}, \quad \text{with which} \quad f_{im} = 1.02 \ \text{mm}. \]

For \((NI) = (NI)_0 = 3900 \ \text{A turns} = 2300S\), saturation has set in—the actual value of \(f_{im}\) will then be

\[
(f_{im})_{\text{actual}} = f_{im} \left[ 1 + 0.14 \frac{(NI) - 1000S}{1000S} \right] = 1.18f_{im} = 1.2 \ \text{mm}.
\]

The curves shown in Fig. 218a cannot be used to calculate \(C_s\) and \(C_c\), as these have been calculated for an object lying at the immersion focus \(F_0\).

17.4.4 The Single-field Condenser-objective

The possibility of obtaining very short focal lengths and low aberration coefficients \(C_s\) and \(C_c\) by situating the object at the centre of lens (where \(B_z = B_{\text{max}}\)) was pointed out by Glaser as early as 1941, but has not been exploited until recently, by Riecke (1962a, b) and Ruska (1962, 1964), because of the practical difficulties of using such a lens. The first half of the field acts as a strongly convergent condenser (Fig. 219d) so that if the diameter, the angular aperture and the intensity of the illuminating beam are to be adjusted with any precision, the entry diaphragm must be centred perfectly and two high quality condenser lenses must be placed in front of the objective.
The excitation of the objective is such that

$$k^2 = \frac{eB_{\text{max}}^2a^2}{8m_0\Phi_0^8} = 3,$$

where $a$ is the half-width of the curve $B(z)$. For $\Phi_0 > 100$ kV, the lens is highly saturated but the curve $B(z)$ only broadens very slowly. We have

$$a \text{ (mm)} = 1.27 + 0.057 I \text{ (kA)} \approx f \text{ (mm)}.$$

In consequence, $C_s$ and $C_e$ do not vary much as $\Phi_0$ is increased (Ruska, 1966) and $C_s$ remains smaller than 1 mm even when $\Phi_0 = 200$ kV. At 100 kV, this lens is better than standard types of objective at 400 kV. Its chromatic aberration is given by $C_c \approx C_s/2$.

A further advantage is that the object is placed half-way between the poles; in the practical design, therefore, the specimen-holder and the apertures can be inserted straightforwardly.

17.4.5 Lenses with Permanent Magnets

No external current source is required by these lenses, and they are therefore very stable. It is, however, difficult to vary their optical characteristics over a continuous range.
For these lenses, we have

$$J = \int_{-\infty}^{\infty} B_0(z) \, dz = 0,$$

so that the image rotation is always zero. In a lens with a single gap, which cannot be adjusted (Fig. 219e), the form of the curve $B_0(z)$ is more complicated than that of the analogous electromagnet, since the above relation ($J = 0$) must be satisfied; the focal length is always greater. The fringe field is suppressed by using two gaps in tandem (Fig. 219f), forming a double lens. To alter the convergence, a soft iron magnetic shunt must be
moved in the vicinity of the permanent magnets, which will diminish the magnetic flux in the axial zone.

The magnets used for this purpose are usually made of a special steel containing aluminium, nickel and cobalt of the "Alnico V" type (Al 8 per cent; Ni 13.5 per cent; Cu 3 per cent and Fe 51 per cent).

The induction $B$ in the gaps of the lens illustrated in Fig. 219f may be calculated approximately as follows.

![Diagram of two-gap lens](image)

Fig. 219f. Two-gap lens, and the corresponding curve $B_o(z)$. $S$ is a mobile shunt of soft iron, which enables us to modify $B_o(z)$ slightly.

The line integral $\int Hdl$ of the magnetic field along a curve $(C)$ passing through the gap, length $l$, and magnetized material, length $L_M$, is given by

$$\int_{A(C)}^{B} Hdl = H_M L_M.$$
If we assume that the permeability of the soft steel is infinite, we have

\[ H_2 l_2 = H_M L_M = l_1. \]

\((H_2 \) and \( H_1 \) denote the fields in the gaps \( G_2 \) and \( G_1 \) respectively, and \( l_2 \) and \( l_1 \) are the lengths of these gaps.) Hence

\[ B_2 = \frac{\mu_0 H_M l_M}{l_2}; \quad B_1 = \frac{\mu_1 H_M l_M}{l_1}. \]

The flux of magnetic induction is conserved over the whole magnet. If \( \Psi_1 \) denotes the flux in \( G_1 \) and \( \Psi_2 \) the flux in \( G_2 \), we can define equivalent surfaces \( S_1 \) and \( S_2 \) for each gap, such that \( \Psi_1 = B_1 S_1 \) and \( \Psi_2 = B_2 S_2 \). If we assume that during its construction the magnet has been subjected to a constant induction \( B_M \) at every point, the flux passing through it is

\[ \Psi_M = B_M S_M. \]

The induction \( B_M \) is related to the magnetic field \( H_M \) by the characteristic \( B = f(H) \) of the magnetic material, which is provided by the manufacturer.

We have therefore

\[ \Psi_M = \Psi_1 + \Psi_2, \]

\[ B_M = \frac{B_1 S_1 + B_2 S_2}{S_M}. \]

We know that

\[ H_M = \frac{l_1 B_1}{\mu_0 L_M} = \frac{l_2 B_2}{\mu_0 L_M}, \]

and finally, we obtain

\[ \mu_0(S_M L_M)(B_M H_M) = B_2^2 S_1 l_1 + B_2^2 S_2 l_2 \]

or

\[ \mu_0 V_M(B_M H_M) = B_1^2 V_1 + B_2^2 V_2. \]

\((V_M, V_1\) and \( V_2 \) denote the volumes of the magnet and of the two gaps, respectively.) In the symmetrical situation \((B_1 = B_2 = B, V_1 = V_2 = V)\), we have \( l_1 = l_2 = s\); we obtain the simple relations

\[ B = \frac{\mu_0 L_M H_M}{s}, \quad (1) \]

\[ B^2 = \frac{\mu_0 V_M}{2 V}(B_M H_M). \quad (2) \]
If $B$ is fixed in a gap of given geometry, formula (2) shows that we must use a magnetic material having as high a value of the product $(B_M H_M)$ as possible, and work at the point on the characteristic $B = f(H)$ at which this product is greatest; $H_M$ is then determined and formula (1) tells us what length of magnet to use.

For Alnico V, for example, we have $(B_M H_M)_{max} = 3.5 \times 10^4$ (m.k.s.a.) = $3.5 \times 10^6$ gauss-oersted. The maximum field on the axis, $B_0$, is obtained in terms of $B_1$ and the dimensions of the gap, $s$ and $D$, from the same relation as we use for electromagnets:

$$B_0 = \frac{B_1}{s \sqrt{1 + 0.45 (D/s)^2}}.$$ 

To alter the convergence of such a lens, we move a magnetic shunt of soft steel near the gap by some mechanical means; this deflects part of the flux in the principal gap and thus causes a small variation in $B_1$ or $B_2$.

17.4.6 Superconducting Magnetic Lenses

Since the development of new superconducting materials (see §9.4) with which high current densities and very high field stability can be attained, lens designs for microscopes and diffraction devices have appeared which take advantage of their properties. These lenses are of several kinds:

(a) lenses without iron;
(b) coils with external magnetic shield;
(c) classical lenses with ferromagnetic pole-pieces and superconducting windings.

Lenses of the first type, suggested by Laberrigue and Levinson (1964 and 1966) and Fernández-Morán (1965) have the advantage that their properties are easily deduced, since $B(z)$ can be calculated and then matched to a Glaser bell-shaped curve, but the field falls off very slowly. It is therefore advantageous to enclose the superconducting coil in a magnetic shield of soft steel or a special cobalt steel. With very high fields, this shield will saturate locally near the axis, but the curves $B(z)$ will still be shorter.

The magnetizing coil can be made in various ways:
(a) Niobium–zirconium wire will allow us to obtain $nI = 30,000$ A in $1 \text{ cm}^2$ cross-section for fields of 20 to 30 kG (Fernández-Morán, 1965, 1966a, b; Ozasa et al., 1966).
(b) With niobium–tin ribbon, appreciably higher fields can be attained (50 to 60 kG); a ribbon 30 to 40 μ thick and 6·4 mm wide can operate with a steady current of 200 A at 30 kG, with 110 A at 60 kG and with 75 A at 90 kG. The behaviour of such a coil can be calculated by using the $I(B)$ characteristic and the formula given in §9.4.

(c) Niobium–tin rings may be used; these are obtained by vacuum evaporation of the compound Nb₃Sn onto thin layers of platinum. Each circular layer supports several concentric rings and the beam passes through a hole at the axis; these layers are then piled together to form a coil. In this case, the current cannot be obtained from an outside source. The assembled discs are now placed in a field $B_{\text{ext}}$ greater than the critical field, produced by an auxiliary solenoid and $B_{\text{ext}}$ is then gradually reduced; when the material becomes normal again ($B = B_c$), high induced currents are produced in the discs so that the flux passing through them remains constant. When $B_{\text{ext}} = 0$, the lens is excited to its maximum field strength. With such a lens, Siegel et al. (1966) and Kitamura et al. (1966) have obtained $B_{\text{max}} = 22·3$ kG, $a = 2·6$ mm, $f = 1·6$ mm, $C_e = 0·59$ mm. The technique used for constructing the rings is described in articles by Hanak (1964) and Petzinger and Hanak (1964).

In all these lenses, in which the field is essentially created by conductors, departures from rotational symmetry can be caused by imperfect positioning of the wires, ribbons or rings (either during construction or after cooling which may bring non-uniform forces into play at the windings). The lenses must therefore be provided with facilities for correcting astigmatism.

To obtain fields that are highly stable in time, it is preferable to use the persistent-current or short-circuit mode; this situation automatically arises with Nb₃Sn rings. With coils, we have only to provide a short-circuit across the input terminals, made of superconducting wire which can be held at a higher temperature than its critical temperature by heating. The coil is supplied in the normal way from an external source and when $i$ has been satisfactorily adjusted, the short-circuit is allowed to cool (and becomes superconducting), and the external source is removed. In theory, the current trapped in the coil will continue unaltered indefinitely. In reality, however, the resistances at the contacts will not be exactly zero and the current will decay very slowly, but the consequent decrease in $B$ is not detectable even after several hours (Fernández-Morán, 1966a, b) at an image with a direct magnification of 20,000.

The exact fine adjustment of the focal length of the lens can be achieved by means of a small auxiliary coil supplied from the outside, which is placed close to the principal lens (Ozas et al., 1966). Another solution is to use a ring transformer with a ferromagnetic core, the secondary of
which (having only a few turns) is connected in series with the winding of the principal lens. By altering the primary which is connected to an external source, the flux in the secondary can be varied as slowly as we may wish, and hence so too can the lens current (Fernández-Morán, 1966a, b).
CHAPTER 18
THE ANCILLARY PARTS OF AN
ELECTRON MICROSCOPE

An electron microscope is equipped with various appendages, which are
devices of the kind which are found in every electron optical instrument
(power supplies and vacuum systems for example). In the transmission
microscope, however, these devices are exceptional in that they must be of
extremely high quality. The provision of the potential for a magnetic micro-
scope barely differs from the corresponding arrangements in a diffraction
camera or an electron image-converter, save in that the degree of stability in
the former must be exceptionally high.

18.1 THE ELECTRICAL SUPPLY

18.1.1 The High Tension Generator

The current which is taken from the supply is very small, at most a few
hundred microamperes; the potential, however, has both to be highly
stabilized and variable, if possible continuously; the maximum value to
which we are restricted usually lies between 50 and 100 kV, and we shall
cite a few details of some of the various solutions which are employed
to-day.

(i) Rectification of the output of a transformer

An X-ray tube generator can be used and the ensemble will consist of a
transformer (50 c/s high tension) and diodes with the aid of which either
half-wave rectification or full-wave rectification can be obtained. The sta-
bility is achieved by holding the frequency of the alternating potential at
50 c/s and placing two saturated iron regulators connected in series, at the
input of the transformer; the output is smoothed with large condensers
(about one microfarad). This instrument is robust and simple, but the
output which it provides is insufficiently stabilized for a modern microscope,
although we should point out that an expert of world-wide renown was still
using a system which had its origin in this type of instrument in 1950

602
(Ruska, 1950). The stabilization can be improved (Bauer, 1943; van Dorsten, 1948) by using an electronic feed-back system. This combination has now, however, given way to systems of the type described in (iii) below.

(ii) A potential multiplier

With a device which multiplies the potential, the potential at the output of the transformer can be reduced and the inverse potentials which the valves have to stand-off are smaller (the diode rectifiers used in television will then be suitable, for example, or alternatively, a chain of silicon or selenium rectifiers can be used). As a particular case, we might mention the potential multiplier of Vastel (1951) which is used in the electrostatic CSF electron microscopes. The maximum high tension is 80 kV, stabilized at best to one part in a thousand, with an output of 100 \( \mu \)A. A 600 c/s oscillator, the output of which is stabilized and variable, provides an alternating voltage with a high enough frequency for a transformer and a filter of small dimensions to be used.

This oscillator can be replaced by a low frequency valve or transistor oscillator, the H.T. of which is easier to stabilize. When the frequency is of the order of a few kilocycles per second, miniature transformers and selenium rectifiers can be used. The final high tension is then stabilized in the same way as it is for the higher frequency systems described below. Stabilities of the order of one part in \( 10^5 \) can be obtained for about half an hour when the high tension is 100 kV. The stability is measured in the following way: the high tension \( V_0 \) is fed into a chain of very large resistances \( (R = 1000 \text{ M} \Omega \text{ for example}) \) which are arranged in such a way that there will be neither continuous discharges nor sparking between the terminals of the individual resistances. A small resistance, \( r \), is placed in series with \( R \); \( r \) for example might be equal to \( R/10,000 \). Between the terminals of \( r \) we find a potential \( \Delta V_0 \approx 10^{-4} V_0 \), and this we connect in opposition to a steady potential \( \Delta V_1 \) provided by a cascade of highly stable batteries. The fluctuations of the resulting potential,

\[
\delta V = \Delta V_0 - \Delta V_1,
\]

represent those of \( V_0 \) apart from a constant factor. The difference \( \delta V \) is traced directly on a pen-recorder, which has been calibrated beforehand.

(iii) High frequency generators

Devices of this type were perfected in the U.S.A. for supplying television tubes. Later, they were used in conjunction with the electron microscope (Vance, 1941) and nowadays a great many different types of instrument are equipped with them. Generators of this type have many advantages; in particular, the use of a high frequency allows us to use very small condensers in the filter circuits to obtain the same ripple. In addition, the
speed of adjustment, which is related to the frequency, is very rapid, and finally, the magnetic screening is very straightforward.

Diode rectifiers like those used in television have to be employed, as semi-conductor rectifiers have too large a parasitic capacity. These valves can be heated with a high frequency potential, however, by using a special oscillator and a special type of transformer with a straight ferrite core, the

![Diagram](https://via.placeholder.com/150)

**Fig. 220a.** Diagram showing a high tension supply of the Cockcroft–Walton type (potential multiplier), giving 50–100 kV in air. $C_1, C_2, \ldots, C_5$: condensers; $D_1, D_2, \ldots, D_5$: diodes; $T_{R_1}$: supply transformer; $R$: resistance chain; $r$: resistance used for adjusting the principal oscillator by feed-back; $T_{R_2}$: transformer for heating the diodes; $P$: primary (connected to earth), supplied by oscillator number 2; $B$: ferrite cylinder; $O$: insulating oil; $T$: perspex envelope; $F'_1, F'_2, \ldots, F'_5$: secondaries, at progressively higher potentials, connected to the filaments $F_1, F_2, \ldots, F_5$ of the diodes; $P_1, P_2$: metal plates with anti-corona rings $A$.

ends of which are left open. The primary is earthed and is wound directly onto a ferrite cylinder; it is enclosed within a perspex column with walls one centimetre thick. The secondaries, which are held at progressively higher potentials, are wound straight onto the perspex tube and connected to the corresponding valves. The insulation between the various stages of the multiplication and earth is provided by the perspex tube (see Fig. 220a).
Again, the high tension column can be built in a very compact form, and immersed in an insulating tank filled with a special oil; the high tension cable is then led away from the upper part, and the supply is fed in at the base of the tank. In this way, "H.T. units" giving 100 kV can be constructed within a cylinder 20 cm in diameter and 35 cm high.

The whole column assembly can, alternatively, be left exposed to the atmosphere, provided we take precautions against continuous discharges by building the column between two metal plates, *P₁* and *P₂*: the lower plate is earthed, and the upper one is connected to the high tension; the potentials of the several parts of the assembly, which increase from bottom to top, effectively correspond to those of the equipotential surfaces within the parallel plate condenser, *P₁ P₂*. A diagram of this kind of generator is to be found in an article by Vastel (1951). An oscillatory high tension is produced in a high-Q pilot-oscillator, and then passes through a stage of power-amplification which also assists the regulation. This potential is then increased by an air-cored HF transformer, rectified, and multiplied by a four-stage Cockcroft–Walton (or Greinacher) multiplier. The stabilization is effected by means of feedback, using a potential divider followed by a differential two-stage direct current amplifier which acts on the grid potential of the power-amplifier. If the H.T. *V₁* increases by an amount +Δ*V₁*, the high frequency potential with which the column is supplied varies in such a way as to reduce Δ*V₁* to zero. The high tension which is thus obtained is very stable and can be varied between 20 and 60 kV (simply by altering one of the potentials of the power amplifier); in this way, in a small space, we can obtain a potential which can be continuously varied between 20 and 60 kV, which has a ripple of less than two parts in 10⁵, and which has a long-term stability of the same order for a ±5 per cent variation of the mains. When the current which is taken is varied between zero and 100 μA, the voltage varies by one part in 10⁴; above 500 μA, however, the potential falls suddenly to a negligible value. This is a most useful phenomenon, especially in the case of the electrostatic microscope, as it automatically prevents electrical break down from destroying the surfaces of the electrodes. This source of potential is, furthermore, hardly dangerous at all; an accidental contact only produces the discharge of one of the condensers of the cascade, of a comparatively small capacity. Moreover, a safety resistance of 5 or 10 MΩ reduces this discharge current to a practically harmless value.

The Metrovic microscope is provided with an analogous device (Haine, Page and Garfitt, 1950); the H.T. ranges between 25 and 100 kV, and is stabilized to one part in 25,000. The very simplified arrangement which is used with the Zeiss A.E.G. gives ± 50 kV with a stability of one in 5000. The reader is referred to a detailed fundamental examination by Mautner and Schade (1947) for the sophisticated theory of the high frequency generator of high tension and to an article by Baldinger (1959) devoted to H.T. generators of the "cascade" type.
(iv) Electrostatic machines

A few attempts have been made (Grivet, 1948) to use Felici's electrostatic machine which has a metallic belt and no regulator. A system of this kind is attractive by virtue of its simplicity and its reduced dimensions (it is cylindrical, 20 × 50 cm), but so far is in its infancy so far as microscopy is concerned; at present, the H.T. which it provides is sufficiently stable to be used for visual scrutiny and straightforward photography, but inadequate if the best resolution is to be attained. Recently, more encouraging results have been obtained by Felici (1957) using a machine with an insulating belt, provided with an electronic regulator.

H.T. supplies of this type are at present in use, which have a range of 20 to 100 kV and can be adjusted to within one part in 10^5, and which yield a few hundred microampères; they take up very little space, and are easy to handle.

(v) Very high tensions

Sources of a steady H.T. between 1 and 4 MV have been produced for many years now by specialist firms, but the stability has not been adequate for the very delicate experiments of electron optics. The exacting demands which nuclear physics has recently made have, however, spurred designers on to study ways of regulating and stabilizing these very high tensions with precision, with the result that it is now possible to use them in electron microscopy.

These very high tension sources are of two types:

"Cascade" models, which resemble those already described, which may be either open to the air (for tensions up to 2 MV; they are then very cumbersome) or enclosed within pressurized chambers (10 or 20 kg of nitrogen or freon) which allows the dimensions to be considerably reduced.

Electrostatic machines, which may use a flexible belt, of the "van de Graaff" type; such machines are capable of giving a very high tension of the order of 5 MV, and for V > 500 kV, they are built within pressurized chambers. Alternatively, they may use a cylindrical insulating conveyor placed within a pressurized chamber (these are the SAMES machines, studied by Felici). We can place several machines of this type in series, and obtain potentials from one to two MV with an output of several mA. The potential can be stabilized to a few parts in 10^5.

A 300 kV microscope which has been described recently (Tadano et al., 1958) employs a van de Graaff generator, while the very high tension microscope at the Institut d'Optique Électronique at Toulouse (Dupouy et al., 1960b, 1961) is supplied by a 1.5 MV cascade generator, similar to the one described by Bas (1960) and by Adler et al. (1960). The system has ten stages, and is supplied with a 400 c/s alternating potential provided by an oscillator; there are three columns instead of the two shown in Fig. 220a,
and these are supplied at the mid-point by a transformer $T_1$, as this considerably reduces both the ripple of the high tension and the regulation effects. Two separate circuits are employed in the stabilization. The potential produced, the oscillator is “coarsely” adjusted, to within about one part in $10^3$, by the first; the transformer $T_1$ is then supplied with a feedback signal which is obtained by extracting the fluctuations of the high tension through a potentiometer, then amplifying these fluctuations and changing their sign (just as in the other systems described above). The stability is about one part in $10^5$, for a few minutes.

The preceding systems will certainly have to be abandoned if voltages of 2 MV and higher are to be obtained; cascade generators have too much energy stored in the chains of condensers, and the accelerating tube can be seriously damaged by a breakdown. In van de Graaff generators, which must be built compactly (with accelerating columns and generator in the same pressurized chamber), the motion of the belt causes troublesome vibrations at the high voltage terminal. We might then consider using electrons from a linear accelerator (Gilroy et al., 1966); there is not longer a high voltage electrode, no electrostatic column and hence no danger of destructive discharges. At the present time, however, the energy spectrum of the electrons emerging from an accelerator is too wide, and a magnetic monochromator would be needed to reduce the dispersion to a reasonable figure (10 to 20 volts).

18.1.2 Slow Variation of the Potential

Feedback devices are excellent for maintaining stability for a comparatively brief interval (a few seconds, for example, which is longer than the time of a normal exposure) but all these devices drift slowly over about ten minutes. This drift is a most inconvenient phenomenon when long photographic exposures are necessary (in electron diffraction, in the velocity analyser, and in the ion microscope), but also in the case of the transmission microscope it is a hindrance, as about 30 seconds are necessary for focusing and exposure, and in any case, the H.T. has to be connected up several minutes beforehand if there are to be no modifications between the focusing and the taking of the photograph. With an electrostatic instrument, this nuisance is reduced.

A simple solution of this problem is to leave the H.T. generator working continuously. A better solution is due to Haine (1950) who suggested that the generator should accelerate an auxiliary electron beam which would pass through a magnetic deflector with 180° deviation (cf. Chapter 22).

Any fluctuations in the high tension act to displace the beam at the exit of the deflector. The beam falls on two identical electrodes which receive equal electron currents for a given H.T. value. If the voltage fluctuates, one of the electrodes receives more charge than the other. If the two elec-
trodos are connected to the two inputs of a differential amplifier, we finally obtain an error potential which is used to regulate the H.T. generator through a feedback loop.

18.1.3 Current Regulation in Magnetic Lenses

As we have seen in Chapter 16, the magnetizing current of a lens must be highly stabilized; the degree of stabilization will depend on the following factors:

*The purpose of the lens*: the objective, for example, must be protected more carefully against fluctuations than the condenser, the stability of which is less important. Regulation of the objective is facilitated by the fact that the current is constant—the current variations which are needed for focusing are small enough not to complicate the structure of the regulator. In projective lenses, on the other hand, wide and carefully controlled variations of the potential have to be available to cover the vast range of magnifications and illuminations.

*The structure of the objective*: the conditions become less stringent when the slope of the curve of focal length as a function of current is reduced by saturation. In the most unfavourable case, however, this current has to be about twice as stable in relative value as the H.T. voltage; in a good commercial microscope of this kind, the lens current is controlled to about one part in 40,000, and the H.T. to one in 20,000.

The solutions which have been adopted have been based upon two properties. The first of these is simply the well-known stability of batteries of accumulators. Up to 1950, therefore, Ruska supplied the lenses of his large research microscopes with a battery of accumulators of 64 volts and a capacity of 100 A hr. The maximum total current was about 2.5 ampères for the whole assembly of lenses. The magnetizing coils are then wound with coarse wire, so that the current density is low (1 A mm⁻²). The heating is therefore very small, and as the yoke is rapidly cooled, the temperature of the coils remains constant after a short running-in period (between 1/₄ and 1/₂ hr). After this time, the coil resistances are constant, as are the resistances of the manganin control rheostats—the e.m.f. of the accumulators which discharge slowly (1/₄₀ of the capacity) is constant enough for the problem to be regarded as solved. The practical success of the solution depends upon the care and attention with which the apparatus is handled. Technically, the success of the method is due to the use of a thick wire with little insulation, but it is also helped by breaking up the winding into smaller parts which are separated by sheets of copper, the sole purpose of which is to conduct the heat away towards the yoke.

Nowadays, however, it is considered inconvenient commercially to use a large battery with all the careful attention which it requires, and we replace it with the aid of electronic rectifiers and regulators.
The power required is rather high, however (between 50 and 250 watts according to the number of lenses and the bulk which they can be permitted) and electronic valve regulators can only provide this at high voltages (100–800 V). A commercial regulator of this kind might employ, for example, a valve rectifier giving 750 V after filtering, the current for each lens being stabilized independently by a separate electronic regulator operating in the way shown diagrammatically in Fig. 220b. The current is then small, of the order of 0.1 A. The windings are of fine wire, and conduct the heat away inefficiently, as the volume which the insulation occupies is appreciable; the heating is slightly worse, and the resistance varies over a long period corresponding to the slow thermal drift, so that stabilization of the current (instead of the voltage) becomes imperative. As a result of recent progress in transistor technology, however, very powerful L.T. supplies (up to 10 kW, 200 V) have been developed which are as stable as generators using valves (see Sauzade, 1960, for example). The coils can be wound with coarse wire with very little insulation when this type of supply is used, and the thermal conductivity is thus better. This conductivity can be still further increased by impregnating the coils with an epoxy resin (araldite) thickened with silica.

![Diagram of a current regulator](image)

**Fig. 220b.** Diagram of a current regulator. If an increase in $E$ produces an increase in $i_0$, the potential at the grid of the triode II which had previously been zero becomes positive. $i_1$ increases and hence the potential at the grid of the triode I is lowered. The current $i_0$ thus tends to diminish.

Good current regulation is assured by using circuitry similar to that of the potential controls; the potential has simply to be stabilized across the terminals of a constant resistance $R$ made of a special alloy, which is placed in series with the load (the windings of the lens, in this case). The current in the objective must be regulated with particular care; examples of this kind of regulator are to be found in the literature (Elmore and Sands, 1950; Dupouy and Perrier, 1960a for H.T. systems with valves. Sauzade, 1960 for transistorized systems). The variations of current with which the necessary adjustments are made are obtained either by altering the feedback circuit or by placing an attenuator with a constant input impedance between the regulator and the lens.
18.2 HIGH TENSION EFFECTS

18.2.1 Breakdown between the Electrodes under a Steady High Tension

In principle, the leads to the various electrodes carry only very small currents, and it is only the cathode lead which takes a steady current of a few $\mu$A to compensate for the charge led away by the beam. No other lead has to carry steady current. Nevertheless, sudden increases in the load on the high tension are often observed when a microscope is in action. This is a sign that a "breakdown" between the electrodes has allowed a discharge to pass directly from one electrode at high tension to the other which is earthed.

Even the slightest of discharge phenomena can be rendered visible by placing a fluorescent screen in the neighbourhood of the region in which the breakdown takes place, as such a screen can be sensitive to current densities of the order of $10^{-11}$ A cm$^{-2}$, but the appearance of discharges which are examined in this way gives no information about the way in which they were formed. On the observation screen, where they are particularly easy to detect, they usually have the form of a circular glow, sometimes intense, but of short duration. The discharges which are produced at the projective lens have a diameter of one or two centimetres, at the objective of about three to five centimetres, while the discharges which occur at the gun fill the whole screen. It is interesting to notice that this glimmering radiance often varies in intensity from one point to the next. In other cases, however, after the instrument has been in operation for some time, the appearance is different—a steady glow of light appears on the screen, which can be of a very feeble intensity but which gradually becomes more intense as the high tension is increased until the screen is blindingly bright. Another method of detection and study is to examine the appearance of the electrodes themselves, either in an experimental electrostatic lens or more simply in a bell-jar within which two plane or spherical surfaces are placed, one of which is earthed while the other is connected to the high tension (Anderson, 1935; Goltz, 1940; Bertein, 1946; and Arnal, 1955). The average current can be measured with a galvanometer, the shape of the impulses can be examined with an oscillograph and the number of X-rays emitted can be counted with a scintillation counter.

In normal conditions (electrodes which have not been outgassed placed in a dynamic vacuum), four classes of discharge can be distinguished.

(i) Below a certain value of the H.T., no discharge occurs and the vacuum is a perfect insulator.

(ii) Suddenly, between 30 and 60 kV, with a spacing of a few millimetres and a source with a rather high internal resistance (40 M$\Omega$), very short
microdischarges appear \(10^{-4}\) sec), several per second, statistically distributed in time; the mean current lies somewhere between \(10^{-8}\) and \(10^{-6}\) A. When the high tension is increased, the rhythm of the microdischarges speeds up (to 10 or 20 per sec), and the mean current increases to several \(\mu\)A. The interelectrode region is the source of a diffuse glow, while at the same time certain metallic parts or certain insulators are struck by charged particles and give a bluish luminescence.

(iii) If we continue to increase the potential, a different kind of discharge appears, which sparkles, and has something of the appearance of ordinary sparks in the air. This type of discharge has the form of a sudden passing glimmer, highly localized, which appears on a metallic or insulating surface. It is exactly as if some kind of self-focusing mechanism were concentrating the discharge into a small beam. Local heating is produced over a microscopic zone (Bertein and Grivet, 1951) and the metal evaporates, which produces an increase in the intensity as a result of thermionic emission and ionization of the vapour. Whereas the microdischarge leaves no microscopic trace on the electrodes, the sparks leave a mark, each of which extends over a region several hundred microns in diameter, and the surface is completely covered with small points each of which corresponds to a point at which the metal has been evaporated.

(iv) If the tension is increased still further, the sparks become more and more frequent, and a new kind of behaviour, arcs, sets in; the current is seen to increase further, after which it becomes stable towards several tens of \(\mu\)A. Permanent bright points appear on the electrodes.

When an insulator is placed between the electrodes in a region in which microdischarges appear, a different mode of behaviour appears, which is characterized, in its early stages at least, by a discharge which is not very intense, but which is continuous, rather than the microdischarges described in the preceding paragraph. A feebly and diffusely illuminated region is observed at the surface of the insulator, between two and five millimetres in diameter, which seems to shift towards the anode, together with small bright permanent curves (which are visible by virtue of a blue fluorescence which appears at the metallic or insulating surface) which are usually very clearly defined; the concavity of these curves faces towards the cathode, their appearance changes with the high tension. For even higher tensions, sparks appear as well.

Despite any amount of research, these phenomena are still poorly understood. For this reason, we shall restrict ourselves to describing some of the features which are best understood among these mechanisms. They play a central role in electron microscopy as they disturb the focusing by lowering the potentials at the electrodes. Further, the charged particles and the flashes of light which are produced fog the photographic plate.
18.2.2 The Parameters which Influence the Breakdown Effects in the High Tension

(i) The role of the vacuum

The dynamic vacuum within a microscope is maintained between $10^{-4}$ and $10^{-5}$ mm of mercury. We might suppose that the best high tension conditions are obtained with the best vacuum, and blame the pump if electrical breakdown occurs too frequently.

In reality, the vacuum plays an unexpected role; the pressure within the apparatus is in general considerably lower than the value at which the results depend upon the quality of the vacuum. It has been demonstrated by van Atta and van de Graaff (1933), Anderson (1935a), Trump and van de Graaff (1947) and by Arnal (1956) that as soon as the interelectrode region ceases to be subject to gaseous ionization, the breakdown mechanism becomes independent of the quality of the vacuum. This condition is fulfilled as soon as the mean free path of the electrons and the ions which are moving through the residual gas within the apparatus is longer than the inter-electrode spacing ($10^{-3}$ mm Hg). Reducing the pressure does not improve the high tension situation in so far as the pressure of the residual gas can slowly affect the concentration of the molecular layers of gas adsorbed at the electrode surfaces. A sudden increase of the pressure may even halt the discharges momentarily, but after a few dozen seconds, they recommence; reducing the pressure has the opposite effect (Bertie, 1946; Arnal, 1956). It even seems to be true that the threshold at which the discharges appear is higher when the pressure of the residual gas is relatively high, about $10^{-4}$ mm Hg (according to Clifford and Fortescue, 1952, the threshold is changed by a factor which can reach two).

(ii) The role played by the condition of the surface; formation

A priori, the state of the surface ought to play an important role if it is cold emission which sets off the discharges, as was believed for a long time. The least irregularity ought to have a harmful effect, as it would produce an increase of the electric field and hence an intense emission of electrons as a result of the point; Anderson (1935), and Arnal (1956) have shown that it has none. Extremely careful polishing is in fact useless and often even harmful, as foreign bodies become embedded in the superficial layer with which the metal of the electrode is coated. An auxiliary bombardment of the electrodes by an electron beam produces only a very slight change of the breakdown threshold. When the electrodes are bombarded with ions on the other hand (Bruck, 1953), breakdown occurs for very small values of the electric field, of the order of 1 kV cm$^{-1}$.

The state of polishing of the surface has been studied by Hadden (1951)
and by Arnal (1955) who used, in particular, electrolytic polishing. The conclusion to which they came is that unpolished surfaces are "neither better nor worse" than surfaces polished electrolytically. It seems that the breakdown thresholds are almost identical, irrespective of the quality of the polishing; whether the surface of any given metal is roughly turned, attacked by an acid, or polished electrolytically, it still has virtually the same characteristics in practice. In electron microscopy, however, the mechanical tolerances (due to ellipticity aberration, for example) require certain surfaces to be polished.

One parameter is, on the other hand, of cardinal importance; this is the superficial state, the nature and the quantity of the various substances within the superficial film of the electrodes, either adsorbed or in the form of inclusions. We know that in effect the surfaces of metals are never physically clean, as various substances (grease, water, oxygen and nitrogen for example) which come from outside can become attached to them in the course of the preliminary cleaning, just as under vacuum conditions within the apparatus, gas occluded within the body of the metal can diffuse to the surface; as breakdown phenomena consist of exchanges between one surface and another, the state of these latter plays an important role. We can appreciate this by watching the electrode surface during the pre-breakdown phase with the aid of an optical projection microscope with a high magnification (Razin et al., 1960). Under the action of an intense electric field ($10^5$ or $10^6$ V cm$^{-1}$), local points of microscopic dimensions appear (a few microns across) composed either of oil molecules, or of small fragments of the metal itself. These points grow, and then break off, thus engendering a microdischarge. After a certain formation time has elapsed, the process becomes more and more rare. Various treatments are applied, to "clean" the surfaces—bombardment with electrons or ions, careful chemical cleaning, or vacuum outgassing. Good results are obtained by producing a discharge in hydrogen at a pressure of 1 mm, with a current density of 0.25 A cm$^{-2}$ for three minutes. The everyday method, however, is simply "formation of the electrodes", which consists in increasing the H.T. very slowly (after reversing the polarity of the electrodes if necessary); small discharges are produced, and then disappear, only reappearing at a higher value of the potential. In this way, electrostatic lenses (with a spacing of 4 mm) can be made to operate at more than 60 kV, whereas the usual figure is only 45 kV.

(iii) The role of the nature of the metal and of the adsorbed gases

The various attempts which have been made to determine the influence of these parameters have produced contradictory results. The superficial state of the metal and the quantity of occluded gas seem to be more important than the nature of the electrodes. Goltz (1940) has compared the
thresholds at which the microdischarges appear in various metals (Fig. 221). He studied the variations in the threshold which result from a first breakdown, and he recommended the use of chrome–nickel steel with chemically cleaned surfaces. Arnal (1955) related Goltz’ observations, and above all the intensity of the microdischarges, to the absorption capacity of the metal in hydrogen. The development of the preliminary discharges would be at least partly due to an exchange of positive ions, $H^+$, and negative ions, $H^-$, with secondary emission coefficients $A^+$, $A^-$ such that $A^+ A^- > 1$. Iovov (1960) too has shown how important is the ionic component of the current. The ions are produced by the adsorbed gas, and are thermally emitted by the surface which is heated locally by the arrival of a heavy particle or an electron which sets off the breakdown. So long as $A^+ A^-$ remains greater than unity, the intensity of the discharge increases, but if the particle exchange remains highly localized at some point on the electrode, the gas adsorbed at the surface of the metal may be temporarily impoverished and $A^+ A^-$ fall below unity; the discharge is halted. The adsorbed layer may then be formed afresh by diffusion of gas from the interior of the metal, and the discharge can recommence after being triggered off anew. For

![Figure 221](image)

**Fig. 221.** The threshold at which the micro-discharges appear: (a) polished steel; (b) polished aluminium; (c) annealed steel; (d) steel covered with contamination.

pressures ($p$) of the order of $10^{-4}$ mm Hg, the high tension conditions would then be improved by a reduction in the number of ions present, which could be the result of a neutralization phenomenon, the ions coming into contact with the gas molecules between the electrodes. When $p > 10^{-4}$ mm Hg, there would on the contrary be ionization within the volume of this gas, an increase in the current, and hence the high tension conditions in the system would be worsened.
The importance of heavy particles in the initiation and maintenance of the preliminary discharges has been displayed by Pivovar and Gordienko (1958) also; these authors entirely eliminate the electron component of the current with the aid of an intense magnetic field parallel to the surface of the electrodes in question—the threshold of the preliminary discharges is practically the same, irrespective of whether or not the magnetic field is present.

By using electrodes made of different metals, it has been possible to demonstrate that it is above all the material from which the anode is formed which determines the characteristics of the microdischarges; although the cathode is probably important as a source of particles once the phenomenon has begun, its nature has very little influence upon the value of the threshold and the frequency of discharges. This is one of the more certain results of these delicate explorations, as it has been confirmed by various authors (Goltz, 1940; Heard et al., 1952; Clifford and Fortescue, 1952; and Arnal, 1955).

(iv) The roles of the potential and of the interelectrode spacing

If the discharges depended only upon the electric field (field emission), we should have a simple relation between the spacing and the breakdown potential, and the phenomenon would depend essentially upon \( V/d \). This simple hypothesis is not at all valid, however, and when the interelectrode spacing is increased, the voltage threshold does not increase in proportion; the gradient of the mean breakdown potential decreases constantly as the distance increases. It is not the field at the electrodes which is relevant, but the total voltage which is applied across them.

For a given fixed interelectrode distance, \( d \), we can increase the potential and determine the thresholds at which the various phenomena appear. These thresholds are rather poorly defined, as the experimental conditions are only reproducible with difficulty; in particular, the state of the surface is never the same between one experiment and another. To attach an order of magnitude to the various phenomena, however, we quote a few of the results. The threshold of the microdischarges, \( V_s \), is tolerably well defined. Anderson (1935) suggested that for copper, \( V_s \) is related to \( d \) by a bilinear formula, of the form

\[
V_s = \frac{144d}{0.97 + d}, \quad 0.2 < d < 4 \text{ cm}.
\]

The experiments of Arnal (1955) led to a law of the form

\[
V_s = k d^{1/3},
\]

in which \( k \) is of the order of 3 kV cm\(^{-1/3}\) for copper. Ionov (1960) sets out with the hypothesis that ions are locally thermally emitted at the points of impact of particles which act as triggers, and offers the following explana-
tion. The local temperature is a function of the power received, and hence of the overall potential between the electrodes, $V$. The ion emission coefficients, $A^+$ and $A^-$, are, however, dependent upon the local electric field $E$, which lowers the extraction potential (Zandberg and Ionov, 1959).

Both for sparks and for arcs, the thresholds are poorly defined. Cranberg (1952) and Heard et al. (1952) proposed a law of the form

$$V_s = k \sqrt{d},$$

for the sparks, in which $k$ is of the order of 40 for stainless steel, and 30 for copper. For the arcs, Anderson (1935) suggested that the empirical law

$$V_s = \frac{590d}{0.77 + d}$$

is obeyed.

(v) The role of the internal resistance of the source of high tension

The importance of this parameter was first demonstrated by Arnal (1955). In general, the electrodes are supplied through a resistance $R$ (the internal resistance of the source, itself, together with a safety resistance) and with respect to earth, they have a parasitic capacity $C$ (of a few dozen picofarads in the case of an electron microscope). The discharges fall into three classes, according to the value of $R$. For small values (tens of k$\Omega$), the microdischarges last between $10^{-2}$ and $10^{-4}$ sec, and have the form of a fairly constant current of a few dozen microamps—they are repeated every 10 or 20 sec.

For medium values of $R$ (up to about 20 M$\Omega$) a current peak is produced by the discharging of $C$, and is followed by a steady period of about the same duration as in the preceding situation but less intense.

Finally, for large values of the resistance ($\geq 20$ M$\Omega$), the signals become ever shorter (down to $10^{-4}$ sec). The current which flows during the breakdown reduces the potential between the electrodes (by about 1000 V) with the result that the discharge ceases by itself; the frequency, on the other hand, increases (to many dozens per second). For the electrodes to be well preserved, it is this final mode of breakdown which is the most advantageous.

18.2.3 Breakdown with a Pulsed Potential

The microdischarge process only reaches its full intensity after an interval of about $10^{-4}$ sec. If the potential is applied in short pulses which last less than $10^{-5}$ sec, therefore, the normal current will not have the time to build up completely. Herreng (1943) has shown, in the course of some experiments with X-ray tubes excited in pulses, that the threshold for sparks becomes 95 kV when the tension is applied in 1 $\mu$sec pulses whereas it had been only 56 kV for a steady high tension.

This advantage which is to be had with a pulsed potential can be put to good use in an electrostatic microscope. The first practical attempts were made by Spivak and Doubinima (1953) in the emission microscope
(2 kV pulses between electrodes 0.08 mm apart). These authors found an improvement in the maintenance of the potential of five to seven times, but only a small magnification was obtained.

More recently, Gardez (1959) has built a very high tension emission microscope; an immersion objective which could not support a potential higher than 40 kV in steady working conditions will support 95 kV if the potential is applied in pulses. (The pulses last about one microsecond, and have a frequency of 200 per second.)

18.2.4 Breakdown along the Insulators

Frequent discharges occur along the insulating supports of certain of the electrodes. Gleichauf (1951, 1951a) has shown that this is not due to the same phenomena as the breakdown between metallic electrodes. The threshold at which microdischarges appear is considerably lowered when an insulator is placed in the interelectrode region—the reduction can reach one half of the threshold which is observed in the absence of the support. The law which governs the appearance of the microdischarges, $V_s = k \sqrt{d}$, is approximately applicable to the breakdowns which occur along the insulators; the constant $k$ depends upon the nature of the insulator—it is of the order of 20 kV mm$^{-1/2}$ for polystyrene and 30 to 40 for perspex and porcelain (Borovik and Batrakov, 1958). The best results seem to be obtained with unglazed zirconium porcelain (Trimmer and Pearlman, 1951) and with araldite considerably thickened with silica (Gribi et al., 1959). The rougher the surface, the better maintained is the potential; this result can easily be verified with porcelain, for example, moulded in such a way as to produce ribs normal to the lines of force of the electric field (and thus to elongate the line of the breakdowns).

In reality, however, the preliminary discharges in the lens always appear much sooner than the formula above would lead us to suspect. The insulators are subjected to a continual bombardment of charged particles which have their origin either in charge exchanges between the electrodes (even though these currents be well below the preliminary discharge threshold, and produce no detectable phenomena between the electrodes) or alternatively in various kinds of contamination (oil or grease molecules which leave carbon deposits after being broken down by the charged particles).

If the charges which are collected on the insulator are too numerous at certain points, very intense local electric fields will be created; these may either penetrate through the insulator, if this latter contains a conductive rod, as in the case of potential leads, or give rise to bright plumes by local ionization of the residual gas, which corresponds to an intense breakdown current and may lead to breakdown in the vacuum.

If the charge density is low, the charges manage to flow towards the metal (as the surface conductivity of the insulator is never exactly zero), following the electric field lines tangential to the insulator. The transition
zone between the metal and the insulator will thus play a very important role; the lines of electric field become progressively detached from the insulator, cross the intervening space and terminate normally at the metal. Further, the surface at which metal and insulator are in contact is never perfectly plane, and little crevices remain where there is no electrical contact. The charges accumulate in this transition zone, and very often flow out into the vacuum where they may ionize the residual gas and also produce emission of secondary particles from the metal. The magnitude of this phenomenon is related both to the number of charges present and to their energy when they strike the metal (and hence to the operating potential); it is essentially the origin of the discharges which limit the H.T. which can

Fig. 222. Two possible ways of joining an insulator to an electrode in such a way as to reduce the risk of breakdown: (a) by metallizing the base of the insulator the electrical contact is improved. (b) by lengthening the path of the charges which flow between the insulator and the metal (Gribi, 1959). The broken line represents lines of electric field strength.
be applied to lenses which have been carefully designed and do not break down between the electrodes.

A first improvement is to ensure that the contact zone between metal and insulator is clearly defined; the end of the insulating support is metallized either by vacuum deposition or by a chemical method, and a proportion of the charges can thus flow steadily across. Alternatively, we can move the metal away from the end of the (non-metallized) insulator, to prevent the field lines which direct the motion of the charges from detaching themselves, as far as possible; Fig. 222a represents this type of solution.

Gribi et al. (1959) have suggested a further improvement which is shown schematically in Fig. 222b, and this has made it possible to increase the operating potential of a lens of normal dimensions from 45 to 50 kV, for which it had initially been designed, to 70 kV. The dotted line represents the electric field line tangential to the flange of the insulator; the shape of the electrode is selected in such a way that this line terminates normally at the bottom of a rounded groove hollowed out in the metal. At $A$, the electric field is lowered, and the local vacuum is better. The threshold at which microdischarges produced by ionization of the gas appear is considerably higher.

18.3 THE VACUUM

18.3.1 Pumps

The majority of microscopes are provided with a primary rotary pump with vanes, together with one or more secondary diffusion pumps. As mercury diffusion pumps require a liquid air trap, it is preferable to use an oil diffusion pump or a rotary molecular pump, and although the latter has the inconvenient property of letting oil into the body of the microscope, many instruments are still equipped with them. Further, it seems that although the pumping rate may nominally be higher, the practical efficiency of oil diffusion pumps is inferior to that of mercury pumps equipped with a good liquid air trap, as the characteristic pumping rate of the latter is higher for certain vapours which condense at the trap; when films are being used, the photographic emulsion and its base provide an abundant source of water vapour and organic vapour.

When the microscope is in use, the vacuum has frequently to be let down and then re-established; it is necessary, therefore, to have a fast and easily controlled pump system. Figure 223 shows an arrangement which is commonly used; it consists of a primary pump I and a diffusion pump II. A system of taps, valves and points at which air can be admitted allows the vacuum to be established or atmospheric pressure to be restored to any part of the arrangement at will.
In the microscopes in regular use today, every effort is made to reduce the pumping time; the instruments are equipped with small "vacuum locks" which allow us to change the specimen or replace the photographic plates (or film) rapidly, without destroying the vacuum elsewhere. The object can be changed in a few seconds in this way (instead of in the one or two minutes which are required in a well-designed instrument without locks) without switching off the filament of the gun or cutting the high tension. Technical details are to be found in von Ardenne (1940), van Dorsten (1948) and von Borries (1949). To obtain the very highest performance and to avoid contaminating the object with fragments of molecules of carbon compounds that the beam has produced, by destroying organic molecules from the pump oils, grease and the joints themselves, it is advisable to use a "clean", very high vacuum, at least in the vicinity of the electron gun, the objective and the specimen holder.

Ultra-high vacuum techniques are being more and more widely used—apparatus made wholly of stainless steel, metal joints, quartz or aluminium insulators, for example—with pumping units consisting of high-speed ion pumps (§13.3.5); without baking, the limiting vacuum of $10^{-8}$ mm Hg is reached after a few hours, but after heating the whole evacuated region to 150 or 200°C, a vacuum of $10^{-9}$--$10^{-10}$ mm Hg can be attained. Work on the nature of the residual gases and the importance of their various partial pressures has shown that they consist principally of nitrogen and carbon monoxide (Hartman and Hartman, 1966).

The interior of the microscope may be divided into two distinct regions, to prevent vapours extracted from the photographic emulsion from rising
towards the object: the zone around the object is a "clean" high vacuum while in the camera, a good vacuum ($10^{-5}$ mm Hg) is adequate. The two regions can be separated in practice by placing a metal plate beyond the lens; a small hole is drilled in the plate around the axis to allow the beam to pass. A simpler solution that is used in many microscopes is to "clean" the region around the object by cooling the adjacent surfaces to the temperature of liquid nitrogen ($T = 77^\circ$K) and connecting this region to a supplementary ion pump with a capacity large enough to maintain a pressure of $10^{-9}$–$10^{-10}$ mm Hg (Grigson et al., 1966, Valdrè et al., 1966).

18.3.2 Measurement of the Vacuum

In normal use, it is not necessary to measure the pressure precisely all the time, but for the detection of accidental leaks, it proves to be most useful to do so.

The crude vacuum can be estimated from the appearance of a glass discharge tube to which an alternating potential of several kV is applied. This gives an indication of the quality of vacuum between 1 and 0.01 mm Hg. At lower pressures, the discharge disappears. The colour of the discharge which depends upon the nature of the gas is useful as a guide to leaks—the doubtful part of the instrument is smeared with alcohol, which passes into the instrument if the part is defective, and produces a striped whitish glow. A Pirani gauge can also be used.

The high vacuum is measured with the aid of a Penning gauge, which gives an output of about 100 $\mu$A for a pressure of $10^{-5}$ mm Hg, or an ionization gauge in which the ion current is effectively proportional to the pressure, one microampère corresponding to about $10^{-6}$ mm Hg. Special ionization gauges (ultra-high vacuum gauges) have been developed for measuring pressures lower than this ($10^{-7}$–$10^{-11}$ mm Hg). With ion pumps, the value of the discharge current in the pump gives the pressure at the pump directly. To detect a small leak which is apparent with this secondary gauge is a matter of logic and patience; the apparatus must be isolated section by section, so that the leak can be localized.

A quicker and more refined method of detecting leaks is now available, as a number of commercial firms have produced leak-detectors using helium. A very fine jet of helium gas is squirted onto the point of the wall to be tested; if the wall is porous, a certain amount of the gas will penetrate through into the evacuated chamber, which is connected to that of the detector. The latter is a simplified but very sensitive mass-spectrometer, adjusted to detect the helium line. Whenever any helium is present in the evacuated chamber, an electric signal is produced, the amplitude of which is effectively proportional to the size of the leak.
A great deal of useful information about the detection of leaks, measurement of the rate of leaking, and the use of one particular helium detector is to be found in an article by Choumoff and Laplume (1957).

18.4 THE GENERAL APPEARANCE OF A MICROSCOPE

We shall conclude with a description of three French microscopes, one of which is electrostatic, the second magnetic and the third is the very high tension microscope recently constructed by G. Dupouy and his collaborators (1960, 1961).

18.4.1 The Compagnie Générale de T.S.F. “M IX”

This is a microscope with three electrostatic lenses, which appeared in 1950–1. The appearance of the exterior is shown in Fig. 224, while Fig. 225 is a cross-section of the optical array. The microscope itself is 75 cm high, but the case and the various appendages increase this to 140 cm.

The maximum magnification is of the order of 50,000. In everyday use, the resolving power is 40 Å at a useful magnification of 25,000 which is adequate for many of the uses to which the microscope is put; if a stigmator is used, a resolution of 20 Å can be attained.

By adjusting the convergence of the first projector (the focal length of which can be varied between 2.5 and 50 mm), the magnification can be reduced to \( G = 2000 \), as we have already explained (§ 15.4.2). The instrument can be used as a diffraction camera, by retracting the contrast diaphragm which is situated above the objective. The illumination is produced by a gun similar to the one designed by Bruck and Bricka (§ 8.3.3). The focal length of the objective is 6 mm, and that of the second projective lens, 2.5 mm.

The object is supported either by a small capsule between 2 and 4 mm in diameter, in the centre of which is an opening between 0.5 and 0.1 mm in diameter, or by a very fine mesh grid with a similar outer diameter.

---

Fig. 224. The CSF microscope, M IX. All the electrical controls are collected together into the right-hand side of the instrument, and all the controls for the vacuum into the left-hand side; the various different taps shown in Fig. 223 are here all controlled by a single cock. The controls for the specimen-holder are to be found on the table. In the centre, reading from top to bottom, we see the gun (off which the cover has been lifted), the window through which the object is inserted, the screen with which the intermediate image can be seen (the image produced by the objective alone), the screen for examining the final image and finally the camera.
Fig. 225. Cross-section of the optical arrangement within the CSF M IX.
These grids or capsules are attached to a specimen-holder which can be moved in two mutually perpendicular directions; the movement is effected with the aid of mechanical demultiplication. The structure of the specimen-holder is illustrated in Fig. 226. Above all, we try to ensure that the movement is smooth, and that the device is insensitive to vibrations, as these affect the resolution very adversely. One solution is to make the holder of the specimen-holder almost an integral part of the objective with the aid of two toroidal joints gripped by two flat discs attached to the upper part of the objective. With other solutions, the object migrates slowly as a result of hysteresis of the springs and thermal expansion of unsymmetrical parts which are heated by the impact of the beam, and this is most inconvenient for long exposures.

18.4.2 The O.P.L. Microscope (Optique de Précision de Levallois)

This microscope, which is provided with a condenser and three magnetic lenses, is a recent development (Fert and Selme, 1957), and is at present in regular production by the French O.P.L. company. It represents the industrial equivalent of Fert’s optical bench (1954), and its originality lies in the construction of the column. This latter consists of a tower of standardized elements, which gives a great practical flexibility (see Fig. 227a). Each element (lens, condenser, ...) is set within a cylindrical section bounded by two plane faces; the upper face contains the groove of an O-ring vacuum seal, and the lower face a centring flange. These faces are ground with great care to ensure that the column, once assembled, be rigid, and that
Fig. 227a. Cross-section of the O.P.L. microscope, built in separate units which can then be stacked up.
the optical elements be aligned. It is hence a simple matter to replace one lens by another, better adapted to the problem in hand. From the electron source to the screen, every essential point can be reached and cleaned without dismantling the column. The instrument can be modified rapidly: by adding an angled element, for example, the gun can be tilted for reflexion microscopy. Figure 227a shows a section of an early version of the instrument, and Fig. 227b a photograph of a recent model.

Fig. 227b. Photograph of a recent O.P.L. microscope.
The magnification can be varied continuously between 2000 and 110,000 at 75 kV and between 2000 and 75,000 at 100 kV. The gun can be supplied with a H.T. of 50, 75 or 100 kV, stabilized to about one part in $10^5$, and gives about ten or twenty $\mu$A. The condenser and the objective are provided with different sets of diaphragms; a device to correct the astigmatism is placed behind the objective. The lens current is regulated to within about one part in $10^5$.

![Image](image.png)

**Fig. 227c.** The surface of a copper phthalocyanine crystal (photographed with an O.P.L. microscope). Periodicity: 9.9 Å.

The theoretical resolving power is 5 Å, and in its present form, the instrument does indeed give a resolution which in practice is less than 10 Å (see Fig. 227c, which shows the surface of a monocystal of copper phthalocyanine, the periodicity of which is 9.9 Å). Special attention has been given to the transitions from microscopy to diffraction and to microdiffraction, which entail the introduction of a special stop into the object plane of the intermediate lens.

Diffraction diagrams can equally well be obtained with a very high resolution ($5 \times 10^{-6}$) by placing the specimen-holder beneath the projective lens; for further information, the reader should consult Chapter 21.
18.4.3 A Very High Tension Magnetic Microscope

We shall now give a brief description of the microscope which has been built by G. Dupouy and his collaborators at the Institut d'Optique Électronique in Toulouse, of which the maximum operating potential is planned to be 1.5 MV. When a very high tension is to be used, numerous modifications are necessary, not to the general optical arrangement, but to the technological features and to the overall bulk of the instrument.

(i) The electron source

The gun is of the ordinary kind, and works at a potential of a few tens of kilovolts in the usual way; it is placed at the top of an accelerating column, and is held at the prevailing H.T. (Φ). Just as in electrostatic accelerators, the column is a tower of ten porcelain rings, separated by metal discs which carry the accelerating electrodes. The latter are in the form of tubes with rounded ends; these tubes are separated by a gap which forms a two-cylinder accelerating lens (see Fig. 227d). The electrodes are held at potentials which are regularly spaced between (−Φ) and earth (Φ = 0), by means of a resistance chain across the H.T. source. The ratio \( \frac{V_{n+1}}{V_n} \) varies, therefore, between the top and the bottom, since

\[ \Delta V = \text{constant} \simeq -\Phi/10, \]

and the lenses are progressively less convergent.

![Fig. 227d. Section of an accelerating tube for very high energy electrons.](image_url)

I: insulator (porcelain or araldite); E: electrode; A: anti-corona ring.
The complete tube behaves like a convergent lens, the focal length of which is determined by the design; it is wise to insert a condenser between the gun and the tube, as we can then adjust the size and aperture of the beam at the base of the column where a diaphragm of small diameter acts as entry pupil for the optical system. The accelerating column is about twenty feet high. The radii of curvature of all the connexions and of the outside surfaces of the metal discs in the column are large enough to prevent continuous breakdown by the corona effect. The high tension installation and the accelerating column are shown in Fig. 227e.

Fig. 227e. H.T. generator of the Cockcroft–Walton type (on the right) and accelerating column (on the left) of the Very High Tension microscope at the Institut d’Optique Électronique laboratory of the C.N.R.S. at Toulouse (Dupouy and Perrier). Above the column are the electron gun and its various adjustment mechanisms, enclosed in a polished metal cover with a large radius of curvature. The whole system is installed inside a spherical Faraday cage.
(ii) The optical system

The optical system consists of a tower of separate elements each with its own clearly defined function, just as in the O.P.L. microscope (see Fig. 227f). There are two condensers $C_1$ and $C_2$, an objective $O$, and intermediate

Fig. 227f. Cross-section of the optical system of the V.H.T. microscope. $D_1$: entry diaphragm; $T$: flexible membrane; $C$: double condenser ($C_1$ and $C_2$); $O$: objective; $N$: deflexion coils for aligning the beam; $O_1$, $O_2$ and $O_3$: the magnetic circuit; $B$: coil; $I$: intermediate lens; $P$: projective; $R$: observation chamber; $D$: observation window (with lead glass), with an eyepiece or a television camera; $F$, $E$: fluorescent screens; $A$: plate-camera; $B$: pedestal with levelling screws; $S$: polar units, which can be moved and centred—the objective polar unit contains the specimen-holder, a stigmator, and various different diaphragms.
sections with windows of lead glass and openings for pumping. The whole assembly is connected to the accelerating column by means of a flexible membrane $T$, and rests on an observation chamber $R$, which is itself placed on a large plate $B$ supported on levelling-screws. The overall height is about eight feet, and, not counting the lead screens which provide protection against X-radiation, the whole assembly weighs about four tons; indeed each lens weighs several hundred kilograms.

The lenses are fixed. To align the various elements, therefore, they are provided with a metal "pole-unit", $S$, which contains the pole-pieces and the various accessories necessary for the role of each lens (specimen-holder,
diaphragms, stigmator, for example); this pole-unit can be displaced laterally by micrometer screws. The vacuum seal between the pole-unit and the rest of the lens is guaranteed by toric joints.

The objective weighs 700 kg and its focal length, \( f \), is 5.5 mm when the current in the windings (28,500 turns) is 0.6 A and 1 MeV electrons are used. It may be used either with an ordinary specimen-holder, or with the special specimen-holder which is described in § 20.3.4; the latter may perhaps enable us to observe living tissue which is surrounded by air at atmospheric pressure.

A major problem is the protection of the operator against X-rays; this has been solved by placing lead screens around the “hot zones” of the microscope (and in particular, at the various levels at which a part of the beam is intercepted by diaphragms: at the entry, the condenser \( C_2 \), the specimen-holder, and the observation chamber); in addition, heavy concrete walls are erected (between the accelerating column and the microscope, and around the room in which the microscope is situated). The various screens and the image which is finally obtained are observed with television cameras. A photograph of the instrument is shown in Fig. 227g; the lead sheaths about the entry diaphragm and the observation chamber can be seen.

18.5 THE OPTICAL ADJUSTMENT OF THE INSTRUMENT

To obtain good images, the possession of a good microscope is insufficient—it has also to be used correctly. In the majority of high quality instruments, the column of lenses can be centred, and hence the mechanical imperfections remedied. The experimenter must be capable of aligning his microscope correctly in the minimum time possible, as this is one of the most basic adjustments to be made to the instrument.

18.5.1 Adjustment of the Gun

In the majority of modern microscopes with a heated cathode, it is easy to adjust the automatically biased gun. The point of the filament has to be centred with respect to the opening of the Wehnelt (or grid), and its height has to be adjusted, and these are operations which can often be performed visually, after withdrawing the gun from the instrument (the RCA microscope). All that is then necessary is to adjust the inclination of the gun to the axis approximately, in order to compensate possible defects of the preliminary centring, and to obtain a true and symmetrical image (on an intermediate screen for example) of the point of the filament for a convenient value of the excitation of the lenses. If the position of the filament can be adjusted while the lens is in use, the grid and the anode remaining centred by virtue
of the construction, it will be necessary to adjust the height of the filament to obtain the correct value of the saturation current and then to adjust the centring until the best possible image of the crossover is obtained on an intermediate screen placed at the same level as the object (the C.S.F. microscope).

18.5.2 Alignment of the Microscope

The purpose of this alignment is, above all, to reduce the displacements of the image which result from variations of the high tension. This is a vital adjustment in the magnetic microscope, as it has become apparent that this "voltage centring" in the magnetic microscope in much more important than the "current centring" which was once foreseen, and with which it does not always coincide (the reason for this lies in the parasitic magnetic fields). As a centring criterion, therefore, we select the point at which the image is least sensitive to variations in the high tension. In an electrostatic microscope, these variations only produce second order changes in the total magnification if all the potentials are obtained from the same potential drop; if, however, the instrument is focused by modifying the polarization of the objective with batteries, these variations can become more troublesome—the instrument must be centred in such a way that the image increases away from its centre, and hence the smallest possible perturbation of the whole ensemble occurs when the polarization of the objective is slightly increased.

In a magnetic instrument, a variation of the high tension produces first order effects; these can be either an increase or a decrease of the size of the image, or above all, a rotation about some point which it is convenient to bring to the centre of the screen if the perturbation of the whole image is to be as small as possible.

We shall describe the alignment procedure in two typical cases; first of all, the case of an electrostatic microscope without a condenser, with two lenses only, and after, that of a magnetic instrument with two lenses and a condenser.

(i) The electrostatic microscope with two lenses, without a condenser

We suppose that the objective and the projective lenses have been centred geometrically onto a common axis during construction. The centring of the instrument consists, therefore, solely in bringing the crossover to this axis.

First, the beam is directed onto the region of the object which is to be examined, which it illuminates brilliantly. With the aid of the focusing potentiometer, the convergence of the objective is then slightly modified. The image which is observed on the screen is then the image of the projection of the object which is produced by the source of illumination on a plane of variable height. Figure 228 shows that the result is that if $S$ is not on the axis, the
image point $I$ which had previously been at the centre of the fluorescent screen is displaced. All that need be done is to adjust $S$ until the sole effect of varying the focus is an expansion or contraction of the image about the centre of the screen. This is a particularly easy operation if the gun can be turned about a point on the axis of the apparatus in the immediate vicinity of the object; if this is not the case, the beam has to be brought onto the object by a lateral displacement of the gun after each alteration of the inclination.

![Diagram](image)

**Fig. 228.** The alignment of an electrostatic two-lens microscope. The crossover $S$ does not lie on the axis, and a modification to the convergence of the objective displaces $I$ to $I'$. The plane $F_0$ is conjugate to the plane $F_p$ of the intermediate image when the objective has its normal convergence. When the convergence is less, it moves to $F'_0$.

(ii) *Alignment with respect to the potential—the case of a magnetic microscope with a condenser*

Here, it is a question of aligning the optical centre of the condenser and the crossover on the axis which joins the optical centres of the objective and the projective lenses. The simplest procedure is illustrated in Fig. 229a,
where the alignment criterion which has been selected is that the image shall be as stable as possible in the presence of variations of the high tension.

(a) A preliminary approximate adjustment of the position of the gun brings the image of \( S \) into the centre of the fluorescent screen, while the condenser is adjusted to give the maximum brightness (the image of \( S \) on the object). The image is vividly brilliant, but the instrument is far from being centred.

(b) The condenser is adjusted to give a very faint illumination (which is then nearly parallel). The point \( I \) of the image which occupies the centre of the screen is then singled out; the H.T. is reduced (or cut if it is fixed), so that the supply capacity discharges. The convergence of the lenses increases: the final image is that of a projection of the object onto a plane which becomes closer and closer to the objective. Figure 229b shows that if the illumination is not centred, the point \( I \) moves away from the centre of the

Fig. 229. The alignment of a magnetic microscope with two lenses and a condenser. When the accelerating potential is normal, the intermediate image plane is \( F_p \), the plane conjugate to \( F_0 \) with respect to the objective. When the accelerating potential falls to a lower value, the intermediate image plane moves to \( F_p \) in order to remain conjugate to the screen with respect to the projective. The point conjugate to \( F_p \) with respect to the objective is \( F_0 \).
screen (to a point \(I'\)), and we take note of the direction in which it moves. A magnification and a rotation of the image around \(I\) is seen at the same time, but with these we shall not be concerned; for the sake of clarity, therefore, the rotations have been suppressed in the figure.

(c) The high tension is restored to its normal value, and the excitation of the condenser to the value which corresponds to maximum brightness. The image of the source reappears at the centre of the screen. We then shift the combination of condenser and gun in such a way that the spot of light is moved a considerable distance from the centre of the screen in the same direction as the point \(I\) had moved previously (Fig. 229c).

(d) The bright spot is brought back to the centre of the screen by moving the gun alone. The source \(S\) and the optical centre of the condenser are now both closer to the projective-objective axis (Fig. 229d).

(e) The cycle of operations just described is repeated until the point \(I\) ceases to move when the high tension begins to fall, and a spot of the maximum brightness is obtained at the centre of the screen by a simple adjustment of the condenser. The complete centring procedure, when performed well, only takes one or two minutes, even in the case of an instrument initially totally eccentric.

18.5.3 Focusing

This is a delicate operation as the depth of focus of the instrument is so large. The best method is based upon direct observation of the Fresnel fringes at the edge of a detail of an object with adequate contrast. The search for the point at which the positive fringe disappears is facilitated by the fact that the contrast is reduced at the exact focus; a focusing microscope is also helpful.

This method is, however, only practicable for rather experienced workers, and the designers have tried to simplify the approximate focusing which is required in everyday use in various ways. With a "wobbler" (Le Poole, 1947) the illuminating beam is made to oscillate, which is equivalent to artificially widening the aperture. If the focusing is imperfect, the image at the screen vibrates, and to obtain a well-focused image, this vibration has simply to be reduced to the minimum (the Philips microscope). Oscillation of the illumination can be replaced by a multiple illumination produced by three sources (one axial, and two lateral). A poorly focused microscope gives a triple image (the Trüb and Täuber microscope). Both of these procedures lead to an image slightly out of focus, as a result of the spherical aberration of the objective.
CHAPTER 19

THE ACTION OF THE OBJECT IN A MICROSCOPE ON THE ELECTRON BEAM; CONTRAST IN THE IMAGE; VELOCITY ANALYSERS

19.1 THE TWO EFFECTS TO WHICH ELECTRONS MAY BE SUBJECTED ON PASSING THROUGH A THIN FILM

19.1.1 Definition of the Mass Thickness $\mu_s$

Since the Lenard window was discovered in 1895, a great deal of work has gone into clarifying the behaviour of fast electrons (1000 eV to 200 keV) as they pass through a thin film of matter. A clear and valuable summary of these fundamental studies is to be found in a very thorough review by Bothe (1933) from which we extract only two essential points.

The first of these is the definition of the fundamental parameter upon which the effects of traversing the solid depend. This is the mass thickness, $\mu_s$; in fact the actual thickness "$d$" through which the electron passes plays a role no different from that of the density $\mu$, and the action of the layer depends upon the product

$$\mu_s = \mu d,$$

in which $\mu$ is the normal specific mass (per unit volume) or density. In addition, the effects which are observed can be classified into three categories according to the value of $\mu_s$, or if the experiments concern a single material only (aluminium was a common choice), according to the value of the thickness $d$.

19.1.2 The Three Types of Scattering. Definition of a "Thin Film"

We can best understand the character of each of the scattering processes by examining the principle on which the theory rests. The scattering, retardation or deflexion which we observe is only the resultant effect of the interactions, in which each electron has been involved, with the nuclei and electrons of the film close to which the electron trajectories happen to

638
have passed. Matter is virtually empty, since the diameters of electrons and nuclei are only of the order of a hundred thousandth of an Ångström unit (a millionth of a millimicron), whilst their mutual separations are a few Ångströms from one nucleus to another and a few tenths of an Ångström from nucleus to electron. It is because the space within the material is predominantly empty that a projectile has a chance of passing through a continuous layer about a hundred Ångströms thick (about 30 atomic layers)—the thinnest which we know how to make—without being either retarded or deflected. As the thickness is increased, the chances of passing near enough to another particle for the beam electron to interact increase, and it is then that the effects of retardation and deflexion appear. The same is true if the atomic number of the material is increased, as the number of planetary electrons increases in proportion, and in addition, the extent of the zone over which the nuclei have an effect increases with their charge. This elementary theory accounts well for the importance of $\mu$, and explains the equivalence of $\mu$ and $d$. Further, it is adequate for us to separate out the three categories.

**Multiple scattering**

In this case, the layer is thick enough for every incident electron to be subjected to a statistically "large" number of interactions. The mean scattering angle, as well as the mean loss of velocity, is considerable. The theory is statistical in nature, as only the statistical trends can be observed with any ease. The most probable energy with which the electrons emerge, $V_m$, is given as a function of the incident energy, $V_0$ by Whiddington's law (1914):

$$V_0^2 - V_m^2 = K \mu,$$

in which $K$ is a constant which is characteristic of the substance; for aluminium, $K$ is of the order of $10^{12} (V \text{ cm})^2 \text{ g}^{-1}$.

**Plural scattering**

Here, the interactions are too infrequent for statistical averages to be meaningful, but too numerous for a simple composition of the diverse effects to be possible. This is an intermediate case, in which the theory is of little help.

**Single scattering**

The layer is thin enough, and made of a sufficiently light material, for each electron to have only one chance at most of interacting with one of the particles in the barrier through which it passes.
In this chapter, we shall only be concerned with single scattering, which occurs when the substance is light, as we can see from Fig. 230, where the curve which describes the maximum thickness for which the scattering remains single is depicted. This maximum thickness for which the scattering is single corresponds to a value of $\mu_s$ of the order of $10^{-5}$ g cm$^{-2}$, or to a thickness which ranges between 100 and 1000 Å according to the nature of the specimen; for the heavy metals, which are used in shadow techniques, the thickness has to be reduced to about 10 Å.

19.1.3 Elastic and Inelastic Collisions

The simplicity of single scattering, the feature with distinguishes it from plural scattering, is more clearly apparent if we take into account the fact that the elementary interactions which are the origin of the scattering, and which are known as collisions since they occupy a very short time in comparison with the period of the planetary motion of the electrons around the nucleus, fall into two categories.

Elastic collisions

These are essentially collisions with the nuclei, as the impact laws of classical mechanics are valid in this case, and give a simple result; if there is
to be an appreciable exchange of energy between two particles in a collision, their masses must be comparable. Here, on the contrary, the disparity is very large, for \( M/m \) is of the order of 50,000 for aluminium and 22,000 for carbon, and the collision simply deviates the electron without affecting its velocity.

**Inelastic collisions**

These are predominant when one of the electrons of the beam encounters one of the electrons in the specimen, as the two participants in the interaction are of the same mass, which is the case most favourable to a mutual exchange of energy—this always entails a loss for the incident electron which was much the faster. To verify this, we need only point out that whereas the energy quanta which are associated with the Bohr orbits (which appear in the experiments performed by Frank and Hertz on absorption in gases) are only a few volts in magnitude, the energy of the incident beam may reach some tens of thousands of volts.

The difference between elastic and inelastic collisions is clearly apparent in the general formula which embraces both cases and expresses the relative value \( \delta V/V \) of the energy loss \( \delta V \) which corresponds to a given angle of deflexion, \( \theta \).

A collision with a nucleus of mass \( M \) and atomic number \( A \) (with \( m \) representing the mass of the electron) is described by

\[
\frac{\delta V}{V} = \frac{m}{M} \tan^2 \theta = \frac{1}{1842 A} \tan^2 \theta \sim 0 \ldots \text{ elastic collision.}
\]

In a collision with another electron, however, we have

\[
\frac{\delta V}{V} = \frac{\tan^2 \theta}{1 + \tan^2 \theta} \sim \tan^2 \theta \quad \text{inelastic collision for small angles.}
\]

In single scattering, where there is only one collision during the passage through the film, the emergent electrons fall into one or other of two classes: those which have been elastically scattered by a nucleus, and those which have lost energy as a result of an inelastic collision with an electron. In the plural scattering region, on the contrary, the situation is considerably complicated by the fact that the successive collisions can belong to different categories.

**19.1.4 Elastic and Inelastic Scattering at Very Small Angles**

The simple models which we have described in the preceding paragraphs—electron scattering at an isolated nucleus or a free electron—only represent the real situation well for very small angles; the formulae of §19.1.3, for example, are valid in the diffraction region for angles of the order of a degree or a few degrees, but they become unreliable if \( \theta \) falls
below half a degree as it does in microscopy ($\theta = 5 \times 10^{-4}$ to $10^{-3}$ radians). In the domain of small angles, we must therefore take into account the binding between the planetary electrons and the nuclei—the theory becomes complicated, and only provides an incomplete description. It does however provide a satisfactory reply to the first of the two important questions which we are prompted to ask.

In the elementary solid angle $d\omega$ in the vicinity of $\theta = 0$, into what proportions are the retarded electrons (inelastic scattering) and the electrons which retain their initial velocity divided? The theoretical results of von Borries (1948) who, like Boersch (1947a), has a made critical study of these delicate points, are set out in Fig. 231; they are quite well confirmed experimentally. Considerably more electrons are retarded than continue without loss of velocity, and in each case, the figure gives the ratio $dN/N_0$ of the number $dN$ of electrons which emerge, in the given category, into unit solid angle, to the number $N_0$ which are incident per unit solid angle.

Further, what is the mean loss of velocity in traversing the material, and what is the spectrum of the velocity loss at the exit? The theory has so far proved to be unable to explain the experimental results; this explains why measuring techniques in this domain are so important. That discrete losses of the order of ten volts exist has been shown experimentally, and has provided an incentive for the construction of accurate and stable velocity filters, with which this category of electrons can be eliminated; as we shall see later, the final result is an increase in the clarity of the images and diffraction patterns, by an improvement of their contrast.

19.1.5 Collision Cross-section

In the preceding paragraphs we have described the scattering and retardation phenomena in a fashion which is adapted to electron microscopy; we have, that is to say, selected certain properties and a particular vocabulary in which to describe them. However, the theory has been considerably developed, and applied to several different situations, and another way of expressing the facts is usually found to be more convenient. It seems useful to introduce the reader to this general form of expression, by defining succinctly the quantities which are found to be useful, and which differ from those which have been used earlier in this chapter.

We are led to define a new quantity, the “effective cross-section”. We consider a parallel incident beam, carrying $N_0$ electrons per unit area of cross-section. We consider the number of electrons, deflected laterally through a mean angle $\theta$, which lie within an elementary solid angle $d\omega$ which is defined by the cones with summit angles $\theta$ and $\theta + d\theta$. The quantity

$$\sigma(\theta) = \frac{1}{N_0} \frac{dN}{d\omega}.$$
homogeneous with respect to area, is known as the "effective differential cross-section" relative to the angle $\theta$. To define the total number of particles scattered into a cone of angle $\theta$, we introduce the integrated form of this expression,

$$\sum_{\theta} = \int_{0}^{\theta} \sigma(\theta) \cdot 2\pi \sin \theta \cdot d\theta.$$ 

Since $d\omega = 2\pi \sin \theta \cdot d\theta$, this quantity is a measure of the total number of electrons deflected through an angle less than or equal to $\theta$. The results of scattering experiments, therefore, will be set out by showing how $\sigma$ varies with $\theta$, and by stating whether or not it is retarded electrons which are involved.

![Graph](image)

**Fig. 231.** The values of the effective differential cross-section in the passage through a flake of carbon (von Borries, 1949).

In Fig. 231, the distribution of the electrons which emerge from a thin film of carbon is shown in terms of the "effective cross-section"; the mass thickness $\mu_s$ is $10^{-8}$ g cm$^{-2}$, and the initial energies lie between 20 and 200 kV. The conclusion of the preceding paragraph is found expressed in a new form—for small angles, the collision cross-section is far larger for the curves which describe inelastic collisions than for those which correspond to elastic collisions. Inelastic collisions are predominant.

More recent information on this topic is to be found in an article by Marton, Simpson, Fowler and Swanson (1962).
19.1.6 Discrete Energy Losses, and the "Collective" Theory

Single scattering, which plays so important a practical role in image-formation, is still imperfectly explained from the quantum point of view; the existence of discrete energy losses, however, clearly suggests that it

![Graphs showing energy loss spectra](image1)

![Graph showing energy loss spectra](image2)

![Graph showing energy loss spectra](image3)

**Fig. 232.** Some examples of the spectra obtained by Voge and Rutheman, showing characteristic losses of the order of 10 eV.

is from this standpoint that the explanation is to be found. First, we shall summarize the clearest characteristics as they appear from several recent critical studies, and the numerous, though not wholly concordant, experimental results (see Fig. 232).

(i) The losses appear in the form of narrow or wide bands.

(ii) The energy loss associated with each band lies between a few volts and 100 V; when the bands are narrow, they can be quite numerous, and the characteristic energies are then usually integral multiples of the energy loss of the first band (the minimum loss).
(iii) It is not possible to relate the appearance of the energy-loss spectrum to the nature of the specimen unequivocally. Fine bands appear for certain metallic films (Al and Mg for example), but also for certain compounds and insulators (SiO$_2$, PbS, Sb$_2$S$_3$ and quartz); the spectra of some compounds are close to those of their base metals (Mg and its salts); some metals, however, have wide bands (Cu, Ag) as have certain non-metals (C).

The state of the surfaces of the films which are used is of fundamental importance, which explains the wide discrepancies between the experimental results. By evaporating an Al film within the body of the apparatus, Powell and Swann (1959) have shown that the spectrum rapidly deteriorates in a few minutes as a result of oxidation.

It is not possible to explain these energy jumps simply as the number of energy quanta which are necessary to extract a bound electron from an inner shell, as the energy in question would be too large, nor as the energy which would be required to transfer a peripheral electron into a valency or conduction band (in the sense in which this expression is used in the theory of metals) as the band would be too wide.

The bands remained without a plausible explanation until 1953, when Bohm and Pines developed a theory of the collective behaviour of electrons in plasmas. This latter is the name which is given to a medium which is electrically neutral but fluid, a completely ionized gas for example or the electrons in which the ions in a metal are immersed; the electrons are free to wander about within the metal and are in fact highly mobile, but it is difficult to understand their motion if only the mutual forces between immediately neighbouring particles are considered. The Coulomb law, which contains the factor $e^2/r^2$, shows that the more distant neighbours (which lie at a distance $R$) of any given electron are sufficiently numerous (they occupy a volume $4\pi R^2 \cdot dR$) to produce a field as important as that of the nearest neighbours, if they act in unison. It is certain that in fact they do not all act concordantly, but this hypothetical case gives an insight into the difficulty which has for so long delayed the calculations. A collective theory must be constructed, which deals with the behaviour of the totality of the electrons. The great achievement of Bohm and Pines was to have attacked this problem directly, and to have succeeded in solving it; starting from the individual interactions, they were able to build up the collective theory. This is a difficult procedure, and the interested reader will find a lucid guide in two articles by Pines (1954, 1955).

Here, we shall simply quote the result: the plasma can be regarded as a continuum, and hence possesses characteristic frequencies of vibration. Nevertheless, it is a very special kind of medium, a mixture of ions and electrons and we must not wholly lose sight of this granular aspect of the structure. This means that the medium possesses a frequency limit below which no vibration can be sustained. This minimum frequency has been known since the basic research of Tonks and Langmuir on ionized gases;
this frequency is the plasma frequency, and is given by

\[(2\pi f)^2 = \omega^2 = \frac{n e^2}{4m \varepsilon_0}.\]

It depends only upon the number \(n\) of electrons per cubic metre; \(e\) is the charge and \(m\) the mass of an electron, and \(\varepsilon_0\) is the permittivity of the vacuum.

Bohm and Pines have shown that such "plasma oscillations" can exist, and consist of a longitudinal oscillation of the electrical charge density (the central force is of an electrostatic nature). The discrete losses which are observed experimentally are each the result of the emission of a plasma oscillation quantum, or "plasmon".

Bohm and Pines' original calculations required a formidable arsenal of mathematical weapons. Since their initial work was performed, however, the physical basis of the phenomenon has been brought out in various other studies (Fröhlich and Pelzer, 1955).

We consider a longitudinal exterior electric field \(E\), frequency \(\omega\), wave number \(k\). Through the Coulomb interaction, this excitation sets up forced oscillations among the electrons; the amplitude of the response is proportional to \(\frac{1}{\varepsilon(\omega)}\), in which \(\varepsilon(\omega)\) is the dielectric constant. If the frequency \(\omega\) is such that \(\varepsilon = 0\), the system can oscillate by itself without outside stimulus. This self-oscillation is precisely what we mean by plasma oscillation. To calculate the plasma frequency \(\omega\), we are led to study the polarization of the electron gas, or what is equivalent, to evaluate its dielectric constant \(\varepsilon(\omega)\). We consider first of all a gas of free electrons, a case analogous to the situation in the upper atmosphere, which was studied by Sir Edward Appleton in the twenties. There will be an opposition between the Coulomb forces due to the applied field, and the Brownian motion of the individual electrons. But at very long wavelengths, the random motion of the electrons can be neglected, so that the screening effect is determined, or more accurately, limited by the inertia of the particles; the dielectric constant is thus given by:

\[\varepsilon(\omega) = 1 - n \frac{e^2}{m \omega^2 \varepsilon_0},\]

in which \(n\) is the number of electrons per unit volume, \(e\) their charge, \(m\) their mass and \(\varepsilon_0\) the dielectric constant of free space. The plasmons are of frequency \(\omega_p = \sqrt{\frac{n e^2}{m \varepsilon_0}}\), the value of \(\omega\) for which \(\varepsilon(\omega_p) = 0\). When the wavelength decreases and hence the wave number increases, the calculated frequency differs more and more from this limiting value, \(\omega_p\). This aspect of the effect has been observed experimentally in the Ruthemmann absorption, by a study of the angular distribution of scattered electrons; the energy of the discrete groups depends upon the scattering angle and
hence upon $k$ (Watanabe, 1955, 1956; Marton, Simpson and McCraw (1955, 1955a).

If the wave number $k$ increases to such an extent that it becomes comparable with the Debye wave number $k_D$, the plasma oscillation disappears. $k_D$ is the reciprocal of the screening radius† which is equal to $\omega_p \langle V \rangle$, in which $\langle V \rangle$ is the root mean square velocity of the random motion of the electrons. This effect has been clearly observed experimentally, and displays how overwhelming is the influence of the individual motions at short distances.

We pass now to the real case of a solid, metal or insulator. A new factor appears, namely the opposition between the Coulomb forces and the binding forces associated with the ions located at the sites of the crystal lattice. We denote the mean random energy of an electron in the lattice field by $\langle \omega_{\text{exc}} \rangle$. If $\langle \omega_{\text{exc}} \rangle \ll \omega_p$, the electrostatic forces are predominant, and the plasma frequency is close to $\omega_p$. This is true of Be, Al, Si and Ge, and for these elements we observe fine lines at the predicted positions. If, on the contrary, $\langle \omega_{\text{exc}} \rangle \approx \omega_p$, the discrete losses are displaced, and above all, considerably widened. This is the case for many of the transition metals and compounds, among other substances. The agreement between theory and experiment is still only qualitative, but it is difficult to improve it at the present stage of development of the subject. The energy losses (in electron volts) are tabulated below for the most intense band.

I. Good agreement

<table>
<thead>
<tr>
<th></th>
<th>Be</th>
<th>C</th>
<th>Al</th>
<th>Si</th>
<th>Ti</th>
<th>Ge</th>
<th>Sn</th>
<th>Pb</th>
<th>Bi</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hbar \omega$</td>
<td>19</td>
<td>25</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>$\Delta E$</td>
<td>19</td>
<td>22</td>
<td>15</td>
<td>17</td>
<td>22</td>
<td>17</td>
<td>12</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

II. More mediocre agreement

<table>
<thead>
<tr>
<th></th>
<th>Fe</th>
<th>Ni</th>
<th>Zn</th>
<th>Ag</th>
<th>Cd</th>
<th>Au</th>
<th>Pt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hbar \omega$</td>
<td>31</td>
<td>35</td>
<td>13</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>$\Delta E$</td>
<td>21</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>20</td>
<td>24</td>
<td>23</td>
</tr>
</tbody>
</table>

Plasmons are present in insulators, just as in metals (Mott, 1955). To take part in a plasma oscillation, an electron has no need to be free; at the frequency in question, it has no time during a single period $2\pi/\omega$ to feel

† The screening radius can be directly defined in the following way: suppose we were to alter one of the equilibrium characteristics of the plasma, in a small region around some point $M$; the plasma would "cure" this defect itself in that a self-modification would occur within a sphere of radius $1/k_D$ and outside this sphere, the "disease" would disappear completely. Debye was the first to consider this self-screening mechanism in the case of a liquid electrolyte.
the crystal lattice binding potential (Nozières and Pines, 1959), which on
its own produces a much slower motion.

In conclusion, we should mention the existence of more complex collec-
tive modes, and in particular, the "surface" plasmons which characterize
vibrations localized in the neighbourhood of the surface of the film (Stern
and Ferrell, 1960, and Ritchie, 1957). The frequency predicted by the
theory for these plasmons is \( \omega_p / \sqrt{2} \) and this seems to be confirmed by experi-
ment (Powell and Swann, 1959).

A very complete study of the characteristic energy losses of electrons in
solids has been published by Marton, Leder and Mendlowitz (1955);
this article collects together the numerous experimental results and summa-
rizes the various theories which have been proposed to explain them.
Articles on characteristic energy losses have also been published by Mend-
lowitz (1960a, 1960b) and Hartl and Raether (1961). In this connexion,
an article by Ferrell (1956) is also relevant; see also Green (1962).

19.2 CONTRAST FORMATION IN AN AMORPHOUS
OBJECT

19.2.1 The Roles Played by the Sensitivities of the Eye and of the Photographic
Plate

We consider two contiguous parts of the image: from the photometric
point of view, each will be characterized by the value of the surface density
of the electron current which bombards the plate or screen. It is the dif-
ference between the values of the current density which gives rise to a
difference between the values of the brilliance (of a fluorescent screen)
or of the blackening (of a photographic plate).

When we examine the fluorescent screen by eye, the difference between
the two regions can be discriminated only if the brilliances of the two
are sufficiently different. Under normal conditions, the relative difference
must be of the order of 5 per cent, as the mean brilliance is not usually
intense enough for the eye to be used in its optimum conditions (where
the limit would be lowered to 1 per cent). The electron intensities would
have to differ by at least 5 per cent for the observer to be able to focus the
instrument correctly by looking at the boundary between the two zones in
question.

A photographic plate records differences of contrast of this magnitude
faithfully, provided the exposure is chosen correctly, that is, in such a way
that the conditions which correspond to the straight section of the photo-
graphic density curve obtain for every region of the image. The emulsion
could record even less marked differences in contrast, and we might hope
to be able to develop the film in such a way as to make them reasonably visible to the eye by copying the primary impression onto a "hard" emulsion. The marginal gain to be had in this way is too small, however, and the cycle of operations too complicated and of too doubtful a fidelity to have seriously caught the attention of experimental workers. It must not be forgotten that it is only in the small details of the object that the microscopist is interested, which correspond to extremely tiny zones of the primary plate; statistical fluctuations produced by the grain are then appreciable, and severely limit the possibility of discriminating between the different amounts of contrast (von Borries, 1949; and Leisegang, 1954). The photographic plate is scarcely more efficient than the combination of eye and fluorescent screen.

19.2.2 Transparent Objects in Electron Optics

Normally, the objects which are used in the electron microscope are thin enough to absorb only a negligible fraction of the incident electrons. Their sole action is to scatter—elastically or otherwise—the incident particles. As there can certainly be no chromatic differentiation with a monochromatic electron beam, it is clear that the contrast which is to be seen (in Fig. 5 for example) is different in origin from the contrast which is formed in an optical microscope image.

We pointed out at the very beginning (§ 1.3) the origin of this contrast: the diaphragm which contains the aperture halts the most acutely deflected electrons, which are the more numerous as the substance has a more pronounced scattering action. A substance which contains heavy atoms (a high value of Z) appears blacker than a substance made of light atoms.

We have also to admit, however, that the contrast is not wholly lost if the stop is removed (Hillier and Ramberg, 1947). To understand this complicated combination of effects, therefore, the exact mode of image formation must be studied in each case.

19.2.3 Image Formation with a Small Aperture

We suppose that a "contrast stop" of small diameter has been placed beneath the object. The useful aperture, $\alpha$, of the instrument is, therefore, well-defined, and it is always desirable that this should be the optimum aperture already defined (§ 16.2). Equally, we shall suppose that the illumination is parallel and we consider a point on the object the effective differential cross-section of which is $\sigma(\theta)$, which corresponds to an element of area $dS$. The number of electrons impinging upon this element of the surface is $N_0 dS$, but only

$$N = N_0 \left(2\pi \int_0^\infty \sigma(\theta) \sin \theta \cdot d\theta\right) dS$$
pass through the diaphragm subsequently to be distributed over the image of area $G^2 \, dS$ (where $G$ is the magnification). The current density at the image will then be

$$i = N_0 \frac{2\pi}{G^2} \int_0^\pi \sigma(\theta) \sin \theta \, d\theta.$$  

As soon as the effective cross-section is known, therefore, we can calculate the contrast, but the cross-section is only known for certain in the two extreme cases of a very thin film (the usual case) and a very thick film (an exceptional case, of theoretical interest only. See Marton and Schiff, 1941; and Hall, 1949 and 1951). In both cases, only the mass thickness is relevant, and we can only determine the thickness of the object if its nature is known; in biology, we have to take into account the possible modifications which may be produced in the object by the bombardment. Two basic characteristics of the microscopic contrast become apparent: the resolving power and the contrast are linked, since they are both dependent upon the angular aperture, and two widely different details will have the same contrast with respect to the background if they have the same effective cross-section.

It is clear that the method of calculation outlined above can only give an order of magnitude. The illumination is not parallel, as often the image can only be made tolerably bright by opening the condenser beyond the optimum value (the illuminating angle is of the order of $5 \times 10^{-4}$). The contrast is then lessened, as an electron which is incident obliquely may be scattered by a detail in such a way that it passes through the diaphragm whereas with parallel illumination it would have fallen outside the opening. We have also to take into account the retarded electrons which "lap" over the image as a result of chromatic aberration. Finally, the collision cross-sections are imperfectly known.

### 19.2.4 Image Formation with a Wide Aperture

A contrast diaphragm with a small opening rapidly gets dirty. Charges collect on the impurities as the beam passes, and the ellipticity aberration worsens. Attempts have often been made to remedy these deficiencies by suppressing the diaphragm, and if this is done, we find that the contrast between two large adjacent regions is weakened but that the contrast between a small detail and the background is hardly affected.

It is to the aberrations that we must ascribe this feature. We consider two objects, one of which, $A$, scatters markedly, while the other, $B$, scatters hardly at all; the beam which emerges from $A$ is wide in angular aperture and will produce a large aberration patch at the image plane as a result of the wide apertures—the brightness will not be very high, therefore. The beam which emerges from $B$, on the other hand, will only produce a small
aberration patch which in consequence will be comparatively intense in brilliance. This qualitative explanation is illustrated in Fig. 233.

This contrast is dependent to a considerable extent upon the focusing, as it is linked to the effect of the aberrations. Although we have represented the images as being uniformly illuminated in Fig. 233, this is only for the sake of simplicity. In reality, the distribution is complicated, and depends upon the focal plane. The contrast in the image is accentuated by a slight defocusing, if only by producing a bright fringe (a negative fringe) which adheres to its border; the image seems clearer, and often, a slightly defocused image is preferable to an image rigorously in focus despite the slightly higher resolution which should in theory be possible with the latter. The effect of overfocusing is identical, and in fact the exact focus corresponds to minimum contrast, which is experimentally the best criterion for identifying it. A slight overfocusing blurs the image, however, as a dark (positive) fringe surrounds the dark details; the increase in contrast is accompanied, in this case, by an appreciable deterioration of the resolution. In conclusion, we should point out that the choice between a precisely focused image, poor in contrast but sure in interpretation, and a slightly defocused image in which new details often appear which must, however, be accepted somewhat circumspectly, must be made according to the type of object which is being studied.
19.2.5 Improving the Contrast

(i) **Shadowing**

It is most desirable in many cases, therefore, to find a way of increasing the contrast between the background and the details in which we are interested.

We have already mentioned the technique with which this can be done. The object is covered with a thin layer of heavy metal (10 Å thick for gold, 4 Å for uranium) evaporated on at grazing incidence. Details with a small mass thickness which alone could not produce sufficient contrast become visible by virtue of their “overlying shadow” which is not covered with metal and therefore appears bright against the dark background which is the image of the part of the object which is covered with metal.

This way of improving the contrast was introduced by Müller (1942). Further details of this shadow technique are to be found in the articles by Williams and Wyckoff (1944), Kahler and Lloyd (1950) and Koenig and Helwig (1950). The main problem is to find a metal which is deposited as a continuous layer and remains thus during observation. The majority of metals recrystallize, and under the electron bombardment these micro-crystals collect into grains and hence create an artificial structure. Uranium and certain germanium and platinum alloys are free of this drawback. Examples of the image of an object shadowed in this way are to be found in Figs. 252 and 256.

(ii) **Dark field operation (strioscopy) and Foucault contrast**

This represents the translation into electron optics of the celebrated experiment performed by Cotton and Mouton with the ultra-violet microscope. Figure 234a–(i) shows schematically an optical arrangement with which the image of the object can be formed by scattered and diffracted electrons alone. The condenser contains an annular aperture with the aid of which we can illuminate the object with a hollow conical beam; a second diaphragm which contains an axial opening of small diameter is placed in the image focal plane of the objective, and removes the direct beam while allowing certain of the electrons which have been deviated from their initial trajectories to continue uninterrupted. The image of the object is then seen as a white region against a black background, and the brightness of the image increases with the scattering power of the object (Fig. 234b).

In another arrangement, which is simpler to construct (Fert and Faget, 1948) a condenser diaphragm in the form of a narrow slit is used (three to five microns wide) while in the plane of the crossover of the objective, a screen is placed parallel to the slit; this screen consists of a very fine metallized thread (0.1 μ) which interrupts the direct beam completely [Fig. 234a–(j)].
Fig. 234a. Two ways of observing against a dark background: (i) with an annular condenser diaphragm $D_0$; the diaphragm with a centrally placed hole, $D_1$, halts the direct beam and allows the scattered electrons ($E.D.$) to pass through; (j) with a condenser diaphragm $D_0$ in the form of a thin slit on the axis; the direct beam is halted by a metal thread parallel to $D_0$—a proportion of the scattered electrons $E.D.$ is transmitted.

Fig. 234b. Dark field image of zinc oxide crystals (Photo: Fert).
Other types of system may be used, although they give considerably dimmer images: for example, the contrast diaphragm beneath the objective may be made eccentric, or if the object is monocrystalline, one or more of the spots of the diffraction diagram may be isolated with this same diaphragm.

This technique which was introduced by Boersch (1936–53) has been developed by Hall (1948) to obtain a resolution of 50 Å with a convenient arrangement in which the obscuring disc is placed between the object and

fig. 234c. Photograph of a diatom, with Foucault contrast (Photo: Fert).

the objective. Ramberg and Hillier (1948) have calculated the theoretical resolution, and even when chromatic aberration provides a limit, it ought to be as good as in the bright field situation. We should mention too the beautiful experiments performed by Möllenstedt (1949, 1950, 1952a, 1952b), who obtained dark field images in which the images formed by the electrons which had been scattered and slowed up were displaced with respect to the image formed by those which had been scattered elastically.

In the course of a thorough examination of electron scattering, Boersch (1947) showed that just as in light optics, the resolution in dark field conditions should be better than the resolution which would be obtained if the
same instrument were used with a bright field. Against a supporting layer of collodion, 100 Å thick, we ought to be able to distinguish atoms heavier than aluminium. At present, practical problems are the reason for the limit to the resolution, and in particular, the need to use stops of very small diameter at the objective in order to diminish the harmful effect of chromatic aberration.

The contrast of certain details which scatter particularly strongly can be considerably increased by removing almost all the electrons in the direct beam and half of those which have been scattered or diffracted by means of a stop in the form of a thin half-plane (a thin sheet of graphite, for example). This is "Foucault contrast" (the Schlieren method), which is characterized by an asymmetrical light distribution; the image displays marked shadowing effects, as if the object had been illuminated obliquely (Fig. 234c).

(iii) Phase contrast

In 1934, Zernike conceived a different way of observing the image from the one used in strioscopy. A thin sheet, which is known as the "phase plate", covers the whole diffraction pattern of the aperture; here, however, the plate is not opaque but instead, introduces an optical path difference between the principal wave and the diffracted waves which is chosen to be equal to ±λ/4 (Zernike, 1942). This enables us to convert the variations of phase in the beam which are created by a transparent object into variations of the contrast at the image. For a long time, a means of transposing this remarkable technique into electron optics has been sought, so successfully has it been applied in light optics. To see how this may be done, we consider the fact that certain studies of electron diffraction (the famous Davisson and Germer experiment, for example, and the calculations made by Hillier and Ramberg, 1947, of the Fresnel diffraction effects) have shown that the potential within a substance is higher than the (relativistically corrected) potential $V^*$ of the incident beam by an amount $\Delta V$ of the order of a few volts; the refractive index of the substance with respect to that of the vacuum is thus:

$$n = \sqrt{1 + \frac{\Delta V}{V^*}} \approx 1 + \frac{1}{2} \frac{\Delta V}{V^*}.$$ 

As the refractive index is greater than unity, the wavelength in the medium is shorter than the wavelength in vacuo. A wave which passes through a thickness $d$ of a material with refractive index $n$ is retarded by a distance equal to

$$\Delta = (n - 1) d = \frac{\Delta V}{2V} \left(1 + \frac{e V}{m_0 c^2} d\right) \frac{1}{1 + \frac{e V}{2m_0 c^2}}.$$
(in which $V$ is the potential applied to the gun) with respect to a wave travelling in the vacuum.

If the path difference $\Delta = \frac{\lambda}{4}$, the phase difference is $\Delta \varphi = n \frac{\pi}{2}$; with collodion, for example, $d$ would be 100 Å for $\Delta \varphi = \frac{\pi}{2}$ and $V = 60$ kV.

The first attempts at using a phase plate were made by Agar, Rewell and Scott (1949); a hole about a micron in diameter was cut in the centre of a thin formvar plate, which was placed at the crossover of the objective. The phase of the electrons which have been diffracted by the object is altered with respect to the phase of the direct beam, as the former pass through the formvar whereas the latter pass freely through the central opening. These attempts were disappointing; it was thought for a time that no accentuation of the contrast could be hoped for from such a system, the blame being cast upon the system itself which was assumed to destroy the phase coherence of the beam, and hence to render any kind of phase contrast impossible. This seemed to be confirmed by the difficulties which were encountered in attempts at electron interferometry with separate beams in which one of the beams was retarded in phase with respect to the other by a material film (Marton, 1953).

Faced with these difficulties, therefore, a method of altering the phase without introducing a real plate has been sought. Locquin (1955) has been able to demonstrate that the contrast of very thin biological specimens without supporting films can be sharply increased (see Fig. 234d) by placing an insulating point in the plane of the aperture diaphragm of the objective; this point becomes charged under the influence of the beam. The mechanism by which this contrast is formed remains without a conclusive explanation, and is still the object of controversy, as no way of explaining it in terms of phase contrast in Zernike's sense of the phrase seems possible. It seems to be due rather to some kind of interference effect, similar to those which can nowadays be straightforwardly produced with the aid of an electrostatic biprism, the principle of which we shall now explain.

A divergent beam of electrons is emitted by a line source $S$, which is made as narrow as possible to improve the coherence (Fig. 234e). The beam passes through an electrostatic system which comprises a diaphragm containing a slit parallel to the source and an extremely fine metallized thread (less than a micron in diameter) which is stretched along the axis of the slit. This thread may be held at an electrostatic potential, $\Phi$, which is positive with respect to the edges of the slit. If $\Phi = 0$, the two beams which are separated by the thread do not subsequently intersect, and only the shadow of the thread is seen framed within the ordinary diffraction fringes which resemble the fringes produced by the edge of a screen (the lenses which follow the biprism provide the necessary magnification). If $\Phi$ is increased, however, the two beams are deflected in opposite directions, approach one another, and begin partially to overlap; they appear to have originated
Fig. 234d. Attempts at phase contrast (Locquin, 1955) photographs of myxomycetis fibres. *Top left*: ordinary photograph with a light background. *Top right*: photograph at the same magnification, after introducing an insulating point into the aperture diaphragm—the contrast is improved and fresh detail appears in the internal region of the fibres. *Below*: photograph at high magnification, showing spiral structures which are invisible with normal methods of observation.
in two virtual sources, \( S_1 \) and \( S_2 \), separated by a distance \( 2l \). In the zone common to both diffraction fringe systems, therefore, interference fringes appear, just as in the Fresnel biprism experiment (already familiar in light optics); the fringes are progressively more closely spaced as \( 2l \) is made larger. The total number of fringes visible in extremely carefully conducted experiments may reach several thousands (Möllenstedt and Düker, 1956, Faget and Fert, 1956, 1957, Faget, Ferré and Fert, 1958); we can thus describe the coherence of the beam by attributing to it a wave-train several thousand wavelengths long, or about a hundred Ångströms. The evolution of a fringe system obtained in this way is shown in Fig. 234f.

If a thin carbon film 100 Å thick is introduced into the path of the beam just beyond the biprism, however, the fringes remain visible; the coherence of the beam, therefore, is not destroyed (Faget and Fert, 1956). Genuine
Fig. 234f. Interference fringes obtained by Möllenstedt with the aid of a device similar to a Fresnel biprism. The phase difference is produced by the electric field of a fine wire, the potential of which increases as we move from top to bottom from zero to seven volts; more and more fringes become visible (19,400 eV).

interference microscopy has thus been achieved, and we can consider afresh the possibility of producing real phase contrast.

In this connexion, we shall now give a brief description of interference microscopy techniques (Möllenstedt and Buhl, 1957; Buhl 1960; Faget and Fert, 1957). The arrangement is as follows: the condenser contains a slit diaphragm and a biprism is placed beneath the objective. The object which is to be examined is placed in an ordinary specimen-holder, so that on the
screen, both the image of the object and the shadow of the thread are to be seen. When the latter is given a positive potential, fringes appear in the zone common to both parts of the beam. At each point, the order of the interference is defined by the geometrical path difference produced by the combined effects of the biprism and the difference in optical thickness between the two corresponding points of the object. If the latter is homogeneous

![Fig. 234g. Interference microscopy: phase difference produced by a graphitic flake which has been inserted into the path of one of the electron beams. The displacement of the fringes depends on both the thickness and the internal potential of the flake (Photo: Fert).](image)

but not uniformly thick, the fringes are sinuous as the difference between any one fringe and the "average" fringe represents the corresponding difference in optical thickness (Fig. 234g). If, on the other hand, the object is of constant thickness, but consists of several constituents, the fringes produced by two neighbouring regions, different in nature, will again be shifted with respect to one another, and will thus indicate the variations of internal potential \( \Delta V \). By measuring the irregularities of the fringe system, we obtain a measure of \( \Delta V \). To make this measurement in practice, the object is left unsupported, so that \( \Delta V \) is obtained directly with respect to the vacuum (Fert and Faget, 1958); the main difficulty lies in the precise measurement of the thickness of the object.

Let us now reconsider phase contrast, with a phase plate. A slight improvement in the contrast has been obtained by Kanaya and Kawakatsu
(1958) with the aid of a collodion or carbon plate, in which numerous holes have been pierced. The direct beam passes through the substance, while certain spots of the diffraction diagram pass unimpeded through the holes. The analogy with the optical arrangement is not always very rigorous, however, and it is only very recently that Faget, Fagot and Fert (1960) have managed to produce the exactly analogous situation. A straight-edged phase plate, accurately parallel to the slit source of the condenser, is placed in the image focal plane of the intermediate lens; to avoid contamination the phase plate is held at a temperature of 200°C. To produce the phase plate, a very fine thread from a spider’s web (between 0.1 and 0.4 \( \mu \) in diameter) is stuck on a thin collodion film, and then covered by evaporation with a thin layer of carbon. The thread is then torn away. In the corresponding zone of the plate is a region thinner than the remainder, and we check that the path difference produced by this difference in the thickness is in fact \( \lambda/4 \) by interference microscopy. The thin zone of the plate must cover the image of the slit source very exactly, so that the diffracted electrons all pass through the thicker region. An example of the improvement in contrast which can be obtained in this way is shown in Fig. 234h.

These experiments confirm that the coherence of the beam is not destroyed by a passage through a thin object. The influence of this passage on the resolution of the instrument is as yet only imperfectly understood, however.
19.2.6 Contrast in Crystals

When the object is crystalline, we can observe the phenomenon of diffraction at the various lattice planes of the structure; just as in ordinary optics, this phenomenon is characterized by the presence of privileged directions along which the waves are in phase and give a diffracted wave. In every other direction, the resultant intensity is zero, as the diffracted wavelets mutually interfere destructively. We shall thus observe intensity maxima for certain relative orientations of the lattice, the incident beam and the diffracted beam. Let us suppose that a parallel beam of electrons falls on a thin crystal which has lattice-spacing \( d \) at one of the Bragg angles of incidence which are defined by

\[
2d \sin \theta = n \lambda.
\]

With \( \lambda = 0.05 \text{ Å}, \ d = 2.5 \text{ Å} \), the first maximum will be observed in a direction which is inclined with respect to the incident direction at an angle \( 2 \theta = 2 \times 10^{-2} \) radians. This angle is in general adequate to prevent the diffracted beam from passing through the objective diaphragm when the illumination is parallel to the axis. Generally speaking, every part of the crystal which is placed in such a way that a Bragg reflexion falls outside the objective diaphragm will appear dark; this explains the presence of dark fringes, which are more often than not sinuous, in the images of thin crystals such as molybdenum oxide. When the orientation of the illuminating beam is altered, these fringes move also.

If, on the other hand, a lattice plane is so oriented that the diffracted beam passes through the objective diaphragm (Fig. 235) and is focused, a bright image of the dark part of the object is seen on the screen (Boersch, 1943); this image is displaced with respect to the ordinary image, and this duplication can easily be seen with needle-shaped crystals (such as those of zinc oxide), and an intense illumination. These "Bragg images" can be distinguished from the paraxial image by the attendant spherical aberration, and by measuring the separation \( d \) between the two "images" we can calculate the spherical aberration constant \( C_s \) from the relation

\[
d = C_s (2 \theta)^3.
\]

During focusing, the light and the dark images move with respect to each other—coincidence between the two corresponds to under-focusing.

For a long time, these "images" were regarded simply as curiosities. A few years ago, however, it was realized that they could provide very interesting information about the structure of the specimen, and an original technique for examining periodic structures has been developed.

By using the direct beam and one of the diffraction spots which correspond to some particular direction of the lattice planes, we obtain an image which can be considered as the projection of the structure and which con-
tains only the first term of the illumination (represented mathematically in the form of an infinite Fourier series), apart from the fundamental. We thus obtain lines separated by the same distance as those lattice planes at which the diffracted waves which have been accepted were diffracted.

![Diagram](image)

**Fig. 235.** The formation of a "Bragg image", displaced with respect to the main image by spherical aberration.

The lines to be seen in Fig. 227c are such a case, for example—this figure may be said to be an "image" of the (201) planes of a copper phthalocyanine crystal with a mean separation of 9.9 Å.

Images of this type were obtained for the first time by Menter (1956) who was studying phthalocyanine crystals.

If, therefore, we eliminate all the diffraction spots save one by means of a contrast diaphragm 30 to 50 μ across, we can study the lattice structure of a crystal and in particular, the faults in the lattice; this procedure is limited, however, by the resolution of the instrument, and this direct method of observation is only applicable when the meshes of the lattice are comparatively great (longer than 6 or 7 Å). The shortest interval which has been observed is that of the (020) planes of molybdenum trioxide (6.9 Å: Bassett and Menter, 1957). This implies that with the instruments available at present, metal lattices are the finest which can be made visible in this way.
The lattices of the large organic molecules can, on the other hand, be observed by diffraction contrast.

An indirect observation technique is now available, however, in which we use the moiré fringes which are obtained when two lattices are superimposed. For some time, fringes had been observed in the images of certain crystals, which were often periodic formed figures of varying degrees of complexity, and could not be explained in terms of simple Bragg diffraction. Several studies have been devoted to them (Boersch, 1942, 1943; Kinder, 1943; Heidenreich and Sturkey, 1945; Pernoux, 1952, Möllenstedt, 1953; and Rang, 1953, 1953a). Overall surveys are to be found in Pernoux’s thesis (1953) and in an article by Menter (1958).

![Diffraction Diagram](image)

**Fig. 236.** Diffraction diagram of a single monocystal (left) and of two identical superimposed monocystals (right) (Photo: Bassett, Menter and Pashley, 1958a).

The fringes which are observed are the interference figures produced by the interaction of the direct wave and the waves which have undergone successive Bragg reflexions at two systems of planes (see below). This can easily be proved by examining the diffraction diagram which corresponds to the zone of the object in which the moiré fringes originate; we obtain a principal diagram of points, which correspond to the upper layer which is the first upon which the beam falls, but each spot is surrounded by a subsidiary diagram produced by the second layer (each beam which has been diffracted at the first layer behaves as a principal beam for the second layer, and is thus partially diffracted a second time (Fig. 236; Bassett, Menter and Pashley, 1958a).
The simplest case to which we have been referring in the previous paragraph corresponds to the superposition of two plane lattices with the same orientation but slightly different mesh-lengths (Fig. 237a). At the image, a new periodicity

\[ D = \frac{d_1 d_2}{d_1 - d_2} \]

is obtained. Alternatively, the periodicity \( d \) of the two lattices may be the same, but they may be inclined to one another at an angle \( \varepsilon \) (Fig. 237a);

Fig. 237a. Optical analogue, showing the formation mechanism of moiré fringes by superposition of above—two different parallel networks, with different periodicities; below—two identical networks, slightly inclined with respect to one another (Menter, 1958).

we now obtain a more complicated figure, in which only the perpendicular fringes are visible when \( d \) is very small, and we now find:

\[ D = \frac{d}{\varepsilon} \]

In these two simple cases, therefore, we have a "moire magnification" of \( \frac{d_2}{d_1 + d_2} \) and \( \frac{1}{\varepsilon} \) respectively.
Fig. 237b. Optical analogue, showing two different ways a single dislocation may appear in a moiré fringe system.

Fig. 237c. The differing appearance of a pair of neighbouring dislocations in a moiré fringe system when the relative orientation of the two real networks (left) varies (Bollmann, 1960).
These two fringe systems can clearly yield combinations of varying degrees of complexity. With the aid of the moiré pattern, a quantitative study of the lattice defects can be undertaken; if for example a dislocation is present in one of the lattices (see Fig. 237b), it appears considerably enlarged in the image. Simple optical models can be used to help interpret the irregularities, often of widely different shapes, which are observed in the fringe system. Figure 237c shows the figures which are obtained when the two fringe systems (i) and (j) are superimposed, and the angle $\varepsilon$ varied (Bollmann, 1960). In Fig. 237d, an example of moiré fringes obtained with flakes of graphite is to be found.

Again, suppose that two identical systems remain together, but are slightly curved; we may obtain dark serpentine fringes, often closed, with widely different appearances according to the inclination of the whole combination with respect to the illuminating beam (Rang, 1958). The oddest patterns of all are obtained when both systems are curved, with
different curvatures, so that an empty space is left between them, shaped like a meniscus or a bubble. The figures then have a centre of symmetry and are illustrated in Fig. 237e (see Pernoux, 1953).

Fig. 237e. The figures at the centres of symmetry, created by bumps or hollows in a lead iodide crystal (Pernoux).

It is in terms of this type of diffraction phenomenon that the fringe systems which we observed in crystals can be explained, and they may have various symmetries, depending upon the relative orientations of crystal and beam (see, for example, Dupouy and Perrier, 1960); here also is to be found the explanation of the structures which appear at very high magnifications in grains of colloidal gold or microcrystals of silver deposited on
mica (Matthews, 1960) or on molybdenite (Bassett, 1960), and we can thus study in particular the epitactic growth of the crystals of an evaporated substance on a suitable support. Figure 237f shows such fringes, obtained with magnesium and molybdenum oxide crystals.

Over the last four years, numerous articles have been devoted to the theory of moiré patterns, and to their applications in crystallography. We mention in particular those of Menter (1958), Rang (1958), Dowell, Farrant and Rees (1956, 1958), Stabenow, Suito and Uyeda (1958), Bassett, Menter and Pashley (1958a,b) and Hashimoto, Naiki and Mannami (1958), all of which were given at the International Conference on Electron Microscopy which was held in Berlin in 1958.

![Fig. 237f. Fringes obtained in molybdenum oxide crystals (left) and molybdenum (right) (Photo: Dupouy and Perrier).](image)

A preparation technique has been developed with which moiré patterns can be produced at will, starting with superimposed thin crystalline layers, which may be of either similar or different natures (Pashley, Menter and Bassett, 1957) and have known orientations. A thorough study of the resolution which can ideally be attained with this method shows that we could hope to observe systems with periodicities of the order of 1 or 2 Å.

19.2.7 Contrast Produced by Defocusing

In § 16.5.3, we described the way in which the Fresnel diffraction fringes at the edge of a screen (either around an opaque particle, or, on the contrary, in the peripheral zone of a hole—in a collodion membrane, for example) can be used to help focus a system and correct the ellipticity astigmatism.
If the image of the edge of the screen is over-focused, the diffraction
diagram is real and lies behind the object; if it is under-focused, the diagram
is virtual and is situated in front of the object. Whereas in ordinary optics
the contrast of the fringes is to all intents and purposes identical in the
two cases, in electron optics the fringes of the virtual diaphragm are far
more sharply contrasted. The first fringe of the under-focused image is
a very bright line, which lies outside the dark image; the contrast of the edge
of the image is accentuated with respect to the general background. The
first fringe of the over-focused image is a dark fringe, which is slightly
separated from the image itself (Fig. 238a). This lack of symmetry between
the contrast in the two cases—which often leads the experimental worker

Fig. 238a. Images of the edge of a hole in a carbon film. In the lower regions,
the image is under-focused; in the upper part, it is over-focused. Notice the im-
proved contrast at the edge of the black zone in the case of under-focusing. Between
the two images, the defocusing is 50 μ.
to settle for a slightly under-focused image—can be understood if we remem-
ber that on the electron scale, a screen never has a clearly-defined edge
($\lambda = 0.04 \text{ Å}$ for 86 kV electrons). If we take the transition zone, across
which the thickness is decreasing, into account, the results of a calculation
of the luminous intensity in the fringe system agree well with experiment
(Fert and Laffitte, 1961).

**Fig. 238 b and c.** Variation of the contrast with defocusing, for a periodic object.

(b) Optical model $\Delta b_1 = 23 \text{ mm}$. (c) Electron image $\Delta b_1 = 30 \mu$. For $\Delta b = \frac{\Delta b_1}{2}$,
the pattern has the same appearance save that the contrast is reversed.

If the image of a periodic system is formed with a sufficiently coherent
illumination, it is again apparent that the contrast varies very rapidly
with the degree of focusing. If, in the simple case of a periodic object
consisting of lines or square meshes of periodicity $a$, the image is longi-
tudinally defocused by an amount $\Delta b$, the figure which is obtained on
the screen is identical to the original figure whenever $\Delta b = \Delta b_1 = \pm p \frac{2a^2}{\lambda}$
(in which $p$ is an integer). If the lattice is close-packed in two dimensions
(hexagonal) the period of the defocusing $\Delta b_1 = \pm 1.5 p \frac{a^2}{\lambda}$ (Fagot and
Fert, 1960). This phenomenon is illustrated in Fig. 238 b in optical micro-
scopy. The object is a small-scale photograph of a periodic system of
opaque discs on a light background; the periodicity is $a = 91 \mu$. With
\( \lambda = 5461 \text{ Å} \), the original form reappears after a defocusing of \( 1.5 \frac{a^2}{\lambda} = 23 \text{ mm} \). While for values of \( \Delta b \) between zero and 23 mm, numerous very different figures appear, which give no hint as to the structure of the object.

**Fig. 238d.** The effect of coherence on the contrast, when the object is periodic (myeline fibres, with a periodicity of about 85 Å).

*Left:* incoherent illumination: \( 2\alpha_c = 6 \times 10^{-3}, 2\alpha_0 = 10^{-2} \).

*Right:* coherent illumination: \( 2\alpha_c = 0.7 \times 10^{-3}, 2\alpha_0 = 10^{-2} \).

\( 2\alpha_c \) represents the angular aperture of the beam illuminating a point on the object, and \( 2\alpha_0 \) the numerical aperture of the objective.

only its periodicity. For \( \Delta b = 11.5 \text{ mm} \), the appearance of the figure has that of the original object, save that the contrast has been reversed. This experiment has been performed afresh, in electron microscopy, with coherent illumination, by Fagot and Fert (1960) and Fagot, Ferré and Fert (1961). Different test objects have been tried. A biological section of regularly distributed grains forms a pattern which resembles the optical model mentioned above; for \( a = 55 \text{ Å} \), and \( \lambda = 0.037 \text{ Å} \) we find \( \Delta b_1 = 1.5 \frac{a^2}{\lambda} = 30 \mu \). A mean defocusing of 15 \( \mu \) gives rise to contrast reversal (Fig. 238c).
If the object has a periodic structure which consists of parallel lines (myeline fibres for example, see Fig. 238 d), we re-encounter the pattern of the original structure when the defocusing is equal to \( \Delta b_1 \); for \( \Delta b = \frac{1}{2} \Delta b_1 \), however, the periodicity is doubled (the parallel lines are twice as numerous). This phenomenon has also been mentioned by Kamiya et al. (1958), in the images which they obtained with crystalline specimens of the mineral serpentine.

In conclusion, we would mention that it is in coherent illumination that all these phenomena appear, that is, when the angular aperture, \( \alpha_e \), of the beam with which the object is illuminated is small with respect to the numerical aperture, \( \alpha_o \), of the objective (\( \alpha_e < 10^{-3} \), for example, and \( \alpha_o = 10^{-2} \)). The angular aperture is reduced with the aid of a condenser diaphragm, which in interference microscopy may be a very narrow slit. The resolution of the microscope is of the same order of magnitude whether the illumination be coherent or incoherent, but the contrast is better in coherent illumination for the reasons already discussed. Nevertheless, the interpretation of the image is liable to become a most delicate affair.

### 19.2.8 Improvement of the Contrast, the Very Low Voltage Microscope

In contrast to the strainings towards very high tensions, which are necessary above all for the examination of thick biological sections, several attempts have been made during the last few years to use very low energy electrons in microscopy, with a view to reinforcing the contrast of thin unshadowed preparations. We now know how to produce ultra-thin sections, and this is an essential condition if electrons of a few kilovolts which have such a low penetrating power are to be used. The resolution of the instrument in theory decreases when the potential \( V \) falls, but is not nevertheless catastrophically reduced by a large reduction of the accelerating potential. From § 16.2.4, we know that

\[
d_{th} \simeq C_i^{1/4} V^{-3/8},
\]

so that if we work with a fixed focal length, \( d_{th} \approx V^{-3/8} \) whatever the potential \( V \); at low values of the potential, however, the focal length of the lenses, and hence \( C_i \), may be reduced, according to the rule:

\[
f \simeq V^{1/2} \quad \text{or} \quad C_i \simeq V^{1/2}.
\]

The resolution, therefore, behaves thus:

\[
d_{th} \simeq V^{-1/4}.
\]

If \( V \) alters from 60 kV to 6 kV, the limit of resolution will only be multiplied by about two. In these conditions, Nixon (1958) hopes to be able to obtain a resolving power of 15 Å.
From the point of view of the contrast, the work of von Borries and Ruska (1940b) shows that for substances with neighbouring atomic masses, it is advantageous to use low tensions to increase the contrast. A detailed study of the scattering of electrons by matter shows that if we are to examine a preparation which rests upon a support of given thickness $e$, there is an optimum potential for which the contrast is greatest; if for example $e = 100 \, \text{Å}$ for a carbon support,

$$V_{\text{opt}} = 3.8 \, \text{kV}$$

(Nixon, 1958).

The work of Nixon (1958), Wilska (1960) and van Dorsten and Premsele (1960) has demonstrated that all the accompanying problems can be satisfactorily solved. It is such difficulties as the reinforcement of the brightness of the gun and the brilliance of the screen at low tensions, the use of sufficiently sensitive emulsions such as the Gevaert Scientia GE 56 (van Dorsten and Premsele, 1960) and the obtaining of an effective accentuation of the contrast which require solution.

If we compare the results of this procedure with the reinforcement obtained if an objective diaphragm is used, we find that a reduction of the accelerating potential from 40 kV to 20 kV improves the contrast to the same extent as a thirty-three fold reduction in the diameter of the opening in the diaphragm; the latter operation is impractical both because of the ensuing contamination of the hole and of the increase in the size of the diffraction disc.

### 19.3 VELOCITY ANALYSERS

The discrete energy losses which we mentioned in §19.1.6 will provide information about the structure of solids when their origin has been more clearly elucidated, and this will be a new and valuable spectroscopic technique. For this reason, many devices known as “velocity analysers” have been developed over the last few years, with which these velocity losses—of the order of 10 eV, we must remember—can be measured.

#### 19.3.1 Magnetic Prism Analysers (cf. Chapter 22)

The apparatus which Voges and Ruthemann (1939) originally used was later considerably improved by Ruthemann (1941, 1942 and 1948). It consists of an evacuated chamber, in the form of a half-ring, which is placed between two coaxial solenoids which produce a magnetic field inside the chamber perpendicular to the plane of the ring. At one extremity is the electron source, a straight tungsten filament, which is situated in front of a narrow slit (0.1 mm), parallel to the filament. The filament is almost completely screened, except for a wide slit which faces the filament—the
screen is held at the same potential (Fig. 239). The accelerating potential can be varied between 2 and 8 kV, and is applied between the filament and the slit. The flat electron beam, the aperture of which is a few degrees, falls upon one of the objects placed on the specimen-holder two or three millimetres from the slit; a remote control is provided, so that one or other of the specimens can be brought into the path of the electron beam from outside.

The velocity analysis is performed after the beam has been deflected through a semi-circle by a magnetic prism. In Chapter 22, the principle upon which such a prism acts is to be found, together with a calculation of the resolving power. To improve this latter, Voges and Ruthemann use a magnetic field which decreases very slowly from the centre towards the periphery. In this way, the spherical aberration is corrected. A field of this kind is produced by two long air-filled coaxial coils with radii \( R_1 \) and \( R_2 \) carrying currents which flow in opposite senses, and by two thick turns of radii \( R_1 \) and \( R_2 \) which lie one on each side of the chamber, and which carry currents which flow in the opposite sense to that of the nearest coil. The total field is constant to about one part in \( 3 \times 10^3 \) along the mean trajectory and falls off slightly towards the exterior by about 5 per cent. After an extremely careful stabilization of the supply currents (a few parts in \( 10^4 \)) and of the accelerating potential, Ruthemann attained a resolution of 1 V in 1200; it would doubtless be possible to improve the focusing by taking
advantage of recent work on $\beta$-ray spectrographs (cf. Chapter 25). The difficulties which have to be overcome are less serious here than for $\beta$-rays, as the dispersion spectrum is very narrow and the mean energy small.

Apertures are placed along the path of the beam, and at the far extremity of the vacuum chamber the beam impinges on a mobile photographic plate on which several spectra can be recorded, and which correspond to a maximum dispersion of 6 or 7 per cent.

![Diagram](image)

**Fig. 240.** The Hillier and Baker (1944) analyser. $C$: gun; $O$: object; $L_1, L_2, L_3$: magnetic lenses; $E$: fluorescent screen; $H$: uniform magnetic field; $P$: photographic plate.

Hillier and Baker (1944) have taken great pains to find out the position and the extent of the region of the preparation through which the electrons pass. Two magnetic reducing lenses (Fig. 240) form an electron probe in the vicinity of the specimen; its diameter can be as small as 20 m$\mu$. This point source is used first of all to throw a highly magnified image of the object on the screen by a shadow effect.

The observer selects the region of the image in which he is interested, and brings it to the optic axis. The probe is then shifted into the object plane (which means that the magnification is infinite and the image blurred); the region of the object which is to be analysed is reduced to the diameter of
the probe. The third lens focuses the divergent beam into a still further reduced image of the probe at the entry to the region in which a homogeneous magnetic field has been established; the latter deflects the beam through 180° and converges it onto a photographic plate. The beam is no longer laminar, as it was in the preceding paragraph, but is axially symmetric. Stops are placed within the last two lenses to reduce the aperture with the object of diminishing the aberrations and increasing the resolving power of the analyser. In principle, the resolution should reach 1 volt in 10,000 but it proves to be useful to increase the probe diameter to 200 μm to avoid contaminating the object. The resolving power is in practice about 1 in 1000.

A far more highly perfected device is at present being used by Marton, Simpson and McCraw (1955a) to study the distribution of the retarded electrons as a function of the scattering angle. A very fine electron beam passes through the specimen—the scattered electrons are then caught by a magnetic field which deflects them through 180° and concentrates them on a fluorescent screen where the brightness is recorded by a photomultiplier. By placing a small diaphragm in front of the analysing field, and making the whole combination of electro-magnet and receiving system rotate about an axis passing through the centre of the object and the screen (Fig. 241), we can collect the electrons which have been scattered through any angle whatsoever. For each position of the system, the magnetic field can be varied, and hence electrons which have lost different amounts of energy can be brought onto the screen. The actual system is remarkable in that it is highly automatic, and hence the angle, the loss and the intensity can all be recorded within the space of a few minutes. The resolution is about one part in 1000 for an accelerating voltage of 20 kV.

---

**Fig. 241.** Analyser used by Marton, Simpson and McCraw (1955a). C: gun; $D_1$, $D_2$: diaphragms; $Sp$: specimen; $E$: fluorescent screen; $A$: magnetic field analyser; $Q$: quartz rod; $P$: photomultiplier; $ZZ'$: axis of rotation.
19.3.2 An Analyser with a Decelerating Electrostatic Field

Boersch (1954) has recently perfected the method of analysis which is due to Lenard. No optical system is required, and the best resolving powers are obtained; Boersch obtained the velocity spectrum of the thermal emission from various different cathodes, and his measurements were repeatable to within $4 \times 10^{-3}$ V for a high tension of some tens of kilovolts—a resolution close to one part in $10^7$.

![Diagram of analyser](image)

**Fig. 242.** Boersch's analyser (1954).
(a) The principle of the instrument, $K$, $W$, $A$: electron gun; $E$: screen; $D$: diaphragm; $C$: Faraday cage; $M$: electrometer valve; $P$: polarization for the Faraday cage.

(b) An example of the spectrum obtained after passage through an aluminium foil.
The apparatus is illustrated in Fig. 242a. The electron beam which is emitted by a tungsten filament is focused by a triode gun and the image of the source is formed on a fluorescent screen in which a small hole has been cut. The electrons which pass through this diaphragm encounter an intense retarding field which fills the space between the screen and a Faraday cage which is held at a potential close to that of the cathode. Only those electrons with sufficient kinetic energy can reach the receiver. The potential of the cage, $\Phi \pm \Delta \Phi$ is slowly and continuously varied about $\Phi$, the value of the negative potential applied to the filament. The current $i$ at the receiver changes from its full value to zero. A galvanometer measures $i$ directly, and the spot is directed onto a photographic plate which is displaced by an amount which is proportional to the variation of the potential $\Delta \Phi$; in this way, the velocity spectrum is recorded directly, whether it be the spectrum of the emitting source, or that of an absorbent film placed between the gun and the screen.

The diagram which is obtained in this way represents the area under the ordinary spectrum, the integrated spectrum; intensity jumps are observed rather than bands. Figure 242b represents the spectrum which is recorded when electrons pass through a layer of aluminium—the positions of the points of inflexion of the diagram correspond to the values of the characteristic losses.

The immense improvement in resolving power and the precision of the measurements owe a great deal to the meticulous adjustment of the optical system of the receiver. When the cage potential is slightly more
negative than that of the cathode, this system is effectively a convergent retarding lens; only a fine beam of electrons close to the axis can pass through this system as over a certain region, radius $r_0$ about the axis, the potential is approximately equal to that of the cathode, while for $r > r_0$, it is more negative and the electrons are reflected. If all the incident electrons which are to be analysed lie within a cylinder of radius $r \leq r_0$, however, they will all meet the same potential barrier. The secondary electrons emitted in the cage, on the other hand, must remain there. These two conditions can be satisfied by making a judicious choice of the dimensions of the retarding electrode and the stops; a detailed discussion of this interesting device is to be found in the article by Boersch.

A variant of this type of analyser has recently been built with a view to studying the losses at very high energies (Blackstock, Birkhoff and Slater, 1955). It consists (Fig. 243) of the electron gun of an oscillograph which gives a parallel beam, a long accelerating tube constructed in sections in the same way as an electrostatic high voltage generator tube within which the field is uniform, a target through which the electrons pass with high energies, a retarding tube similar to the other tube, and finally the actual analyser itself. The analysis can thus be made on electrons with energies of a few hundreds of volts either with the Boersch device, or with an electrostatic analyser analogous to Ruthemann's magnetic analyser.

In conclusion, we should mention another type of filter-analysers, described by Simpson and Marton (1961); the electrons are slowed down in an ordinary three-electrode immersion objective after passing through the anode which is provided with a stop $D_1$. The potential of the grid is adjusted to give an extremely reduced image of $D_1$ in the plane of the plane cathode $K$, in the centre of which a small hole has been cut; the potential of $K$ is close to that of the filament. The electrons are then again accelerated in an objective lens symmetrical to its predecessor, and with the same cathode $K$. The anode diaphragm $D_2$ is the image of $D_1$ produced by the complete system, and the emergent beam is then identical to the incident beam. At $D_2$, a fluorescent screen is placed, followed by a photomultiplier. The "sharpness of the cut-off" can be defined as the voltage $\Delta \Phi$ necessary to reduce the intensity of the beam from 90 to 10 per cent of the current transmitted with the lens positive. This quantity depends upon the diameter $d$ of $D_2$—if $d = 1$ mm, $\frac{\Delta \Phi}{\Phi} = 5 \times 10^{-4}$ and if $d = 0.32$ mm, $\frac{\Delta \Phi}{\Phi} = 5 \times 10^{-5}$. If we use the lens as a velocity analyser, the order of magnitude of the energy losses which we can measure is appreciable. This quantity depends upon the maximum energy $E_m$ of the electrons in the retarding plane, for which the optical properties are conserved. For $d = 1$ mm, $E_m$ may reach 1 per cent of $\Phi$; for $d = 0.32$ mm $E_m$ would be about $0.5 \Phi$. 
19.3.3 Analysers with an Electrostatic Lens

For some years now, we have been aware of the abnormal way in which an electrostatic einzel lens behaves when the potential of the central electrode is more negative than that of the cathode. This is the behaviour which we designated transgaussian in § 8.2. Möllenstedt (1949) has found a means of taking advantage of the high chromatic sensitivity of this type of lens in the transgaussian region.

![Diagram](a)

![Diagram](b)

Fig. 244. The principle of Möllenstedt's analyser (1949).
(a) The appearance of the trajectories when the central electrode is made steadily more negative.
(b) The ordinate of the trajectory at the fluorescent screen when the potential at the centre of the central electrode tends towards the cathode potential which is chosen as origin.

We consider a very fine monochromatic electron beam, rigorously parallel to the axis, which approaches the lens at a distance $r_0$ from the axis. When the bias of the central electrode is slowly varied, becoming more and more negative, the convergence of the lens first of all increases so that the trace of the beam at the screen $E$ (Fig. 244) moves further away from
the axis. The convergence then passes through a maximum and the ray returns towards the axis, intersects it, and moves away from it again, then returns and subsequently oscillates more and more rapidly; finally, the lens is transformed into a mirror. The displacement of the spot from $A$ to $B$ during the first oscillation corresponds to only a very slight change in the potential of the electrode, and in the vicinity of the axis the displacement is to all intents and purposes proportional to the change in potential. Conversely, if we set the potential of the central electrode at the value which corresponds to telescopic behaviour of the lens (when the emergent ray is parallel to the incident ray), and slowly vary the accelerating potential of the incident electrons, this effect is most sensitive and the trace of the emergent beam is displaced across the screen. After calibration, even the velocity dispersion in an electron beam can be discerned. In reality, it is a laminar electron beam and a cylindrically symmetrical analysing lens which we use, so as to obtain a band spectrum on the screen. It might seem tempting to use a higher order transgaussian trajectory, as the chromatic sensitivity of the lens increases very rapidly with the order of the emergent ray; the difference in velocity between two successive oscillations is too narrow for the needs of the analysis, however—what we should like in fact is a virtually linear dispersion over a zone 100 or 200 V wide in the vicinity of the accelerating potential. Further, the resolving power of the instrument would not be increased as the size of the spot increases as well as the dispersion.

The essential features of an electrostatic analyser are, therefore (see Fig. 245):

(i) An electron source as monochromatic as possible. The standard tungsten filament can be replaced by an oxide coated cathode, as the velocity dispersions are respectively 0.3 to 0.5 V and 0.1 V.

(ii) A circular or slit objective which produces a beam with a given angular aperture in front of the analysing lens.

(iii) A specimen-holder which is placed directly behind the objective; the specimen-holder can be retracted, and the analyser can then be calibrated simply by varying the initial velocity of the beam in ten-volt steps with an auxiliary group of batteries.

(iv) A slit with sides as nearly parallel as possible, between two and five microns wide. An adjustment with which the slit can be moved laterally in vacuo to adjust its distance from the axis is provided.

(v) The analysing lens, which is placed immediately after the slit. This is a slit lens, the edges of which must be strictly parallel to the edges of the analysing slit if the trace of the beam on the screen is to be undistorted. The outer electrodes of the lens are in reality circular holes, to give a slight longitudinal magnification to the trace.

(vi) The receiver system, which consists of a retractable fluorescent screen which is used for making adjustments, and a photographic plate which is
equipped with a special type of shutter with the aid of which the plate can be exposed twice—one half records a calibration spectrum while the other half records the unknown spectrum.

(vii) A very carefully smoothed H.T. supply; to first order, the behaviour of the instrument is independent of fluctuations in the supply, in the same

![Diagram](image)

**Fig. 245.** Basic features of an electrostatic analyser. *C*: gun; *E*: screen and photographic plate; *O*: objective; *F*: slit; *L₁*: objective; *L₂*: analysing slit lens; *P₁*: supplementary gun polarization, for fine adjustment; *P₂*: Wehnelt polarization.

way as an electrostatic microscope, since the polarization potentials change in the same ratio. Here, however, the energy loss which is to be measured remains constant and does not vary with the fluctuations of the potential. The result is a broadening of the lines due to the action of the residual alternating potential which must, therefore, be smoothed out with extreme care.

The calibration spectrum is obtained by taking successive photographs on the same plate of the bands which correspond first to the accelerating potential $\Phi$ and then to potentials $\Phi - k \Phi$ (which correspond to slower electrons); the latter are obtained by applying a series of carefully measured potentials to the cathode, in steps of 10 V for example.

Figure 246 represents a spectrum obtained in this way (Gauthé, 1955) and in the upper part, the spectrum produced by a beam which has passed through a thin film of silver. The value of the characteristic loss can be read off immediately. In a slightly improved form (Klein, 1954),
the resolving power of this instrument can be 1 V in 35,000 with a tungsten cathode and a one or two micron slit, and 1 V in 70,000 with an oxide-coated cathode (Möllenstedt, 1952a, 1952b).

The most marked defect of this analyser lies in the widening of the image of the slit which is observed when we move away from the axis, while the dispersion only increases slowly; the consequence of this effect is a pronounced reduction in resolving power for bands far from the origin. Möllenstedt and Dietrich (1955) were able to remedy this drawback, bringing all the bands successively to the centre, by compensating for the velocity loss produced by the specimen. The fluorescent screen is provided with a narrow slit which is placed in the path of the rays which emerge parallel to the axis (Fig. 247) and which subsequently produce the band which corresponds to electrons which are not slowed up. On the normal fixed potential $\Phi_n (-U$ in the figure) of the analysing lens, they superimpose a correction potential which varies linearly with time $\Delta \Phi(t)$ (or $\Delta U$). The emergent rays which correspond to the various energies which are produced by the retarding process sweep across the screen and successively cross the slit $F_2$. If the photographic plate which is placed beneath $F_2$ is moved with the same velocity as the trace of the beam, every point of the spectrum will be recorded with the same degree of definition. The slit $F_2$ can then be widened to increase the brightness; a still better arrangement is obtained if the photographic plate is replaced by a transparent fluorescent screen followed by the photoelectric cell of an electron multiplier.

Lippert (1955) has made a theoretical study of the optical properties of einzellens working in transgaussian conditions, and of the final resolving power of this type of analyser as a function of the thickness of the central electrode and of the inter-electrode distances. If the size and angular aperture of the incident beam are sufficiently reduced (although
at the expense of the luminosity), it is possible to attain a resolution of the order of one part in $10^8$ with the lens which has the optimum dimensions. An analyser of this type, working at a maximum potential of 70 kV, has been described by Kushnir *et al.* (1960). The incident particles may be either electrons or lithium ions. In the spectra which are obtained, the losses which are

![Diagram]

Fig. 247. An improved arrangement, with which a resolving power which is constant over the whole extent of the spectrum can be obtained (Mollenstedt and Dietrich, 1955).

found with Li$^+$ ions are of the same order of magnitude as those which electrons display on passing through such gases as helium, argon, nitrogen and oxygen.

Another very high performance version has been described by Dietrich (1958a, b, c); a great deal of practical advice is to be found in his articles. (See also Metherell and Whelan, 1965, 1966.)

### 19.4 VELOCITY FILTERS

The slowing up of the electrons by the sample results in a reduction of the contrast at the image. With a velocity analyser, however, this can be overcome, as we can separate the retarded electrons from the remainder and collect them before they reach the image. This is the task of the “velocity filters” which are used in microscopy and diffraction.
19.4.1 Grid Lenses

The principle of the filter is simple: all that need be done is to insert a potential barrier into the path of the beam which is sufficiently high for only the unretarded electrons to be able to get past, and to accelerate the beam again before forming the image. Putting this principle into practice is, however, a matter of great difficulty. The barrier must be of the same height over the whole extent of the beam; the electrons, which have very low velocities in this region, are highly sensitive to space charge effects and to parasitic electric or magnetic fields. Finally, the barrier behaves as a lens, and we must be sure that this additional lens neither upsets the ordinary focusing nor introduces new aberrations.

Fig. 248a. Filter with a grid, with which the contrast in an electron diffraction photograph can be heightened (Boersch, 1949).

The simplest solution is to place a sheet of mesh which is held at the same potential as the cathode behind the first lens of the microscope or diffraction camera (Boersch, 1949, 1953, see Fig. 248a). The shadow of this grid need not be visible at the image, but the secondary electrons which are produced by the impacts of the ions which are created by the reflected electrons will perturb the final image. The best results are found in electron diffraction. When the filter halts any electron which has been retarded by more than 4 V, the continuous background of the diagrams vanishes almost completely. An interesting fact is observed—the Kikuchi lines entirely disappear under these conditions, a clear demonstration of the fact that they are a consequence of retarded electrons only.

A form of mesh filter has recently been described by Beaufils (1959); it consists of two closely inter-related parts (Fig. 248b). The upper part is a
retarding immersion objective, the cathode of which is held at the same potential as the microscope filament, and consists of a plane grid (mesh-length: 25 μ). The focus of the objective coincides with the crossover

![Diagram]

**Fig. 248 b.** The grid filter (Beaufils). *P*: intermediate projective; *F*: crossover, common to *P* and to the decelerating immersion objective *Ob*; *A*: anode; *W*: Wehnelt; *K*: cathode with several electrodes. to ensure that the equipotentials be plane and parallel; *G*: grid (at the source potential); *I.C.*: electrostatic image-converter (*K*: cathode; *A*₁, *A*₂, *A*₃, *A*₄: accelerating electrodes).

of the projection lens which is placed immediately above. Under these conditions, the electrons arrive normally at the grid. They then pass through the second part of the system which consists of a particular type of immersion objective which resembles those employed in image converters (§ 14.2.3, Fig. 174) and which produces a twice magnified image on the final fluorescent screen. The useful surface of the grid corresponds to a circle 3 cm in
diameter. When the grid is connected to the cathode, the potential at the centre of the holes is equal to \((\Delta V + 4)\) volts where \(\Delta V\) is the contact potential between the metal of the grid and that of the cathode.

19.4.2 The Filter Lens

In an ordinary unipotential lens, the potential barrier on the axis is as high as 90 percent of the cathode potential, and it is very difficult to raise it even a few volts further as the aberrations increase rapidly. We can,

![Diagram](image)

Fig. 249a. Filter with a lens, with which the contrast in an electron image can be improved (Boersch, 1953).

however, lower the potential of the central electrode or reduce the diameter of the central hole. Boersch (1953) chose the latter solution, and placed a thin film of platinum in which a small hole a few tenths of a millimetre in diameter had been pierced, in the mid-plane of the lens. This film was gripped between two 3 mm diaphragms (Fig. 249a) which act to: 2 potential distribution like immersion lenses. The quantities must be determined extremely precisely, as it is necessary that the lens shall work in the trans-gaussian region at the third maximum of the convergence—at this point, the spherical aberration and the distortion are minimized. The working point corresponds to an axial potential of 3 V for an H.T. of 27 kV. As the aberration constants are higher than those of the other lenses, the filter lens can only be used as an intermediate lens or as a projective lens, after the objective has reduced the aperture of the beam considerably. As all the useful electrons are to surmount the same potential barrier, a field diaphragm of a few hundredths of a millimetre must be introduced to avoid edge effects.

The filter increases the contrast in the image considerably, especially for thick objects (such as biological sections, for example). This technique has been gradually improved over several years, in particular by Forst (1958) who employed a very thick central electrode instead of a thin diaphragm.
Fig. 249b. Filtered images of magnesium oxide crystals. *Above*: Elastically scattered electrons ($\Delta E = 0$). *Below*: Electrons which have suffered the first characteristic loss ($\Delta E = 12.5$ eV).
19.4.3 The Magnetic Filter

A velocity analyser allows us to separate in space beams of electrons which have lost different amounts of energy on passing through the object; at the exit of the analysing unit (a magnetic lens or prism), a beam with known characteristics can be selected. Using this technique, and a stigmatic magnetic prism, Castaing and Henry (1962) have been able to obtain images of the specimen, produced first by the unperturbed but filtered beam (all the electrons which have made inelastic collisions are eliminated) and then by the electrons which have suffered a characteristic energy loss (Fig. 249b).
CHAPTER 20

SPECIMEN PREPARATION†

20.1 THE PROBLEM OF PREPARATION

Electron images are formed in conditions very different from those which are encountered in optical image formation; so marked is the difference that when, towards 1939, the electron microscope had reached a sufficiently high degree of perfection to be appreciated and employed by microscopists, its usefulness was still in doubt. Opinions were no less pessimistic as regards examining a surface (it is difficult to thin down a metal film sufficiently for it to be observable by transmission and the reflexion image is very distorted) or a biological section (the vacuum and the electron bombardment destroy the sections; it is difficult to obtain a thin enough section; before shadowing was discovered, it was impossible to see a particle less than two or three hundred Å in diameter). It was not enough to build a good instrument—methods of specimen preparation specially adapted for the electron microscope had also to be discovered. Although many problems are still without a totally satisfactory solution and the methods employed are in constant evolution, anyone who nowadays wishes to use an electron microscope has a vast battery of preparation techniques at his disposal, which are capable of providing a basis for the examination of the majority of specimens. What seemed to be virtually an absolute limit in 1940 was in reality only a complex mass of problems. Ingenious methods of overcoming these have been developed, which both contribute to the conditions in which good electrons microscope images are formed and leave the structure of the specimen untouched.

Each new application of the electron microscope requires, generally speaking, a new technique to be developed or one or more existing techniques to be adapted or combined. For this reason, we shall describe no more than the basic ideas, and refer the reader to the bibliography for further details. The recent advances in electron microscopy are due above all to the development of preparation techniques and to the use of reliable and straight-

† Very detailed discussion of many aspects of this subject is to be found in Magnan's *Traité de Microscopie Électronique*, Hermann. Paris, 1961.
forward microscopes, rather than to the introduction of ultra-high quality instruments. To obtain useful results, it is in no sense indispensable to be able to work continually close to the theoretical resolution limit, any more than it is inevitably necessary to use an immersion objective in light optics.

20.2 SUPPORTING THE OBJECT

20.2.1 Objects which Can Be Used Without a Support

A certain number of objects, mainly metal oxides (such as those of Mo, Zn, Cd and Mg) are capable of maintaining themselves on the rim of the specimen-holder with no other support than their cohesive forces. They hang from the rim of the opening in a diaphragm. This is the case with zinc oxide, mentioned in Chapter 1 (Fig. 2). Such specimens (in particular, the globules of carbon which are obtained by the combustion of a drop of benzine) are excellent as test objects with which the quality of an instrument can be judged; since the different particles lie in different planes, every photograph contains a region in focus.

20.2.2 Supporting Grids

In an optical microscope, the object is placed on a glass plate. In an electron microscope, the problem with which we are faced is to find a support which is strong enough to carry the object during the examination and which is transparent to the particle being used in the instrument without possessing a structure which could produce contrast. If the object is large, a grid can be used; this method is used for thin sections, wires, some diatoms, and generally speaking for any object larger than the mesh-length of the grid. These grids are either constructed by electrolytic deposition or by weaving the wires, but the latter have the disadvantage of an undulating weft. A method of obtaining grids with a high transparency coefficient and a very short mesh-length photographically has recently been described (Challice and Sutton, 1952). With such grids as these, it ought to be possible to examine with ease a large number of kinds of very small object without any other support, even down to the larger bacteria. An organic supporting film liberally pierced with holes can also be used (Jaffe, 1948). Grids are commercially available with which the coordinates of the various squares of the network can be determined; this means that it is possible to mark some particular part of the object, and make a comparative study in the optical microscope and the electron microscope for example (Bishop, 1949a; Hyam and Nutting, 1952; and Nankivell, 1953).
20.2.3 Organic Supporting Films

The most interesting objects are too small to be examined by the above technique, and must be supported on a transparent film between 150 and 200 Å thick which can be obtained by evaporating a few drops of a very dilute solution of varnish in a suitable solvent. This film is then placed on a grid which is in turn supported by the specimen-holder. For further details of this preparation technique, the reader is referred to a general examination by Drummond (1950), and to the articles by Schuster (1946) and Leroux (1953). The thickness of the film can be obtained by measuring the area of the membrane into which a known volume of the solution is transformed.

The table on p. 695 summarizes some recent attempts. In Fig. 250 the transparency of the supporting film to electrons is shown as a function of energy and thickness (Weber and Fraggstein, 1954).

20.2.4 Mineral Supporting Films

Organic supporting films are not very strong and tear under the action of electrostatic forces; they are easily melted under the effect of bombardment. For this reason, mineral films which are obtained by evaporation in a vacuum are found to be preferable. The installation required for vacuum evaporation is almost identical to the apparatus required for shadowing.
(i) The apparatus

This is shown in Fig. 251. The vaporized atoms move in a straight line and settle in a thin film on a surface which lies in the path of the flow. A vacuum of about $10^{-4}$ mm Hg is necessary. A wise precaution is to hold the oven at a temperature slightly below the temperature at which the metal evaporates for about a minute, in order to outgas completely the various parts which are to be held at a high temperature; the preparation is screened off during this period.

The evaporation oven can be constructed from a tantalum gutter (5 mm wide and a few tenths thick) or a heated filament which is bent into a helical or conical form, and which may be smeared with aluminium oxide, or beryllium oxide for substances which evaporate at a high temperature (Olsen, Smith and Crittenden, 1945; and Dunoyer, 1951). In the case of uranium or platinum, Williams and Backus (1949) recommend that the filament should be doubly folded into the shape of a hairpin. The nature of the heating filament in any given case (W, Ta, Mo, Cr, Chromel) depends upon the material which is to be evaporated (Caldwell, 1944). It is sometimes possible to avoid corroding the filament too rapidly by plating it with platinum; this involves immersing it in platinum chloride, and subsequently heating it gently in air. As only small amounts of metal are to be evaporated it is pointless to use filaments larger than 0.5 mm in diameter (heated with 12 to 20 A) and we thus reduce the thermal damage which may be inflicted on the object. If the metal to be evaporated forms an alloy with the filament, or is not particularly volatile, a diameter of 0.75 mm and a heating current of 50 A may be necessary.
<table>
<thead>
<tr>
<th>Organic varnish</th>
<th>Concentration %</th>
<th>Solvent</th>
<th>Evaporation over</th>
<th>Properties</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parlodion</td>
<td>1 to 3</td>
<td>Amyl acetate</td>
<td>Water</td>
<td>Not very resistant</td>
<td>Schaeffer and Harker</td>
</tr>
<tr>
<td>Collodion</td>
<td></td>
<td>Butyl acetate</td>
<td></td>
<td></td>
<td>(1942)</td>
</tr>
<tr>
<td>Nitrocellulose</td>
<td></td>
<td>Dioxane</td>
<td>Glass (spreads poorly over water)</td>
<td>Resistant but not very flat</td>
<td>Ellis (1947)</td>
</tr>
<tr>
<td>Formvar 15/95</td>
<td>0.2 to 0.4</td>
<td>Ethylene dichloride</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(polyvinyl formol)</td>
<td></td>
<td>Warm xylene</td>
<td>Water or glass</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkathene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorinated rubber</td>
<td>2 to 3</td>
<td>Toluene</td>
<td>Water or glass</td>
<td>Insoluble in alcohol; resistant to bases and acids</td>
<td>Davoine and Pernoux</td>
</tr>
<tr>
<td>Methyl-cellulose</td>
<td></td>
<td>Amyl acetate</td>
<td></td>
<td></td>
<td>(1951)</td>
</tr>
</tbody>
</table>
The best ovens operate with either induction heating (Otis, 1933; and Picard and Joy, 1951) or heating by electron bombardment (Brochard, Giacomino and Jacquinot, 1951). These devices, which are noticeably more complicated, allow us to lay down a finer deposit.

(ii) The materials evaporated

The principal materials used are:

(a) Silicon monoxide, SiO, which can easily be evaporated on a very thin collodion film which is then burnt or dissolved. The film, which has no structure, can only be used at thicknesses less than 20 Å (Hall, 1948). SiO is preferable to silica which is less easy to evaporate.

(b) Beryllium, evaporated onto a glycerine surface (Hast, 1948). A 20 Å layer is sufficiently resistant, but tends to crystallize; Kaye (1949) remedied this by using an Al–Be alloy (40–60).

(c) Silicon, evaporated in a crucible of beryllium oxide (this is very solid, and chemically inert).

(d) Carbon films are excellent supports. They can be obtained by vacuum discharge (König, 1948, 1951) or with an auxiliary ion gun (Möllenstedt and Hubig, 1954) as well as by evaporation in a vacuum (Bradley, 1954). The films which are used are about 40 Å thick.

(iii) The thickness which is evaporated

It is difficult to estimate the thickness of the evaporated layer, as it is not easy to hold the conditions constant and to obtain perfectly reproducible results. Some workers use a known quantity of the substance, which is, for example, measured by using a known length of wire (Howard and Berntson, 1952) which is wholly evaporated; others cover the base with a large volume of the substance, and determine the mass \( M \) which is evaporated as a function of time and heating current (Robillard, 1949; and Bishop, 1949a, 1949b).

The thickness of the film can be calculated if we know the evaporation law of the oven (see the various ways of performing the calculation in Holland and Steckelmacher, 1952). With a point filament emitting isotropically in all directions, and a small plane surface inclined at an angle \( \alpha \) to the direction of evaporation, the thickness is

\[
\varepsilon = \frac{M \sin \alpha}{4\pi d^2 a},
\]

in which \( M \) is the mass evaporated, \( d \) the distance, and \( a \) the density.

This formula does not give very precise values, as the density of the layer may be very different from the density of a macroscopic quantity of the same substance; this, however, is not very important as it is only the
mass thickness $\mu_c = a \varepsilon$ which appears. Williams and Backus (1949) suggest that the calculated thickness should be multiplied by $\frac{3}{4}$ to take into account the efficiency of evaporation which does not usually exceed 75 per cent. Further details are to be found in the articles by Rang (1952), Schoen and Davis (1951) and Nandy (1954).

### 20.3 Direct Observation of the Object

#### 20.3.1 Thin Biological Sections

If the object is small in size (powdered oxide, for example) all that need be done is to distribute a few grains over a support. If the object that we wish to examine is thick, however, we must cut a thin section which will be transparent to electrons. Further, the large depth of field of the instrument will produce a superposition of the images of various planes of the object at the screen, so that interpretation of the details in the image of a relatively thick object will be a very delicate operation.

Sections which are to be used in an electron microscope must be about 100 Å thick; ordinary microtomes cannot be used therefore (thin sections for optical examination are a few microns thick) and we shall now say a few words about the refinements which have proved to be indispensable.

(i) *The double-sectioning method (Sjöstrand, 1951)*

This method gives very fine sections (a few hundred Ångströms thick) with which a very high resolution (20 to 50 Å) can be attained. A thick section of the object is first of all cut (0.5 to 1 μ); this section is then enveloped in collodion, and the majority is removed with a special microtome (see Fig. 253). The part which remains is then examined—it can be extremely thin, and need not be at all deformed by the knife-blade. Unfortunately, this is a most delicate technique.

(ii) *Modified standard microtomes*

The present tendency is to use ordinary microtomes which have been modified in such a way as to reduce their travel; in particular, this can be done by thermal expansion (Newman et al., 1949, 1949a; Latta and Hartmann, 1950; and Eden et al., 1950). Sjöstrand (1953, 1953a, 1954 and 1955) has built an extremely high precision microtome with which sections between 100 and 200 Å thick are normal, and the finest sections can be as thin as 70 Å. The object is fixed eccentrically at the end of a stout rotatable rod, and only passes the knife-blade once in each cutting cycle. The same result (with a single cut) is obtained in Haanstra's microtome, where
Fig. 252. An example of a biological object examined after shadowing.

Biological Interpretation: The particles of bacteriophage attach themselves to the cell at the top of the figure by their tails; this cell is "receptive" (an E. coli F cell); they do not, on the other hand, attach themselves to the cell at the bottom which is "not receptive" (an H Fr cell). Further, a connecting bridge between the two cells can be seen, which allows genetic material to be transferred from one type of cell to the other.
the object oscillates to and fro; as the object returns under the knife, the support of the latter contracts slightly by magnetostriction, so that the blade is slightly separated from the object (Sjöstrand, 1954).

![Diagram](image)

**FIG. 253. Method of obtaining the double section.**

The cutting-edge of the knife can be an ordinary razor-blade (von Ardenne, 1948; and Bernhard and Gautier, 1951); alternatively, a splinter of diamond, synthetic sapphire, or tungsten carbide can be used, or more simply, a glass splinter prepared in the way described by Latta and Hartmann, (1950), or Lépine and Croissant (1953). Although many workers prefer to use steel knives sharpened in the laboratory (Richards, 1950; Hillier, 1951; and Sjöstrand, 1953) the present tendency would seem to be a return to glass knives, which consist of the edge of a freshly broken wedge. A trough full of liquid is attached to the knife support, in such a way that the cutting-edge comes into contact with the meniscus. The sections are laid out on the liquid, and are thus gathered automatically.

Finally, we should mention that it seems that the methods of cutting biological sections have become refined enough for materials to be examined on the molecular scale; the examination can now be made with a resolution of 20 Å, but an improvement of this down to 10 Å would advance our understanding of the fundamental cellular processes considerably.

### 20.3.2 Thin Metal Sections

One method of examining metals consists of thinning down a metal disc until the thickness is small enough for the transparency to be adequate. The first efforts of this kind were made by Heidenreich (1949) who set out with a metal disc the same size as the grid of the specimen-holder, a few tenths of a millimetre thick; the two faces of this disc were then successively thinned down at the centre by electrolytic polishing until a few small holes started to appear. In favourable cases, certain zones close to the holes are thin enough to be examined by transmitted electrons. The method has been simplified by Saulnier and Mirand (1958), who employ an electrolytic polishing device which is available commercially; the specimen is first of all thinned down to a few hundredths of a millimetre, and then polished electrolytically on the two faces successively, in order to get rid of beaten-in layers. Finally, it is further thinned down, using the same
bath at high intensity for a few tens of seconds. Fragments of the specimen are torn away and remain in the bath where they are collected; some of these are between 0.3 and 0.5 mm in diameter and a few hundred Å thick, and can be used directly.

In another method, the electrolytic polishing is followed by a final thinning down by cathode sputtering, produced by bombarding the two faces of the specimen with ions (Castaing and Laborie, 1953, 1954; Castaing and Lenoir, 1954), or alternatively, the whole thinning-down process is

---

**Fig. 254.** A sample of the alloy Al–Cu (4%), aged at 200°. (a) Oxide replica (Castaing). (b) Evaporated aluminium replica (Baillie). (c) Direct transmission after thinning down (Castaing).
FIG. 255a. Diagram of the reflexion microscope (Fert, Marty and Saporte, 1955). The electrons fall on the object at an angle $\theta$; the ion gun to be seen on the right is used for cleaning the surface; $S_1 S_2$ is a stigmator.

Fig. 255b. Microphotograph of an electrolytic copper surface: the two magnifications differ by a factor 10, the illumination angle is 2° and the observation angle 6° (Fert, 1954).

conducted with ions (Castaing, 1955a, b). The surface of the specimen is perfectly clean, and any source of artificially added substance is eliminated. The full resolution of the electron microscope can therefore be fruitfully employed in examining the metal; finally, the various phases which are apparent in the micrograph can be identified by localized electron diffraction (Castaing, 1955a, b, 1956). The thinning down can be performed within the microscope itself, and it is possible to watch the process continuously while it is actually in progress (Castaing, 1954a, b). Specimens which are non-
metallic and poor conductors, and which cannot be polished electrolytically, can be thinned down and examined at a very high resolution by the same process. It seems likely that the standard methods of observing replicas will have to be replaced by ionic thinning down in cases where a very high resolution is required.

20.3.3 Observation of Surfaces by Reflexion

It is possible to build the electron optical counterpart of the metallographic microscope in which the surface of the specimen is examined by reflexion. The first attempts to do this were made by Ruska (1933, 1940) and later by von Borries (1940). The illumination fell upon the specimen at grazing incidence, and good images were obtained. The method was then abandoned.

Fig. 255 c. Perlitic casting. The magnifications differ by a factor 2.5, the illumination angle is 2° and the observation angle 23° (Fert, 1955).
for a considerable time, but has now returned to favour, and a number of instruments have been specially constructed for examination by reflexion (Menter, 1952; Cosslett, 1952; and Fert, 1952, 1952a and 1954). The resolving power is limited by the chromatic aberration of the objective and by the velocity dispersion of the scattered electrons; Haine (1953) estimated his limit to be about 70 Å, but various other workers have reached only 300 Å. The angle at which the illuminating beam is incident must be very acute (a few degrees); the instrument possesses the special property of being highly sensitive to variations of relief in the specimen—differences between the various levels which are smaller than 20 Å are perceptible, provided their lateral extent is greater than the limit of resolution.

The majority of reflexion microscopes work with a scattered beam which emerges at a very acute angle as well. The scattered intensity is then appreciable, and the image is bright. The image does possess, however, the serious drawback of being considerably contracted in one direction, and since it is not easy to recognize the surface of the object at so acute an angle, the image is difficult to interpret. Fert and his collaborators (1955) avoided this problem by using a much larger angle of emergence (about 24°) but at the expense of a considerably longer exposure which is then about half a minute; in this way, he obtained excellent images with a resolving power of 300 to 400 Å, easy to interpret in that the object is comparatively normal (the ratio between the magnifications in the two directions is about 2·5; see Fig. 255c).

20.3.4 Observation of Living Biological Sections

When biological sections are introduced into the vacuum which is maintained within a microscope, the cells dry out—only dead tissue can be examined. Further the penetrating power of electrons accelerated through a few tens of kilovolts is so small that in practice one is obliged to use extremely thin films, at most a few hundred Ångströms thick. If living matter is to be studied, therefore, we must employ a special specimen-holder within which the preparation experiences normal conditions of pressure and humidity, and work with very high energy electrons, which are capable of passing through living cells of the order of a micron in thickness, the substance on which the preparation is supported, and the layer of humid air which surrounds it, without absorption and also without excessive scattering. The experimental work of Stojanowa (1958) has shown that the intensity of a beam of 80 kV electrons, incident on two collodion membranes 100 Å thick, strengthened by the evaporation of a layer of carbon, together with 0·01 mm of air at a pressure of 600 mm Hg, is reduced by three-quarters.

Recently, Dupuy and Perrier (1960, 1961) have succeeded in transforming into reality one of the oldest of the biologists’ dreams as regards the electron microscope, namely, the formation of an electron image of
living cells without killing them either before or during observation. They use a special type of magnetic microscope (which has already been described, in § 18.4.3) which is operated at potentials between 600 kV and 1.5 MV. The specimen-holder which is used is illustrated in Fig. 255 S1; it has the form of an extremely short cylindrical chamber (the height of the cylinder can be varied between one and ten microns), the plane end-faces (\(F_1\) and \(F_2\)) of which contain holes \(O_1\) and \(O_2\) each 0.1 mm in diameter which are in turn closed with parlodion membranes strengthened with carbon. These membranes constitute the windows which are transparent to the electrons; an extremely fine tube \(T\) provides communication with the outside atmosphere (If the volume of the reservoir of air \(C\) which surrounds the axial zone is adequate, the tube can be dispensed with when the observations last only for a short time.) When the chamber is placed in a vacuum, and the pressure is atmospheric, the windows have only to withstand a force of 8/100 g wt. The cells to be examined (bacteria) are placed on one or other of the two windows, and the chamber is then assembled under an optical microscope with the aid of which the thickness of the "strip" of air can be measured; the chamber is then placed in the gap between the pole-pieces of the objective.

The bacteria which are being studied are subjected to the action of the electrons and also to the fairly intense X-radiation which is produced by the electrons in the part of the beam which falls on the entry diaphragm of the chamber. The first experiments seemed to show that when the instrument is well adjusted, the bacteria remain alive despite the radiation. Figures 255 S2 and 255 S3 are photographs of bacteria taken at relatively
Fig. 255 S3. Photograph of anthracomorphous bacillus, showing the spores
(Photo: Dupouy and Perrier).

low magnification and high tension (650 and 750 kV respectively), during the first attempts to adjust the instrument. This method ought to be capable of providing valuable information on the evolution with time of the constituents of living cells, with a resolution which is of the same order of magnitude as that of ordinary microscopes.

20.4 INDIRECT OBSERVATION OF THE OBJECT

20.4.1 Replicas

It was Mahl (1940, 1941) who thought of an artifice with the aid of which the surface of a massive object could be examined without thinning and without the oblique examination which reflexion entails; his idea therefore marks one of the major advances in the history of electron microscopy. The process consists in examining by transmission a moulding of the surface, a sufficiently thin moulding for the electrons to pass through easily, which is known as a replica or imprint of the surface. This moulding can be
examined with all the resolution of which the microscope is capable—the only limit to the fineness of the details of the surface which we can see is constituted by the fidelity of the replica. We must obtain as fine and as faithful a replica of the surface being examined as possible, by covering this surface with a thin film which settles into all the irregularities, and then withdrawing this film without deforming it. After this, we can extract all the information which this replica is capable of giving with the aid of a good microscope.

We shall classify the various replica techniques under two main headings, according as the replica is obtained directly or in two stages. We shall only describe the most typical methods, without entering into the host of variations which various authors have suggested.

20.4.2 Direct Replicas

(i) Oxide replicas

This was the first method of replication which was used (Mahl, 1940, 1941 and 1942), and so far as fidelity is concerned, it is probably the best which is known at present. Unfortunately, it has the serious drawback that it can only conveniently be applied to aluminium (Heidenreich and Shockley, 1947; Brown and Honeycombe, 1951; Chaudron and Bussy, 1953; and Bussy, 1954) and aluminium alloys (Geissler and Keller, 1947; Castaing, 1949; and Castaing and Guinier, 1949a, 1949b). After electrolytic polishing, and chemical attack, the surface of the specimen is oxidized anodically, in such a way that a film of aluminium oxide between 200 and 400 Å thick is formed on its surface. This film has then to be detached; various ways of performing this operation without tearing the film have been proposed, and the reader will find details of these in the articles already mentioned, and in the articles by Walkenhorst (1947), Hass and Scott (1950), Gulbransen (1948), Mahl and Nielsen (1948), Tsou and Nutting (1950) and Yamaguchi (1951 and 1952).

(ii) Plastic replicas

After the surface of the specimen has been subjected to an attack the purpose of which is to display the variations of its internal structure as variations in the relief, it is wetted with a dilute solution of some plastic substance (a solution of 0.5 per cent formvar in dioxane, for example). When the solvent has evaporated, the specimen is covered with a thin plastic skin (less than 1000 Å thick) in intimate contact with its surface. To obtain the desired replica, we have simply to detach this plastic film. This delicate operation is the only difficulty encountered in this method, which is otherwise very rapid and well adapted for a systematic long-
drawn-out study. We can withdraw the film either wet, easing it away with
tweezers after immersing the specimen in a liquid (warm water, for example),
or dry; in the latter case, a specimen-holder grid is placed against the sur-
face of the specimen which is covered with the plastic film, and an adhesive
ribbon is then pressed against the whole combination. The ribbon is gently
drawn away, bringing with it the grid and the replica (Schaeffer, 1942;
Nutting and Cosslett, 1950). This latter method is, however, more rough,
and there is a risk of introducing foreign materials if the film is not very
thick (Habraken, 1954).

20.4.3 Indirect Replicas

In this case, the replica is made in two stages: first, a thick mould of the
specimen is taken, which is easy to withdraw; a thin replica of the face
of this first imprint which had been in contact with the specimen is then
taken. This second replica can then be isolated with ease, simply by dissolv-
ing the intermediate mould. We shall mention two methods which are very
much in favour at present.

(i) The evaporated aluminium replica (Hass and McFarland, 1950; and
Fert and Dargent, 1952)

The surface to be reproduced is covered with a relatively thick layer of
aluminium by evaporation in vacuo, and this layer is removed with a band
of adhesive ribbon. The whole combination is then subjected to anode
oxidation. The aluminium layer is oxidized over the face which had been
in contact with the specimen (we have found that the aluminium layer has
to be covered with a collodion film before the adhesive ribbon is struck
on if we are to avoid oxidation of the back face, and hence the formation
of a double layer of oxide). The oxide coated aluminium is detached from
the adhesive ribbon simply by immersing it in water, slightly acidified with
nitric acid if necessary. The layer is then placed in a solution of mercuric
chloride which dissolves away the aluminium but leaves the aluminium
oxide replica. The replica is robust, and gives a very fine image with good
contrast; this is a valuable method for many types of object.

(ii) The carbon replica (Bradley, 1954)

A thick film of formvar is deposited onto the surface to be studied, and
then pulled away dry. A thin film of carbon (about 50 Å thick) is deposited
onto the face which contains a mould of the specimen, by evaporation
in vacuo. This is done by placing two bars of hard graphitic carbon, each
sharpened to a point, in the vacuum chamber with their points just in con-
tact. If a current (of 30 or 40 A) is passed between the points, an intense
local heating is produced and hence a slight evaporation; the current is cut when the deposit has reached a suitable thickness. The formvar is then dissolved in chloroform, and a fine carbon replica remains. Despite the intermediate formvar replica, the resolution is excellent; according to Bradley (1955), this indicates that the poor resolution which is obtained

![Image](image_url)

**Fig. 256a.** Carbon replica of a broken surface of mild steel (the magnification is 20,000; Plateau).

![Image](image_url)

**Fig. 256b.** Evaporated aluminium replica of the surface of an Fe-Cu alloy which has undergone thermal attack (magnification: on the left, 90,000, and on the right, 20,000; Plateau).
with a normal formvar replica is not, as has been generally believed, due to the infidelity of the mould, but to the perturbations created by the electrostatic field produced by charges which attach themselves to an insulator. The carbon replica can also be used as a direct imprint (when the carbon film is evaporated directly onto the specimen itself); in this case, however, the specimen must be dissolved to detach the replica. Bradley ob-

tained very handsome replicas of mineral crystals in this way. If the carbon film is thick enough, stereoscopic views of large crystals can be obtained. Some photographs of replicas are reproduced in Figs. 256.

20.4.4 Contrast in the Replica

The image which is obtained depends essentially on the mode of replication which has been used and if errors of interpretation are not to be made, the mechanism of contrast formation must be well understood in each individual situation. We have seen (§19.1.1) that the shadowing technique confers upon the image the appearance which the surface of the
object would have if it were strongly illuminated at a small angle of incidence; the shadowing layer therefore creates the main contrast, and the replica method which has been used is not very important. The image is very "plastic" and can be interpreted immediately. The shadowing can, however, reduce the resolution slightly, and this is not always desirable; we shall study the contrast in a replica which is not shadowed, therefore, and we shall examine three typical cases—a plastic replica, a replica of the "evaporated" type, and a chemical replica.

(i) The plastic replica

This is the case of the direct unshadowed formvar replica. The lower part of the plastic layer settles more or less perfectly against the surface to be reproduced. If the relief of the object is not too pronounced, the upper surface of the layer is virtually plane (Fig. 257). The contrast results from variations in the thickness of the plastic film, which is thin where the original surface was ridged, and thick where the surface had deep valleys—the positive, therefore, will be dark in the hollows and bright at the projections. The appearance is analogous to that of the bottom of a lake, when seen from a certain height.

(ii) The "evaporated" type of replica

In this category, we place all the replicas with a constant thickness, such as the silica replica, the carbon replica, and the evaporated aluminium replica, although in the last case, the mould proper is obtained chemically. The film of uniform thickness settles against the surface to be reproduced.
The contrast results from the variations in the apparent thickness which are due to changes of slope of the surface (Fig. 258). The thickness of the material through which the electron beam passes is \( e \sec \theta \). The transmission varies inversely as the local slope; on the positive, a dark region corresponds to a steep incline, without any indication as to its direction. The appearance of the positive can be related to the appearance which the surface would have if it were observed with a metallographic microscope at small aperture; *a priori*, nothing distinguishes hollows from projections, so that a group of hemispherical projections would appear as a group of dark rings with clear centres, and would be indistinguishable from a group of hollows of the same shape.

(iii) *The chemical replica*

The metallographic oxide replica is representative of this case. If the specimen is homogeneous (the pure metal), the replica falls into the preceding category (replicas of the "evaporated" type), and the contrast is produced according to the same laws; we have a simple replica of the "shape", to deal with. If, on the other hand, the specimen is heterogeneous (for example, a precipitate), the precipitates may not be attacked in the oxidization process; they remain in their place on the film of oxide, and show up black on the positive. Often, however, the precipitates are dissolved during the oxidization (for example, anodic oxidization). The replica then possesses the property of itself displaying the structure of the specimen, without a preliminary attack being necessary. The various cases which may appear
for precipitates of different sizes are illustrated in Fig. 259 (Castaing, 1949), when the anodic oxidation potential is not too high. Any very small precipitates remain included within the film of oxide, and show up black on the positive if their scattering power is higher than that of the oxide. The presence of larger precipitates is betrayed by holes or (clear) slits in the film of oxide. Coarse precipitates are dissolved by the oxidization, and their edges are covered by the film. If the anodic oxidization potential is high even the small precipitates produce holes (punctures) in the oxide film; the effect on the positive is a number of white points which are not, unfortunately, always easily distinguishable with certainty from the accidental punctures which are statistically distributed over the surface of the replica.


The intense electron bombardment to which an object being examined in a microscope is subjected may raise the temperature of the object considerably. Since the thermal capacity of the region being bombarded is
very small, thermal equilibrium is reached very rapidly. The equilibrium
temperature of any given point on the object depends upon:

(a) the energy absorbed locally; this in turn depends upon the intensity
and the accelerating potential of the beam, and upon the thickness, the
density and the mean atomic number of the region in question;

(b) the rate at which the calories absorbed are dispersed, by radiation
and by conduction.

To calculate the equilibrium temperature rigorously is difficult; we
mention the approximate calculations of von Borries (1948) who takes
only the radiation into account and of von Borries and Ruska (1939)
who make allowance for conduction but only in the case of a supporting
film. An exact calculation would require the percentage of electrons which
lose a given energy, the percentage of electrons absorbed, and the impor-
tance of secondary emission all to be known. According to von Borries,
the minimum electron intensity with which the image can be focused with
a resolution of 50 Å raises the temperatures to 50°C when the accelerating
potential is about 100 kV and the object has a mass thickness of 10^{-6} g cm^{-2};
for a resolution of 10 Å and an object of mass thickness 10^{-5} g cm^{-2}, the
temperature rises above 500°C. If the accelerating potential is increased,
the light yield at the screen falls, while at the same time the energy losses
in the object diminish; for a given brightness, therefore, there is an opti-
num potential which is close to 100 kV. Stefan’s law shows that the radia-

![Diagram](image)

Fig. 260. A simple model for studying the heating of the object in an electron
microscope.

tion losses are scarcely affected by cooling the chamber. On the other hand,
should the conduction losses become predominant, the temperature of
the object settles down to a value which is a given amount above the tem-
perature of the vacuum chamber. It is advantageous, therefore, to diminish
the latter temperature.

A special case in which the conduction losses predominate is provided
by the examination of thin metallurgical specimens by transmission (Hei-
denreich, 1949; Castaing, 1954a,b). For the study of light alloys in particular,
it is important that the temperature of the specimen should not exceed about fifty degrees, as otherwise the structure may be modified. It is a straightforward matter to calculate the temperature increase of a metallic object which is uniform in thickness, if we neglect radiation losses. If the object is of thickness $e$, and is supported along a circle of radius $R$ by the specimen-holder which constitutes a thermal sink at the temperature $t_0$ and

\[ t_m = t_0 + \frac{W}{4\pi JC} \left( 1 + 2 \log_e \frac{R}{r} \right). \]  

(Castaing, 1951).

\( J \) is the mechanical equivalent of heat, and \( C \) the thermal conductivity. If we suppose that \( W_0 \) is the electron energy which is incident upon the specimen, we can assume for the purposes of an approximate calculation that this energy is absorbed, roughly speaking, exponentially according to a law of the form \( W = W_0 \{ 1 - \exp(-\sigma e) \} \); for small values of \( e \), this law reduces to \( W = W_0 \sigma e \). For the heating, therefore, we find

\[ t_m - t_0 = \frac{W_0 \sigma e}{4\pi JC} \left( 1 + 2 \log_e \frac{R}{r} \right). \]

As an example, we shall calculate the heating at the centre of a layer of aluminium (\( C \approx 0.5 \)), cooled along a circle of radius $R = 1$ mm and subjected to an electron bombardment of intensity $1 \mu A$ and accelerating potential $50$ kV, which is uniformly distributed over a circle of radius
$r = 0.05$ mm in the centre of the disc (these are approximately the working conditions in a R.C.A. microscope at a magnification of 20,000). The coefficient $\sigma$, which is defined by the electron absorption according to Lenard and by the retardation of the electrons according to Thompson and Whiddington, is about 1000 in this situation (Zworykin et al., 1949). The heating which is found at the centre, therefore, is about $13^\circ$. This slight heating corresponds to an electron intensity of $12$ mA cm$^{-2}$ at the object, or $3 \times 10^{-11}$ A cm$^{-2}$ at the image when the magnification is 20,000 (straightforward focusing). If the brightness of the image is increased by a factor 100 (to $3 \times 10^{-9}$ A cm$^{-2}$ which is a very brilliant image), which can be done by concentrating the electron beam into a circle with radius $r = 5 \mu$ (condenser “at the focus”), the heating at the centre is still only about $22^\circ$ (Castaing, 1956).

We see that metallic objects which are good thermal conductors remain relatively cool under the impact of the beam in ordinary illuminating conditions; this advantage has another aspect, however—the contamination is very rapid so that a given region can only be usefully examined for less than a minute. In these conditions, it is clear that the shadowing technique will be particularly valuable to reduce the thermal load of a specimen of insulator. These considerations are, however, only applicable to specimens which are sensibly uniform in thickness. The observation of thick and absorbent particles which are spread over a thin supporting film and which are in practice cooled by radiation alone can heat them to such an extent that they melt. Forestier, Haasser and Uhl (1950) have systematically studied the melting of the various elements as a function of electron intensity (Fig. 261), and the results agree with Mendeléeff’s columnar classification.

Apart from simple thermal effects such as recrystallization, melting or partial sublimation, the effect of electron bombardment upon the specimen can equally well be a chemical transformation; we mention the transformation of $\text{Ca(OH)}_2$ into $\text{CaO}$ and of $\text{MoO}_3$ (transparent flakes) into $\text{MoO}_2$ (jagged crystals; Pernoux, 1953). Organic substances too undergo fundamental transformations—according to König (1948) collodion supporting films are transformed into graphite during bombardment. It is not easy to be confident, therefore, that the object which is being observed is the same as the one which was inserted into the microscope. The simplest way of reducing these transformations to a minimum is to conduct the observation with the minimum electron bombardment, and hence at as low a magnification as is tolerable. Alternatively, we should mention the method suggested by von Ardenne (1940, 1948) in which a jet of hydrogen at low pressure is directed onto the specimen: (this should also reduce the contamination—see Castaing and Descamps, 1954); a diaphragm is placed above the region of the specimen in which we are interested, and is retracted only for taking photographs.
CHAPTER 21

THE DIFFRACTION CAMERA

21.1 ELECTRON DIFFRACTION

21.1.1 Description of the Phenomenon

The first verification of de Broglie's theory of the wave-nature of the electron is to be found in the interpretation of the precise though complicated experiments of Davisson and Germer (1927), who studied the reflection of slow electrons at a nickel crystal. Shortly afterwards, a striking confirmation appeared when first Thomson (1928–9) and later Ponte (1929) showed that it was possible to obtain diffraction diagrams with electrons which are similar to the diagrams obtained with X-rays. To do this, a very fine

Fig. 262. The diffraction diagram of zinc oxide (Lafourcade). 716
beam of fast monoenergetic electrons is fired through a very thin polycrystalline layer. Figure 262 is a reproduction of the diagram obtained by transmission through polycrystalline zinc oxide; it is to be noticed that analogous diagrams can be obtained by "reflecting" the electron beam from a plane surface at grazing incidence.

21.1.2 Translation of the Bragg Theory

We shall now forget the corpuscular aspect, and henceforward, the electron beam—which is supposed parallel and monoenergetic—will be characterized by the wavelength $\lambda$ of the associated wave which we have already defined in § 15.2. It will be the relativistic version of the formula which relates $\lambda$ to the energy, $V$, of the electrons, which will be useful, in view of the extreme precision of the experiments and the high potentials which are used in certain diffraction cameras, namely

$$\lambda = \frac{h}{\sqrt{2emV}} \left(1 + \frac{eV}{2mc^2}\right)^{-1/2} = \frac{12.25}{\sqrt{V}} (1 + 0.9786 \times 10^{-6} V)^{-1/2},$$

where $\lambda$ is in Å and $V$ in volts. At 100 kV, the relativistic correction has reached 5 per cent; an accelerating potential of 60 kV corresponds to a wavelength of 0.05 Å.

We consider a crystal within which is a family of lattice sites two of which are drawn in Fig. 263a. The electron wave is scattered by one of the atoms $A'$ of one of these planes, and we examine the resulting phenomena in a direction which is inclined to the incident beam at an angle $\theta_i$. The
intensity in this direction will be appreciable only if the various waves diffracted by the atoms at the receiver are in phase; otherwise, the waves will be destroyed by interference, so numerous are they.

This phase coherence must first of all be assured for all the waves diffracted in the selected direction by the atoms in any one plane. This condition leads to the standard reflexion law, which we have already taken into account in Fig. 263a. Further, the coherence must still be maintained for a pair of atoms which lie in different lattice planes distance \( d_i \) apart. This requires that \( HA' + A' K \) should be a whole number of wavelengths, which implies

\[
2d_i \sin \alpha_i = 2d_i \sin \frac{\theta_i}{2} = n \lambda.
\]

Diffraction will only be found in a direction making an angle \( \theta_i \) with the incident direction if electrons are incident at an angle \( \alpha_i \) on a crystal which has lattice planes a distance \( d_i \) apart.

![Diagram](image)

Fig. 263b. Diffraction by a polycrystalline specimen.

If the specimen is a single crystal, the diagram which is obtained consists of a regular pattern of spots which form a geometrical design the appearance of which depends both upon the nature of the crystal and the direction of the incident electrons. If the specimen is composed of a large number of small randomly orientated crystals, however, the selective reflexions give rise to diffracted beams in the form of hollow cones, of semi-angle \( \theta_i \), which produce rings of radius \( R_i \) on a screen at a distance \( L \) (Fig. 263b); this is the analogue of a Debye–Scherrer diagram for X-rays. By measuring \( R_i \), we can calculate the corresponding value of \( d_i \). Since the angles \( \theta_i \) are very small (at most a few hundredths of a radian), we have

\[
d_i = \frac{\lambda L}{R_i}.
\]
If the rings which correspond to large values of $d_i$ are to be observed, $L$ must be large; for $d_i = 10\ \text{Å}$ and $\lambda = 0.05\ \text{Å}$, we should have $R_i \geq 1\ \text{cm}$ if $L \geq 2\ \text{m}$ which would give a prohibitively large instrument. We shall see how this difficulty can be surmounted electron optically.

Another diffraction pattern, which only appears when the object is thick, is known as the "Kikuchi line". Inelastic scattering in the thick object produces a homogeneous continuous background, which appears to be traversed by bright or dark lines and bands which cross one another and form regular geometrical patterns. These patterns which are in general obtained with a highly convergent illumination, are also valuable in determining crystal structures (Wilman, 1948a, 1948b).

21.1.3 Comparison of X-rays and Electron Rays

It is of some interest to make a careful comparison between X-rays and electron rays; the essential characteristics are as follows:

(i) Wavelength

For accelerating potentials between 50 and 100 kV, the wavelengths which are associated with electrons lie between 0.054 and 0.036 Å. The X-ray wavelengths which are normally used in crystallography, and which correspond to the $K_x$ lines of the various elements, are far longer, as the following table shows.

<table>
<thead>
<tr>
<th>Element</th>
<th>Pt</th>
<th>W</th>
<th>Mo</th>
<th>Cu</th>
<th>Fe</th>
<th>Cr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda (\text{Å})$</td>
<td>0.19</td>
<td>0.21</td>
<td>0.71</td>
<td>1.54</td>
<td>1.93</td>
<td>2.29</td>
</tr>
</tbody>
</table>

The line which corresponds to the shortest wavelength will be the $K_x$ line of uranium, $\lambda = 0.1\ \text{Å}$.

(ii) Penetrating power

Electrons have a vastly smaller penetrating power than X-rays. For example, an aluminium sheet 3 cm thick only absorbs about 9/10 of the intensity of a beam of hard X-rays ($\lambda = 0.2\ \text{Å}$), while 5 µ of the same material are sufficient to halt 100 kV electrons completely. In both cases, the absorption increases with the wavelength and with the atomic number of the element through which the beams pass. Transmission electron diffraction can only be observed, therefore, with specimens of a mean thickness less than 0.1 µ. A smaller quantity of the substance is required; in reflexion, only surface layers can be studied effectively.
(iii) **Monochromatism**

The X-radiation which is emitted by a tube consists of a number of lines superimposed on an intense continuous background. To obtain monochromatic radiation, we must either use filters, or selective reflexion at a crystal, which produces a marked reduction in the intensity of the beam. It is very easy to obtain a monochromatic electron beam, on the other hand, in which all the power dissipated by the source corresponds to precisely the chosen wavelength.

(iv) **Power**

The efficiency of the energy yield of an X-ray tube scarcely exceeds one part in a thousand, and the radiation is very diffuse. To obtain a beam which is both very fine and very intense, therefore, is an extremely difficult task. Further, the selective reflexion coefficients are of the order of $10^{-4}$. Exposures of several tens of minutes are often necessary to obtain X-ray diffraction diagrams. In electron diffraction, conversely, the intensity is much higher, and the diagram can often be recorded in a few seconds. This fact is due to two fundamental causes: firstly, the possibility of concentrating electrons into a fine and very intense pencil, with the aid of a suitable electron optical arrangement, and secondly, the fact that the interaction between matter and electrons is considerably more powerful than the interaction between matter and X-rays.

(v) **Refractive index**

In both cases, the refraction of the radiation as it passes through a crystalline medium can be neglected in a first order approximation, as the refractive index is very close to unity. For X-rays,

$$
\mu - 1 = - \frac{N e^2}{2\pi m(v - v_i)^2},
$$

where $N$ represents the number of electrons per unit volume of the substance, $v$ the frequency of the radiation and $v_i$ the frequency of the absorption line of the material nearest to $v$; $\mu - 1$ varies between $10^{-6}$ and $2 \times 10^{-4}$ for various substances. For electrons, we must take the mean electrostatic potential $\Phi$ of the crystal lattice into account; $\Phi$ is of the order of 10 V and the refractive index is effectively given by

$$
\mu = 1 + \varepsilon, \quad \text{where} \quad \varepsilon \simeq \frac{\Phi}{V},
$$

and $V$ is the accelerating potential. For the usual accelerating potentials, therefore, $\varepsilon \simeq 10^{-4}$. We have already encountered this refractive index in the section devoted to phase contrast (§19.4.3).
If we take this refractive index into account, the angle which is defined by the uncorrected Bragg relation

\[ \sin \frac{\theta_i}{2} = \frac{n \lambda}{2d_i}, \]

must be replaced by an angle \( \varphi_i \), such that

\[ \sin^2 \frac{\varphi_i}{2} = \sin^2 \frac{\theta_i}{2} - (\mu^2 - 1). \]

We shall see later on how this refraction can be made apparent and \( \Phi \) measured.

### 21.1.4 The Different Diffraction Techniques

Two arrangements are in use, in one of which we work with the transmitted beam, in the other, the reflected beam.

In the first case, the beam passes through a specimen which is in the form of a very thin film (a few hundred millimicrons or a few thousand Ångströms at most). The specimen is usually polycrystalline, with the crystals randomly orientated, and we need not worry greatly whether or not the beam is perfectly normally incident upon the preparation; if we are dealing with a single crystal, however, or orientated organic molecules, the angle of incidence must be defined exactly.

In the second case, the beam falls at grazing incidence upon the specimen, which can be either massive, or a thin film resting on a massive support. The mechanism of this phenomenon is the same as that of the preceding case, as in spite of appearances, this is once again transmission diffraction. So small is the penetrating power of the electrons that they are diffracted while passing through the microscopic surface irregularities. We can only collect half-rings, as the other halves are masked by the object.

Later on, we shall examine in detail the problems raised, in particular by transmission diffraction; the techniques are less refined and the instruments simpler in the reflexion case.

### 21.2 The Properties of a Diffraction Camera

#### 21.2.1 Resolving Power

(i) **Definition**

The object of every piece of research in this field is to measure the radii of the rings which are obtained on a photographic plate (or the distance between the spots in the case of monocrystals) and to separate closely neighbouring rings when the specimen is homogeneous.
It is natural to adopt the conventional definition of the resolving power of a recording instrument, which is given by the expression

$$\mathcal{R} = \frac{R}{\Delta R}$$

in which $\Delta R$ is the smallest distance between two neighbouring rings which can just be discriminated on the photograph. The ratio $S = \frac{\Delta d_i}{d_i}$ is often used as a definition of resolving power in terms of the lattice constants. $\mathcal{R}$ can be deduced from $S$ by taking into account the precision with which $L$ and $\lambda$ have been measured, as we shall show later. It should be noticed that $\mathcal{R}$ defined in this way is proportional to $R$ and $\lambda$ and inversely proportional to the lattice spacing $d_i$ and to $V$; we cannot give an estimate of $\mathcal{R}$ without quoting the values of $\lambda$ and $d_i$ to which it corresponds.

(ii) The general expression for the resolution

We consider the sketch of the most simple type of diffraction camera with a lens, shown in Fig. 264. The electron beam emerges from a source $S$ of radius $\rho_s$, passes through a weak lens $L_1$ which forms the image $S'$ of $S$ on a fluorescent screen (or a photographic plate). The specimen is placed immediately after the lens, a distance $L$ from the screen. The diffracted beams which correspond to lattice-planes of the same family also converge onto the screen, giving very fine rings. The thickness of these rings, $\Delta R$, depends upon two factors, very different in nature:

(a) the radius of the spot on the screen $\rho_s$ when there is no diffraction; this depends upon the quality of the optical system, and can be reduced at will by improving this system.

(b) a source of natural broadening due to the passage through the specimen, about which the experimenter can do practically nothing.

In modern instruments, it is finally this second phenomenon which limits the resolution $\mathcal{R}$ to values below $10^3$ whereas the optical system alone
is capable of giving $R = 10^4$. If $\varrho_s$ is the contribution of the specimen to the broadening of the spot, and $r$ the overall radius of the central spot

$$r = \varrho_s + \varrho_e.$$ 

The limiting distance between two rings which can just be separated is given by

$$\Delta R = 1.7 r,$$

if we assume that the intensity is distributed uniformly over the spot. If we round off the coefficient, and take

$$\Delta R = 2r,$$

we find

$$R = \frac{R}{2r} = \frac{\lambda L}{2r d_i}.$$ 

The distance between the specimen and the screen, $L$, can, as we shall see, be artificially increased by introducing electron lenses. $L$ will then represent the “equivalent length” of the instrument, which is defined by the ratio (see Fig. 265)

$$L = \frac{R_i}{\theta_i},$$

in which $R_i$ and $\theta_i$ are the values of $R$ and $\theta$ which correspond to planes separated by $d_i$; $L$ can be ten times the actual distance between the object and the plate.

![Fig. 265. Definition of the equivalent length.](image)

The quotient $Q = \frac{L}{2\varrho_s}$ is often known as the “quality factor” of the instrument—it depends only upon the optical part of the system, and gives a method of comparing different models. In the ideal case of a specimen which produces a negligible broadening of the spot ($\varrho_e \ll \varrho_s$) we shall have

$$R \approx Q \frac{\lambda}{d_i}.$$
(iii) The broadening of the spot which the specimen produces

This broadening is due to a host of phenomena, some of which are inevitable.

The first of these is provided by the perturbation to which the diffracted beam is subjected in passing through the specimen; the electrons may be deviated from their trajectories by elastic collisions, and in instruments where the specimen is followed by a lens or several lenses, electrons which have undergone inelastic collisions will be affected by chromatic aberration. These phenomena resemble the ones which were analysed in connexion with the formation of image contrast in microscopy (cf. Chapter 19).

The width of the ring also depends upon the size of the microcrystals involved, just as the width of the band which is produced when monochromatic light is diffracted at a grating varies inversely as the number of lines of the grating; in the electron case, the width is proportional to the number of diffracting planes, and hence to the width of the lattice, or finally, to the thickness of the crystal. As the Bragg angles concerned are very small, the effect of the transverse dimension \( t \) of the crystals is by far the most important. This enlargement term is effectively equal to

\[
\delta \approx 0.9 \frac{L\lambda}{t},
\]

and this formula is identical to the corresponding X-ray expression. Although this relation has been obtained for the case of a single crystal or for microcrystals all of the same size, it is equally valid as a statistical mean when the crystals are unequal in size, provided their dimensions are spread in a Gaussian distribution about \( t \).

Finally, the refraction of the diffracted beams which is due to the fact that the refractive index is slightly different from unity may produce duplication of the rings, which has the appearance of broadening if the resolution is inadequate.

There is virtually no way of reducing the effects of these phenomena save by an increase in the H.T. There are other defects, however, which can be considerably lessened by good experimental technique: for example, dispersal of the electric charges which have accumulated on the specimen, and a satisfactory distribution of microcrystals in the preparation. During the bombardment, the object becomes charged due to the inadequacy of the secondary emission which is produced by fast electrons; if the specimen is a good insulator, the effect is even more pronounced. These charges can often be eliminated simply by mounting the specimen on a conducting grid in the transmission arrangement, but with reflexion the specimen is often massive, and a special discharging device has to be incorporated into the instrument; an auxiliary gun sprays slow electrons (200 to 1000 V) over the whole surface of the object which is discharged by the abundant
secondary emission which results, as the multiplication coefficient is much
greater than unity, at low potentials. The ionization of the residual gas
also has a favourable effect. These charges, which vary both in number and
in distribution with time, deviate the diffracted rays by a mechanism exactly
the same as the electron scattering at the lattice. We must also prevent
the photographic film from becoming charged by placing it on a metal
support, although these charges are less harmful since their effect occurs
at the end of the trajectory.

Further, the width of the rings depends a great deal upon the preparative
technique. In reflexion, the surface must be perfectly polished so that the
dimensions of the residual irregularities shall be of the same order as the
depth of penetration of the electrons. In transmission, the thickness must
be uniform; if the specimen is in powder form, the distribution must be
as perfect as possible so that there are no opaque aggregates, and so that
every orientation is present, and hence complete diagrams will be obtained.
Examining the specimen in the microscope is a valuable help in obtaining
suitable specimens.

In conclusion, we should point out that for different substances, the actual
broadening varies widely with the degree of perfection of the crystals.
For example, magnesium oxide rings are about three times finer than the
rings produced by a layer of aluminium evaporated onto the same support
with crystals of the same size (Hillier and Baker, 1946).

(iv) The radius of the spot produced by the lens when the beam is per-
fectly monoenergetic

We call \( r_e \) the radius of the source \( S \) used, that is the radius of the cross-
over of the gun, and consider the most usual arrangement (Fig. 266); here the specimen is placed after the lens. The spherical aberration of this

\[
\triangle = C_s r_e^3 = C_s \left( \frac{r_e}{l} \right)^3,
\]

Fig. 266. Aberrations: the object follows the lens.
in which $\alpha_t$ is the semi-aperture of the beam which produces the rings, $r_a$ is the radius of the opening in the diaphragm, and $l$ is the distance from the diaphragm to the screen, which is the same as the distance from the lens to the screen. $l$ is usually in the vicinity of $2f_0$, so that

$$C_s \alpha_t^3 = \frac{C_s}{2f_0} \frac{r_a^2}{(2f_0)^3}.$$

$C_s/2f_0$ is of the order of 25 for a magnetic lens when $2f_0 = 20$ cm, and about 10 cm in diameter when we use an unshielded lens. The linear magnification $G$ of $L_1$ is always close to unity, so that

$$q_s = G q_c + C_s \alpha_t^3 \approx q_c + C_s \alpha_t^3.$$

![Fig. 267. Aberrations: the object precedes the lens.](image)

In some instruments, the specimen is placed before the lens $L_1$ (see Fig. 267). This is not a very good design, because extra aberrations appear, namely coma and astigmatism. The coma term, which is the more important, is given by

$$K \beta r_a^2,$$

in which $\beta$ is the angle at which the ring in question would be seen from the image nodal point of $L_1$, and $K$ is the coma constant. This term varies proportionally with the radius $R$, therefore, and hence this arrangement should be rejected.

We can always reduce the angle $\alpha_t$ to reduce the spherical aberration to a negligible value. When $f_0$ is of the order of 10 or 20 cm, $G = 1$ and $r_a = 0.1$ mm, the spherical aberration disc is of the order of a fraction of a micron in diameter; if the distance between the lens and the screen is 20 to 40 cm, the effect is negligible, and

$$\Delta R = 2\rho = 2\rho_c.$$

$q_c$ must be reduced to the uttermost; to do this, the source $S$ is obtained either by powerfully reducing the image of the crossover obtained with a simple triode gun, or more simply by placing a stop of very small diameter immediately after the anode of the triode gun. With the former, $q_c$ can easily
be reduced to below 10 \( \mu \). For \( L = 50 \text{ cm}, \lambda = 0.05 \text{ Å} \) (\( V = 60 \text{ kV} \)), and \( d_i = 1 \text{ Å} \), we should have
\[
R_i = \frac{\lambda L}{d_i} = 2.5 \text{ cm}
\]
when \( 2q_e = 5 \mu \), so that if we take the complete optical arrangement into consideration,
\[
\mathcal{A} = 5000 \quad \text{and} \quad Q = \frac{L}{2q} = 10^8.
\]
These conditions do not lead to a prohibitively large loss of brightness, and the exposure time remains of the order of a minute.

We can, therefore, find a good enough optical system for the resolution to be limited by the structure of the object itself. In certain problems, we wish to measure precisely the contribution of the object to the enlargement of the spot; in such a situation, we prolong the exposure and work in conditions such that the intrinsic diameter of the spot is negligible, which can give \( \mathcal{A} \approx 10,000 \).

(v) The radius of the spot provided by the lens when chromatic aberration is present

The accelerating potential may contain slight ripple; or it may display small, fast fluctuations \( \Delta V \) due, for example, to microdischarges. These defects result in two new enlargement terms.

The wavelength of the associated wave varies slightly, which produces an increase \( \Delta R_i \) in \( R \):
\[
\frac{\Delta R_i}{R} = \frac{\Delta \lambda}{\lambda} = \frac{1}{2} \frac{\Delta V}{V}.
\]
To obtain \( \frac{\Delta R_i}{R} \approx 10^{-4} \), the H.T. must be stabilized and the residual alternating component suppressed in such a way that \( \frac{\Delta V}{V} \leq 5 \times 10^{-5} \), which is at the limit of present-day possibilities. In the more usual situation, where a resolution of 100 to 200 suffices, it is easy to construct an H.T. supply such that \( \frac{\Delta V}{V} \) is of the order of \( 10^{-3} \), and hence, such that this first source of enlargement can be neglected. (We refer the reader to \$ 18.1 where information about H.T. sources is to be found.)

The magnification \( G \) of the lens varies with the electron velocity; for a magnetic lens, \( f \) is proportional to \( V \). A variation in \( f \) results in a broadening \( \Delta R_2 \). It can easily be verified that this term is far smaller than its predecessor.
In the case where the object lies before the lens, the two terms are of the same order of magnitude but of opposite signs. We shall later examine the attempts which have been made to obtain an arrangement which will be achromatic even for large variations of $V$.

21.2.2 Fidelity

As any ring is defined by the corresponding diffraction angle $\theta_i = \frac{R_i}{L}$, measurement of the radius $R_i$ of the ring must give the value of $\theta_i$ unambiguously, and hence the lattice-spacing $d_i$ also:

$$d_i = \frac{\lambda}{\theta_i} = \frac{\lambda L}{R_i}.$$ 

We must ensure that the quantities $\lambda$ and $L$ are stable over a period of time, and that the given experimental conditions can be attained exactly.

This fidelity is affected by a number of factors:

(a) variations in the potential, $V$, with time, and the immensely difficult problem of obtaining the same value of the potential $V$, in two different experiments separated by a fairly long interval; the latter, which is the more serious, may be the result of inadequately accurate measurement of the high tension, or of uncertainty about the value of $\lambda$.

(b) variations in the equivalent length $L$ which, as we shall see, depends upon the convergence of the lenses and of the optical system which follows the specimen in many instruments. Here again, the lens convergence must remain constant with the passage of time, and take the same value in two successive experiments if we apply the same magnetizing current. Nevertheless, hysteresis of the magnetic framework can produce slight variations. Similarly, an unexcited lens has a slight convergence as a result of the remanent field, the value of which may fluctuate with time.

In simple diffraction cameras, this objection does not occur in transmission, since $L$ and the distance from the object to the screen are equal (the object is placed immediately behind the lens). In reflexion, however, there is doubt about the position of the point of impact of the electron beam which is nearly tangential to the surface being examined. Considerable effort has gone into attaining this long-term fidelity, but although the excitation currents in magnetic lenses can be measured to within about 1 per cent it is very difficult to obtain measurements which are reproducible to better than 1 per cent. The fidelity is considerably better during a series of consecutive experiments throughout which the various parameters can be treated as constants. We can get round the problem by verifying the "constancy" of $C$, defined by $d_i = C/R_i$, before each important piece of research is undertaken, by looking at a known material which is perfectly stable in time.
The layers of evaporated gold and aluminium which have often served as test specimens in the past are in fact a very poor choice, as the lattice-spacing varies with the conditions under which the layer is prepared. Thin films of sodium chloride or lithium fluoride evaporated on collodion are to be preferred. We can also use an object-support designed to hold two specimens, and superimpose a part of each of the two diagrams which correspond to a known specimen and to the specimen being studied. An efficient though laborious solution is to supply two identical diffraction cameras working in parallel with the same potential, and to take the photographs of the known and the unknown specimens simultaneously, one in each of the two instruments. At each moment, therefore, we can determine the "constancy" of the instrument in question (Yamagushi, 1953).

In certain situations, where we are studying the structural modifications which a mechanical or thermal treatment produces on the specimen by examining the variations which result in the radii of the rings, a precise calibration of the apparatus is unnecessary as the form of the original diagram is known at the beginning.

21.2.3 The Accuracy of the Measurements

(i) The accuracy of the measurements made on the recorded diagram

This depends upon three quantities.

(a) The error in measuring $R$ on the plate; this can be reduced to a negligible amount after a suitable enlargement of the photograph. On some diagrams with poor contrast, in particular for the rings close to the overexposed central zone, a difficulty in determining $R$ arises. The distribution of brightness in the photograph is of the form shown in Fig. 268, in con-

![Fig. 268. The distribution of luminous intensity across a recording of the diagram, with an intense continuous background.](image-url)
sequence of the very intense continuous background; the radius of the ring no longer corresponds to the distance between the highest point of the peak and the axis, and the error increases as the radius of the ring is reduced. In this situation, we take the average of the distances from the points of inflexion $A$ and $B$ to the centre as the value of the radius. There is an advantage in moving the first rings away from the central zone, by using a projective lens which magnifies the diagram sufficiently. Experiment shows that the over-exposure of this central zone is due in fact to parasitic rays, and is hardly magnified at all by the lens.

(b) The error in $L$, which is appreciable in reflexion; with a good instrument, this error can be reduced to a few tenths of a millimetre, say to $\frac{\Delta L}{L} \leq 10^{-3}$.

(c) The error in $V$ which can be eliminated by bracketing the measurement on the unknown sample between two auxiliary measurements on a test-object. If this is done carefully, the length of the lattice-spacing can be measured to better than 1 per cent, with a good instrument; the ultimate precision which is obtainable in practice in favourable cases lies in the region of 0.5 per cent.

(ii) Measurement of the thickness of the rings

With this measurement, a supplementary factor intervenes—the thickness is a function of the exposure time, since the blackening of the photographic emulsion varies rapidly as a function of the incident charge $Q$. Figure 187 of Chapter 15 is an example of a photographic density curve. The exposure time must be measured carefully, and we must work on the linear part of the photographic density curve; the width of the ring is given directly. Here again the presence of a continuous background can introduce errors (Lafourcade, 1954) and make the measurement uncertain, and we have to employ a special procedure to reinforce the contrast of the rings (Brockway and Bartell, 1954).

(iii) Electron intensity measurements

The intensities of the various rings or spots can be measured from photographs, but the absolute accuracy is poor, especially when the peaks are superimposed on a strong continuous background (5 or 10 per cent). It is virtually impossible to compare the intensities of the diffracted beams with that of the direct beam using this method, since the central zone of the diffraction diagrams is violently over-exposed. Furthermore, the continuous background caused by electrons scattered inelastically at the object is often troublesome.
The electron current can be recorded at every point of the diagram either by moving a Faraday cage provided with a very small entry diaphragm across the region where the diagram is produced or, alternatively, by shifting the actual diagram by means of deflexion coils over a slit or hole cut in the observation screen. The current can then be measured with a Faraday cage or a crystal scintillator followed by a photomultiplier, and the current density along a diameter (or a chord) of the diagram is then recorded. Here again, the continuous background causes difficulties. It is possible, however, to reduce this parasitic current to a very large extent by placing a filter lens between the diaphragm and the measuring device, which will allow only elastically scattered electrons to pass (Denbigh and Grigson, 1965). If the blocking potential is varied, it becomes clear that the rings (or spots) consist of two classes of electrons: elastically scattered electrons and a certain fraction of electrons that have lost energy inelastically (and in particular, those that have undergone one of the characteristic losses). Despite this inelastic scattering, the latter have therefore retained some degree of coherence. Filtering thus enables us to reduce the width of the peaks and to increase the resolution and accuracy of the traces on the photographs; in particular, we reduce the error in determining the size of the micro-crystals by diminishing the uncertainty in the width of the peaks.

21.2.4 The Interest in Using Very High Tension

In conclusion, we shall summarize the reasons why very high tension is of interest in diffraction.

With a very high tension, we can, first of all, examine thicker specimens (the majority of substances in powder form which are used in industry have grains larger than a micron in size).

The intensity of the continuous background of the diagram is decreased as fewer electrons are scattered.

As a result of the high penetrating power of the electrons, the specimen can be examined at even more closely grazing incidence. We can thus examine monomolecular layers.

Finally, the enlargement term of the rings which depends upon the transverse dimension of the crystals and which is proportional to \( \lambda \) decreases when \( V \) is increased, as does the broadening due to refraction.
21.3 SOME PRACTICAL EXAMPLES

21.3.1 General Summary

After thirty years' evolution, specialized instruments using this technique have been developed in all the important fields of research.

(i) The search for high resolving power

The earliest diffraction cameras had no lenses and were faithful replicas of their X-ray counterparts. An extremely fine electron beam was obtained with the aid of an electron source (hot or cold) and a long collimator containing a series of diaphragms with very small holes (a few hundredths of a millimetre); the beam passed through the specimen, after which the diffracted beams struck the plate a distance \( L \) away.

The search for ever higher resolutions led to the use of smaller holes which produced an inconvenient diminution of the brightness, and exposures which were far too long. No other way of maintaining the degree of accuracy was available, except a considerable tightening of the tolerances on the variations of \( V \). The development of electron optics has provided a way out of this dilemma; the designers have introduced a focusing lens with which the current density incident upon the plate can be considerably increased, and a real image of the source \( S \) can be formed upon it. The source at first consisted of a small hole in a diaphragm placed directly behind the gun, but later a powerful condenser behaving as a reducing lens was introduced, and with this a still finer source was obtained.

(ii) Direct magnification of the rings and accentuation of the contrast

The resolving power can also be increased by lengthening \( L \); further, high values of \( L \) are necessary if the rings of small radius which correspond to large values of \( d_i \) are to be examined conveniently (these large values are encountered in the interesting case of organic crystals). We can, however, scarcely make the instrument longer than 60 cm without considerably slowing down the pumping time whenever the specimen is changed, and making the magnetic screening difficult. By placing a supplementary lens \( L_2 \) after the specimen, however, we can eliminate this inconvenience, and give \( L \) a value which can sometimes be longer than two metres. In Fig. 269, the lens \( L_1 \) forms an image of \( S \) and hence forms a diagram \( d \), near the object focal plane of \( L_2 \); the latter forms a magnified image, \( d_2 \), of \( d \) on the screen. In these conditions, the equivalent length is far greater than the actual object–screen distance. By altering the convergences of \( L_1 \) and \( L_2 \) it is clear that \( L \) can vary between wide limits.
This arrangement does not improve the resolution, however, since $L_2$ increases the radius of the spot as well as the radius of the rings—the resolution $\mathcal{R}$ obtained is governed by $L_1$. The essential advantage is that the rings close to the centre of the over-exposed central halo, which is produced by the direct beam and the electrons scattered by the object, are moved away outwards. Many of the scattered electrons have suffered considerable losses of velocity and as a result of the chromatic aberration of $L_2$, they are spread out, which reduces the density of the continuous background at the centre; on the other hand, the diameter of the over-exposed zone is practically independent of the size of the spot, and even if the latter is made larger the zone remains the same size—the reasons for this are poorly understood.

A large proportion of the retarded electrons can be removed with the aid of a velocity filter (cf. § 19.4) which increases the contrast on the diagram considerably. This filter lens also increases the resolution, as it reduces the enlargement term due to the object itself by removing the electrons which have been retarded after selective reflexion, and which would not be focused at exactly the same point as electrons which had not been slowed down.

(iii) *The microscopy of the specimen*

The region of the specimen which is illuminated can be made very small, and we can for example pick out a single crystallite in a heterogeneous body. We can thus make a chemical and a crystallographic analysis of the specimen. It is extremely useful to be able to chart accurately the crystallites of which the specimen is composed, or to note the appearance of the particles by which the diagram is being formed, in order better to recognize the various components of a mixture. We must find a way of allying the two techniques of microscopy and diffraction. This can be done in two ways. Either we build a high resolution diffraction camera, equipped with the necessary accessories for examining the object by shadow microscopy, or we use an electron microscope which can be transformed into a high quality diffraction camera; this type of instrument can only rival
the other type described above if the traditional position of the object in the optical system is changed, and if the structure is modified until it is closely similar to that of a diffraction camera (cf. Duncumb, 1962).

This solution has the advantage of allowing us to obtain genuine micro-diffraction, that is, to obtain the diagram of a microscopic zone of the specimen with a good resolving power. This allows us to identify the different phases in the case of an object which is finely heterogeneous (a thinned down metal alloy, for example).

21.3.2 High Resolution Diffraction Cameras

Some typical models are described in the articles by Hillier and Baker (1946), Picard, Smith and Reisner (1949), Allen, Mackle and Sutton (1951), Le Poole (1947) and Lees (1954). The more specialized instruments described by Le Poole (1954) and Corbet et al. (1955) will be examined later. The pumping systems, the electrical supplies for the gun and the lenses, and the design of the specimen-holder and of the camera with which the image is recorded are closely similar to the corresponding parts of a microscope, which have already been described; we shall only elaborate on the general arrangement of the optical system and its possibilities, and the features peculiar to the diffraction. We should mention that the potentials employed very often exceed 50 kV, which suggests that magnetic lenses should be used. The reader is referred to Chapters 17 and 18.

(i) The electron source

The working potential is selected between 50 and 100 kV—only a few special instruments operate at very high voltages, 100 to 400 kV or even 1 MeV (Papoular, 1955). The source is normally a heated filament, but may also, more rarely, be a cold discharge. The gun resembles a three-electrode microscope gun, with autobias to increase the stability of the output. The “point” source can then be obtained in two ways.

In simple diffraction cameras, an anode diaphragm with a hole of 50 to 80 μ is sufficient for extremely fine rings to be obtained, provided that it is followed by a weak lens with magnification close to unity (Lafourcade, 1954).

To increase the resolution, one or two powerful reducing lenses are placed in front of the gun, and with these a genuine electron probe about a micron in diameter can be obtained; this reduction of the diameter results in a considerable increase of the angular aperture of the emergent beam; suitably placed stops eliminate any electrons which have been too widely deviated and which might well strike the walls and the pole-pieces of the lenses, and after reflexion, alter the appearance of the diagram. The diaphragms
do, however, reduce the brightness considerably, and a gun capable of a larger output (100 μA) becomes a necessity.

To give some orders of magnitude, we quote the results obtained by Castaing (1951). With a suitable optical system, we can easily attain an intensity of $10^{-7}$ A in an electron probe one micron in diameter, when the angular aperture of the beam is of the order of $10^{-2}$. If we reduce this aperture to $10^{-3}$ by placing a diaphragm in front of the diffraction lens, the intensity which reaches the central spot on the screen falls to about $10^{-9}$ A. When the object is present, the maximum diameter of the spot is 100 μ and the thickness of the rings is of the same order. The current density in the rings (or spots) again far exceeds $10^{-11}$ A cm$^{-2}$, which is the value which corresponds to a highly magnified electron microscope image, with an exposure time between a half and one minute.

(ii) The optical system itself

Several different lens systems have been proposed, but they can all be traced back to two basic combinations:

(a) a so-called “diffraction” lens $L_1$ forms the image of the source directly on the screen. This lens is very weak (focal length between 10 and 25 cm) and is often placed mid-way between the source and the screen. The object is placed after the lens, both so that the process of focusing does not affect the size of the rings, and to reduce the aberrations. The aperture is reduced by a factor of several thousand, by a diaphragm which precedes $L_1$, and the distance between the specimen and the screen, $L$, can vary between 25 and 70 cm (see Fig. 267).

(b) alternatively, the lens $L_1$ is followed by a projective lens $L_2$, which gives a magnified image $d_2$ of the small diffraction diagram $d_1$ which is produced by $L_1$ (Fig. 269). The size of $d_1$ depends upon the distance between the object and $L_2$; if this is too short, $L_2$ must be a strongly convergent lens, but the aperture of the beams must be reduced to prevent the spherical aberration from being too large. If the diagram is large, other aberrations also intervene, namely coma and astigmatism. To obtain the optimum resolution, we must find a compromise. $L_2$, however, must always be perfectly corrected for distortion, and it is preferable that it should consist of two lenses in tandem wound in opposite senses, to annul the rotation of the diagram.

The first system, which is the simpler, always gives the better resolution (Hillier and Baker, 1946). When the final spot is a few microns across, the resolution can in theory reach 10,000 (Tanaka and Hashimoto, 1953). A diaphragm, with an aperture of at most 50 μ, is also required. In the majority of cases, however, a diaphragm of about 0.1 mm, which gives higher brightness and leads to an exposure time of 10 or 20 sec, will be adequate. The
second arrangement, as we have already mentioned, gives us a means of varying \( L \). We have

\[
L = \frac{ab - b^2}{f} - af
\]

with the notation of Fig. 270b. By altering \( f \) and readjusting \( L_1 \) for the focusing, we can make \( L \) vary between 0 and 2 m, for example, for values of \( a \) between 25 and 50 cm.

(iii) **Shadow microscopy**

On considerably increasing the convergence of \( L_1 \), a very fine electron probe is obtained in the vicinity of the specimen (Fig. 270a). When \( L_1 \) is the only lens, a highly magnified shadow of the zone which is illuminated

![Fig. 270a. Shadow microscopy; the principle.](image)

is obtained directly, and the magnification is given by the ratio between the distances between the object and the screen and the object and the probe. \( G \) could represent an infinite magnification if the probe were a point, but the effect of the penumbra which is produced by the finite extent of the beam limits the useful magnification. If we neglect the spherical aberration, the resolution at high values of the magnification is bounded by the presence of the diffracted rays. At every point on the object, the incident ray is transformed into a conical pencil of semi-aperture \( \theta_i \) (for \( \theta_i \), we can take
the angle which corresponds to the last very intense ring on the diffraction diagram of the object; the lattice spacing is then \( d_j \). The size of the diffraction spot is \( \delta_i = 2a \theta_i = \frac{2a \lambda}{d_i} \), where \( a \) represents the distance between the object and the screen (Fig. 270b). It is necessary that the size of the penumbra \( \delta_s = 2 \theta_s \frac{a}{b} \) should be less than \( \delta_i \) (\( b \) represents the distance between the probe and the object, and \( \theta_s \) the radius of the probe). This means that

\[
2 \theta_s \frac{a}{b} \leq \frac{2a \lambda}{d_i},
\]

or

\[
G \leq \frac{a \lambda}{\theta_s d_i}.
\]

For \( a = 20 \text{ cm}, \ d_i = 2 \text{ Å}, \ \text{and} \ \lambda = 0.05 \text{ Å}, \) for example, we obtain \( G \leq \frac{5 \times 10^2}{\theta_s} \) (\( \theta_s \) measured in microns).

If we include the effect of spherical aberration, which distorts the image, we find that the limit is of the same order. In the best instruments, the resolving power scarcely exceeds 200 Å at 50 kV with a magnification of 10,000 and a probe about 100 Å in radius. For most cases, this is adequate, the image being simply a means of control. When we have a two-lens system available, it is better to move the probe further away from the object and obtain a high magnification by exciting \( L_2 \).

(iv) *Shadow microdiffraction*

We dispense with \( L_2 \) and vary the excitation of \( L_1 \) continuously between the values which correspond to Figs. 267 and 270; if a diaphragm with a very small aperture (20 to 50 \( \mu \)) precedes the lens, we pass through a series of intermediate stages (Fig. 271) which are used to obtain the diagram of very small regions of the specimen, and in particular, of individual crystallites. The size of the spot at the screen increases in proportion as the illuminated area decreases, however, and the resolving power falls extremely rapidly. With the notation of Fig. 271, we have

\[
d_0 = 2r_a \left(1 - \frac{l'}{v}\right) + 2\theta_c m'_1,
\]

\[
d'_a = 2r_a \left(\frac{l + l'}{v} - 1\right) + 2\theta_s,
\]

in which \( m'_1 \) is the magnification of \( L_1 \), \( \theta_c \) is the radius of the source \( S \), \( \theta_s \) is the radius of the spot focused on \( E \), and \( r_a \) is the radius of the aperture diaphragm.
When \( g_s \) is very small, and the values of both \( g_c \) and \( g_s \) are a few tens of microns, it is easy to obtain an illuminated area of a few tenths of a micron with a resolving power of the order of 10 or 100 (Hillier and Baker, 1946); this is often sufficient to examine spot diagrams. The central spot encloses the direct shadow image of the illuminated zone, and the diffraction spots are images of the same zone against a dark background.

(v) The specimen-holder and other accessories

If a diffraction camera is to operate both in transmission and in reflexion, the specimen-holder must be highly refined; it must be mobile in every direction and rotatable in several ways so that we can adjust the angle of incidence of the beam easily by tilting the object. There are still more complicated specimen-holders which allow the specimen to be heated or cooled, over a temperature range of \(-180\)°C to \(1000\)°C. This is necessary in metallurgical studies.

The observation chamber has, therefore, to be large enough to contain the numerous devices with the aid of which the object can be subjected to various mechanical and thermal treatments:

(a) the mechanical apparatus which is required for cutting, grinding and polishing under vacuum;

(b) an ion gun with which the surface can be slowly etched and progressively cleaned; this is nowadays a standard technique (Trillat, 1954; Fert, 1954);

(c) ovens, for preparing thin films of various substances by evaporation;

(d) a gun to provide slow electrons for discharging the specimen. A gun of this kind has to provide current densities of the order of 0.4 to 5 mA cm\(^{-2}\), without contaminating the surface of the specimen.

(vi) Recording the diagram

The plate cameras in everyday use resemble those which are used in microscopy, but recently, a recording technique has been developed with
which the evolution of the specimen can be filmed during any kind of treatment (Boettcher and Thun, 1954; and Trillat, 1954). The Debye diagram is formed on a screen in which a very fine diametrically placed slit has been cut, beneath which a photographic film slowly passes. A ring is recorded as a pair of straight lines; as soon as the structure of the specimen alters, the distance between the lines varies and new lines appear, the intensity, width and position of which vary with time, depending on the crystallographic modifications of the specimen.

The electron density distribution in the diagram (see §21.2.3) can also be recorded by invoking all the resources of modern electronics for measuring low currents (of the order of $10^{-15}$ A with scintillators and low-noise particle photomultipliers with no windows, followed by counters). Either the measuring device is moved across the diagram (Lennander, 1952) or the diagram is shifted electrically across a diaphragm in front of the measuring instruments (Goodman, 1961). If this scanning is very rapid, the current density curve can be fed into an oscilloscope directly; this is the principle underlying the scanning diffraction devices (see §21.3.3), which thus allow one to observe very rapid changes in the structure of the specimen.

(vii) Recent efforts: pulsed diffraction cameras

Attempts have also been made to render the lenses of a diffraction camera nearly achromatic; in this way, we should hope to obtain high resolution without having to eliminate the fluctuations in the H.T. supply which could be tolerated up to a relative value of 5 or 6 per cent. Lenses which are insensitive to big variations in potential have been described in the literature (Cosslett and Jones, 1950; Le Poole and Salvat, 1950; Bernard and Salvat, 1954; and Le Poole, 1954). In the simple case of a system with a single lens, the principle is as follows—the specimen is illuminated in parallel light with the aid of a suitable system of condensers, and the lens is placed after the object (Fig. 272a). If the accelerating potential increases by $\Delta V$, the diffraction angle $\alpha$ is decreased, and the corresponding rays intercept the lens closer to the axis. The increase $\Delta V$ will also produce an increase of the focal length of $L_1$, so that the angle $\beta$ through which the beam is deviated by $L_1$ will be smaller. For certain values of $f, a$ and $b$, these opposed effects exactly compensate one another. We have the relations:

$$\beta = \frac{h}{f} = \frac{a \lambda}{f}, \quad R_2 = R_1 - \beta b = \lambda(a + b) - \frac{\lambda a b}{f},$$

but

$$\alpha = \frac{\lambda}{2d} = \frac{C_1}{V^{1/2}} \quad \text{and} \quad f = \frac{C_2}{V}.$$
That the effects balance out is expressed by $\frac{dR_2}{dV} = 0$, so that the achronomatism condition is

$$f = \frac{3ab}{a+b}.$$  

But $R_2 = \frac{2}{3} R_1$ and the sensitivity is reduced.

Fig. 272. An achromatic diffraction camera. The principle of (a) the single-lens situation and (b) the two-lens case.

We can establish relations which are analogous, though more complicated in form, for the two-lens case. These are given in detail in the article by Corbet et al. (1955). By a judicious choice of the geometrical parameters of the arrangement of these lenses, we can not only suppress the rotation of the diagram, but also make the distance $L$ partially independent of the position of the object and cancel the distortion for the mean operating potential $V$. Since the distortion reappears for potentials on either side of $V$, the lenses must be chosen in such a way that this aberration is very small. The distortion, which has the effect of making the outer rings too large (cushion distortion), also broadens these rings when we are working with a variable potential. A magnetic system which fulfils these specifications and, in addition, makes the radii of the rings to some extent independent of the variations in the excitation of the lenses has been studied by Le Poole (1954).
Another achromatic two-lens arrangement (Fig. 272b) is equipped with an instrument (Corbet et al., 1955) with which the specimen can be studied simultaneously by microscopy. A diffraction camera of this kind is illustrated schematically in Fig. 273. There are three diffraction lenses with which either direct or magnified diagrams can be obtained. The equivalent length can vary between 65 cm and 450 cm, while the distance between the object and the screen is 60 cm. With the two condensers, the diameter and the parallelism of the incident beam can be adjusted; by altering the convergence of these condensers, and thanks to the possibilities introduced by the two lenses \( L_4 \) and \( L_5 \), shadow microscopy is possible with a magnification up to 5000. The gun is supplied with pure alternating potential, but the grid is autabiased in such a way that only a small fraction of each period in the neighbourhood of the maximum (100 kV) reaches the gun. The pulses are sufficiently narrow for the potential variation during each pulse, \( \Delta V \), to be of the order of 5 or 6 per cent. The currents are measured to about one part in \( 10^3 \); the equivalent length is independent of the position of the object provided the axial displacement of the latter is not more than \( \pm 5 \text{ mm} \). In these conditions, the resolution is of the order of 1000, and the accuracy with which \( d_i \) can be measured reaches 0.5 per cent, which is equal to the accuracy of the best ordinary instruments.

21.3.3 The Scanning Electron Diffraction System

This apparatus, which was suggested by Grigson (1961, 1962), contains no optical elements beyond the specimen; with a system of deflexion coils, the diagram can, however, be moved rapidly (the maximum velocity corresponds to \( 50 \times 10^{-3} \text{ s} \)) across a hole 20 or 50 \( \mu \) in diameter cut in the fluorescent screen, behind which is a photomultiplier. The electron density distribution along a diameter is transmitted to an oscilloscope and can be photographed. With an auxiliary Faraday cage, the curves so obtained can be calibrated.

In a considerably improved version, Denbigh and Grigson (1965) have placed a retarding field filter behind the slit. If the filter is supplied with
Fig. 273a. Diagram of the scanning electron diffraction unit (1) with magnetic energy analyser; (2) with electrostatic energy filter.

A steady retarding voltage that can be adjusted in magnitude, electrons that have lost large amounts of energy can be progressively eliminated so that finally, only elastically scattered electrons are collected. If, furthermore, the voltage of the high tension electrode of the filter is modulated with a weak alternating potential, we obtain a small modulated component in the transmitted current, corresponding to electrons lying within a band of energies of width $\Delta V$ centred on a given energy $V$ (which can be shifted across the whole spectrum by altering the steady voltage, $V_F$, at the filter). The diagram corresponding to a given energy will then be obtained by using an a.c. amplifier at the receiver. This electrostatic filter, which is not simple to use in practice, can also be replaced by a magnetic filter (Grigson, 1965), consisting of a deflecting magnet followed by a thin slit.
Fig. 273b. Growth of an 80–20 Ni–Fe film on carbon. Successive contours recur at thickness increments of approximately 10 Å. Sweep frequency 0.025 c/s; 38 kV electrons, two intensity scales (Grigson, 1965).

Fig. 273c. Diagram from a thin film of aluminium: improvement of contrast and resolution by velocity filtering (1) diagram without filtering; (2) all electrons that have lost more than 18 V are eliminated (so that electrons that have suffered the 15 V characteristic loss are transmitted); (3) the loss barrier is now raised to 3 V; the continuous background has virtually disappeared (Denbigh and Grigson, 1965).

and a photomultiplier. Just as in the preceding case, only those electrons corresponding to a predetermined energy $V$ in the whole diagram will be collected at the photomultiplier.

With this scanning electron diffraction instrument, rapid alterations in structure can be followed, or the continuous growth of a thin film evaporated inside the instrument. Figure 273a illustrates the design of the device and Fig. 273b is an example of the diagram obtained during the growth of a Ni–Fe film on a carbon layer; Fig. 273c shows an example of contrast enhancement by electrostatic filtering.
21.3.4 Diffraction in an Electron Microscope

The high quality of the optical system and the electrical supply in electron microscopes might lead us to hope for a good resolving power when the instrument is used as a diffraction camera. The demands which are made upon the optical system in microscopy and in diffraction are rather different, however, and it often awkward to transform a very good microscope into a good diffraction camera. The introduction of supplementary lenses which are used in diffraction and not in microscopy, and *vice versa*, has proved to be indispensable in instruments which are to serve both ends (and which are known as "universal").

At first, the microscope was transformed into a diffraction camera by moving the specimen, but the present tendency is to transfer from one mode of operation to the other without touching the specimen. Nevertheless, if we require pictures with very high resolution ($\mathcal{R} \geq 1000$ for example), the only solution is to fall back upon the arrangements peculiar to diffraction cameras, and to a specialized instrument.

(i) *Transformation of a microscope into a diffraction camera*

In a two-lens microscope (von Borries and Ruska, 1940, and Picard and Reisner, 1946), a diffraction camera can be produced by removing the pole-pieces of the lenses and introducing an extremely small diaphragm into the condenser, all without disturbing the object. The projective lens $L_2$, excited

![Diagram](image)

Fig. 274. Diffraction in an electron microscope with a mobile specimen-holder. (a) The instrument as a microscope. (b) The instrument as a diffraction camera.
alone, may serve as a diffraction lens, but the combination of $L_1$ and $L_2$ is equally practicable. The necessity of removing the pole-pieces arises from the fact that we must avoid intercepting the outermost diffracted beams.

In other instruments, the projective lens $L_2$ is followed by a unit which is designed specially for diffraction work, and contains a second specimen-holder $P_2$ and a weak diffraction lens $L_3$. When the second specimen-holder is not in place, the instrument behaves as a microscope (Fig. 274a) and the lens $L_3$ is switched off. To obtain a diagram, the object has to be withdrawn from the instrument and placed on $P_2$. The objective $L_2$ then produces a fine source, while $L_3$ gives the desired diagram on the screen (Fig. 274b). Shadow microscopy too is possible (Hillier, Baker and Zworykin, 1942; and Rühle, 1950).

(ii) The direct passage from microscopy to diffraction; high resolution micro-diffraction

In the majority of recent microscopes, we can pass from one mode of observation to the other with comparatively few operations, namely, variation of the excitations of the lenses and of the apertures in the stops, which can be effected from outside. Van Dorsten et al. (1950) describe an instrument which has a condenser $L_0$ and four lenses. The microscope proper works with three lenses (Fig. 275). As we have explained in §15.5, $L_1$ is the objective and $L_3$ and $L_4$ are projective lenses; $L_2$, which is weak, is used only in diffraction, to image the diagram produced by the objective $L_1$ in front of $L_4$. The 40 $\mu$ contrast diaphragm $D_1$ which is situated behind $L_1$ is replaced by a 1 mm diaphragm which does not intercept the diffracted rays. In the plane $P'$, where the intermediate image of the object is formed, a zone of any desired area can be isolated with the aid of a variable stop $D_2$. If $L_1$ has a magnification of 100, we can select a 1 $\mu$ zone of the preparation with a 100 $\mu$ aperture. From Fig. 277a, where the electron paths through the system are shown, we can see that the rays which produce the diagram really are those which correspond to the portion of the
Fig. 276. Microdiffraction (Photos: Fert). *Above*: the image of the zone being studied. *Below*: the corresponding diagram.
object defined by this aperture (Fert and Ito, 1955). An example of this remarkable technique is given in Fig. 276. Unlike the shadow micro-
diffraction case, there is no loss of resolution when the illuminated area is reduced. The diaphragm $D_2$ usually consists of two interlinked parts which describe a square hole (Fig. 277b). The area can then be varied simply by moving the two parts relative to one another. A high degree of mechanical precision is always necessary.

To obtain high resolution, the condenser must be readjusted. In general, the passage from microscopy to diffraction is possible in all three-lens microscopes. The third lens $L_3$ works at a fixed excitation in conditions of minimum distortion, and the convergence of the weak second lens $L_2$ can be varied between wide limits (Grivet and Regenstreif, 1950; Tanaka and Hashimoto, 1953). Instead of forming a second image of the object still further magnified with $L_2$, we simply reduce the convergence considerably in order that the image $d_2$ of the diagram $d_1$, which is formed by the objective close to its image focal plane, shall be formed near the object focus of $L_3$. This mode of operation has been described in § 15.5. Under these conditions, the magnification of $L_2$ is always less than or equal to unity, and $d_2$ is small enough to be magnified entirely by $L_2$ working normally. The aperture of the diffracted beams is, however, larger than the aperture of the beams which originate in points on the second intermediate image of the object in microscopy, and the various aberrations (spherical aberration,

![Diagram](image)

**Fig. 277.** Microdiffraction. (a) The ray paths. (b) An example of a field dia-
phragm.
coma, ...) are accentuated. We have thus to make every effort to obtain an extremely fine source \( S \) in front of the object, either by introducing a very small diaphragm if the object is illuminated directly by the beam which is emitted by the gun, or by a very different adjustment of the condenser if such a lens already exists.

(iii) The reflexion diffraction camera used in conjunction with an emission microscope

This combination has the same advantage in reflexion as microdiffraction in transmission, namely, that the spot can be very exactly localized within an extremely narrow and well-defined region of the object, together with the added advantage that both the diffraction diagram and the surface being studied can be examined simultaneously on two separate fluorescent screens.

The instrument is shown diagrammatically in Fig. 277 S 1. The electron beam of the diffractograph falls on the surface under examination at almost grazing incidence, and produces the picture which corresponds to the illuminated region on the lower fluorescent screen. The object forms the cathode of an electrostatic immersion objective with its axis horizontal, which produces a magnified image of the surface on a second fluorescent screen by means of the electrons which are emitted by the surface either by thermionic emission or by secondary electron emission from an auxiliary ion beam. The diffraction camera beam is stopped right down and thus illuminates only a very small area of the surface; at the same time, a few slow secondary electrons are liberated, which are collected by the immersion objective and give rise to a bright spot on the image of the surface. By shifting the object, we can check the position of the spot and bring it exactly onto the detail which is to be examined.

A device of this type has been built by Arnal (1958 a, b). The accompanying thermionic emission microscope is founded on the instrument described in \( \S \) 13.4; at a distance of 50 cm from the object, it gives an image magnified 70 times. The object is earthed in order to avoid disturbing too radically the diffraction camera beam which crosses the cathode–grid gap of the immersion objective. The anode of the latter, together with all the interior of the surrounding chamber and the fluorescent screen, is held at an accelerating potential of 30 kV.

The diffraction camera beam is only slightly disturbed by the relatively small accelerating field of this type of objective; to reduce both the cross-section at the level of the plane in which the object lies, and the elongation (due to the grazing incidence) of the zone which is illuminated, the beam is stopped down by means of a slit, \( 100 \times 2 \mu \). The spot at the object can be reduced to \( 8 \times 32 \mu \) in this way without the brightness of the diagram becoming insufficient.
Fig. 277 S1. Diagram of the optical system of a composite instrument, diffraction camera and emission microscope (Arnal). G: gun; C: condenser; O: objective; P: projective; S: slit; D: diaphragm; Sp: specimen; E₁: fluorescent screen and camera (diffraction diagrams); Ob: immersion objective; E₂: fluorescent screen (image of the surface).

This is a most valuable instrument for the metallurgist, and it has in particular been used to study the phase transformations in titanium, near 950°C, after a lengthy heating at 900°C. The crystalline structure of a specimen is almost always perfect in these conditions, and we can examine a diagram which displays Kikuchi lines. The size of the spot is much less than that of the titanium crystals, and can easily be directed onto the zones which are in the process of transformation (Fig. 277 S2). Similarly (Sorel, 1959, 1962), we can follow the progress of surface oxidization, or analyse the epitactic coating on a massive specimen.

21.4 SOME APPLICATIONS

We shall simply review these briefly; for further information the reader is referred to the bibliography established by Rymer (1953).
Fig. 277 S2. Examples of the application of the above instrument. (a) The surface of a specimen of α-titanium; the bright spot (top left) shows the position of the spot of the diffraction camera \((G = 70)\). (b) The diagram corresponding to the crystal struck by the beam of the diffraction camera (Kikuchi lines), [021] axis. (c) The surface of a specimen of β-titanium \((G = 70)\). (d) Diagram corresponding to the central crystal for the preceding specimen, [311] axis (Photo: Mrs. Sorel).
21.4.1 Geometrical Crystallography

(a) Determination of the lattice constants, with the possibility of identifying the sample. The latter may be a very thin layer at the surface of an object being observed by reflexion (a layer of oxide or of lubricant, or a polishing layer for example).

(b) Measurement of the changes of the lattice constants during a mechanical deformation of the specimen.

(c) Information about the type of deformation which has occurred (slip with rotation, for example) can be obtained from the diagrams of mono-crystals (Wilman, 1951).

(d) The size of the crystals can be calculated from measurements on the thickness of the rings (Lafourcade, 1954), and the form of individual crystallites can be obtained from the shape of the spots on the diagram.

(e) Kikuchi reflexion diagrams are produced only when the crystal lattices are perfect; this gives a means of estimating the "finish" of a surface, and its degree of order.

(f) Determination of the structure of organic crystals, and of the positions of the large molecules in a crystal—the latter may display preferred orientations.

21.4.2 The Study of the Internal Potentials of Crystals

The mean electrostatic potential within a crystal differs from the electrostatic potential of the surrounding vacuum by about 10 or 15 V. This potential difference produces a refraction of the electron rays, the refractive index being slightly different from unity. This refraction can easily be seen when we work with reflexion, as the diffracted beams may be displaced several millimetres away from the position which the simple theory predicts. In transmission, with an instrument capable of very high resolution, it has been possible to detect a duplication of the rings as the diffracted rays are refracted either one side or the other of the Bragg angle, according to the orientation of the faces of the microcrystal at which the rays enter and leave. The distance between the two "components" of a ring does not in general exceed 100 μ for $R \sim 1$ or 2 cm. By studying this separation, we can obtain data about the value of the internal potential.

Onto the mean potential is superimposed a spatially varying potential which is closely related to the crystal lattice. The amplitude of this variation and the thickness of the crystals can be determined by making measurements on the interference fringes produced by neighbouring diffracted rays. We can obtain either fringes of equal inclination, by illuminating the crystal with a convergent beam (shadow micro-diffraction), or fringes of equal thickness which appear when the crystal is irradiated with a parallel beam. The reasons for the existence of these fringes have been explained in Chapter 19.
CHAPTER 22

ELECTRON PROBE DEVICES

22.1 ELECTRON PROBES AND MICRO-PROBES

The name "electron probe" is given to the real image of a real electron source, and in particular, to the crossover which lies at the exit of an electron gun. The simplest example is the spot of a cathode ray oscillograph, but, except in extremely unusual circumstances, the diameter of this spot is not less than a few tenths of a millimetre.

The progress which has been made in electron optics during the last twenty years has been such that we can now obtain very small probes indeed, which may be as little as a few hundred Ångströms in diameter; these are often known as "micro-probes". By operating an electron microscope in the opposite sense, we could clearly obtain a probe with a diameter of the same order of magnitude as the resolution of the instrument. The current in such a probe would, however, be inadmissibly tiny (10^{-14} to 10^{-16} A, for example). The search for high brightness in the probe is always accompanied by efforts to compress the greatest possible intensity into the smallest possible diameter; the reasons for this will be apparent when we come to consider applications later in this chapter.

In the first section, which is devoted to a type of instrument which is particularly useful in metallurgy, the micro-probe analyser, we shall describe the general method by which a micro-probe may be obtained, the problems encountered, and the inherent limitations. Next, we shall give a brief description of two other types of micro-probe apparatus: the X-ray projection microscope and the scanning microscope. Finally, we shall say a few words about some recent techniques in which the kinetic energy carried by an intense electron beam is used to work metals; a very small intense probe may be used for drilling, and probes of several kilowatts for welding or melting the refractory metals, under vacuum.

22.2 MICRO-PROBE ANALYSERS

22.2.1 The General Principle

An electron probe device, with which an almost point-by-point chemical analysis of a massive metal specimen can be carried out without destruction of the latter, has been studied and built by Castaing (Castaing and Guinier, 752
1949, 1950 and 1951; Castaing, 1951, 1954 and 1955). The impact of electrons onto any kind of anticathode results in X-radiation which contains the X-rays which are characteristic of the various elements from which the anticathode is formed. A spectrographic analysis of the X-rays which are emitted therefore gives us information about the chemical composition

Fig. 278 S1. Castaing’s electron probe microanalyser: general appearance.

of the extremely small volume which the probe has irradiated. For massive specimens, the resolution of the instrument is limited by the depth $d$ to which the electrons penetrate, and there is no point in producing a probe smaller than $d$, in diameter. (When the specimen is thin, an electron microscope preparation for example, we try to obtain a probe the diameter of which is equal to the thickness of the preparation—a few hundred Å.) For electrons, accelerated through some tens of kilovolts, we shall have
\( d \approx 1 \mu \). We shall describe in detail Castaing’s electron probe, the exterior appearance of which is shown in Fig. 278 S1. The instrument contains four main parts:

(a) the electron optical system which produces the probe;
(b) a metallographic reflexion microscope, with which the specimen can be examined and which can be sighted on the point to be analysed;
(c) the specimen-holder with its displacement mechanism;
(d) X-ray spectrographs for quantitative analysis.

### 22.2.2 Electron Optics: Obtaining the Probe

The optical system must be capable of concentrating as high an intensity as possible into a circle less than a micron in diameter; in theory, we have simply to reduce the image of a heated cathode to the required degree with the aid of two very convergent lenses. Figure 278 S2 is a diagram of the optical system which is employed. An ordinary triode gun emits electrons which are accelerated through a potential \( V \) which can be varied between 10 and 40 kV, and which are then incident on a first diaphragm, some tens of microns in diameter; this diaphragm lies in the plane in which the crossover (or “circle of greatest contraction”) of the beam is formed.

![Diagram of electron optics](image-url)
$D_1$ acts as entry pupil, and creates a boundary to the useful surface $S$ of the cathode, for which we can define a brightness $B$ such that

$$B = R \frac{i_0 eV}{\pi kT};$$

$i_0$ is the emissive power of the cathode, $eV$ is the acceleration energy, $R$ is the efficiency of the gun and $k$ is Boltzmann's constant. The diameter of $D_1$ is much smaller than that of the crossover, in order that we should be able to obtain as high a uniform brightness as possible. Measurements which have been made with $V = 30$ kV, $T = 2700^\circ$ K, and $i_0 = 2$ A cm$^{-2}$ give $B = 58,000$ A cm$^{-2}$ sterad$^{-1}$, or an efficiency of $R \approx 70$ per cent. If the diaphragm $D_1$ is not used, the efficiency falls to about 10 per cent. The first lens $L_1$, a condenser, produces a very reduced image $C_1$ of the cathode, but the aperture of the beam is considerably increased; a second diaphragm $D_2$ which is placed at the entry to the second lens $L_2$, the objective, reduces this aperture to a value $\alpha$ of about $10^{-2}$ radians. An image $C_2$ of $C_1$ which is still further reduced is obtained in the neighbourhood of the focal plane of $L_2$. Were the aperture $\alpha$ infinitely small, the diameter of $C_2$ would be the Gaussian diameter $d_0$. To increase the intensity, however, $L_2$ must accept a value of $\alpha$ as large as possible, and the aberrations will thus increase the diameter of the probe. The dominant aberration is the spherical aberration of $L_2$ ($L_1$ is used at a very small aperture, of the order of $10^{-3}$); if we denote the spherical aberration constant of $L_2$ by $C_s$, we can write

$$d_{\text{real}} = d_0 + C_s \alpha^3.$$ 

The brightness $B$ of the successive images of the cathode remains constant, and we can show that the total intensity $I$ in the probe of diameter $d$, which varies as $\alpha^2$, reaches a maximum if $\alpha = (d/4 C_s)^{1/3}$, that is, when the diameter of the aberration disc is equal to one-quarter of the final diameter $d$. For this optimum value of $\alpha$, the intensity $I$ is given by the expression:

$$I = \frac{9\pi}{64} \frac{eV}{kT} R i_0 \frac{d^{8/3}}{(4 C_s)^{2/3}}.$$ 

For $C_s = 3.6$ cm and $d = 1.2 \mu$, we ought to obtain a probe current of the order of 1.07 $\mu$A with $V = 30$ kV.

Experiment shows, however, that another aberration which increases the diameter $d$ must be taken into account; this is the aberration which is due to the residual ellipticity of the openings in the electrodes of $L_2$, which results in an astigmatism which we cannot neglect. Fortunately, this aberration can be wholly corrected by a device placed outside the lens, as Bertein foresaw as early as 1946. Using a weak multipole lens, which is placed beneath $L_2$, and employing a new method of eliminating astigmatism, Castaing has been able to produce a probe with a diameter which is close to the minimum value predicted by the theory, and to obtain a current $I$ of 0.77 $\mu$A in a 1.2 $\mu$ probe.
To increase $I$, it will be advantageous to employ lenses with very little spherical aberration. This is one of the reasons why magnetic lenses are regarded as superior to electrostatic ones for this purpose. In this type of analyser, however, designed as it is for solid specimens, the lens $L_2$ must have its focus external, and will hence have a rather long focal length; the spherical aberration constant is therefore rather high. Very asymmetric lenses are employed, in which the exit opening is extremely small (Mulvey, 1958), as this is the arrangement which allows us to work as close as possible to the specimen (the latter, which may be ferromagnetic, must be outside the field). The contrary is true in analysers for thin specimens working in transmission conditions, as the specimen can now be immersed within the lens, just as in microscopy, and we can thus use very short focal length lenses (Duncumb and Melford, 1959). Further improvement is possible by introducing spherical aberration correctors, which resemble those which have been proposed in electron microscopy (Archard, 1959. See Chapter 7). This last solution has not as yet been incorporated into any instruments of this type on account of the difficulty of the adjustments to be made.

### 22.2.3 Observation of the Object and of the Point of Impact of the Probe

A genuine metallographic microscope is closely linked to the electron optical system and must satisfy several requirements: it must be placed far enough from the specimen for the scattered electrons—which would rapidly contaminate the glass surface—not to be received, the magnification and the numerical aperture must be adequate, an axial hole must be pierced to allow the incident electron beam to pass, and there must be no insulating surface in the way of the beam as electrostatic charge would perturb the probe.

All these conditions are satisfied in Castaing’s apparatus by a special mirror objective, which consists of two concentric spherical mirrors. The focal length is 12 mm, the aperture 0.48, and the resolving power about 0.7 μ. By using a system with two mirrors and two semi-transparent plates inclined at 45° to the optic axis we can throw the light from a lamp placed outside the vacuum enclosure onto the object, and capture the rays reflected by the surface of the object in an eyepiece (which is provided with a graticule) which is also placed outside (Fig. 278 S3). The total magnification is 400.

With this microscope, the object can be observed while the instrument is in action, and the point of impact of the probe rigorously localized. This point should coincide with a point of intersection of the cross-hairs of the graticule. It is visible on the surface of the object as a black point (which is due to contamination by carbon molecules), which appears after a few minutes bombardment.
22.2.4 The Specimen-holder

The specimen must be cylindrical with diameter between four and eight millimetres, and about the same height. The specimen-holder is such that the perfectly straight edge of a tantalum plate can be brought onto the axis of the probe; the shadow of this edge on the fluorescent screen which lies at the bottom of the vacuum chamber is a guide to the correction of the astigmatism of $L_2$. Further, forty examples of pure elements, which are placed on one side or the other of the sample being analysed, can be placed successively under the probe without having to modify any of the adjustments.

The movements, parallel to and across the axis, are effected with the aid of screws of extremely fine thread, with graduated heads.

22.2.5 The Spectrographs

The X-radiation which is emitted by the object is analysed with the aid of two spectrographs, each provided with a curved crystal and a counter, which are placed within the vacuum. The crystal and the counter rotate on
the same focusing circle 25 cm in radius, and the radiation reflected by the crystal is concentrated onto the centre of the window of the counter. At each instant, the angle at which the X-radiation is incident on the crystal—and hence the reflected wavelength—is indicated on a series of drums graduated in degrees and minutes, with an accuracy of better than a minute of arc. One of the spectrographs contains a quartz plate and a Geiger–Müller counter, and is used to detect radiation with a wavelength between 0·6 and 4·5 Å; with this spectrograph, all the elements between chlorine and molybdenum can be analysed using the K lines, and all the elements heavier than molybdenum with the L line. The other spectrometer detects soft X-rays (4 to 10 Å), and makes it possible to analyse light alloys (of aluminium or magnesium). It contains a curved mica reflector, 50 cm in radius, and a special counter with a Mylar entry window; a gas current flows continually (90 per cent Ar, 10 per cent CH₄).

The spectrograph chambers are separated from the probe region by exit windows which consist of beryllium plates 0·3 mm thick; these windows can, however, be retracted when the whole system has been evacuated, so that as little radiation as possible is absorbed between the object and the counter. The arrangement of the spectrographs on either side of the specimen is shown in Fig. 278 S4.

22.2.6 Chemical Analysis

It is possible, therefore, after choosing an interesting region of the object, to bring up the probe, and to measure very accurately the wavelength and the intensity of the lines emitted by a small region one or two microns in diameter around the point at which the probe is incident. We first make a qualitative analysis, scanning the spectrographs over the whole possible range of wavelengths, and after this we can set about making a quantitative analysis. We note the intensity of a line characteristic of some particular element which has been identified in the zone being analysed, and then place the representative of this element in the pure state under the probe. To a first approximation, the ratio of the two measurements gives the concentration of the element at the point in question. To make a second, more accurate, approximation, various corrections are necessary. The greater part of the X-radiation is created in the volume being analysed, but before reaching the spectrograph, it undergoes a certain amount of absorption within the specimen itself; if the latter contains highly absorbent substances, the measurement may be falsified. A correction must be applied to the measured intensities, therefore, the “absorption correction”, if we are to obtain the proper values of the intensity with which to form the ratio. The necessary correction curves have been obtained empirically by Castaing and Descamps (1958), and show that for the light elements it is better to
Fig. 278 S4. The main tube of the microanalyser. 

- **a**: adjustment of the height of the object (focusing of the eyepiece).
- **b**: displacement of the object.
- **c**: window through which the object is introduced.
- **d**: levers, for retracting the X-ray exit windows.
- **e**: means of introducing film for recording the Kossel lines.
- **f, g**: hoods for the evacuated spectrographs.
- **h**: opening for introducing film (diffraction).
- **k**: eyepiece.
use a small accelerating potential (10 to 15 kV). It is for these elements that the analysis is least accurate.

These corrections assume that the emergent angle of the useful rays is well defined and equal to 18°; this requires the specimen to be perfectly plane, and hence forbids the use of electrolytic polishing; a very careful mechanical polishing is to be recommended. A further correction, for the fluorescent radiation emitted by the specimen, can also be made, as certain atoms of the specimen may be excited by the primary X-radiation. Correction formulae are available (Castaing, 1951; Castaing and Descamps, 1955) with which this source of error can be eliminated.

22.2.7 Some Applications of the Micro-probe

The most usual field of application of the probe is metallurgy, but we might equally well analyse any type of specimen, conductor or insulator (provided the latter is first coated with a thin conducting layer so that the charge in the incident beam can flow away). Examples are to be found in the numerous articles by Castaing and his associates, and in an article by Mulvey (1958); particular examples are the chemical analysis of steels, the deposits produced by rubbing one metal against another, and inclusions in alloys.

The problems of intermetallic diffusion can be studied most elegantly with a micro-analyser; a tranverse section of the zone in which the two metals in question are in contact is polished metallographically and the concentration of the two constituents is established point by point starting from an arbitrarily selected origin in one of the pure metals. This method has been successfully applied to a number of different systems: U–Zr, U–Mo and U–Ti [Fig. 278 S5(a)]. The equilibrium diagram can be established on the basis of accurate measurements of the concentration. A chart of the distribution of manganese in a Hadfield steel (C = 1·88%, Mn = 12·5%) is shown in Fig. 278 S5(b); this chart was obtained by a point-by-point analysis (Philibert and Bizouard, 1959). A rapid study of the surface of the specimen can be made by scanning this surface with an electron probe controlled by a suitable deflection system placed above the objective (Cosslett and Duncumb, 1956); the scanning is synchronized with that of an oscilloscope, the intensity of which is modulated by the counter which collects the X-radiation characteristic of any given element. To increase the resolution of the system, we could alternatively leave the probe stationary —this would allow the spectrograph to be used always at its optimum adjustment—and employ some mechanical means of shifting the specimen.

A thorough analysis of the various possibilities of electron probes is to be found in an article by Castaing (1960).
**ELECTRON PROBE DEVICES**

![Graph showing uranium-titanium diffusion at 950°C during 24 hr](image1)

- **Interface after diffusion**
- **Penetration, μ**

**Fig. 278 S5.** Examples of the uses of the micro-probe. (a) The diffusion curve, traced in a direction normal to the zone of contact between uranium and titanium, after heating to 950°C for 24 hours. (b) The distribution of manganese in a special steel (Philibert and Bizouard, 1959).

### 22.2.8 Other Recent Types of Micro-probe

One of the problems which appear when we employ electron probes is that of determining the exact location of the point of impact. We have already seen how the observation in the instrument described above is
effected with the aid of a special mirror objective which is placed inside the last magnetic lens. Alternatively, we may construct a mechanical system to very high precision, with which we can rotate the specimen-holder through 180° about a horizontal axis in such a way as to be able to examine the specimen through a side-window with the aid of a high-resolution optical microscope (Mulvey, 1958, 1959).

The scanning technique which was suggested by Cosslett and Duncumb (1956) and which was inspired by the technique of scanning microscopy (see below) makes it possible to obtain the image of the surface of the specimen rapidly, relative to a given characteristic X-ray line. The probe can then be directed onto one of the details which were visible on the image, by using the afterglow which lasts about 30 sec.

A way of perfecting the method is to collect the secondary electrons which are emitted by the surface in a scintillation counter, and thus obtain the modulation signal; the image is much brighter (Duncumb, 1958).

The scanning is produced by a double system of coils, which are placed above the final reducing lens.

With a recent version due to Nixon (1960), ultra-thin specimens can be analysed and simultaneously examined by transmission microscopy at high magnification. The principle is as follows: the image of the preparation is focused onto the final fluorescent screen in an ordinary microscope; in the centre of this screen is a small hole, which lies on the optic axis of the instrument. Through this hole a second electron beam is directed up the instrument, and forms a micro-probe in the object plane; if the system is perfectly aligned, this probe falls on the region the image of which we are examining. The X-rays which are produced by the probe are collected in a special proportional counter which is placed near the specimen-holder. With this instrument it should be possible to obtain probes one to three-hundred Ångströms in diameter with a current of $10^{-11}$ A.

22.3 THE SCANNING ELECTRON MICROSCOPE

22.3.1 The Principle of the Instrument

We have already mentioned the principle behind this instrument in the preceding section. The electron probe "scans" the surface which is to be examined, propelled by a system of crossed coils which are placed before the final reducing lens, with the same action as the spot of a television camera (Fig. 278 S 6). A receiver which is placed near the surface collects continuously the secondary electrons which are emitted by the surface. The signal which is obtained is amplified, and then fed to the modulation electrode of a television tube, adjusted to scan in synchronism with the
probe. In this way, we obtain a highly magnified "image" of the surface; the resolution is in theory equal to the diameter of the probe, provided all parasitic signals have been thoroughly eliminated, and the specimen is perfectly flat. Just as in a microanalyser, the probe diameter must be

adjusted to be of the same order of magnitude as the penetration of the electrons into the specimen being examined. It is therefore advantageous to employ low energy primary electrons but this in turn necessitates a very careful elimination of the alternating parasitic magnetic fields which would tend to enlarge the spot. Further, the role of the diffraction aberration becomes the more important as $\lambda$ increases. Finally, at very low voltages, a broadening of the probe is perceptible as a result of space charge. We are thus forced to make a compromise between these three factors: potential, intensity and spot diameter. A satisfactory compromise is obtained with potentials of only a few kV, with which we can obtain a probe intensity of about $10^{-9}$ A in a probe of a few hundred Ångströms diameter.
22.3.2 The Origins of the Contrast

The secondary electrons are of two kinds:
(a) primary electrons reflected at high energies (between 50 eV and the initial energy of the incident beam);
(b) genuine secondary electrons, emitted with energies of less than 50 eV.

We can collect the electrons of either category to produce the modulation signal but the slow secondaries provide more information about the structure of the surface being studied. The fast electrons are scarcely deflected at all, since the electric fields in the neighbourhood of the surface are weak, and the electrons hence follow almost rectilinear paths. The slow electrons are, on the contrary, strongly influenced by these fields, and they can be collected more easily in a receiver placed near the specimen and slightly polarized to attract the secondaries.

The origins of the contrast which is obtained when these electrons are employed are several:
(a) The relief, which affects the rate of emission of secondaries, in consequence of the different orientations of the elementary constituents of the surface of the specimen. Certain regions are hidden from the receiver, and seem darker. This effect of relief is particularly striking if we collect only the fast (reflected) electrons, the trajectories of which are far more acutely influenced by the angle of incidence.
(b) The chemical nature of the specimen. Since the emission coefficients of different elements differ widely at low bombardment potentials, we can, for example, pick out the segregations or clusters in an alloy.
(c) The crystal orientation, since even for the same metal, the secondary emission varies (if relatively little) with the nature of the crystal faces being bombarded; the presence of this effect has been displayed experimentally.
(d) The way in which the metal has been worked (Davoine and Bernard, 1956). If a metal has been drawn, the contrast of certain regions is increased; this effect is perhaps due to a superposition of the two preceding processes during the deformation of the surface, and to the creation of new crystal faces by the impact of the beam.
(e) The contact potential differences between neighbouring areas, or "potential contrast"; these are such junctions as the boundary between two monocrystals of the same metal or between two semi-conducting strips doped with different impurities ($p$-$n$, $p$-$n$-$p$, junctions . . .). We can completely interchange the brightness of the different strips by biasing the specimen (Everhart, Smith, Wells and Oatley, 1958). The contact potential difference (which may vary between a few tenths of a volt and a few volts) is emphasized by a transition zone, two microns wide, which gives rise to intense electric fields which eliminate the vast majority of the secondary electrons, and this zone has a highly contrasted appearance.
22.3.3 Recent Developments

The most important improvements have been made in the domain of the receivers of secondary electrons, with a view to increasing the useful signal to the uttermost, and eliminating the background noise which the reflected electrons produce. The receiver may be an electrode system similar to that of a photomultiplier. We cannot, however, use the same type of layers to perform the multiplication, as the vacuum is relatively poor. We make do with copper–beryllium dynodes (Bernard and Davoine, 1958) which are periodically reactivated. Alternatively the receiver may be a grid, polarized to one or two hundred volts, which attracts the secondary electrons. The latter pass through the grid and are accelerated through a few kV, and finally fall on a fluorescent screen. The light is transported to a photomultiplier with the aid of a perspex light-pipe.

In other attempts to obtain micro-probes the correction of astigmatism and the use of lenses with little spherical aberration have been considered; it should be possible to obtain probes 100 Å in diameter (Smith, 1960).

Descriptions of these devices are to be found in the articles by Davoine (1958), and Smith (1960), together with the characteristics of the associated electronic circuits; recent applications are described in the articles by Thornley (1960), and Davoine, Bernard and Pinard (1960) and also in the articles already quoted.

Figure 278 S 7 shows two aspects of the same zinc sulphide crystal; image (a) was obtained by collecting the secondary electrons, whereas image (b) was formed by collecting the light emitted as fluorescence by the crystal onto a photomultiplier (3000 to 6000 Å).
22.4 THE X-RAY PROJECTION MICROSCOPE

22.4.1 The Principle of the Instrument

Figure 278 S8 is a schematic cross-section of an X-ray projection microscope. The micro-probe $S$ is formed just beyond the final reducing lens $L_2$, at the surface of a thin target, $F$. Those of the X-rays, which are created in the zone upon which the electrons fall, which are emitted through the target, pass through the object under examination, $O$, and produce a projection image of the latter on the screen, $P$. The linear magnification is equal to the ratio of the distance $SO$ to the distance $SP$ when the object is thin. If the object is thick, however, the magnification varies according to the position on the axis of the point being examined; the zone close to $S$ will be the most magnified.

The more intense the electron current in the probe, the brighter will be the image. As we shall see later, however, the resolution improves as the diameter, $d_z$, of the spot is reduced and the operating potential $V$ is lowered; it improves, therefore, as $i$ is diminished ($i$ varies with $d_z^{8/3}$ if we take spherical aberration into account as we have done earlier in this volume, and also with $V$). Once again, we are forced to have recourse to a compromise, with the result that the operating potentials ordinarily used are of the order of 5 to 20 kV.

22.4.2 Theoretical Resolution and Visibility

The resolution of the instrument is essentially related to the diameter of the source and to the diffraction, always supposing the optical properties of the lenses to be perfectly stable in time.

The image-forming mechanism is the same as in every conical projection system. The non-zero transverse dimensions of the source give rise to
regions of shadow $AB$ and to penumbral zones $AC$ and $BD$ (see Fig. 278 S9). If we consider some feature of the object of width $d$, we find:

$$AB = \frac{ad - ds(a - b)}{b}$$

$$AC = BD = \frac{ds(a - b)}{b}.$$

We shall have

$$AB = 0,$$

for $d = d_0$, where

$$d_0 = \frac{a - b}{a} ds.$$

We regard $d_0$ as the diameter of the smallest perceptible detail in the object. Generally speaking, we require high magnification and hence $b \ll a$ so that

$$d_0 \approx d_s.$$

The smallest perceptible detail is the same size as the source.

Fig. 278 S9. Formation of the projection image. $d_s$: source; $d$: object; $AB$, shadow; $AC$ and $BD$: penumbrae.

If we now consider two neighbouring points, diameter $d_0$, the illumination in the penumbra ($u_i$) varies linearly with the distance to the centre, and we shall regard two points as just separated when the total illumination between the two spots is about 75 per cent of the brightness at the centre of each spot (Fig. 278 S 10) for the resolving power $\delta_i$ in image space; we thus find:

$$\delta_i = 1.25 u_i,$$

and transferring back into the plane of the object

$$\delta_0 = 1.25 u_i;$$
when the linear magnification $M$ is $a/b$, we obtain:

$$\delta_0 = 1.25 \frac{d_s(a - b)}{b M},$$

and hence

$$\delta_0 \approx 1.25 d_s.$$

Fig. 278 S10. Resolution of two neighbouring details.

This justifies the rule of thumb which is normally adopted:

$$\delta \approx d_s.$$ 

In reality these penumbral regions are overlaid by Fresnel diffraction fringes. Only the first fringe, which lies at a distance $x_1$ from the edge of the geometrical shadow (Fig. 278 S11), is troublesome, as it is very bright.

Fig. 278 S11. Formation of the Fresnel fringes.

The fringes are the result of interference between the direct wave $SB$ and the diffracted wave $AB$; the maximum of the first fringe lies at a distance $x_1$ from the line $SA$, where

$$x_1 = \left[ \frac{a(b - a) \lambda}{b} \right]^{1/2}.$$ 

Transplanting back into the object plane, in the case $b \ll a$, the equivalent distance

$$x_0 = \frac{x_1}{M}$$
is given by

\[ x_0 \simeq (b \lambda)^{1/2}, \]

in which \( \lambda \) is the wavelength of the radiation employed.

The "perceptibility" will be limited by diffraction except when the condition

\[ x_0 \leq \delta_0 \simeq d_s \]

is satisfied, and this will impose an upper limit on \( b \):

\[ b < \frac{\delta_0^2}{\lambda}. \]

If for example we fix the resolution (\( \delta_0 = 0.1 \mu \), say) and \( \lambda = 1 \text{ Å} \), we find \( b < 100 \mu \) which is still feasible; if \( \lambda = 10 \text{ Å} \), however, \( b < 10 \mu \) and this is less easy to achieve.

To obtain the resolving power, we must determine the intensity distribution between two neighbouring points of the image. To a first approximation, we arrive once again at the general formula for the resolving power when the limit is provided by diffraction:

\[ \delta = \frac{0.625 \lambda}{x_0} = 0.625 \frac{b \lambda}{d_0} \]

(where \( \alpha_0 \) is the angle at which the object element is "seen" from a point on the source: \( \alpha_0 = \frac{d_0}{b} \)).

We can summarize these results as follows: the limits of perceptibility \( \delta_p \) and of resolution \( \delta_R \) will be reduced by diminishing \( b \) and \( \lambda \) and are given approximately by

\[ \delta_p = (b \lambda)^{1/2}, \]

and

\[ \delta_R = 1.25 (b \lambda)^{1/2} \simeq \delta_p, \]

when the limits are a consequence of diffraction.

If we choose a small enough value for \( b \), we arrive at the simple relation

\[ \delta_p \sim \delta_R \sim d_s. \]

Diffraction will thus certainly limit \( \delta_p \) and \( \delta_R \). Since we cannot go on reducing \( b \) indefinitely with this method, we can scarcely hope to obtain a resolution better than 100 Å. Usually, however, it is the dimensions of the source which limit the resolution.

### 22.4.3 The Diameter of the Probe and the Thickness of the Target

As we saw in § 22.1, the maximum intensity \( I_s \) which can be attained with a probe of diameter \( d_s \) varies in the following way:

\[ I_s \simeq \frac{d_s}{C^{2/3}} \cdot V \cdot i_0, \]
since the spherical aberration of $L_2$ imposes a limit on the aperture of the beam which can be usefully employed.

Here, we can use very short focal length lenses, as the potentials are low and the probe need not necessarily be outside the lens, unlike the situation in the micro-analyser where it is the reflected X-rays which we collect. If $d_s$ is reduced, $I$ falls very rapidly, so that when we try to obtain a very high resolution, the brightness is extremely feeble; we shall no longer be able to focus the image directly (focusing consists of bringing the probe onto the target), and it has proved necessary to invent special ways of focusing (see below).

If we are willing to accept a comparatively poor resolution, $d_s$ can be increased, but this in turn cannot be continued indefinitely, as the rate of dissipation of heat in the target is limited. The incident power is almost wholly converted into heat, and increases as $d_s^2$ if the current density is constant; the rate at which a very thin film can conduct away this heat increases with $d_s$ only, however, since the heat flux is wholly radial. The melting point $T_M$ of the target is quickly reached, and in general, we avoid temperatures higher than $\frac{3}{4} T_M$. For preference, therefore, we use metals which are good conductors, or the refractory metals: tungsten and copper, silver and gold. The thickness of the target is then fixed by the way in which the incident electrons are scattered within it. In fact, the electrons are only completely stopped inside the metal after they have collided enough times to convert all their kinetic energy into thermal energy. It is usually agreed that the large majority of the incident electrons is scattered into a cone of semi-angle $30^\circ$ to $35^\circ$; this leads us to give the X-ray source a diameter $d_s$ which is equal to the thickness $h$ of the target, when the spot diameter is negligible.

When a very fine source is required, we have to reduce both $h$ and $d_s$. Langner (1956) has calculated the influence of spherical aberration, astigmatism and X-ray scattering in the target on the size of the final "source" $d_s$, when the geometrical spot diameter is negligible. With gold and a potential of 10 kV, the value of $d_s$ and hence the resolution attainable in practice, is 400 Å when $h = 0.1 \mu$.

Reducing $h$ in turn implies a reduction of the X-ray intensity emitted by the target, however. If we require a large resolving power, and at the same time a reasonably bright image, we are obliged to try to increase the intensity of the probe; all things considered, this can only be done, for a given probe diameter, by increasing the specific emission of the cathode (since we are trying at the same time to reduce $V$, which lessens the brightness), and by correcting the spherical aberration very perfectly.

All these questions are gone into very thoroughly in the book by V. E. Coslett and W. C. Nixon entitled *X-ray Microscopy* (Cambridge, 1961).
22.4.4 Instrumental Details

The optical system is identical to that of a microanalyser. A system with a single reducing lens could be used, but the two-lens arrangement is more flexible; the second lens is then always operated close to maximum convergence (and thus to the minimum value of $C_s$) and the diameter of the probe which is finally obtained is adjusted by varying the magnification of the first lens. In such conditions, the axial position of the probe varies very little, and only the slightest of alterations to the excitation of the second lens is necessary to bring it back onto the target.

Once again, magnetic lenses are preferable, as we can obtain at the same time very short focal lengths (with electrons accelerated through a few kilovolts) and values of $C_s$ much smaller than those we should be able to reach with electrostatic lenses. Suppose, for example, $C_s$(mag) = 1 mm, and $C_s$(electr) = 32 mm; for the same diameter $d_s$, we have $I$(mag) \( \simeq 10 \) or for the same intensity, $d_s$(electr) = 2.4 $d_s$(mag).

The first instrument of this kind (von Ardenne, 1939) was electrostatic, although the second version—von Ardenne, 1956— is magnetic. An electrostatic instrument of a slightly different kind has been built by Marton et al. (1957); the electron source is a very fine point, which works by field emission.

![Figure 278 S12. The X-ray microscope of the Laboratoire d'Optique Électronique du C.N.R.S. at Toulouse (Dupouy and Verdier).]
and gives a high brightness; a single reducing lens forms the image of this source on the target. The diameter of the final spot is about 5 μ. With an electrostatic instrument with two lenses, Newberry and Summers (1954) have been able to attain a resolution of 1 μ; the focal length of their lens \( L_2 \) was 1 mm and the spherical aberration constant \( C_s \) was 2.5 mm. The exposure time was of the order of ten or twenty minutes.

Most projection microscopes are, nevertheless, magnetic, and all are based on the structure of the instrument built by Cosslett and Nixon (1951, 1952 and 1953). The instruments which have been derived from this early form are described in articles by Cosslett and Pearson (1954), Cosslett, Nixon and Pearson (1956), Siegel and Knowlton (1957) and in Verdier's thesis (1960). Figures 278 S12 and 278 S13 respectively show the apparatus

![Fig. 278 S13. Image of a diatom, direct magnification \( G = 55 \), exposure time 10 minutes, H.T. 7.7 kV.](image)
built by Verdier, and a photograph of a diatom taken with it. With this type of instrument, a resolution of 0.1 \( \mu \) has been reached (Nixon, 1955), and this will certainly be improved upon.

Although the final resolution of this type of system is a hundred times poorer than that of a transmission electron microscope, slow variations of potential and mechanical vibrations must be most carefully eliminated, as the exposure times (for high resolution) are extremely long, frequently lasting several minutes.

When the diameter of the X-ray source is less than 0.1 \( \mu \) it is often impossible to focus the probe by direct observation of a test grid on the fluorescent screen, as we should normally do if the intensity were adequate. Various methods have been developed which employ the secondary electrons emitted by the target, and with the aid of one of these methods, the focusing can in fact be carried out. In Nixon's solution (1958), the electrons which are emitted by the thin target in the forward direction are used (the "forward-scattered electrons"); a large proportion of the primary electrons pass through the target and form the projection image of a very fine grid on the fluorescent screen just as in shadow microscopy. This image is bright enough for the probe to be focused. The X-ray image is formed with the aid of the X-rays which emerge from the source at an angle of about 60° with the axis of the instrument [see Fig. 278 S14(i)]. The resolution which is obtained is of the order of 0.1 \( \mu \). Here, however, both target and object are inside the vacuum—this limits the applications of the instrument, but does, on the other hand, allow us to use very soft X-rays.

Ong Sing Poen and Le Poole (1958) have suggested another method in which the electrons which are scattered back towards the face of the target on which the beam is incident, are used. The other face of the target can, therefore, be in air. These back-scattered electrons represent about 1 per cent of the whole beam; they travel back through the lenses in the opposite direction, and by tilting the optic axis very slightly so that the image lies in the same plane as the crossover but a little to one side of it, we can examine the image on a fluorescent screen [Fig. 278 S14(ii)]. Here again, a resolution of 0.1 \( \mu \) has been reached even though the X-ray image is not visible at the final screen.

In conclusion, we should mention that the difficulty can of course be avoided by replacing the fluorescent screen by one of the image intensifiers which are described in Chapter 13.

To end this section, we shall say a few words about the projection technique which has recently been developed by Pattee (1958) and which he calls "Microfluoroscopy". The preparation is spread directly on the fluorescent screen and is illuminated by the beam of X-rays which emerges from a source about 100 \( \mu \) away. The screen is about one micron thick, and can thus be examined with the aid of an optical microscope with a high magnification (\( \times 1000 \)) and a very wide objective (and hence a very shallow depth
**Fig. 278 S 14.** Focusing aids used by (i) Nixon, (j) Ong and Le Poole.
of field). The grain-size of the screen must be smaller than the desired resolution, and it is for that reason that evaporated screens free of structure are employed (Feldman and O'Hara, 1957). A resolution of about 0.25 μ can be obtained in this way, namely, that of the optical microscope, and we can use a source with a greater diameter, \( d_s \), and hence a high brightness.

22.5 MACHINING BY ELECTRON BOMBARDMENT

22.5.1 The Electron Beam as a Heat Source

When an electron beam of intensity \( I \) and potential \( V \) falls on a target, the energy \( V I \) which had been carried by the beam is transferred to the target almost entirely in the form of heat. Even for comparatively high potentials, of 100 to 150 kV, all this energy is distributed into an extremely thin film of matter, a few microns thick. This is far thinner than we could conveniently obtain with high frequency heating, and apart from this, we can localize the zone being heated far more precisely. It is natural therefore, that with the development of guns capable of giving a very high intensity (see Chapter 12, §12.2.3) and capable of concentrating a power of 10 to 50 kW into a few square millimetres, methods should have been developed with which this energy can be used for melting or for evaporating metals in vacuo. (There is the added advantage of the chemical cleanliness of the vacuum.)

Another advantage of heat sources obtained in this way is that it is possible, even when the average power is relatively low, to produce enormous energy densities, which may exceed those which even the most excellent electric arc could provide by a factor of a thousand or ten thousand; this is thanks to the high quality electron optical systems which can now be constructed. If we work in pulsed operation (which enables the current density emitted by the cathode to be increased, and hence the brightness of the source also), it is possible to obtain an energy density of \( 10^8 \) or \( 10^9 \) W cm\(^{-2} \) in a probe a few tens of microns in diameter. This produces an almost instantaneous evaporation of the zone bombarded.

22.5.2 The Heating of a Target Bombarded with Electrons

Several authors have recently calculated in an approximate fashion the temperature distribution and the energy balance in the zone under bombardment (Steigerwald, 1958; Bas, 1960), not to mention earlier work which had been performed in connexion with the heating of the anti-cathodes in X-ray tubes.

We consider first of all the steady state situation. A beam of diameter \( 2r_0 \) and uniform density falls on a thick target (Fig. 278 S15) for which
\( \hbar \gg r_0 \). We assume that the beam arrives at the bottom of a hole which it has itself made. The incident energy is transformed into heat within a thickness \( e \); this zone is warmed up, melts, and is partially evaporated (the vapour pressure increases exponentially with temperature). A small proportion of the heat is radiated away, but the rest is conducted away into the target. If, in the steady state, we suppose that the lower face of

![Fig. 278 S15. Punching by electron bombardment.](image)

the target remains at a temperature \( t_0 \), an analysis of the energy balance reveals that the conduction outweighs the latent heat, and a **considerable zone must be melted**. The temperature distribution is represented by the curve in Fig. 278 S16a (Bas, 1960), calculated for the case of an alumina

![Fig. 278 S16a. The form of the temperature distribution in the neighbourhood of the bombarded zone. \( t_m \): maximum temperature; \( t_f \): melting point; \( t_0 \): temperature at the lower edge of the piece being machined (Bas, 1960).](image)
target; the temperature at the centre \((t_m = 2260^\circ C)\) is higher than the melting point \((2050^\circ C)\). The radius of the melted zone is of the order of \(1-2 \, r_0\) if \(t_0 = 0^\circ C\), but is of the order of \(2 \, r_0\) if \(t_0 = (1/2) \, t_m\). The metal is evaporated at about \(v = 0.1 \, \text{mm sec}^{-1}\) at the centre, but at \(r = r_0\), this rate of evaporation, \(v\), is practically zero.

The molecules which have been evaporated leave the target with a certain momentum \(mv\), and there is in consequence a reaction which manifests itself as a pressure on the melted zone which hollows out the centre of the latter, spreading the molten matter out towards the cold surrounding region.

We should mention too that there is a very steep radial temperature gradient at the edge of the beam, so that for targets which are poor conductors, there is a danger of cracking the target.

The evaporation rate is higher, for a given incident energy, when the thermal conductivity \(\lambda\) is lower (and hence \(t_m\) higher), and when the metal evaporates more easily. It is possible to calculate the temperature \(t_1\) for which \(r_m = 0.1 \, \text{mm/sec}\); this temperature is in most cases higher than the melting point \(t_f\) but the difference \((t_1 - t_f)\) which varies between very wide limits according to the nature of the target, may even take negative values (for chromium, \(t_1 - t_f = -225^\circ C\)). The tendency to evaporate which a metal displays can thus be characterized by the factor \(\beta = \frac{\lambda}{t_1}\), which is expressed in W cm\(^{-2}\)—the smaller \(\beta\) is the more easily can the metal be machined. For everyday substances, Bas (1960) gives values of \(\beta\) between 190 for \(\text{Al}_2\text{O}_3\) and 7700 for tungsten. The metals which are good thermal conductors—silver, copper, and aluminium—follow immediately behind tungsten. With these figures, an energy density of \(\sigma = 7.7 \times 10^6 \, \text{W cm}^2\) would be necessary for \(\lambda \, t_1 = 7700\) (tungsten) and \(r_0 = 10^{-3} \, \text{cm}\). Experiments on alumina show that the energy density which is necessary is even greater, but nevertheless can be reached.

Provided we have beams of high enough energy density at our disposal, therefore, we can work the most refractory substances in steady state conditions, whether it be a question of boring, welding or melting. The zone which has been worked is, however, surrounded by a region which has melted and resolidified, and by a zone which has been disturbed to a greater or lesser extent from the structural point of view by the sharp temperature increase.

To bore or drill a metal, with machine tolerances of the order of a micron, we avoid this difficulty by using an intense beam in very short bursts (Steigerwald, 1958, 1961). If a beam of \(10^9\ \text{W cm}^2\) and 100 keV falls on a tungsten surface, the temperature of the layer which stops the electrons is raised to 3650 K in \(10^{-8}\) sec, and a point one micron from the edge of the beam will reach its melting point after \(7 \times 10^{-6}\) sec. If, therefore we apply the energy in short bursts, about a microsecond long and far enough apart, the zone which lies 1 \(\mu\) away from the rim of the vaporized
region will not reach its melting point. This is the principle underlying the "electron drill", with which extremely narrow machining tolerances can be satisfied by moving the probe about over the target which is to be cut or punched (Schleich, 1958; Opitz, 1960; see Fig. 278 S16b).

Fig. 278 S 16b. Cutting out with an electron beam. The pipe is made of a special steel; the wall is 2 mm thick, and the cutting speed is 30 cm min⁻¹ (with a Carl Zeiss drill) (Photo: Schleich).
22.5.3 Various Applications of Intense Probes

We shall describe briefly these applications in order of increasing complexity of the optical system.

(i) Melting and welding

Processes of this kind were developed in France by Stohr in 1957, to solve the welding problems created by nuclear reactor technology. Here, every effort is made to achieve very intense beams well focused into a crossover a few tenths of a millimetre or a few millimetres across. In general, the "Pierce" optical system is adequate; the gun has a wide tungsten cathode, the back of which is heated by electron bombardment, and is provided with an accelerating electrode and an electrode for concentration. The whole arrangement is planned to give the smallest possible crossover a few centimetres from the anode, and at high intensities we are forced to work with potentials of several tens of kilovolts. The region in which the gun is situated is separated from the working region, so far as the vacuum is concerned, by the opening in the anode, and is provided with a separate pump to ensure that the vacuum is good despite the ejection of gas from the heated metals, and the evaporation of the latter.

In this way, probes 0.8 mm in diameter can be obtained at 60 kV and a heating power of 5 kW. If we are willing to sacrifice to some extent the fineness of the probe, we can increase the power to 40 or even 100 kW, and thus melt refractory metals in considerable quantities in a crucible to form alloys.

Probes of a few tenths of a millimetre are employed for welding together the edges of metal plates which are very easily oxidized; the performances which can nowadays be obtained with electron welders are better than those of argon arc torches, as regards speed, fineness of the joint and above all, penetration. The melted zone can easily reach two to five centimetres, so that comparatively thick plates can be welded together in this way. A good weld is ensured by scanning the beam laterally, with an amplitude of one or two millimetres, over the line along which the two edges which are to be soldered together are in contact; the plates are meanwhile moved parallel to this line. We can, for example, weld together the edges of two aluminium plates 25 to 50 mm thick in a single operation, at a speed of 30 to 50 cm/min (Stohr, 1961; Fig. 278 S17). When the plates are made of materials which are good thermal conductors, a zone a few millimetres wide is found on either side of the weld which has been thermally disturbed. When, on the contrary, the substances are poor conductors of heat, stainless steel or the nickel alloys for example, the "penetration-to-width" ratio of the weld can reach 25; with 9.5 mA and 130 kV stainless steel plates 12 mm thick can be welded together at a speed of 35 cm/min. In these substances, a crystallographic analysis shows that the thermally disturbed
zone in the neighbourhood of the weld is wholly absent, and that the weld itself is perfectly homogeneous and is composed of micro-crystals.

For refractory substances, such as tungsten, the depth of the melted zone is adequate for plates one or two millimetres thick to be welded, provided the potential is very high: 130 to 150 kV (Hoffmann, 1961).

If the instantaneous power is increased still further, we can cut metal plates just as with an ordinary blow-lamp, but we can obtain far narrower cuts than with the latter.

In this type of application, the radiation is particularly intense, and the working region has to be screened with a layer of lead thick enough to intercept all harmful radiation.

Numerous articles on the results which have been obtained with electron welders are to be found in the Proceedings of the Third, Fourth and Fifth Symposia on Electron Beam Processes (Boston, 1961, 1962 and 1963). Figure 278 S18 shows one of the high power welding installations, designed by J. A. Stohr, which is at the Centre d'Études Nucléaires de Saclay (France).
(ii) Boring and drilling

As we have already seen, it is possible even with steady state conditions, to bore relatively fine holes by evaporating away the centre of the melted zone, but if we require the very tightest tolerances, we are forced to use pulsed operation. The power supplied to the target is less than in the preceding case, but the probe is very tiny and the energy density may be considerable.

The first device of this type was described by Schleich (1958). A diagram is given in Fig. 278 S19. The various parts are as follows:

- a "Fernfokus" gun;
- a magnetic system for aligning the beam;
- a stigmator;
a highly convergent reducing lens;  
a deflexion system, which can be supplied with a variable potential,  
for shifting the probe;  
a target-holder, equipped with a high-precision displacement system;  
an optical microscope for examining the target.

![Diagram of an electron drill](image)

**FIG. 278 S 19. Diagram of an electron drill (by courtesy of Carl Zeiss, Oberkochen).**

The natural setting of the gun corresponds to a bias beyond cut-off;  
when the system is in action, the gun is regularly triggered by applying  
a series of positive pulses to the grid—both the width and the frequency  
of these pulses can be altered (over ranges of 2 to 70 $\mu$sec and 50 to $10^4$ c/s  
respectively). The current into the probe can be varied between zero and  
eight milliamperes, for accelerating potentials between 25 and 150 kV,  
and probe diameters from 10 to 40 $\mu$.

With this machine, it is possible to bore round holes from 20 to 200 $\mu$  
in diameter, and slits one millimetre long and 40 $\mu$ wide in plates 0·4 mm  
thick. A "width-to-depth" ratio of 1/50 is attainable; the tolerance is a  
few microns.
Figure 278 S20 shows several cruciform holes which have been bored through a stainless steel plate, which can serve as a plastic-wire drawing plate or "spinnert". Between each pulse of electrons, the probe is moved to a different point of the shape which is to be cut out, to reduce to the uttermost any heating of the piece being machined. It is possible to programme beforehand the voltages which are to be applied to the probe deflector, in such a way as to obtain most complicated cross-sections. Figure 278 S21 gives a magnified view of a narrow slit. Such a device as this has a vast range of applications in cutting out the elements of miniature electronic circuits (resistive films deposited on a ceramic material and semi-conductors; see, for example, Opitz, 1960, Steigerwald, 1961).
(iii) The use of micro-probes for recording

A micro-probe a few hundred Ångströms across can equally well be used as a heat source with which to form an impression on a structure-free collodion film. A collodion film 500 or 1000 Å thick is decomposed into carbon by the probe, and is at the same time thinned down by the heating. We can therefore machine the plate, and subsequently examine it with a high magnification electron microscope; signs (dots, dashes, letters) which are "written" on the film can be rendered more easily visible by means of the same types of shadowing technique as are used in microscopy (Möllenstedt and Speidel, 1960). The practical arrangement is as follows: a diaphragm a few microns across is placed on the specimen holder of a magnetic microscope, and in this way we can obtain a probe 100 Å in diameter in the focal plane of the objective. The collodion target is placed in this plane, and is sufficiently transparent to electrons for an image of the diaphragm to be visible on the fluorescent screen of the microscope. The probe is moved about by shifting the diaphragm, and can be followed on the final screen. If we compel the final image to follow an arbitrarily chosen figure drawn on the screen, the probe will trace out the same pattern on the collodion film on an extremely reduced scale.

This technique can be used to rule extremely fine lines or groups of parallel lines, which can subsequently serve as micrometers or as test
objects for determining the resolving power in electron microscopy (the lines may be 140 Å thick and 1000 Å apart, for example).

Finally, we should mention that this technique may be used to construct very small printed circuits. An insulator is coated with a thin layer of conducting material, the appropriate parts of which are then cut away by the electron beam (Fig. 278 S 22).

Fig. 278 S22. Cutting away the surface of a resistance with an electron beam. A layer of CrNi has been evaporated onto a ceramic support. The line width is 10 μ. (Drill by Carl Zeiss).
CHAPTER 23

THE ELECTRON (ION) PRISM

23.1 THE HISTORY OF THE MASS SPECTROSCOPE

Goldstein's work on electric discharge phenomena led him in 1886 to the discovery of *electrically positive rays* ("canal" rays). These rays could be deflected with an electric or magnetic field. In 1910, J. J. Thomson working at Cambridge made use of this property to break them down into their several components (by the parabola method). In particular, he found that two components were present in neon. From this observation stemmed the discovery of isotopes, and its importance was clearly displayed by F. Aston who, in 1919, built the first spectrograph (Fig. 279).

![Fig. 279. Aston's spectrograph.](image)

Aston's instrument "focuses energies", a notion which we shall understand more clearly later on; at about the same time, in America, Dempster (1918) was building an instrument which "focused directions", and soon after, he discovered the isotopes of magnesium. During the next few years, Aston established the existence of a large number of isotopes, and improved the accuracy of his spectrograph (1927). He pointed out the differences between the atomic masses and the anticipated integers; the differences, which were also studied by Costa (1925) in France, are now recognized to be a highly significant feature of the energy in nuclear physics.

In 1929, Dempster and Bartky suggested the method which is known as double focusing, by which the quality of these instruments is considerably improved; it was at this point that the high quality spectrographs which are associated with the names of Bainbridge, Jordan and Mattauch, to name but a few, came into being. Meanwhile, instruments with electro-
mechanical methods of recording were being developed—these are the devices known as “spectrometers”—with which quantitative analyses could be made, and which came to be built both in the laboratory and industrially under the stimulus of Nier, Bleakney and Hipple; spectrometers are tending to play a most important role both in ordinary chemistry and in nuclear chemistry.

23.2 THE PRINCIPLE OF MASS SPECTROSCOPY:
THE SPECTROGRAPH AND THE SPECTROMETER

23.2.1 The Principle

Mass spectroscopy is a method of analysis based on the fact that ions of the same kinetic energy but different in mass can be separated according to their different masses by a static magnetic field. The features fundamental to the separation process, which of course takes place in a very high vacuum, were all present in essence in Aston’s first instrument (Fig. 279). The substance to be studied is first ionized, partially transformed, that is, into ions, in an ion source. This source must be designed in such a way that a narrow beam of positive ions is emitted through the exit slit, S, all with nearly the same energy, or more concisely, in such a way that the ions are homogeneous. This energy is, as usual, measured by the accelerating potential \( V \). The beam then passes through an analysing field, the vital part of which is a uniform magnetic field of induction \( B \) which is created in the gap of an electromagnet. When it leaves this field, the beam will have been separated into several components which correspond to the various masses, \( m, m', \ldots \). The intensity of each component is recorded by a receiver which may be either a photographic plate or some electrical device linked to a pen-recorder; in this way, we obtain the “spectrum” of the substance, which has then only to be interpreted.

23.2.2 The Two Types of Instrument: Spectrographs and Spectrometers

The analogy with the optical spectroscope which analyses a beam of light is obvious; in the present case, the analysing field replaces the glass prism—its action on the trajectories lies in the domain of the optics of charged particles, although here the particles are not electrons but ions. The ions, each of mass \( m \), should produce as fine as line as possible to ensure that we obtain the optimum separation. The ions arrive, however, in a bundle of rays with different inclinations. Let us suppose to begin with that they are all of the same energy, \( V \); ion optics ensures their angular focusing, that is, ensures that they converge to the line image of \( S \). It is in the neighbourhood of this image, therefore, that we should place the receiver (Fig. 280).
In reality, however, these ions of mass $m$ have various energies, ranging between $V$ and $V + v$ in value; they will not, therefore, all converge to exactly the same line. The image $S$ displays chromatic aberration in the electron optical sense (§ 7.4). This is a defect which can if necessary be eliminated at the expense of a more complicated optical system; when this is done, however, ions of mass $m$ but of slightly different energies and inclinations all converge to the same line—we say that there is both energy and angular focusing or simply double focusing. This distinction will be an important one in defining the two basic types of instrument which we are about to describe.

(i) Spectrographs

The receiver consists of a photographic plate $II$, while the field remains fixed; this method is well-adapted for measuring the positions of the lines and hence for determining the values of the masses $m, m', \ldots$. For extreme accuracy, the lines must be very fine, which implies double focusing; the analysing field therefore becomes more complex as a supplementary electric field $E$ must be added (although Aston's first instrument did contain an electric field, there was in fact only single focusing, which was velocity focusing).

(ii) Spectrometers

Here, the "collector" consists of an electrode placed behind a slit $S'$ (Fig. 280); on this slit, we successively bring the ions of each of the various masses present by varying either $B$ or $V$ (magnetic or electric scanning), and we measure the corresponding currents, that is the line intensities.
It is easy to make quantitative analyses with this type of instrument; only angular focusing is required here, and this can be achieved with a magnetic field.

A spectrometer is a simpler instrument than a spectrograph, therefore, and the wide variety and the interest of their chemical applications have led to the construction of industrial models. Spectrometers also exist for separating out isotopes—in essence, the structure of these instruments is the same, but they are considerably more bulky. We quote a few normal orders of magnitude (they will be higher for the separation of isotopes).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of the trajectories</td>
<td>50 cm</td>
</tr>
<tr>
<td>Size of the ion source and the collector</td>
<td>3 cm</td>
</tr>
<tr>
<td>Width of the slits S, S'</td>
<td>&lt; 1 mm</td>
</tr>
<tr>
<td>Accelerating potential V</td>
<td>a few kilovolts</td>
</tr>
<tr>
<td>Magnetic induction B</td>
<td>a few kilogauss</td>
</tr>
<tr>
<td>Ion current carried</td>
<td>(i &lt; 10^{-9}) A</td>
</tr>
</tbody>
</table>

23.2.3 Preliminary Survey

The analysing field is the characteristic feature of a spectrocope; examining this field, we return into the domain of electron optics, the essential object of this book. In consequence, it is with the study of this field that we shall begin. After summarizing the simplest properties of the field, we shall consider the more general problem of the optics of systems with a symmetry plane, and more precisely, of the sectors with which we are effectively concerned in mass spectroscopy. We shall study their angular focusing action (on ions \(m\) of the same energy) and afterwards, the possibility of double focusing. We shall describe some of the more representative instruments, and examine their resolving power, which is the fundamental characteristic of any instrument.

23.3 Elementary Examples of Analysing Fields: Uniform Fields

23.3.1 Equations of Motion

We consider an ion of mass \(m\) and charge \(e\) which we shall suppose positive, moving with velocity \(u\); the equation of motion in an electric field \(E\) and a magnetic field of induction \(B\) has the form

\[
m \frac{du}{dt} = e(E + u \times B).
\]  

(23.1)

In the left-hand side of this equation is implicit the assumption that the motion can be described by classical mechanics, which is easily justifiable
since the ion velocity $u$ is always considerably smaller than the velocity of light in mass spectroscopy. We obtain the magnitude of $u$ directly from the conservation of energy relation which is deduced from (23.1), and takes the simple form

$$\frac{1}{2}mu^2 = -e\Phi,$$  \hspace{1cm} (23.2)

where $\Phi$ is the potential from which $E$ is derived, measured relative to the potential of the ion source where the ions are stationary. The minus sign expresses the fact that positive ions can only move through regions of negative potential. The motion of a charged particle is, we should remember, also governed by a principle of least action; the effect of a field is expressed by attributing a certain fictitious optical index $n$ to each point of space. $n$ is proportional to $|\Phi|$, and hence independent of $m$ in the case of a purely electrostatic field; in the presence of a magnetic field, on the other hand, $n$ does depend upon $m$. Given a beam of monoenergetic ions of various masses, therefore, a magnetic field is required to split it according to the different masses. More precisely, the parameter which is characteristic of each ion is $m/e$ (from equation (23.1) above); an ion of mass $m$ doubly charged behaves like an ion of mass $\frac{1}{2} m$ singly charged. For this reason, we shall deal henceforward with singly charged ions.

### 23.3.2 The Uniform Magnetic Field, the Dempster 180° Spectrometer

To begin with, we recapitulate the very simple and well-known case of particles moving through a uniform magnetic induction $B (E = 0)$; in what follows, we shall adopt the convention that $B$ acts in the vertical direction, and we shall only be interested in motion in the perpendicular or horizontal directions. This motion has an initial horizontal velocity and can easily be made precise by projecting equation (23.1) on the normal; the trajectories are circles, described at a constant angular velocity:

$$\omega_0 = \frac{e}{m} \cdot B.$$  \hspace{1cm} (23.3)

The radius $R$ depends upon $u$ and hence upon the accelerating potential, which is constant in this case, $|\Phi| = V$ so that

$$R = \sqrt{\frac{2m}{e}} \cdot \frac{V}{B},$$  \hspace{1cm} (23.4)

or if we characterize the (singly charged) ion by its molecular mass $M$ and measure in grams, volts, gauss and centimetres, we find:

$$R = 143.6 \sqrt{\frac{M}{e}} \cdot \frac{V}{B}.$$  \hspace{1cm} (23.5)
If, for example,
\[ V = 1600 \text{ V}, \]
\[ B = 1600 \text{ gauss}, \]
\[ M = 16, \]
we find that \( R = 14.36 \text{ cm} \).

We idealize the actual situation in a spectroscope by assuming that the field \( B \) is bounded laterally by a vertical plane \( II \); a horizontal ion beam of small aperture enters normally through the slit \( S \) which lies in the plane \( II \) (Fig. 280); the ions describe semi-circles inside the field and the masses separate out into well-defined circles of different radii \( R \). A simple geometrical examination shows that in addition, the trajectories which correspond to particles of the same mass converge to virtually the same vertical segment at the point where they leave the field. We say that there is "angular focusing", and it is obviously sensible to place the receiver at the point where this focusing occurs, namely in the plane \( II \), whether the receiver be a photographic plate or a collector \( S' \). This was the principle of the first spectrometer (Dempster, 1918); the magnitude of the mass \( M \) incident at any moment can then be obtained from the relation (23.5). We shall simply point out that \( M \) is proportional to \( B^2 \) when the scanning is magnetic and to \( V^{-1} \) when it is electric. If we are recording photographically, equation (23.5) gives the value of \( M \) as a function of the abscissa \( X(=2R) \) of the line: \( M \) is proportional to \( X^2 \).

### 23.3.3 \( B \) and \( E \) Uniform Crossed Fields

We consider the horizontal motion in a region in which a uniform vertical field \( B \) and a uniform horizontal field \( E \) act.

(i) *A particular solution—the Wien velocity filter*

One particular solution of the problem consists of uniform rectilinear motion at a velocity \( u_0 \) defined by
\[ E + u_0 \times B = 0. \]
The vector \( u_0 \) forms an orthonormal triad with \( E \) and \( B \) (Fig. 281a); its magnitude is given by
\[ u_0 = \frac{E}{B}. \]  \hspace{1cm} (23.6)

If we consider a pencil of ions which emerges from a slit in the initial direction \( u_0 \), the only ions which continue in a straight line and hence can be extracted with the aid of slits such as \( S_2 \) are those with a velocity \( v_0 \), irrespective of their mass. This arrangement is known as Wien's velocity filter (Fig. 281a). If we place such a filter at the exit \( S_1 \) of a very hetero-
geneous ion source, we obtain at the slit $S_2$ a beam which is homogeneous in velocity for every mass, and hence suitable for spectroscopic analysis (see, for example, Bainbridge, 1932). In reality, of course, this beam is not limited to a single ray $S_1S_2$, but has a small finite aperture.

![Fig. 281a. Wien’s velocity filter.](image)

If, on the other hand, we place a Wien filter at the exit of a source of ions which are homogeneous in energy, the ions will be separated according to their velocities, and hence to their masses; this we speak of as *velocity spectroscopy* (Oliphant, Shire and Crowther, 1934).

(ii) Arbitrary trajectories; Bleakney and Hippel’s cycloid spectrometer

We return to equation (23.1), and by writing out the components in the three directions we obtain three linear second order differential equations, the right-hand sides of which result from the existence of an electric field $E$.

![Fig. 281b. Focusing the cycloidal trajectories.](image)

From the standard theory of differential equations, we know that every possible trajectory is described by a linear combination of two solutions: the general motion in the case $E = 0$ together with a particular value of the trajectory when $E \neq 0$. 
Alternatively, we can say the general solutions represent the superposition of uniform circular motion at an angular velocity \( \omega_0 \) (equation (23.3)) onto a rectilinear uniform motion at a velocity \( v_0 \) (equation (23.6)). The trajectories are therefore cycloidal in form with a spatial periodicity

\[
D = \frac{2\pi v_0}{\omega_0} = \frac{2\pi m}{e} \frac{E}{B^2}.
\]

If we consider a beam which emerges from \( S \) (Fig. 281 b), all the ions of a given mass will converge into this point a distance \( D \) away, irrespective of their initial directions and their energy; we have, therefore, focusing which is theoretically perfect, but it is not always easy to take practical advantage of this fact. Bleakney and Hipple (1938) built a spectrometer on this principle, and more recently the Consolidated Company have made such an instrument available commercially.

### 23.4 Ion Optics in Fields with a Symmetry Plane

The fields which we have just examined are important special cases of a field-structure whose optics we are about to study. The essential characteristic is a horizontal symmetry plane (the plane of the figures). By this, we mean that at any two points symmetrically placed with respect to this plane, the electrical potential has the same value, and the scalar magnetic potential takes equal and opposite values (with symmetrically placed pole-pieces).

As a result of this symmetry, there are plane trajectories which lie in the plane of the figure, as the electric and magnetic fields in this plane are horizontal and vertical respectively. The complete ion beam will travel in the neighbourhood of the symmetry plane; it begins as a superposition of a number of components of various energies and masses, the geometry of which we are about to study.

#### 23.4.1 The Beam \((m, V)\); the Gaussian Approximation to the Trajectory Equation

A beam which is composed of ions of a given mass \( m \) and initial energy \( V \) remains \textit{a priori} in the form of a compact bundle around a "mean" trajectory \( T_0 \), which lies in the symmetry plane and which we shall designate the optic axis (Fig. 282). This clustering together has been studied in a masterly fashion by Cotte (1938) in very general conditions; we shall
sketch out the method which he used, making the simplifications which the symmetry of the present system allows.

We take a curvilinear coordinate system, in which an arbitrary point \( N \) has the coordinates \( x, y, z \), defined in the following way: we erect the plane which passes through \( N \) and is normal to the mean trajectory \( T_0 \), intersecting the latter in \( n \); \( z \) is then the curvilinear abscissa of \( n \), measured in

Fig. 282. An optical system with a symmetry plane: a beam of particles of fixed mass \( m \) and energy \( V \).

the sense of the motion along \( T_0 \). \( T_0 \) thus forms a curvilinear axis, \( Oz \). \( x \) and \( y \) are respectively the horizontal and vertical Cartesian coordinates of \( N \) from \( n \) in the normal plane, and the relative orientations are such that the triad \( x \ y \ z \) is right-handed. To be able to write down the equations of motion, we must know the metric in this coordinate system, the expression for the distance \( ds \), that is, between two infinitely close points \( N, N' \). If \( R(z) \) is the radius of curvature of \( T_0 \) at \( z \) (measured along \( Ox \)), and hence negative in the case illustrated in Fig. 282), we obtain the simple expression

\[
\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \left(1 - \frac{x}{R}\right)^2 \mathrm{d}z^2.
\]

The electric field is defined by the scalar potential \( \Phi(x, y, z) \), the origin of which is chosen in such a way that \( |\Phi| \) represents the energy of the ions. The magnetic field is defined by the induction \( \mathbf{B}(x, y, z) \), the vector with components \( (B_x, B_y, B_z) \). In the calculations, however, the vector potential \( \mathbf{A}(x, y, z) \) of the field will be required at a certain point; we represent the covariant components of \( \mathbf{A} \) by \( A_x, A_y, A_z \), and \( \mathbf{B} \) is calculated from \( \mathbf{A} \) by the relation

\[
\mathbf{B} = \text{Curl} \mathbf{A},
\]
or in this particular case†

\[
\begin{align*}
B_x &= \frac{1}{1 - \frac{x}{R}} \left( \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right), \\
B_y &= \frac{1}{1 - \frac{x}{R}} \left( \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right), \\
B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}.
\end{align*}
\]

(23.7)

To find the trajectories, we start from the principle of least action (§ 6.2.3) which takes the following form when we take into account the fact that it is with positively charged particles that we are concerned (and \( \Phi \) is negative):

\[
\delta \int L \, dz = 0, 
\]

(23.8)

\[
L = \sqrt{\Phi} \frac{dx}{dy} + \sqrt{\frac{e}{2m}} A \cdot \frac{ds}{dy}.
\]

(23.9)

Writing \( \frac{dx}{dy} = x', \ldots \), we have

\[
L = L_1 + L_2,
\]

\[
L_1 = \sqrt{\Phi} \sqrt{x'^2 + y'^2 + \left( 1 - \frac{x}{R} \right)^2},
\]

\[
L_2 = \sqrt{\frac{e}{2m}} (A_x x' + A_y y' + A_z).
\]

From equation (23.8), the Euler–Lagrange equations can be deduced:

\[
\frac{d}{dz} \left( \frac{\partial L}{\partial x'} \right) - \frac{\partial L}{\partial x} = 0,
\]

(23.10)

\[
\frac{d}{dz} \left( \frac{\partial L}{\partial y'} \right) - \frac{\partial L}{\partial y} = 0,
\]

which are differential equations in which time does not appear and from which the required trajectories can be obtained.

Since we are only interested in trajectories close to the axis \( T_0 \), we shall assume that the quantities \( x, y \) and their derivatives \( x', y' \) are sufficiently small for powers higher than unity to be negligible. As a result, the fields can be expressed as expansions of second degree which simplify, by reason

† Cf. Angot: Compléments de Mathématiques (2nd ed., page 130).
of the symmetry, to the following forms:

\[
\begin{align*}
\Phi &= q + q_1 x + q_2 x^2 + q_3 y^2 + \cdots < 0, \\
B_x &= b_1 y, \\
B_y &= b + b_1 x + b_2 y^2, \\
B_z &= b_3 y.
\end{align*}
\] (23.11)

The coefficients \(q, q_1, \ldots, b_1, \ldots\) are functions of \(z\) alone; the fact that \(b_1\) appears both in \(B_x\) and \(B_y\) is a consequence of

\[\text{Curl } B = 0.\]

It is convenient at this point to write down the expressions for the various derivatives which appear in the equations (23.10) in terms of the quantities (23.11), at least to the first order

\[
\frac{\partial L_1}{\partial x'} = \frac{1}{1 + \cdots} \frac{|\Phi| x'}{|q|} + \cdots,
\]

\[
\frac{d}{dz} \frac{\partial L_1}{\partial x'} = \left( x' \frac{\partial}{\partial x} + y' \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \frac{\partial L_1}{\partial x'} = \frac{\partial}{\partial z} \frac{\partial L_1}{\partial x'} + \cdots
\]

\[= \frac{1}{|q|} x'' - \frac{q'}{2 |q|} x',\]

\[
\frac{\partial L_1}{\partial x} = \frac{\partial}{\partial x} \left[ |\Phi| \left( 1 - \frac{x}{R} + \cdots \right) \right] = - \frac{|\Phi|}{R} \left( 1 - \frac{x}{R} \right) \frac{\partial \Phi}{\partial x}
\]

\[= \frac{2 \frac{q}{R} - q_1 + \left( 3 \frac{q_1}{R} - 2 q_2 \right) x}{2 |q| - \frac{q_1}{|q|} x},\]

\[
\frac{d}{dz} \frac{\partial L_2}{\partial x'} - \frac{\partial L_2}{\partial x} = \frac{e}{2m} \left( x' \frac{\partial A_x}{\partial x} + y' \frac{\partial A_x}{\partial y} + \frac{\partial A_x}{\partial z} \right)
\]

\[- x' \frac{\partial A_x}{\partial x} - y' \frac{\partial A_x}{\partial y} - \frac{\partial A_z}{\partial x} \right),
\]

\[= \frac{e}{2m} \left[ - y' B_z + \left( 1 - \frac{x}{R} \right) B_y \right] = \frac{e}{2m} \left[ b + \left( b_1 - \frac{b}{R} \right) x \right].\]

Substituting these various expressions into (23.10) we obtain a pair of linear second order differential equations for \(x(z)\) and \(y(z)\) the right-hand sides of which can certainly be put equal to zero, through the very choice of the axis \(Oz\) itself, as this latter is a particular solution, \(x = 0, y = 0.\)
These equations are the equations of the required trajectories

\[
2 q \dot{x} + q \dot{x} + \left(3 \frac{q_1}{R} - 2 q_2 \right) + \sqrt{\frac{2e}{m}} \left| \frac{q_1}{2} \right| b + \left| q \right| \left( \frac{b}{R} - b_1 \right) \right| x = 0 ,
\]

(23.12)

\[
2 q \dot{y} + q \dot{y} + \left(-2 q_2 + \sqrt{\frac{2e}{m}} \left| q \right| b_1 \right) y = 0 .
\]

23.4.2 Consequences: the Structure of a Beam \((m, V)\) and Angular Focusing

We might say that these equations define the trajectories close to \(T_0\) in the Gaussian approximation, and they do in fact represent a generalization of the ordinary differential equation for electron trajectories close to the optic axis in axially symmetrical electrostatic lenses (§5.2.2). Important consequences stem from these equations. First, the two equations (23.12) are independent: they give \(x(z)\) and \(y(z)\) respectively; we can therefore consider the two projections separately. From the linearity of the equations, we can deduce quite simply (§4.3) that for each projection, there are angular focusing properties. The most important concerns \(x(z)\); this we shall call horizontal focusing, and we shall discuss it in the following way. The trajectories which originate in a point, or more generally, in a vertical "object" focal line \(P\) (a slit, for example) all converge onto a vertical "image" focal line \(P'\) (which will be a narrow band); the corresponding planes (which are perpendicular to \(Oz\)) are conjugate: the Lagrange–Helmholtz relation can be applied, in the form (cf. §4.3.3)

\[
\text{angular magnification} \times \text{linear magnification} = \sqrt{\frac{q_{\text{object}}}{q_{\text{image}}}}.
\]

The case in which both the objects and their images lie in field-free regions is important, as the useful parts of the trajectories involved are straight lines, and we can perform the same geometrical constructions as in an ordinary centred system.

The focusing of \(y(z)\) which we call vertical focusing represents a similar correspondence between the horizontal focal lines in the frontal planes. In general, this new pair of conjugates does not coincide with the preceding pair, but in mass spectroscopy, no interest attaches to this vertical focusing in the standard approximation at least, as the effects due to the field boundaries are neglected. In these conditions, the field structure is further simplified, as \(\Phi\) depends upon \(x\) and \(z\) only, and \(B\) is uniform and vertical, so that the forces which act upon the particles are always horizontal; it is clear that the vertical focusing can only occur with virtual elements; rays which are initially horizontal remain horizontal.
23.5 SECTOR FIELDS; ANGULAR FOCUSING; THEIR APPLICATION IN SPECTROMETERS

The fields which are used in spectrometry are in general bounded by plane vertical faces $L, L'$, if we neglect edge effects. They can be regarded as occupying a sector of angle $\Theta$, the axis of which we denote by $K$ (Fig. 283).

![Diagram](image)

Fig. 283. Perpendicular magnetic focusing by a sector field.

and hence they have become known as sector fields. Their optical properties have been examined by Herzog (1934), Mattauch and Herzog (1934), Cartan (1937) and Mattauch (1940). In general, these properties can be obtained directly without having recourse to the general theory which we have just outlined; the latter is useful, however, for defining a priori the (horizontal) angular focusing properties.

23.5.1 The Uniform Magnetic Field $(B)$; Ordinary Spectrometers

The optic axis $T_0$ is a circular arc $C$ of radius $R$ inside the sector smoothly joined to two straight lines which represent the incident and emergent particles.
(i) "Perpendicular" focusing: the case in which $T_0$ at the entry to the sector is perpendicular to $T_0$ at the exit

This is clearly the case if $T_0$ is incident normally upon the face of the sector at a distance $R$ from the point $K$. We shall first of all establish the image focus $F'$, which is defined by the intersection of $T_0$ with a neighbouring ray, incident from the same direction (Fig. 283a). The two trajectories $HH'F'$ and $H_1H'_1F'$ can be regarded as two successive positions of the same rigid figure which consists of a circular arc of radius $R$ and a tangent to it. In the displacement involved in changing from one figure to the other, the centre of the circle $K$ moves to $K_1$, and $H'$ to $H'_1$ so that the instantaneous centre of rotation is situated at the intersection of the corresponding normals $KN$ and $H'z$. As the emergent rays intersect at this point, this is the required focus $F'$; alternatively, we can say that $F'$ is obtained by drawing a line through $K$ parallel to the direction of incidence.

The optical properties of the system can easily be obtained; if we attach a second sector to the first, similar to it but with a sectoral angle which may be different, and turned back in the opposite sense, we obtain an overall system (Fig. 283b) which clearly gives the image $P'$ of a point object $P$ which lies on the axis. The interrelation is simple: $P$, $P'$ and $K$ are collinear, which we can express as a conjugacy relation between the distances $PH = l$ and $P'H' = l'$. Let us consider in particular the pair of conjugate points $A$, $A'$ which lie on the external bisector of the sector; the corresponding angular magnification is $-1$ on account of the symmetry, and hence $A$ and $A'$ are "antiprincipal points" (between which the linear magnification is $-1$). Once the foci $F$, $F'$ and the antiprincipal points $A$, $A'$ of the system are known, the image conjugate to any given object can be constructed in the normal way.

Spectrometers with sector fields

In a spectrometer, the focusing is usually perpendicular and symmetrical, and is produced by a magnetic sector; the two slits $S$ and $S'$ will be situated at the antiprincipal points (Fig. 284). The ion beam falls on the entry face

Fig. 284. Symmetrical perpendicular focusing; a sector spectrometer.
at a fixed distance $R$ from $K$, and the collector receives ions the mass of which satisfies equation (23.4). The sectoral angles which are used are $\Theta = 90^\circ$ (e.g. Hipple, 1942) and $\Theta = 60^\circ$ (Nier, 1940 and 1947). Dempster's instrument (Fig. 280) represents the extreme case for which $\Theta = 180^\circ$.

(ii) "Oblique" focusing: the case in which the axis $T_0$ is arbitrary

Once again, we can use the simple geometrical reasoning due to Cartan (1937). First, we establish the image focus $F'$ in the case in which the optic axis is normal to the entry face alone (Fig. 285). The two neighbouring rays $HH'F'$ and $H_1H'_1F'$ by which this focus is determined are related by an infinitely small rotation about an instantaneous centre $I$ which lies at the point of intersection of the normals at $J$ (the centre of the arc $HH'$) and $H'$, at the entry and exit faces respectively. The focus $F'$ thus lies at the foot of the perpendicular dropped from $I$ onto $H'z$. To obtain the diagram which gives us the geometrical construction for conjugacy on the axis, we have simply to place a similar sector next to the first, turned back in the opposite sense. With such a construction, we can, for example, obtain the angular focusing of a mass quite different from $m$ which corresponds to the adjustment of the spectrometer shown in Fig. 284.

23.5.2 Fields with an Electric Component

An electric sector field will be, by definition, the field of a cylindrical condenser, restricted to the wedge defined by two meridional planes $L$ and $L'$. The potential will be equal to the potential outside ($-V$, say) along a
circle \( C \) of radius \( R \). The field varies as \( 1/r \), and is described by the potential

\[
\Phi = E_0 R \log \frac{r}{R} - V,
\]

in which \( E_0 \) is the value of the field on \( R \), positive if it is attractive. More generally, the case of a sector field which is at once both electric and magnetic has been considered theoretically by Henneberg (1934) and by Herzog (1934); such a field is formed by the superposition within the same sector of a uniform magnetic induction \( B \) and a cylindrical electric field \( E \) (Fig. 286).

There are ions \((m, V)\) for which the circle \( C \) is a particular trajectory inside the sector. They can easily be defined, as the velocity \( u \) corresponds to \( V \) in such a way that the normal projection of equation (23.1):

\[
\frac{m u^2}{R} = e(E_0 + u B),
\]

leads to

\[
R = \frac{2V}{\sqrt{\frac{2e}{m} V B + E_0}}.
\]

(23.13)
This relation is a more general version of equation (23.4) which concerns purely magnetic fields, but it should be realized that in practice, circular trajectories are the exception. Given ions \((m, V)\) which satisfy this condition, we shall now examine their angular focusing properties in the neighbourhood of the circle \(C\) which plays the role of the optic axis \(T_0\), in the case where the field is unbounded.

We shall apply the trajectory theory in curvilinear coordinates; the various parameters have the following values:

\[
x = r - R; \quad \varphi = -V; \quad q_1 = E_0; \quad q_2 = -\frac{E_0}{2R}; \quad \text{and} \quad b = B.
\]

Equation (23.12) for the focusing takes the simple form

\[
x'' + \frac{1 + q^2}{R^2} x = 0,
\]

in which we have written

\[
q = \frac{E_0 R}{2V}.
\]

For the trajectories close to \(C\), therefore, we have

\[
x = x_0 \cos \left\{ \sqrt{1 + q^2} (\theta - \theta_0) \right\},
\]

where \(\theta = z/R\) is the azimuthal coordinate about \(K\). An object point \(P\) therefore gives successive images \(P', P''\) at angular intervals \(\theta_f, 2\theta_f\) where

\[
\theta_f = \frac{\pi}{\sqrt{1 + q^2}}.
\]

Two important special cases are: the purely magnetic field \(q = 0\)—we have Dempster focusing, \(\theta_f = \pi\)—and the purely electric field \(q = 1\) (from equation (23.13)). The angular interval between successive foci is then

\[
\theta_f \equiv \frac{\pi}{\sqrt{2}} = 127^\circ 17'.
\]

From these results, we could deduce the optical properties of sector fields \(\Phi\) in the perpendicular focusing case by combining the sectors (Fig. 286a). As an example, we shall calculate the image focus \(F'\); an incident trajectory \(T\) which is parallel to the axis \(T_0\) is described by the following expression within the sector, with respect to \(C\):

\[
x = x_0 \cos \left( \sqrt{1 + q^2} \theta \right);
\]

it emerges at an angle \(\varepsilon\) given by

\[
\varepsilon = -\sqrt{1 + q^2} \frac{x_0}{R} \sin \left( \sqrt{1 + q^2} \Theta \right)
\]
from which we can calculate the distance $F'$ to the point of intersection with $T_0$ immediately:

$$H'F' = \frac{R}{\sqrt{1 + q^2}} \cot(\sqrt{1 + q^2} \, \Theta).$$  \hspace{1cm} (23.16)

### 23.6 SECTOR FIELDS; MASS AND ENERGY DISPERSION; DOUBLE FOCUSING; SPECTROGRAPHS

#### 23.6.1 Dispersion of the Beam for $\Delta m, v$

Of the whole beam of ions which emerges from the source, we have examined one component, which consists of the particular family of ions $(m, V)$ and which we assume to be focused perpendicularly. To obtain the structure of the whole ensemble, we shall now consider the adjacent components, $m + \Delta m$ in mass and $V + v$ in energy. The emergent rays are \textit{a priori} shifted relative to the first case, and we shall simply calculate the extent of this shift at the level of the image $P'$, which is described by the transverse vector $d = P'Q'$ (Fig. 287). The optic axis of the rays $(m, V)$ within the sector was the circle $C$, radius $R$; the optic axis of the rays $(m + \Delta m, V + v)$ is another circle $C'$, with radius $R + \Delta R$ which we shall now compute.

In the case of a purely magnetic field, $\Delta R$ can be obtained by differentiating the general expression (23.4) for the radius $R$:

$$\frac{\Delta R}{R} = \frac{1}{2} \left( \frac{\Delta m}{m} + \frac{v}{V} \right).$$  \hspace{1cm} (23.17)
When the field is purely electric, the problem cannot be solved in this way. 
$C'$ is the circle along which ions of the same energy as the incident ions, 
$V + v - E_0 \Delta R$, would travel; the equation of motion along the normal 
takes the form:

$$\frac{m u^2}{R + \Delta R} = 2 \frac{e(V + v - E_0 \Delta R)}{R + \Delta R} = e \frac{E_0}{R + \Delta R} \frac{R}{E_0}$$

($\Delta m$ does not appear), whence (with $R = \frac{2V}{E_0}$),

$$\frac{\Delta R}{R} = \frac{v}{R E_0} = \frac{1}{2} \frac{v}{V}.$$ (23.18)

As soon as $\Delta R$ is known, the mean trajectory $T'$ of the dispersed ions can 
easily be established—$T'$ is incident normally along $PH$, and subsequently 
is described by the following expression with respect to $C'$:

$$x = -\Delta R \cos(\sqrt{1 + q^2} \theta);$$

$T'$ leaves the sector at an emergent angle

$$e' = \sqrt{1 + q^2} \frac{\Delta R}{R} \sin(\sqrt{1 + q^2} \Theta),$$

and the disparity $\delta$ at the level of $P'$ which gives the distance $l'$ to $L'$ can 
easily be evaluated, and is found to be

$$\delta = \left[ 1 - \cos(\sqrt{1 + q^2} \theta) + \sqrt{1 + q^2} \frac{l'}{R} \sin(\sqrt{1 + q^2} \Theta) \right] \Delta R$$

$$= K_m \frac{\Delta m}{m} + K_v \frac{v}{V} .$$ (23.19)

Applying (23.17) and (23.18), we obtain the two most important special 
cases:

(i) **$B$ only ($q = 0$)**

$$K_m = K_v = \frac{1}{2} \{ R(1 - \cos \theta) + l' \sin \theta \},$$ (23.20)

and if in addition the focusing is symmetrical (Fig. 284), we can easily 
show that

$$K_m = K_v = R .$$

(ii) **$\Phi$ alone ($q = 1$)**

$$K_m = 0 ,$$

$$K_v = \frac{1}{2} \left\{ R(1 - \cos \sqrt{2} \theta) + l' \sqrt{2} \sin(\sqrt{2} \theta) \right\}.$$ (23.21)

$K_m$ expresses the dispersion of the beam as a function of the mass, and $K_v$ the 
chromatic aberration, which is *a priori* harmful.
23.6.2 Double Focusing

An arrangement for which \( K_v = 0 \) but \( K_m \neq 0 \) is achromatic, and hence focuses with respect both to angle and to energy; it produces, therefore, double focusing, and the ion beam converges to \( P' \) irrespective of variations in the energies of the ions. This is the case in a sector for which \( q = -1 \), which defines a particular combination of \( B \) and \( \Phi \)—we might have deduced this from the fact that for such a sector, the radius \( R \) of \( C \) takes a stationary value \((\partial R/\partial V = 0)\).

![Diagram](image_url)

**Fig. 288.** Dempster’s double focusing spectrometer.

We shall now demonstrate that the same property can be attained in a way which is advantageous from the practical point of view; rather than superimposing the magnetic and electric fields, we juxtapose them in the appropriate fashion. We suppose that the ion beam which is emitted at \( P \) passes through two successive sector fields, the first of which is electric (denoted by the suffix 1), and the second magnetic (suffix 2) as shown in Fig. 288. The object slit \( S \) produces an intermediate image \( P' \) and a final image \( P'' \) which alone is important. The beam is doubly focused if ions \((m, V + v)\) form rays of the type \( SQ'P'' \). In these conditions, as we see from Fig. 288, \( P'Q' \) is none other than the chromatic aberration of the electric sector, \( \delta_1 \) say; but \( P'Q' \) is also the chromatic aberration, \( \delta_2 \), of the magnetic sector which would focus \( P'' \) into \( P' \) in the same geometrical arrangement (reversing the sense of the motion and that of \( B \)). We see,
therefore, that $\delta_1 = \delta_2$, so that
\[ R_1(1 - \cos \sqrt{2}\Theta_1) + l_1' \sqrt{2} \sin \sqrt{2}\Theta_1 = R_2(1 - \cos \Theta_2) + l_3 \sin \Theta_2. \] (23.22)
This is the double focusing condition when the angles both of incidence and emergence are right angles.

Other arrangements can be found in which the sign of one of the members is changed. An intermediate diaphragm $D$ (Fig. 289) can be used to restrict the beam to an energy band which is narrow enough to ensure double focusing although the latter in fact only exists in first order. The condition for a given mass $m$ is not in general valid for other masses $m'$ (except through continuity if the latter are close to $m$). In other words, given the beam which emerges from $S$, the position $II_v$ at which “energies” are focused only coincides with the position $II$ of the angular focus for a particular mass $m$ (Fig. 288).

![Fig. 289. Mattauch's double focusing spectrometer.](image)

23.6.3 Application to Spectrographs

We shall mention a few spectrographs in which there is double focusing. The general theory of this type of instrument is to be found in Mattauch and Herzog (1934); see also Bainbridge (1953).

(i) Dempster's instrument (1935)

This is shown schematically in Fig. 288, and is defined by
\[
\Theta_1 = \frac{\pi}{2}, \quad \Theta_2 = \pi, \quad l_2 = 0,
\]
\[ R_1 = 8.5 \, \text{cm}, \quad R_2 = 10 \, \text{cm}. \]
Equation (23.22) leads to \( l'_1 \), and from conjugacy, we have \( l_1 = 1 \) cm. The two positions \( II \) and \( II_v \) are markedly separated from one another which is most inconvenient.

(ii) Bainbridge and Jordan's instrument (1936)

This instrument is defined by

\[
\begin{align*}
\Theta_1 &= \frac{\pi}{\sqrt{2}} = 127^\circ 17', \\
\Theta_2 &= \frac{\pi}{3}, \\
R_1 &= R_2, \\
l_1 &= l'_1 = 0, \\
l_2 &= l'_2 = R_2 \sqrt{3}.
\end{align*}
\]

(23.22) can be verified straightforwardly and \( II \) and \( II_v \) remain close to one another.

(iii) Mattauch's instrument (1936); see Fig. 289

Here, the optical arrangement is such that \( l'_1 = l'_2 = \infty \); \( S \) and \( P'' \) lie at the object and image foci of the two sectors respectively, and the double focusing condition takes the simple form:

\[
\sqrt{2} \cdot \sin \sqrt{2} \Theta_1 = \sin \Theta_2
\]

(23.23)

(in which \( l'_1 \) and \( l'_2 \) are infinitely large and equal). A particular solution is

\[
\begin{align*}
\Theta_1 &= \frac{\pi}{4\sqrt{2}} = 31^\circ 50', \\
\Theta_2 &= \frac{\pi}{2},
\end{align*}
\]

and from conditions (23.16) which provide us with the focus of the first sector we find \( l_1 = \frac{R_1}{\sqrt{2}} \). It is to be noticed that the relation (23.23) does not depend upon the values of the two radii, and is thus valid for any mass provided the arrangement is such that all the images for \( \Theta_2 = \pi/2 \) can be received; this can be achieved by halting the field at the "oblique" plane \( II \) (Fig. 289) in which the receiver, a photographic plate, is placed. It is this spectrograph due to Mattauch which has the great advantage of providing double focusing simultaneously for all masses (\( II \) and \( II_v \) are identical). The distribution of the masses \( M \) can easily be determined as a function of the abscissa \( X \) of the bands. \( M \) is proportional to \( X^2 \) as in the simplest example (Fig. 280).

(iv) Double focusing spectrometers

Spectrometers for measuring masses have also been constructed on the lines described above (Nier and Roberts, 1951; Hannay, 1954). Apart from the advantage of electrical recording, they possess that of needing double focusing only for the particular mass arriving at the collector, and hence for a single point \( S' \) alone.
23.7 THE VARIOUS PERTURBING EFFECTS

So far, we have taken for granted the validity of certain approximations and assumed certain ideal conditions, which we must now examine critically. We shall find that some minor adjustments will have to be made to the results, and we shall restrict the discussion to the most important case, namely, the purely magnetic sector.

23.7.1 The Spherical Aberration

The assumption that the rays \((m, V)\) which emerge from \(P\) converge onto a point \(P'\) (see Fig. 284, for example) is only a "Gaussian approximation": the approximation with the aid of which the equations of motion were established. We shall now apply more rigorous optical laws; the emergent rays are tangent only to an element of the caustic surface through \(P'\)—their distance at \(P'\) remains second order infinitesimal. In other words, if \(\alpha\) is the semi-aperture of the beam, the image \(P'\) is spread out into a spot by the transverse spherical aberration \(\delta_0\) which is proportional to \(\alpha^2\) (Fig. 290). For 180° focusing (Fig. 280), it is no problem to show that

\[
\delta_0 = 2R(\cos \alpha - 1) \\
\approx -R\alpha^2.
\]  

(23.24)

The same is true for symmetrical arrangements with normal incidence. For example, with \(R = 15\) cm, we have

\(\alpha = 1^\circ, \quad \delta = 0.05\) mm,
\(\alpha = 0.1^\circ, \quad \delta = 0.5\) μ.

Electric fields would give results of the same order of magnitude. \(\alpha\) is made as small as is necessary with the aid of a supplementary slit in a diaphragm, which prevents the aberration from affecting the resolving power.
It is, however, easy to correct the aberration $\delta_0$ of a purely magnetic field in a spectrometer. We have, therefore, the possibility of admitting beams with large apertures, and hence of higher intensity. All that is necessary is that the analysing field should be bounded not by plane faces but by faces $L, L'$ of the appropriate shapes. We select the object point $P$ and the image point $P'$ in advance, and hence obtain the entry face $L$ (Fig. 291a); the circular portions of the trajectories are then defined, and hence so are the exit points $H'$ which must be selected, and which define the shape of the face, $L'$ at which the particles emerge.

The solution to the problem is very arbitrary, therefore, and we extract such solutions as are most convenient. Smythe, Rumbaugh and West (1934) showed how wide parallel beams ($P$ at infinity) could be focused with the aid of a field bounded by two circular cylinders (Fig. 291b), and subsequently, other solutions have been studied by Bainbridge (1947), Hintenberger (1948), Kerwin and Geoffrion (1949) and Mayne (1952) in particular. Alternatively, plane faces can be retained (with oblique incidence), and the spherical aberration reduced not to zero but to third order infinitesimal.
In conclusion, we should mention that for the beam to be focused into a perfect point, \( P' \), it is not necessary that the object \( P \) should be a point, and hence \( S \) a very narrow slit. \( P' \) may even be produced by a caustic which is itself produced by a wide slit source (Cassignol and Nief, 1951); Cassignol (1953) has shown how the spherical aberration of the whole instrument in question can be corrected empirically.

The aberration has also been considered for combinations of magnetic and electric sectors, with a view to perfecting double focusing spectrometers (Johnson and Nier, 1953).

23.7.2 The Influence of Edge Effects of the Analysing Field

The magnetic field \( B \) can be considered as being uniform between the ideal faces \( L \) and \( L' \); it always extends beyond this region, however, in the form of a fringing field or edge effect \( \beta(z) \) (at the edges of the gap), the effects of which will be twofold.

(i) Displacement of the horizontal focusing (Coggeshall and Muskat, 1944; Ploch and Walcher, 1950)

We consider one of the rays of the beam which is being focused perpen-

![Fig. 292. The influence of the overlap field on the focusing.](image)

-dicularly (Fig. 292); the incident portion \( PH_1 \) is not perfectly rectilinear, but satisfies the equation

\[
\frac{m u^2}{q} = - m u^2 x'' = e u \beta(z),
\]
in which \( z \) and \( x \) are measured with respect to the initial direction \( PH \). At \( H_1 \), therefore

\[
x' = -\frac{\eta}{R},
\]

where we have written

\[
B \eta = \int \beta(z) \, dz.
\]

\( \eta \) is to a certain extent the width of the region of uniform field which integrates to the same value as the fringe field; \( \eta \) is of the order of magnitude of the width of the gap. It is exactly as if the incident ray \( PH_1 \) had been replaced by a ray \( P_1 H_2 H_1 \) emerging from a well-defined object \( P_1 \) and the actual field replaced by an ideal field \( B \) (without boundary effects) the boundary \( L_1 \) of which was situated a distance \( \eta \) from \( L \); the position of \( P_1 \) relative to \( P \) can equally well be calculated, but the shift is smaller. We proceed in the same way at the exit; the two points \( P_1, P'_1 \) are collinear with \( K_1 \) according to the properties of ideal focusing, and Fig. 292 shows the resulting appearance of the focusing of \( P \) at \( P' \). To obtain the image of \( P \) at \( P' \), we must move the thin edge \( K \) with respect to the line \( PP' \) through a distance of the order of magnitude of \( \eta \), and hence of the gap (subject to possible empirical adjustment).

(ii) The appearance of vertical focusing

Outside the symmetry plane, the fringe field has a horizontal component. In particular, we consider its projection on the principal normal to the mean ray \( T_0 \) of the beam: \( \beta_x = b_1 y + \cdots \). If this beam is oblique with respect to the faces \( L \) and \( L' \), \( \beta_x \) is not zero and the equation of the vertical component of the trajectories is, from (23.12),

\[
y'' - \sqrt{\frac{2e}{mV}} b_1 y = 0.
\]

The trajectories are slightly curved, therefore, and hence slightly focused vertically. A horizontal incident beam does not remain exactly horizontal if the focusing is oblique. This property, which was underlined by Cotte (1938) in his general theory, is of great interest in \( \beta \)-ray spectroscopy; here, we attempt to accentuate the phenomenon (with mixed fields) so that the beam is focused simultaneously in the horizontal and the vertical directions.

A point object \( P \) would then give a point image \( P' \) (Siegbahn and Svartholm, 1946 and 1950). This seems to be of little interest as far as mass spectroscopy is concerned.

23.7.3 Space Charge

We shall summarize only briefly the other possible perturbing influences. The effect of space charge (mutual repulsion between the ions) is to oppose the convergence; the effect is perceptibly larger for an ion beam than it is
for an electron beam of the same energy and current density \( i \). In fact, since \( i = q u \), we can see that the charge density \( q \) is higher for ions, as heavier particles move more slowly for a given energy, and thus the velocity \( u \) is smaller.

In mass spectroscopy, the perturbation is always negligible, at least with ordinary values of the current, \( i \), in the region of \( 10^{-6} \) A; for beams in which \( i = 10^{-6} \) A, however, the effect of space charge is no longer negligible, and is reflected in an increase in the distance between the object and the image (Walcher, 1949; and Robinson, 1949). The intense beams which are used in spectrometers for the separation of isotopes (\( i \) of the order of 1 mA) require particularly careful attention. Higher voltages may be used, for example (30 kV or more), or the space charge may be destroyed by admixing slow electrons (Bernas, 1953; Bernas, Kaluszyner and Druaux, 1954).

Finally, we must realize that various types of aberration may result from imperfections in the construction of the instrument: the alignment may be defective, the vacuum inadequate (the ions will be scattered by the residual gas), the conductors may be covered with a film of impurity which is liable to become charged, the source of H.T. or the fields may be unstable and produce fluctuations of the beam. There are no diffraction effects however, as the wavelengths associated with ion beams are so very short (due to the rather large masses).

### 23.8 The Resolving Power of Spectrometers and Spectrographs

#### 23.8.1 The Width of the Line; the Resolving Power of the Field

The problem is to obtain a measure of the ability of the various instruments to discriminate between ions of different masses. If \( m + \Delta m \) is the mass closest to \( m \) which can be distinguished by the analysing field, we define the resolving power of this field for the mass \( m \) to be

\[
\left| \frac{m}{\Delta m} \right|
\]

The object is the exit slit \( S \) at the ion source, of width \( s \). We shall first of all calculate the width \( s_1 \) of the line on which ions of mass \( m \) fall in the plane of observation which is conjugate to \( S \). If \( G \) is the corresponding optical magnification, the main component of the width is given by the Gaussian approximation, \( Gs \); to this, however, we must add the chromatic aberration which is due to the energy spread \( v \) of the beam. If the aperture \( 2x \) is suffi-
ciently small and the construction sufficiently careful for there to be no other causes of aberration,

\[ s_1' = Gs + K_v \frac{v}{V}. \]  

(23.27)

The band defined by particles of mass \( m + \Delta m \), however, lies a distance \( K_m \frac{\Delta m}{m} \) away. We usually agree to call the bands separated if this distance is equal at least to the width \( s_1' \); this produces the following expression for the reciprocal of the resolving power of the field:

\[ \left| \frac{\Delta m}{m} \right| = \frac{G}{K_m} s + \frac{K_v}{K_m} \frac{v}{V}. \]  

(23.28)

23.8.2 Spectrographs

The resolving power of spectrographs with double focusing is obtained by putting the second term equal to zero. If, for example, we apply this formula to the spectrograph of Bainbridge and Jordan, where the focusing is magnetic (radius \( R_2 \)) and symmetrical (antiprincipal points), we have \( K_m = R_2 \) (from equation 23.20) and \( G = 1 \), so that

\[ \left| \frac{m}{\Delta m} \right| = \frac{R_2}{s}. \]  

(23.29)

The other arrangements give less straightforward calculations, but the preceding relation gives the order of magnitude. We should remember, however, that only in Mattauch's instrument is the double focusing property valid for all values of the mass. For example, we obtain a theoretical resolving power of about 40,000 if \( R_2 = 20 \text{ cm} \) and \( s = 5 \mu \); this demands a semi-aperture \( \alpha \) of the order of 0.1° or 0.2° if the spherical aberration is not to appear (see § 22.7.1).

23.8.3 Spectrometers

For purely magnetic analysing fields, \( K_m = K_v \) and from (23.28) we find

\[ \left| \frac{m}{\Delta m} \right| < \frac{V}{v}. \]  

(23.30)

As an examination of types of ion source shows, the relative heterogeneity \( v/V \) can hardly be less than \( 10^{-3} \); the resolving power of the field can barely exceed 1000, therefore, a far lower value than the resolution which can be obtained with double focusing in spectrographs. As a result, the width of the object slit \( s \), which always appears as the ratio \( s/R \) (or a term of this order) only begins to appear in the expression for the resolving power for widths of about \( 1/10 \text{ mm} \), and there is no advantage in choosing it any narrower.
The resolving power of the field (23.28) will not, generally speaking, be that of the apparatus. The spectrometer here receives the useful part of the beam in the slit $S'$ of width $s'$ of the collector; this slit must be at least as wide as the emergent beam, and we may even choose an appreciably wider value, $s' > s'_1$ for reasons which we shall see later. The resolution criterion for the whole spectrograph therefore requires that masses $m$ and $m + \Delta m$ should not fall on $S'$ simultaneously, so that the resolution of the spectrometer itself is less than that of the field alone (23.28). The former is defined by

$$K_m \left| \frac{\Delta m}{m} \right| = s'$$

if $s \ll s'$, but if this condition is not fulfilled, we can easily take account of $s$; in the usual cases, then,

$$\left| \frac{m}{\Delta m} \right| = \frac{R}{s'}.$$

If, for example, $R = 15$ cm and $s' = 0.5$ mm, the resolving power is about 300. This value is for the most part sufficient in the everyday problems of mass spectrometry.
CHAPTER 24

THE INDIVIDUAL PARTS OF A
MASS SPECTROSCOPE

24.1 THE ION SOURCE

We shall first of all study the ion sources which are designed to transform the substance to be analysed into positive ions; so far as is possible, the representatives of each mass \( m \) will move at the same velocity (and thus have the same energy \( V \)), and the whole bundle of ions will emerge through a slit \( S \) in a narrow beam which is the ion object for the analysing field. Every source, therefore, consists of two more or less distinct parts:

(i) The source itself, at which the ionization process is initiated, and which is as nearly equipotential as possible so that the ions are closely homogeneous.

(ii) A mechanism for collimating and accelerating the particles through a potential \( V \), with which an ion beam of the required energy is produced.

Various designs have been proposed, suitably adapted for solids, gases and vapours. We shall only consider ion sources for mass spectroscopy, however.

24.1.1 Sources Produced by Discharge in a Gas

(i) *The ordinary luminous discharge (for gases)*

The earliest of the sources of positive ions is the ordinary luminous discharge in a gas at a pressure \( p \) of the order of \( 10^{-2} \) mm Hg (the dimensions of the envelope being considerably larger than the mean free path). The cathode is usually perforated by a narrow canal (less than a millimetre in diameter) through which a beam of positive ions which comes from the anode can pass; these are Goldstein's "canal" rays (see Fig. 293a).

This type of source possesses several defects which limit its usefulness. First, the ions are not created in an equipotential region, as their production takes place in the same region as their acceleration; this has the fatal result of producing an ion beam which is extremely heterogeneous in velocity. The energy spread \( v \) is an appreciable fraction of the mean energy \( V \) which is obtained at the exit; if \( V \) is of the order of a few thousand volts, \( v \) is about 1000 V. We cannot use a source of this kind in spectroscopy
therefore, without either inserting a velocity filter which eliminates the
majority of the rays, or using double focusing which has the advantage
of retaining a wider band of energies. Further, the ions which are obtained
are widely assorted, as the phenomena which occur at the discharge are
extremely diverse; single or multiple ionization may occur, with or without

![Diagram](image)

Fig. 293a. A discharge source: the canal rays.

rupture of the molecules, and charge exchange, condensation, and chemical
reactions between ions and molecules are all possible. A discharge
in oxygen, for example, produces many ions for each isotope: \( \text{O}^+, \text{O}^{++}, \text{O}_2^+, \text{O}_3^+ \ldots \). If nitrogen too is present, combinations such as \( \text{NO}^+ \)
will also occur. The spectra, therefore, are rich in bands, which can be
used as reference positions and for the measurement of masses \( m, m' \ldots \)
by the doublet method in spectrography.

On the other hand, such a source is not suitable for quantitative analyses,
for spectrometry; the spectra become too complex, and—a more important
objection, this—the superposition rule is no longer valid. The spectrum
of a mixture is not obtained by superimposing the spectra of its constituents.

The intensity \( i \) of the ion beam which emerges from \( S \) is much less than
1 mA. Various improvements have been made with a view to increasing \( i \)
or to working at a lower pressure \( p \) (the superposition of an auxiliary
magnetic field or use of an HF potential); the sources become more complex,
however, and they are not used for spectroscopy. They are used to supply
particle accelerators, giving intense beams of atomic ions. Several studies
have been devoted to these discharge processes—an analysis of this work is
to be found in the book by Ewald and Hintenberger (1953).

(ii) The gas arc

An arc is a discharge in which the cathode is rendered incandescent,
either spontaneously, or by auxiliary heating which results in a sharp
increase in the intensity of the current and reduces the potential across
the terminals (1 A, 50 V for example). The current \( i \) is larger and the ions
are considerably more homogeneous (v = 10 V, for example); acceleration is in this case a separate operation. Arrows are too complex in construction and too brutal in action to be used in spectroscopy; the only instruments in which they are used are those for separating isotopes where an ion current of a few millamps is used (Koch and Bendt-Nielsen, 1944; Bernas, 1953), and as ion sources for accelerators such as cyclotrons and synchrocyclotrons.

24.1.2 High Frequency Sparking (for Solids)

A general method for producing ionization of solids is furnished by high frequency spark discharge which was first applied to mass spectroscopy by Dempster (1935). The spark flies between two tapered electrodes or between an axial electrode A and a plate in which an opening has been cut (Fig. 293b). A is composed of the material to be ionized, or of a metal tube in which the material is enclosed in powder form. The pair of electrodes L and S take care of the collimation and the acceleration. The ions which are obtained have a noticeable energy spread v, although less wide than for canal rays—a velocity filter or double focusing is required. Here again, a wide diversity of ions is produced, which gives the method its spectrographic importance, as it complements to a certain extent the luminous discharge—the two phases involved are the solid phase and the gaseous phase respectively.

24.1.3 The Surface Ionization Source (for Solids)

From certain solids, we can instigate evaporation of positive ions by placing the solid over an incandescent anode. This phenomenon occurs for elements which are readily transformed into ions, and which are thus
electropositive. We denote the ionization potential by \( V_i \), and the potential barrier at the metal support by \( V_s \); the positive emission varies with \( (V_s - V_i) \); the following formula, known as the Langmuir formula (1925), expresses this more precisely:

\[
\frac{n'}{n} = \exp \left[ \frac{F(V_s - V_i)}{RT} \right]
\]  

(24.1)

\( n \) and \( n' \) are the numbers of neutral atoms and positive ions which are evaporated, respectively; \( F = 96,500 \) coulombs = 1 faraday.

A tungsten filament, for example \( (V_s = 4.55 \text{ V}) \), is perfect for obtaining a beam of Cs\(^+\) ions \( (V_i = 3.86 \text{ for Cs}) \); for alkali metals, alkaline earths and rare earth metals, oxidized tungsten \( (V_s = 8 \text{ V}) \) must be used. These sources were first explored by Dempster (1918) and Moon and Oliphant (1932). Alternatively, we can heat not the element which is to be ionized but one of its compounds—the halides (Koch, 1936) or the alumino-silicates (Blewett and Jones, 1936); this technique has been developed extensively by Couchet (1954). Intense sources \( (1 \text{ mA cm}^{-2}) \) have been studied by Kunsmann (1927), and employed in particular by Smythe, Rumbaugh and West (1934) in their wide beam instrument designed to separate isotopes (Fig. 291). The theory of the emission from compounds is still virtually unknown, but the practical operation is quite satisfactory.

These surface ionizers have a very distinct character. It is clear, first of all, that they are very selective, and do not give parasitic ions since the ionization phenomenon is strictly localized at the anode. Further, as the latter is to all intents and purposes equipotential, the velocity spread \( v \) is very small, as it is equal to the spread of the initial thermal energies, about 0.2 V. The ion beam can be formed in the same way as before (Fig. 293b, electrodes \( L \) and \( S \)).

Certain negative ions can be evaporated from a hot cathode—in this case, it is with electronegative elements that we are concerned (halogens); these have been studied by Hintenberger (1951) in particular.

24.1.4 Sources Produced by Electron Bombardment (for Gases and Vapours)

The ions emitted by this type of source are produced by electron bombardment and the range of applications is very extensive; sources of this kind were first used by Dempster (1916) and later developed, principally by Bleakney (1932), Tate and Smith (1934) and Nier (1937 and 1947).

(i) Description and properties of an ordinary source

The standard model is due to Nier; diagrams of the horizontal \( xOz \) projection and the vertical \( yOz \) projection are shown in Fig. 294. The substance to be studied is a gas which arrives through the tube \( G \) or a
vapour which is produced by heating the appropriate solid in a small adjacent compartment (Bernas and Nier, 1948; Louw and de Villiers, 1954; and Cameron, 1954). The gas penetrates at very low pressure \( p \) (10\(^{-4} \) to 10\(^{-5} \) mm Hg) into a small ionization chamber \( C \), where it is ionized by an electron beam which is emitted by a heated tungsten ribbon \( t \). The accelerating potential \( V_e \) of the electrons is applied between \( t \) and \( C \)—it must, of course, exceed the ionization potential of the gas being analysed, and is usually set between 50 and 100 V. The current \( I_e \) of the beam is a few hundred microamps. We shall see later that ions which are created in this way have a very low kinetic energy \( (V_e) \); a fraction of these ions has then to be extracted from \( C \) through the slit \( S_e \), and accelerated and aimed by the field of the collimating and accelerating electrodes \( L \) and \( S \). This can be done with the aid of an auxiliary electrode \( r \), for example, which is held at a positive potential \( V_e \) of a few volts with respect to the walls of \( C \); \( r \) produces a field which has the desired effect upon the ions without sensibly modifying the electron beam. This latter is considerably more perturbed by the divergent lens effect produced by the slit through which the beam enters \( C \). To prevent the beam from spreading, a vertical magnetic field

Fig. 294. Nier's source.
of a few hundred gauss is often employed; the effect of such a field, we recall, is to wind the electron trajectories in helices around the lines of force, and hence to guide them in a direction which is on the average vertical. This operation is controlled with the aid of the current $I$, which is incident upon a collector electrode or trap $t'$ which is placed at the exit from $C$ opposite $t$. Finally, since the chamber $C$ is pierced with several openings, it is important that the areas of these latter should be as small as possible; it must be possible to maintain the internal pressure $p$ without harming the vacuum outside which is at most of the order of $10^{-6}$ mm of mercury (cf. § 23.4.2).

A source built in this way can be easily and reproducibly adjusted. As the chamber $C$ is almost equipotential, the velocity spread $v$ is only a few volts; this spread is due to $V$, and to the initial kinetic energy $V_0$ of the ions which is always very slight—in general, it is a question of the mean thermal energy of the particles (about $4 \times 10^{-2}$ V at normal temperatures). On the other hand, the pressure $p$ is low enough here for each ion of the beam to be the result of a single process; unlike the case of a luminous discharge there are no "reactions" in $C$. The spectra obtained obey the superposition law and are susceptible to quantitative analysis, and therefore to spectrometry. We shall see that the ion currents $i$ which are obtained are of the order of $10^{-3}$ A at most, which is normally perfectly adequate.

(ii) High intensity sources

More intense sources have been produced. The procedure known as oscillating electrons represents a large advance, by extending the length of the path of the ionizing beam considerably (Finkelstein, 1940). An application of this procedure is to be found in the source due to Heil (1943), which we can describe by supposing that in Fig. 294, the trap $t'$ is replaced by a second emissive ribbon at the same potential as the first ribbon, $t$. The reflecting electrodes are distributed on one side and the other in such a way that each ionizing electron traverses the chamber $C$ a large number of times. The increase in intensity can also be obtained by arranging that a higher proportion of ions emerges from the ionization chamber; this can be done by directing the electron beam along the axis and concentrating the resulting ions by means of suitable electrodes at the exit slit (Cassignol, Geller and Moreau, 1953).

Finally, other types of intense source can be obtained by intensifying the ionizing beam $I_e$ and increasing the pressure $p$; the behaviour then tends towards that of arcs, however, and the spectra no longer obey the superposition law. We might mention as examples the sources of Walcher (1944) and Bernas and Nier (1948) which can be used for isotope separation—the currents $i$ can reach a milliampère.
(iii) Various ionization processes

It is useful to introduce the reader at this point to the various ionization phenomena which are produced at the source by the electron bombardment; for further details, the article by Hagstrum and Tate (1941) or the book by Robertson (1954) should be consulted. For the present, we shall examine a simple example, namely, carbon monoxide. This diatomic gas produces the following ions (e = electron):

\[
\begin{align*}
\text{CO} + e & \rightarrow \\
\text{CO}^+ & + 2e \\
\text{CO}^{++} & + 3e \\
\text{C}^+ & + O + e \\
\text{C}^+ & + O + 2e \\
\text{C} & + O^+ + 2e \\
\text{C} & + O^- 
\end{align*}
\]

(24.2)

(the reaction CO + e → C + O + e can obviously not be detected here). The frequency of each of these reactions, that is, the probability with which each occurs, can be calculated; this probability is often defined by the number of times \( n \) that the reaction in question is produced by an electron over a distance of 1 cm of its trajectory at a pressure of 1 mm of mercury. At the pressure \( p \) of the ionization chamber \( C \), the order of magnitude of \( n \) is at most 10; only a small fraction of the electrons is useful, therefore—we shall restate this conclusion more precisely a little later when we are examining the yield of a source.

The determination of \( n \) is one of the most important applications of the mass spectrometer; in particular, we study the variation of \( n \) as a function of the energy \( V_e \) of the ionizing electrons. For the reaction mentioned above, the following results are obtained: (i) and (ii) are ionizations in the proper sense of the word, while (iii) (iv) and (v) are ionizations with simultaneous destruction of the molecule; these five reactions can occur beyond a threshold potential \( V_t \), which is the ionization potential or more generally, the potential at which the corresponding ions first appear. For (i) and (ii) the values of this potential are 14 and 41 V respectively. The variation of probability as a function of \( V_e \) is of the form labelled \( a \) in Fig. 295. Reaction (vi) is a capture accompanied by destruction of the molecule, and its probability curve is different in form; the function has a maximum (or “resonance”) for \( V_e = 9.5 \) V (see Fig. 295, curve \( b \)). CO thus gives the positive ions which have the following integral masses when \( V_e \) is of the order of 100 V, and the various isotopes have been extracted: 12 (C\(^+\)), 16 (O\(^+\)), 14 (CO\(^++\)) and 28 (CO\(^+\)). If we were to analyse the negative ions in the spectroscope (and for this, we should have simply to change the sign of the potential \( V_t \)), we should obtain only 16 (O\(^-\)) which is a considerably less characteristic result. Here we have the first justification for the preference
which is accorded to positive ions; another reason is that the reactions which result in negative ions have markedly smaller probabilities than the others.

![Diagram of ionization curve](image)

**Fig. 295.** The two types of ionization curve.

The simple example of CO gives some indication of the complexity of the spectra which will be produced by polyatomic molecules, the hydrocarbons for example. This complexity is sometimes still further increased by the fact that some ions are metastable, by which we mean that they are not extant throughout the whole passage through the spectroscope; it is the ions which result from their dissociation which are responsible for the appearance of lines which seem to correspond to non-integral masses in the spectra (Hipple, Fox and Condon, 1946; and Hipple, 1947).

The initial kinetic energy \( (V_e) \) of the ions which are produced is necessarily equal to the thermal value for straightforward ionizations, such as (i) and (ii), by virtue of the elementary laws of collision in mechanics. This is no longer true when the ionizations are accompanied by breakdown of the molecules, as the ions which are produced in certain reactions may possess a supplementary kinetic energy; this is at most of the order of one or two volts, however, and hence only slightly disturbs the homogeneity of the complete ion beam.

### 24.1.5 Focusing and Accelerating the Ions in a Source

As an example, we consider a source due to Nier (Fig. 294), which behaves very much like every other source of fairly homogeneous ions. The problem is, we recall, to produce an ion beam of energy \( V \) and small aperture \( 2\alpha \), which emerges through an exit slit \( S \). The latter is earthed, as this helps the insulation of the ionization chamber \( C \) which is held at the
positive potential \( V \). With the aid of an electrode with a slit \( L \) which is supplied by a potentiometer, the concentration onto \( S \) of the ions which are produced in \( C \) can be improved; simple ion optics shows how this can be done.

(i) The existence of lenses

All the slits are vertical; we shall always be working in conditions where the optical system possesses a "horizontal" symmetry plane, that of the analysing field. The optics is here further simplified (Fig. 296); on the one hand, the focusing is only horizontal since the fields depend only upon \( x \) and \( z \) so long as the slits are narrow enough; otherwise we should expect to find vertical focusing (Bertein, 1950). On the other hand, the source possesses a vertical symmetry plane also, so that the optic axis \( Oz \) is rectilinear; the coordinate system \( Oxyz \) reverts to the Cartesian form. If we use the same notation as in the preceding chapter, the focusing equation (23.12)—in the Gaussian approximation—is reduced to:

\[
2 \phi x'' + \phi' x' + \phi'' x = 0.
\]

(24.3)

This could be demonstrated directly very simply by modelling the analysis on that of the optics of axially symmetric systems; the Laplace equation gives

\[
\phi_2 = -\frac{\phi''}{2}.
\]

The slits play the role of a lens, a role which has been analysed to various approximations (Bertein, 1950, 1951; and Vauthier, 1950, 1954). A particularly simple approximation to the potential distribution on the axis, \( \phi(z) \), is
the distribution created by the juxtaposition of the uniform fields, \( E_0, E_1, E_2, \ldots \) which would be produced by diaphragms without openings; this in fact is Gans' method (§ 4.5.2). In such conditions, the rays are composed of parabolic arcs, joined at known angles of refraction (Davisson and Calbick, 1932) at each diaphragm. We can, however, put less confidence in this kind of analysis for the trajectories within the chamber \( C \); here, we shall simply remark that \( C \) forms a kind of immersion objective (§ 8.3 and Chapter 13) with the extractor field \( E_0 \), which creates an ionic object at \( S_0 \). \( L \) has then to image \( S_0 \) on \( S \), a requirement which we can analyse by establishing a general conjugacy relation:

\[
\frac{E_1}{1 - V'_1} - \frac{E_2}{1 - V'_2} + E_1 - E_2 = 0, \tag{24.4}
\]

in which \( E_1 \) and \( E_2 \) are the fields on one side and the other, \( V'_1 \) and \( V'_2 \) are the potentials at the object and the image due to these fields; the unit of potential is the potential of \( L \), and the origin is the zero of these energies.

In the present application,

\[
V'_1 \simeq 0, \quad V'_2 = \frac{V}{V - V_2}.
\]

An examination of the linear magnification \( \gamma \) shows that it is possible to select the position of \( L \) in such a way that the passage of the rays through \( S \) is assured, when the width of the two slits \( S_0 \) and \( S \) are comparable and of the usual order of magnitude, namely, a few tenths of a millimetre.

With the aid of the Helmholtz–Lagrange relation, however,

\[
\gamma \frac{\alpha}{\alpha_0} \sim \sqrt{\frac{v}{V}} \sim \frac{1}{30}, \quad \text{for example.} \tag{24.5}
\]

We can foresee that when the concentration is strong (\( \gamma < 1 \)) there is a risk of the angle \( \alpha \) being too large for the analysing field to produce good focusing. The sources of Nier (1947)—see Fig. 300—include stops which are placed behind \( S \) precisely to limit the aperture of the beam.

(ii) *The influence of the magnetic field of the source (of induction \( B_0 \))*

The vertical field exerts a transverse force on the ions, parallel to the axis \( O \alpha \); the result is simply a slight horizontal displacement of the trajectories and hence of the focusing. Since the slit \( S \) is narrow, the ions could miss it; they can, however, always be brought back to \( S \) by applying a small potential difference between the two half-plates of which \( L \) is composed. We should notice that this phenomenon depends upon the mass \( m \) of the ions, so that we shall find discrimination.
24.1.6 The Yield from a Source and Its Perturbations

(i) Calculation of the ion current, \( i \), at the exit, \( S \)

We shall always talk in terms of a source of the standard form. We suppose that a given gas enters the ionization chamber at a pressure \( p \). If the electron beam is of intensity \( I_e \) and the mean length is \( l \), while \( n \) is measured with respect to the total ionization, the ion current created in \( C \) is

\[
I = n I_e l p. \tag{24.6}
\]

For everyday orders of magnitude: \( l = 1 \, \text{cm}, \, p = 10^{-4} \, \text{mm Hg}, \, n = 5 \) and \( I_e = 100 \, \mu\text{A} \), we find \( I = 5 \times 10^{-8} \, \text{A} \).

It is, however, clear that the majority of the ions does not reach the slit \( S_0 \), so that the current \( i \) which emerges through \( S \) is only

\[
i = K I \quad \text{where} \quad K \ll 1. \tag{24.7}
\]

\( K \) is difficult to calculate, as the process of extraction of the ions from \( C \) is complex. It seems reasonable to take \( K \) of the order of \( \frac{1}{100} \) at most, which would give \( i \approx 5 \times 10^{-10} \, \text{A} \).

(ii) Discrimination

The formulae (24.6) and (24.7) above can be applied equally well to the total ion current and to the current of any one particular type of ion.

If ions of one mass are emitted by the source more easily than those of another mass—if the ratio \( K = i/I \) depends upon the nature of the ions, that is—we say that there is discrimination within the source. This discrimination may have several origins, as \( K \) is an overall measure of a complex type of transmission.

In spectrometry where the potential \( V \) is scanned, the value of \( V \) which corresponds to reception is higher for light ions than for heavy ions; the same is true, therefore, of the values of the field \( E_0 \) by which the extraction is for the most part produced. The discrimination, in consequence, favours light ions (Coggeshall, 1944; Washburn and Berry, 1946). We can estimate the importance of this phenomenon by comparing the results of an analysis of \( \text{CCl}_4 \) obtained by the two methods of scanning (Blears and Mettrick, 1947).

The units are arbitrary (see the table on page 826).

When the scanning is magnetic, the magnetic field \( B_0 \) of the source is a cause of discrimination, at least in the normal situation where it is held constant; the mass \( m \) for which the half-plates have been corrected tends to be favoured with respect to the others. This defect can be overcome by using only relatively low values of \( B_0 \), of about 100 or 200 gauss.
Another type of discrimination affects ions which are formed with an initial kinetic energy, as this energy clearly modifies the trajectories within C, and hence the extraction phenomenon.

In fact, however, we shall see later on that if it is important to reduce discrimination in the measurements of the relative isotope abundance as far as possible, this is no longer true in actual chemical analyses. In the latter case, the important thing is that we should work in constant conditions, and hence we must struggle to overcome the parasitic processes which are more harmful than simple discrimination.

(iii) *The various parasitic effects*

It is the fluctuating nature of these effects which is dangerous. We classify them as hysteresis, "memory" effects, and ageing; we shall mention only the most important.

The substances of which the source is composed tend to be outgassed in the course of operation, because of the relatively high temperature at the ribbon $t$ (which can reach $150^\circ$) and the electron bombardment of the surfaces. Parasitic lines appear in the spectra.

The gas to be analysed may be thermally modified (dissociation) in the vicinity of $t$. This phenomenon varies in time with the ageing of $t$ or its possible replacement.

The gas may affect the emissive power of the ribbon itself (by modifying the work function); the hydrocarbons may increase it by one-third, while oxygenized substances have an even clearer reducing effect. Adjustment of the source electrically considerably attenuates these perturbations.

There is a risk of the metal surfaces becoming charged, by being covered with films of impurity; the trajectories are modified, especially in C.

If we are to operate in the constant conditions which are necessary for analyses, these defects must be reduced as much as possible if fluctuations which clearly exceed 1 per cent are to be avoided. The metal parts should be made of tantalum, stainless steel, or nichrome and the insulators of pyrex or melted silica. The source and the analysing tube must be dehydrated (250 to $300^\circ$), especially after exposure to air.
It is recommended that the emissive ribbon should be conditioned and for further details, the reader should consult the book by Barnard.

24.2 COLLECTION OF THE IONS

24.2.1 The Photographic Plate (Spectrographs)

The ion beams which emerge from the analysing field can be made visible with the aid of a fluorescent screen composed of willemite or zinc sulphide, provided their energy ($V$) is at least of the order of 5 kV. For recording the spectra in the spectrograph, the photographic plate has been in regular use since its introduction by Thomson (1911) in his parabola method. As a result of the very low penetration by the ions, useful emulsions contain very little gelatine, and are of very fine grain (for example, Schumann or Ilford plates). The potentials $V$ should be relatively high, for example 20 kV or more. The exposures last at least a few minutes.

Once the law which describes the blackening of the plate had been determined experimentally, we could in principle measure the intensities $i$ and hence arrive at spectrometry proper—this would be less straightforward than making electrical measurements at the collector, however. We might mention that the darkening is not strictly proportional to the product $iV$, which measures the number of ions received (Lichtblau and Mattauch, 1941).

24.2.2 The Collector (Spectrometers)

By scanning either the magnetic induction $B$ or the potential $V$, the various beams are brought successively into a small Faraday cage $C$ which is placed behind a slit $S'$. $C$ thus captures the current $i$ which corresponds to each mass in turn (Fig. 297). The width $s'$ of $S'$ must be chosen to be at least equal to the width of the line $s_1'$ and in fact, we use a decidedly higher value, so that the beam $i$ is wholly received not only for a particular value of $B$ (or of $V$) but for a small range of neighbouring values; in these conditions, each peak which is recorded has a flat top, which simplifies the interpretation.

The resolving power of the instrument is affected by $s'$. The slit $S'$ is earthed, and the same is effectively true of $C$ (see below).

The addition of an intermediate stop $D'$ which is held at a negative potential of the order of 20 V prevents the secondary electrons which are produced at $C$ by ion bombardment from leaving, as their departure would falsify the measurements of the current $i$. Another diaphragm $D''$, held at an appreciable positive potential (smaller than $V$, however), improves the spectrum by eliminating the continuous background which is produced by ions which have been scattered by the residual gas. Since the energy of
these scattered ions is less than $V$, $D''$ creates a potential barrier which prevents them from reaching the collector, without disturbing the useful ions at all (Nier, Stevens et al., 1947).

A more drastic means of clarifying the spectra consists in holding the whole collector at the same potential as the source; the most convenient arrangement is to connect both to earth, and the analysing tube will then have to be connected to the high tension ($-V$). This arrangement has been used to study the ions which are produced by the dissociation of metastable ions which have been formed in the source ($\S$ 24.1.4. See Hipple, Fox and Condon, 1946; and Hipple, 1947). It transpires that all such ions have an energy smaller than $V$, just as do the scattered ions.

Rather than try to prevent the emission of secondary electrons, we can, on the contrary, take advantage of it as a means of amplifying the value, $i$, of the current received considerably; it is possible to achieve this by placing a multiplier in front of the Faraday cage (Allen, 1939, 1950; Nier, Roberts and Franklin, 1949; and Barnett, Evans and Stier, 1954). The incident current will, of course, still have to be amplified by a valve circuit, but the value of electron multiplication resides in the fact that the importance of the background noise is reduced, to such an extent that the limiting value of $i$ which can be detected is reduced from $10^{-15}$ to $10^{-17}$ A. This is a well-known phenomenon, which was first put to use in television where the particles to be detected are photons. In the present case where ions replace the photons, the theory is the same (Zworykin and Morton, 1955) but to obtain long-term stability seems to be difficult as the flow of ions damages the first plate of the photomultiplier.

24.3 THE ELECTRICAL CIRCUITRY OF A SPECTROSCOPE

The circuits are fed from an alternating supply, in commercial instruments at least. Arrangements for rectifying and smoothing the potential must be provided therefore, to avoid fluctuations prejudicial both to the
resolution and to the sensitivity. These arrangements are not markedly different from the ones we have described in connexion with the microscope (Chapter 18).

24.3.1 Measurement of the Ion Currents \( i \) (Spectrometers)

The ion current which falls on the collector is at most of the order of \( 10^{-10} \) A, and it is usually measured with the aid of an electrometer, with which currents down to \( 10^{-15} \) A can be detected. The current \( i \) flows through a very high resistance \( R, 10^{11} \Omega \) say, which is connected to the collector \( C \) (Fig. 297); we have simply to measure the potential across the terminals \( r = Ri \). For this, \( R \) is placed at the input of an electrometer valve \( L \), which is a vacuum tube in which the control grid \( g \) is highly insulated by means of

![Amplifier for measuring ion currents](after Bertein, Vastel and Beaussier)

various special features of the construction and operation: the presence of a supplementary grid, an under-heated filament, low potentials to avoid ionization of the residual gas (10 V at most), and a special structure to avoid both the emission of very soft X-rays by the plate and photo-electric emission by the grid in the presence of this weak radiation.

These are serious inconveniences in making a direct reading with a galvanometer, for example, of the measurement obtained at the exit of \( L \). Furthermore, the characteristics of \( L \) are very curved, and hence there is a complete lack of proportionality. Finally, parasitic capacities and the capacities
of the screening produce a capacity $C'$ shunted across $R$; the time constant, $C' R$, of this combination may be more than a second which makes readings on the peaks excessively slow. These inconveniences are remedied by placing a D.C. amplifier $A$ after the electrometer valve stage; $A$ has a high gain, of the order of 1000, and $R$ is arranged in such a way as to ensure unity feedback. The principle is shown schematically in Fig. 297; a practical example is illustrated in Fig. 298. In these conditions, a galvanometer connected to the exit indicates the potential $v = R i$, and the reading is therefore proportional to $i$; further, the time constant is now negligible. We can show moreover that the collector $C$ remains virtually at earth potential.

The smallest current which can be detected with an extremely carefully constructed amplifier is of the order of $5 \times 10^{-16}$ A. The spectrometer is therefore capable of displaying peaks, the relative heights of which are only separated by one part in $10^5$ or $10^6$; this is the explanation of its extreme sensitivity, and its ability to detect the constituent parts of a mixture even in feeble traces. In this continual striving after sensitivity, it is particularly important to reduce the phenomenon of drift which is peculiar to D.C. amplifiers. On this point, the study by Bonne and Margoloff (1951) is to be consulted.

Very low current measurements are tending more and more to use vibratory modulators which allow the steady potentials to be replaced by alternating potentials, and the methods of amplifying alternating currents to be used, as these are considerably better so far as the stability is concerned. We should remember, too, that it is possible to produce the first amplification with the aid of a multiplier placed before the collector; this procedure, which is very sensitive, is advantageous where we are making measurements on very small quantities of isotopes.

24.3.2 Continuous Recording of the Results

The measuring circuitry is completed by a recording device, which must be able to accept spectra in which the various peaks may be of very different heights (Figs. 299 and 300). For this purpose, it is possible to devise a means of recording photographically the spots of a number of galvanometers of different sensitivities, functioning simultaneously. Alternatively, we can use a pen-recorder which is coupled in such a way as to record automatically the sensitivity of each of the peaks (Washburn, Wiley and Rock, 1943; Surdin, 1949; and Bonne and Margoloff, 1951). To obtain the scale on which the masses are measured, which is necessary for interpreting the spectrum, we can use a mass-marker, which is based upon a continuous measurement of the induction $B$ during the magnetic sweep (Hipple, Grove and Hickam, 1945). It usually takes a few minutes to record a spectrum containing masses between 24 and 100.
Fig. 299. The spectrum of xenon (C.S.F.).

Fig. 300. Spectra of hydrocarbons (Inst. du Pétrole).
24.3.3 Other Circuits

To obtain the acceleration potential \( V \) and the current \( i_B \) with which the electromagnet \( (B) \) is supplied requires two carefully filtered rectifying circuits; the stability often reaches one part in 10,000 (for fluctuations of 10 per cent at the mains). A still higher stability can be obtained by an original procedure in which an auxiliary spectrometer tube is used (Nier, Roberts and Franklin, 1949). The values of \( V \) and \( B \) which are produced must be variable; in analytic spectrometers with a high resolving power, typical ranges would be 500 to 3000 V and 1000 to 6000 gauss respectively \( (i_B \sim 100 \, \text{mA}) \).

![Diagram](image)

**Fig. 301.** Supply circuit for the ion source (after Winn and Nier, 1949).

Sources which depend upon electron bombardment require a variable supply for heating the emitter \( t \). Essentially, it is a question of stabilizing the electron beam \( I_e \) which produces the ionization. The ways of doing this differ appreciably depending upon whether the feed-back acts upon the filament supply (Ridenour and Lampson, 1937) or upon the electrodes of the supplementary control (Winn and Nier, 1949; Caldecourt, 1951); the latter method has the advantage of not causing thermal variations in \( t \) which might influence the spectra (Fig. 301). There are, however, always parasitic fluctuations in a source (§ 23.1.6) which are too complex to be completely removed by straightforward regulation.
24.4 THE VACUUM AND THE GAS CIRCULATION

24.4.1 The Vacuum

Within a spectroscopic tube, the vacuum must reach at least $10^{-6}$ mm Hg. If the residual pressure is higher, some of the ions will be scattered, and both the sensitivity and the resolving power will be reduced; further, the residual gas takes part in the ionization at the ion source, and the result is a number of parasitic lines or peaks. These can be avoided by a careful secondary pumping system: a mechanical pump together with one or more diffusion pumps operating in parallel at various points along the tube, a liquid air trap or a trap consisting of solid carbon dioxide in acetone, the minimum number of demountable joints, and no grease in the secondary vacuum system. Periodic dehydration is recommended (250 to 300°C).

Despite the extreme care which is given to producing the vacuum, the very high sensitivities of mass spectrosopes result in a weak parasitic residual spectrum, which must be taken into account; the spectrum contains in particular H$_2$O, CO (in certain metallic apparatus), the various mercury isotopes (Hg$^+$ and Hg$^{++}$), or traces of the hydrocarbons, depending upon whether the diffusion pump works with oil or mercury.

24.4.2 Circulation of the Gas which Is to Be Analysed

There is also, moreover, an essentially dynamic vacuum as the substance to be analysed has continually to be circulated by the pumping system. As a concrete example, we shall consider the case which occurs the most frequently in which we have a gaseous mixture (Fig. 302). The gas is contained within a flask $B_0$, which is sealed to the entry duct which has previously been evacuated by an independent pumping unit. By opening the taps, some of this gas is made to flow at low pressure into the reservoir $B$. A steady flow is released from $B$, passes through a very narrow opening or leak $F$, and reaches the ion source; here, the gas moves into the region of the vacuum proper, the tube $T$ and the associated pumps. The spectroscopy can then begin.

The composition of the gaseous mixture is defined by the partial pressures $P, P', \ldots$ of its various constituents $G, G', \ldots$ in the flask $B$. The result of the ionization is that the partial pressures in the chamber $C$ which forms the ion source are now $p, p', \ldots$; with the present notation, the overall pressure in $C$ is therefore not $p$ but $p + p' + \cdots$. From the point of view of the analysis, it is of course important to examine the ratios $p/P, \ldots$; we distinguish between the possible types of flow.
(i) *Flow which is everywhere molecular*

This case occurs when the mean free path of the molecules is everywhere appreciably higher than the transverse dimensions of the gas circuit; this simply requires the pressure $P$ to be less than about $10^{-2}$ mm.

For each constituent $G, \ldots$ the type of flow which is known as molecular is then established independently. This flow is the result of the corresponding pressure $p$ just as the passage of electricity through a conductor is a result of the electric potential difference; we can make use of this analogy to construct the equivalent electrical circuit diagram (Fig. 303). The quantity $Q$ of gas $G$ is measured, for example, by its volume in litres, referred
back to N.T.P. Each section of the piping is characterized by its pumping rate \( U (l. \text{ sec}^{-1}) \) which plays the role of conductance. \( B \) is thus represented by a condenser with a charge \( Q \) at a potential \( P \); this condenser is discharged through a series of resistances which represent the leak \( F (1/U_F) \), the slits in the ionization chamber \( C (1/U_C) \), and the piping of the actual vacuum, the trap \( T \), and the secondary pump \((1/U_\sigma)\).

We shall summarize the conclusions which can be drawn from an examination of the equivalent circuit diagram; first of all, the pressures which are desirable at the various levels, \( P, p, \rho_0 \)—say \( 10^{-2}, 10^{-5} \) to \( 10^{-4} \), and \( 10^{-6} \text{ mm Hg} \)—determine the relative orders of magnitude of the pumping rates; these latter can easily be attained, however: \( U_F \ll U_C \ll U_\sigma \) (for example, \( U_\sigma = 20 \text{ l. sec}^{-1}; \ U_C = 1 \text{ l. sec}^{-1}; \ U_F = 10^{-3} \text{ l. sec}^{-1} \)). The leak will simply be a circular hole about \( 1/10 \text{ mm} \) in diameter.

If we compare now the various constituents \( G, G' \ldots \), the pumping rates in principle vary as \( M^{-1/2} \) when we pass from one constituent to another (\( M \) is the molecular mass), and ratios such as \( U_F/U_C \), \( U_F/U_\sigma \), \ldots are thus independent of the type of gas. The same is therefore true of the pressure ratios:

\[
\frac{p}{P} = \frac{p'}{P'} = \ldots
\]  

(24.8)

In other words, the (relative) composition in \( C \) is the same as that in \( B \). The speed \( U_\sigma \) often does not follow an \( M^{-1/2} \) law, however; certain gases are condensed at the liquid air trap, which has the effect of sharply increasing the corresponding speed. In fact, the repercussion on \( p \) is only slight if \( U_C \ll U_\sigma \), and even though the equalities in equation (24.8) are unreliable, the ratios \( p/P, \ldots \) at least are constant provided the trap is constantly topped up, and we shall see later on that this is adequate for the majority of analyses. Since the pressure \( P \) in the reservoir \( B \) is so low, it is possible for \( B \) to become discharged to an appreciable degree in the course of an analysis, and this is one of the inconveniences of the molecular regime. The \( M^{-1/2} \) law implies that light gases will be discharged more rapidly than heavy gases—the proportion of heavy gases in the mixture increases steadily, and we say that there is a fractionation.

(ii) Partially viscous flow

It is more convenient that the pressure \( P \) in \( B \) should be of the order of at least a few millimetres or centimetres of mercury. The flow is then viscous (Poiseuille) as far as the leak \( F \); beyond \( F \), the pressure must drop sharply to the value \( p \), and the flow becomes molecular. Here, then, the resistance of the leak \( F \) must be considerably higher than in the preceding case. A flow pattern of this kind is very complex, and has been studied by Honig (1945)
in particular. It is true, however (Nier, 1948), that if $F$ consists of a long capillary tube (12 cm $\times$ 0.15 mm for example), the gas flows out of $B$ as a unity, without fractionating, so that the composition within $B$ remains unaltered. In the other part, however, the light gases flow more rapidly beyond the leak (where the regime is molecular) and instead of (24.8) we shall have

$$\frac{1}{\sqrt{M}} \frac{p}{P} = \frac{1}{\sqrt{M'}} \frac{p'}{P'} = \ldots$$

(24.9)

The composition in $C$ is thus no longer the same as that in $B$, and further, the ratios $p/P$, $\ldots$ are no longer constant. Some suitable method for interpreting the results of the analysis is necessary. Adjustable leaks have been adapted for this type of flow (Fig. 304).

![Adjustable leak](image)

**Fig. 304.** Adjustable leak (Ewald, 1950)
A steel needle (6), fixed at the end of a brass rod (2), slides in a very narrow channel pierced in a lead plate (7); its displacement is driven by the knob (1), and read accurately on a scale (3). The space between the joints (4) is filled with a mixture of vacuum grease and oil, through the opening (5). A bellows, soldered on the piece (10), is tightened on the flange of the needle valve output by the screw-nut (9).

(11) Gas inlet.

### 24.5 The Various Types of Measurement in Mass Spectroscopy

The measurements fall into two distinct categories; these are concerned either with the positions of the bands and in this case, it is the atomic masses which are determined spectrographically, or with the intensities of the bands, with the object of making quantitative analyses by spectrometry.
In the latter category, we shall also have to consider separately the situations where we are concerned with two isotopes (measurement of relative abundance) and with ordinary chemical analysis.

### 24.5.1 Determination of the Atomic Masses (Spectrography)

The bands here must be extremely fine, and hence one of the double focusing spectrometers described in Chapter 22 must be used, in particular the Mattauch instrument. Figure 305 shows as an example the spectrum obtained in the vicinity of the mass \( M = 20 \) together with the corresponding photometric curve. After identifying the ions which correspond to the various lines, their molecular masses \( M \) are determined from their positions \( X \)

---

**Fig. 305.** The multiplet which is obtained in the neighbourhood of mass 20 (after Mattauch, 1938), and its photometric recording.
on the photographic plate $II$. In principle, we have simply to apply the relation which is known \textit{a priori} to relate $M$ and $X$, namely:

$$M = K X^2 \quad \text{(Mattauch's instrument).} \quad (24.10)$$

The origin of the abscissae $X$ and the theoretical value of $K$—the situation is complicated by the presence of fringe fields—are not known sufficiently precisely for the rough method above to be satisfactory. It is better to determine the values of $M$ little by little, by making measurements upon small spectral regions, checked section by section.

\[\text{Fig. 306. Spectrometry: the spectrum of the isotopes of tungsten (after Inghram, 1946).}\]

The most favourable ion sources will be those which emit a wide variety of ions (spark sources, for example).

We shall examine this point more precisely by describing a typical method which is known as the "method of doublets". We recall first of all that with the standard convention according to which the atomic weights are defined with respect to $^{16}\text{O} = 16$ (which is known as "the physicists' scale") the various masses $M$ are all very nearly whole numbers; the ions which are

\[\dagger\text{ In the so-called "chemists' scale" which is usually employed in chemistry, the mass 16 is given to the mixture of isotopes of which atmospheric oxygen is composed.}\]
associated with the same integer thus produce bands which are very close
together, and hence form a doublet or, more generally, a multiplet in the
spectrum. Close to \( M = 20 \) for example (see Fig. 305), we find \(^{40}\text{A}^{+},^{20}\text{Ne}^{+}\), \(^{18}\text{OH}^{+}\) and \(^{18}\text{OD}^{+}\). To illustrate the principle of the method of deter-
mining a mass, we consider the example of \(^{13}\text{C}\), given the masses of H and
\(^{12}\text{C}\), namely

\[
H = 1.008130, \quad ^{12}\text{C} = 12.003861.
\]

It is easy to obtain a spectrum which contains the bands corresponding
to \(^{12}\text{C},^{13}\text{C}\), and \(^{12}\text{CH}\), the last two of which form a doublet. The mutual
separations are measured with a microscope, giving

\[
\begin{align*}
( ^{12}\text{C}, ^{12}\text{CH}) &= \Delta X_1 = 9.9624 \text{ mm,} \\
( ^{12}\text{CH}, ^{13}\text{C}) &= \Delta X_2 = 0.0427 \text{ mm.}
\end{align*}
\]

If we designate the masses by their actual chemical symbols, we obtain
from equation (24.10) the relation

\[
\frac{\sqrt{^{13}\text{C}} - \sqrt{^{12}\text{CH}}}{\sqrt{^{12}\text{C}} - \sqrt{^{12}\text{CH}}} = \frac{\Delta X_2}{\Delta X_1}, \tag{24.11}
\]

and hence

\[
^{13}\text{C} = 13.007583.
\]

Neither the position of the origin of \( X \) nor the value of the constant \( K \)
has appeared. Measurements of this kind do, of course, require special
precautions; on this point we refer the reader to the book by Ewald and
Hintenberger. The accuracy is not sufficient to confirm the correctness
of all the digits which are normally quoted—only four or five places of
decimals, according to the various different cases, can be guaranteed.

### 24.5.2 Quantitative Analysis; Relative Isotope Abundances (Spectrometry)

Figure 306 shows the spectrum which is obtained for the isotopes of
tungsten. In these analyses, the use of an ion source depending upon elec-
tron bombardment is recommended, above all for gaseous mixtures. Suppose,
for example, we have to determine the relative abundance of the barium
isotopes \(^{10}\text{B}\) and \(^{11}\text{B}\), that is, to determine the ratio of their concentrations
\([^{10}\text{B}]:[^{11}\text{B}]\) in natural specimens. We analyse spectrometrically the gaseous
compound \(\text{BF}_3\) which is a mixture of the two constituents \(^{10}\text{BF}_3\) and \(^{11}\text{BF}_3\)
which we shall designate by \(G\) and \(G'\). The relative abundance we require
is equal to the ratio of the partial pressures in the analysing flask \(B\), that is

\[
\frac{[^{10}\text{B}]}{[^{11}\text{B}]} = \frac{P}{P'}.
\]

It is natural to hope that this can simply be obtained as the ratio of the
heights of two homologous peaks, \(i, i'\), which are received at the collector;
this would imply that
\[
\frac{i}{\bar{P}} = \frac{i'}{\bar{P}'}.
\] (24.12)

This is true provided none of the isotopes is privileged with respect to the remainder, that is, provided there is no discrimination along the chain of processes which leads from \( P \) to \( i \), namely:

(a) \( P \) (flask) \( \to p \) (ionization chamber): gaseous flow,
(b) \( p \to I \) (ionization chamber): ionization of the gas,
(c) \( I \to i \): passage through the ion source. (24.13)

If there is to be no discrimination, the process (a) must be a molecular flow; we can, however, use a partially viscous flow provided a correction factor is introduced which takes account of the molecular masses according to (24.9). So far as (b) and (c) are concerned, the ionization probabilities must be equal as must the transmission probabilities in the passage through the ion source [the coefficient \( K \) of relation (24.7)]; these requirements are satisfied in practice, as the relative separations between the masses of the isotopes are small, and the magnetic scanning has an averaging effect (see the case of Cl in the table given in \$24.1.6\). The same would not be true for a mixture of H and D; in such a case, the standard method of analysis described in the next section would be suitable.

Save for this case, the ratio of the abundances is equal to the ratio of the heights of the peaks; we can obtain this latter ratio to an accuracy of better than \( 10^{-3} \) (as far as \( 0.2 \times 10^{-3} \)) if a double collector is used so that two isotopes can be received simultaneously and their heights compared by an opposition method (Nier, Ney and Inghram, 1947).

### 24.5.3 Ordinary Chemical Analysis (Spectrometry)

Here the problem is different; given an arbitrary mixture of gases \( G, G', \ldots \) it is no longer a question of obtaining "homologous" peaks which can be assigned to the various different components of which we need only measure the relative heights. We have to measure the individual spectra of each of the constituents beforehand—we suppose them to have been determined at N.T.P. (Fig. 307). These spectra, like the spectrum of the mixture, are defined by the heights of the peaks produced by a certain number of ions with molecular masses \( M_1, M_2, \ldots \); the results may be tabulated thus:

<table>
<thead>
<tr>
<th>Masses</th>
<th>( M_1 )</th>
<th>( M_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectra</td>
<td>( G )</td>
<td>( i_1 )</td>
</tr>
<tr>
<td></td>
<td>( G' )</td>
<td>( i'_1 )</td>
</tr>
<tr>
<td>Mixture</td>
<td>( f_1 )</td>
<td>( f_2 )</td>
</tr>
</tbody>
</table>
Here, the question of discrimination does not arise; the results can be interpreted very straightforwardly if the processes which lead from $P$ to $i$ proceed independently for each of the constituents, which implies a molecular flow. In these conditions, the spectrum is formed by superposition;

![Spectra of iso-butane and n-butane](image)

**Fig. 307.** Spectrometry: the spectrum of normal butane and of isobutane (after Hipple, 1942).

if $P$ and $P'$ are the partial pressures of the constituents, that is, the unknown quantities we are seeking, then

\[
\begin{align*}
    j_1 &= P_1 i_1 + P'_1 i'_1 + \cdots \\
    j_2 &= P_2 i_2 + P'_2 i'_2 + \cdots \\
\end{align*}
\]

(24.14)

If there are $n$ constituents, we have simply to write down this system of $n$ linear simultaneous equations for the $n$ lines $M_1, M_2 \ldots$ which have been measured, and to solve for $P, P', \ldots$ When the streaming is partially viscous, and the gases in the mixture flow as a single unit, this method can be used provided the spectra of the constituents are not determined under normal conditions but for equal linear flow velocities. This can be achieved by making measurements in which an internal reference gas or standard is used. A great deal of information about this point is available, primarily in articles dealing with industrial applications (mixtures of hydrocarbons).
Analyses in which six components are present can be performed without difficulty, and if we use an electronic computer to solve equations (24.14), even more constituents can be handled. The accuracy obtained depends upon a great many circumstances; the fundamental limiting factor is the lack of perfect constancy with time. At present the order of magnitude of the error in each constituent is 0·5 per cent.

24.6 AN OVERALL VIEW OF PRESENT-DAY INSTRUMENTS, AND THE FIELDS IN WHICH THEY ARE EMPLOYED

The fundamental part of a spectroscope is its analysing field, and in studying this we have had occasion to review the main kinds of instrument. We shall now point out the essential features of the models at present in use; so far as the detailed description and the details of the construction are concerned, we shall give a few references to typical examples. Detailed bibliographies are to be found in the articles by Roth (1949), Reis et al. (1950) and Dibeler and Hipple (1952) and above all, in the article by Reis (1955).

24.6.1 Instruments for Measuring Masses Accurately

The number of spectrographs is still very limited; we recall the particularly advantageous design of Mattauch’s instrument. A recent model has been developed by Ewald (1946, 1950 and 1951) which gives a resolving power of 30,000; the analysing tube is built from solid bronze and steel. To achieve the desired properties, the construction must be extremely precise and numerous vacuum-sealed control units are required. We should mention in particular the possibility of employing an ion source using either gas discharge or sparking where the exit slit consists of two small parallel cylinders which can be altered in orientation bodily (Fig. 308).

Under this same heading, we ought also to mention the double focusing spectrometer of Nier and Roberts (1951), which is appreciably more rapid and convenient to use, but with which it does not at present seem possible to obtain the same resolving power.

24.6.2 Analytic Spectrometers

Spectrometers (with single focusing) are considerably more numerous, as a result of the extreme simplicity of their construction and the wider range of their applications; for some years now, they have not only been constructed individually in the laboratory but have also been made commercially.
Fig. 308. Mattauch's spectrograph, as built by Evauld. 1: the support for the photographic plate 3, 2: the yoke of the electro-magnet 4: brass wall of the vacuum chamber 5, 6, 7, 9: means of adjusting the direction of the beam in vacuo 8: leads for the deflexion prism 10, 17: pump connexions 12: insulation between the (earthed) instrument and the source 13: ion source, to which the gas is admitted through 14, 15: vacuum cock 16: adjustment for the intermediate diaphragm.
Bleakney's instrument (1932) already displayed a notable evolution in comparison with the first version of Dempster (1918); instruments of truly "modern" conception appeared still later, however: Nier (1940) and Hipple (1942)—see Fig. 309. The sector fields are $60^\circ$ and $90^\circ$ respectively, and a large proportion of the instruments used to-day is derived directly from these models. There are, however, always partisans of the $180^\circ$ field, as the arguments over the advantages and disadvantages of the various angles cannot be said to impose any one choice rather than another.

Present instruments tend to be all metal, and easily demountable; they can be separated into several categories according to their various qualities which in turn determine their applications.

(i) *Analytic high resolution spectrometers*

Here, the resolving powers range between about 150 and 400; at present, there are more than 200 of these instruments in the U.S.A., and the resolving power of the most accurate instrument commercially available can reach 400
(Brown, Mitchell and Fowler, 1941; Graham, Harkness and Thode, 1947; Nier, 1947; Bertein and Vastel, 1949). Since these instruments are of commercial importance we shall briefly describe a typical example.

We consider the standard spectrometer due to Nier, the particularly simple structure of which is represented in Fig. 309. The sector is $60^\circ$ and the radius $R$ of the trajectories is 15 cm. The analyser consists of a copper tube 3 cm in diameter which is flattened transversally to 13 mm in such a way as to fit into the gap of the electromagnet. A wider part made of inconel carries the ion source, and is coupled to the pump circuit by a spherical joint. The diffusion pump is not made of metal, here, but of pyrex; the pumping rate is about seven litres per second at the spherical joint. The vacuum is measured with a Penning gauge (a current of 1 $\mu$A for a vacuum of $10^{-6}$ cm Hg).

The ion source (Fig. 310) can easily be removed, as it is fixed to a cover made leakproof by a Teflon joint; a thorough baking is possible, since the fixing clamps are water-cooled. The electrodes of the source are made of nichrome, and its magnetic field is 150 gauss. A double collector is employed for collection of the ions as this permits two neighbouring peaks to be received simultaneously and compared by a null method. The focusing is adjusted by shifting the analysing tube with respect to the electromagnet which is held fixed.
(ii) Spectrometers with a medium resolving power (50 to 80)

These spectrometers are particularly suitable in industry for controlling chemical reactions or separations (as "monitors"); they are much lighter, more robust, and less expensive. They may be fitted with a permanent magnet instead of an electromagnet, and with an oscilloscope on which the spectra can be displayed directly. This class of instruments is at present passing through a period of particularly rapid development (Siri, 1947; Walcher and Gifford, 1954).

(iii) Simplified spectrometers, for detecting leaks

It is these which are at present the most numerous; they are used to detect the presence of gases such as helium in tubing or in some given space. They are, of course, equally convenient for analysing sufficiently light gases (Nier, Stevens, Hustrulid and Abbott, 1947; Cassignol, Geller and Moreau, 1953).

![Diagram of the "Diatron" spectrometer](image)

**Fig. 311.** The "Diatron" spectrometer (Consolidated).

It is interesting to notice that the focusing principle, which is theoretically perfect, in which crossed fields are used (cycloidal trajectories—Bleakney and Hipple, 1938), has recently been taken up again under the name of the "diatron"; the corresponding resolving power reaches 150 despite the extremely small dimensions, of the order of centimetres (Fig. 311).

24.6.3 Spectrometers for the Separation of Isotopes

The possibility of separating isotopes in considerable quantities with a spectrometer is related to the existence of very intense ion sources which were not available to the early workers who first envisaged isotope separation (Aston, 1927; Morand, 1927). Progress has been made in this direction, and there are now several separators which furnish the isotopes which are
required for the increasing number of applications of radioactive isotopes; the instruments at Oak Ridge in the U.S.A. ("calutrons") produce more than 200 different isotopes. In France, we mention the separator of Bernas (1953; see Fig. 312), with which masses of 10 to 100 mg can be

![Diagram of a spectrometer](image)

Fig. 312. A spectrometer for the separation of isotopes (after Bernas, 1953). 
E = ion source; C = electro-magnet of the source; D = air-lock; F = accelerating electrode supported by the copper cylinder G; H = pyrex tube; I = isolator rods; K = cam for adjusting the inter-electrode distance EF; L = high tension plug; M: demountable part of the pole-pieces; P = collector; O = pyrex window; N = baffle of the collector chamber; R = window for studying the undeflected beam; S = cooling circuit of the main-magnet coils; T, A = vacuum pumps; B = carriage; Y = polarized grids for neutralizing space-charge.

obtained in 24 hours, using ion beams of several mA; the instrument consists of a spectrometer with a sector field of 60°, and larger dimensions than the analysing instruments: the radius of curvature of the trajectories, $R = 50$ cm; the width of the gap is 7 cm; $B_{\text{max}} = 12,000$ gauss; $V_{\text{max}} = 50$ kV. An instrument of the same type has been constructed in Amsterdam by Kistemaker.

24.6.4 Present-day Applications

In conclusion, we shall list the various fields in which the mass spectrometer is employed:
(a) To make accurate measurements of the atomic masses—to obtain the mass defects involved in balancing the energy in nuclear reactions.
(b) To measure relative abundances—to study the evolution of reactions by the indicator technique.
(c) To perform analyses by isotopic dilution (in biology) and to determine the geological time-scale.
(d) To make chemical analyses, as industrial monitors; in particular, to study the chemistry of the hydrocarbons.
(e) To study molecules: determination of the ionization potentials or ion threshold potentials, and hence calculation of binding energies; to study dynamic chemical processes involving metastable ions and free radicals.
(f) To study nuclear reactions.
(g) To detect small traces of matter—leaks in vacuum systems, impurities in metals.
(h) To separate isotopes.

The diversity of these applications explains the considerable importance of mass spectrometers both in ordinary chemistry and in the most modern kinds of research.

24.7 MASS SPECTROSCOPES WITH A HIGH FREQUENCY FIELD

We shall conclude this study by outlining briefly the principle of a class of instruments which also separate ions according to their masses, but in a manner very different from that of the normal spectrometer. In these instruments, a high frequency electric field is employed—for a given supply frequency, ions of a certain mass will be given an oscillatory motion, the frequency of which matches that of the supply, and will pass through the instrument. Only these ions fall on the collector. The remaining ions cannot resonate, and the instrument is so designed that they are rejected either towards the walls or towards the source.

The omegatron (Goudsmit 1948, Hipple and Thomas 1949) is a small cyclotron. The ions are produced in the gap of an electromagnet by an ionizing electron beam, as in an ordinary source. If the frequency of the supply to the "Dees" is \( f \), the only ions which can resonate are those with a mass \( M \) given by

\[
M = \frac{eB}{2\pi f};
\]

\( e \) is the ionic charge and \( B \) is the magnetic induction. These ions trace out a spiral and fall on the collector which is situated on the periphery. Ions with masses other than \( M \) remain near the centre. This instrument has been used to determine the masses of the light elements with very high
The individual parts of a mass spectroscope

precision. An instrument which is comparable but which is based upon the principle of the linear accelerator has been built by Bennett (1950). This is a linear omegatron, which does not require a magnetic field. The high frequency electric potential is applied to a series of equidistant grids, and the only ions which can pass through the instrument are those for which the time of transit between the grids is equal to half the period.

The omegatron (Wagener and Marth, 1957) is extremely sensitive and still gives useful signals at pressures of $10^{-10}$ mm; this is the instrument which is best suited to the study of the composition of the residual atmosphere in evacuated apparatus and sealed-off tubes. Recently, ordinary spectrometers have equalled this performance in this application, but at the cost of special construction. The metal spectrometer tube is welded to the glass entry and exit tubes; both the taps and the joints are metallic (either lead or aluminium). We can therefore bake the whole apparatus at more than 450°C, and draw off the residual gases by means of an absorption effect; this latter can either be produced by means of titanium vapour pumps or by actual absorption in an ionization gauge. The research performed recently by Bayard and Alpert (1950) has led to the construction of ionization gauges which are capable of simultaneously measuring pressures of $10^{-11}$ mm Hg accurately, and providing an efficient "pumping" action as soon as the pressure falls below $10^{-6}$ mm Hg.

Another invention which also allows these very high vacuums to be attained in practice is the "ion pump", in which the residual atmosphere is ionized and drawn off by electrical forces, and this whole class of instruments opens up a new and interesting field of research; the phenomena associated with absorption by surfaces can in particular be measured with certainty.

Recently Paul and his collaborators (1953 and 1958) have proposed a type of instrument depending upon the principle of strong focusing which we described in Chapter 10. These lenses are convergent in the plane $xOz$ and divergent in the plane $yOz$. With a chain of such lenses we can produce a mean overall convergence in all directions—a result of the mutual compensation—and hence direct a beam of particles along the axial zone (§ 10.5). Paul suggested that the electrodes of a single but very long lens should be fed with an alternating potential. It is clear that the particles will then be subjected to a focusing field for one half-period and a defocusing field during the subsequent half-period. So far as the particles are concerned, the situation is exactly the same as if they were passing through a continuous chain of strong convergence lenses supplied with fixed potentials.

First of all, we set out the equations of motion for a particle (mass $m$, charge $e$ and velocity $v = \dot{z} = \sqrt{\frac{2e\Phi_0}{m}}$) in a field with quadrupole symmetry, independent of $z$ but varying sinusoidally with time at a frequency $f$. We assume that the length of the trajectory which the particle covers in each period is short in comparison with the length $L$ of the lens.
We have
\[ \Phi(x, y) = \frac{\Phi_1}{a^2} (y^2 - x^2) \cos \omega t, \]
in which \( \pm \Phi_1 \) represents the peak potential applied to the electrodes; the latter are all tangent to a circle of radius \( a \) (see Chapter 10). \( \omega \) is the angular frequency, \( \omega = 2\pi f \).

The equations of motion take the form:
\[ m \ddot{x} + \frac{2e \Phi_1}{a^2} \cos \omega t \cdot x = 0, \]
\[ m \ddot{y} - \frac{2e \Phi_1}{a^2} \cos \omega t \cdot y = 0, \]
\[ m \ddot{z} = 0, \]
in which we have assumed that \( \dot{z} = v = \) constant. These equations can be written in the form:
\[ \frac{d^2x}{d\tau^2} - (a_x - 2q \cos 2\tau) x = 0, \]
(for the \( x \) component); with \( a_x = 0 \),
\[ q = \frac{4e \Phi_1}{ma^2 \omega^2}, \quad \text{and} \quad \tau = \frac{\omega t}{2}. \]

The solution is of the form:
\[ x = A e^{\mu \tau} \sum_{n=-\infty}^{\infty} C_n e^{in\tau} + B e^{-\mu \tau} \sum_{n=-\infty}^{\infty} C_n e^{-in\tau}. \]

If \( \mu \) is real or complex, the amplitude increases indefinitely; if, on the contrary, \( \mu \) is purely imaginary, the solution will be finite for all values of \( \tau \). The trajectory will be stable and can pass through the lens. The first ranges of values of \( q \) which correspond to stability are
\[ 0 < q < 0.92, \]
and
\[ 7.5 < q < 7.52, \]
but \( q \) is independent of the velocity of the incident ions. The only parameter of interest is the mass \( m \); only those particles which have a mass greater than \( m_0 \) (\( m_0 \) is the solution of the equation \( q = 0.92 \)), can pass through the lens, and the first zone thus constitutes a high-pass mass filter. The second domain represents a band-pass filter. For values of the mass which correspond to instability, the amplitude of the oscillations of the trajectories will grow exponentially; the resolving power of the combination is better in proportion to this growth and hence improves as the number of oscillations is increased. If the electrode length \( L \) is fixed, therefore, we should use slow ions \( (v/f \ll L) \).
If we employ the second zone, the values of $\Phi_1$ become inconveniently high. We can alternatively construct a band-pass filter with the aid of the first zone only, by superimposing a steady quadrupole voltage $\pm \Phi_G$ on the high-frequency voltage. This gives:

$$\ddot{x} + \frac{2e}{ma^2} (\Phi_G + \Phi_1 \cos \omega t) x = 0,$$

$$\ddot{y} - \frac{2e}{ma^2} (\Phi_G + \Phi_1 \cos \omega t) y = 0,$$

and now we require that

$$a_x = -a_y = \frac{8e\Phi_G}{a^2 m \omega^2} q = \frac{4e\Phi_1}{a^2 m \omega^2},$$

the ratio $u = a/q = \pm 2\Phi_G/\Phi_1$ depends only upon the ratio of the applied voltages.

In the first domain, $u$ has an upper limit, $u_{\text{max}} = \frac{2\Phi_G}{\Phi_1} = 0.333$; the domain of stable operation is infinitely narrow at this point, and corresponds to $q_1 = 0.706$.

To each mass, therefore, corresponds a ratio $u$ or a frequency $f$, for which the value of $q$ lies within the selected band in the neighbourhood of $u_{\text{max}}$.

For a constant field we can calculate the mass resolving power:

$$\frac{m}{\Delta m} = \frac{0.75}{1 - (u/u_{\text{max}})}.$$

If we vary $\Phi_1$ (or the frequency $f$) and it is with a source of slow ions of various masses that we are concerned, the different masses in question appear successively at the detector—the system behaves as a mass spectrograph.

When the beams are intense, the role of space charge is important and the equations of motion are in consequence far more complex.

The accuracy of the simple theory can be demonstrated with rubidium ions (Paul and Raether, 1953) when the current is weak. For $f = 2.55$ Mc/s, $\Phi_1 = 1000$ V, $a = 1$ cm and $L = 50$ cm, the positive rubidium ions of masses 85 and 87 are separated with the best resolution for $u = 0.328$. Very good resolution has been obtained with light ions (lithium, oxygen, nitrogen) by Gaudaire (1960), with a relatively short instrument (50 cm).

Increasing $L$ (to several metres), we can reach a resolving power $m/\Delta m$ of several thousand. Paul and his fellow workers have also shown how the same system can be transformed into an isotope separator giving a current of several milliamperes (Paul, Reinhard, and von Zahn, 1958). The electrical supply becomes considerably more complex.
CHAPTER 25

β-RAY SPECTROGRAPHS WITH ELECTRON LENSES

25.1 INTRODUCTION

β-rays are emitted directly by radioactive substances, and are created indirectly when a secondary target is bombarded with γ-rays. The energy spectrum of the β-radiation is a rich source of information about the nuclear structure since the position of the lines in the spectrum is related to the separation of the nuclear energy levels, and the intensity of the lines is a function of the transition probability between the various levels (See Siegbahn, 1955; and Deutsch and Kofoed-Hansen, 1959). For this reason, there is an immediate interest in constructing spectrographs which are both bright and dispersive, if by these terms we understand the electron optical counterparts of their ordinary optical meanings.

Very many kinds of spectrograph have been successfully tried; we shall not be concerned with those in which there is only a plane of symmetry, which resemble glass prisms from the point of view of symmetry; we shall consider only those instruments in which actual lenses are employed, and in which there is axial symmetry.

There would, in principle, be no obstacle to prevent us from developing a theory which was also valid for instruments containing only a plane of symmetry; we have in fact already described the main properties of such devices when we were dealing with mass spectrographs, and there is very little we can add to this treatment, without invoking the considerably more complicated theory such as is to be found in the reviews by Siegbahn (1955, Chapter III: Beta-ray spectrometer theory and design) and by Bainbridge (1953, Chapter V: Charged particle dynamics and optics). Here, we shall only remark that there are objections to using iron to shape the magnetic field when high accuracy is sought: hysteresis in the yoke introduces uncertainty into the relation between current and field intensity. This small defect is less harmful in mass spectrographs because the mass numbers are integers; this explains why iron-free prisms are found in the domain of β-ray spectroscopy, although they are unknown elsewhere (Moussa, 1952).

For lenses, on the other hand, the task is far simpler since Glaser (1941) and Glaser and Lammel (1943) have developed a whole battery of formulae which describe very adequately the properties of magnetic lenses.
In the $\beta$-ray spectrograph, we make use of the extreme dispersion of electron lenses, that is, the fact that the image position is highly dependent upon the emission velocity of the particles; this velocity is inversely proportional to the de Broglie wavelength, which explains why the phenomenon is called dispersion. The energies in question range between very wide limits, from a few tens of kilo-electron volts up to 25 MeV. Electrostatic lenses are out of the question, for their convergence would be too small in the megavolt region. We use magnetic lenses, therefore, which remain suitable throughout the whole energy range.

The following highly simplified outline explains the principle behind the operation which was first suggested by Kapitza (see Tricker, 1924); it recalls the optical method of focal separation used by Wood (see Tricker, 1924) to study infra-red spectra.

### 25.2 THE GENERAL PRINCIPLE

The principle upon which the instrument depends has been discussed earlier, and can be understood by examining Fig. 313. We shall first of all consider the theoretical behaviour of the instrument when it is provided with a point source; this is not the usual case, but it can be reached with intense conversion lines.

![Diagram](image)

**Fig. 313.** The principle of the spectrograph.

A point source at $A$ emits $\beta$-rays in all directions. The aperture diaphragm, $D_0$, restricts the angular aperture of the beam which reaches the lens, $L$, to $2\alpha$; rays which are homogeneous in energy, $\varphi_1$, are converged into a point image $A_1'$, by the action of $L$. Rays corresponding to a different energy $\varphi_2$ (greater than $\varphi_1$) will, however, meet again in a different point $A_2'$.
(beyond \( A'_1 \)) as the convergence of the lens is less for faster particles. If a diaphragm \( D_s \), in which a small hole has been pierced, is placed at \( A'_1 \), it will separate the rays of energy \( \varphi_1 \) from all the others provided that it is protected from direct radiation by a lead screen, which is generally placed at the centre of the lens.

### 25.3 ENERGY SELECTION BY AN APERTURE

#### 25.3.1 Optical Theory

If the image \( A'_1 \) were a point, the selection would be rigorous; the lens \( L \), however, is not perfect, and the image of the point \( A \) is in reality a small circular spot, the radius \( \varrho_s \) of which is given by the following formula:

\[
\varrho_s = G C_s \lambda^3, \tag{25.1}
\]

with the usual notation, namely:

- \( G \) is the magnification of the lens for the pair of points \( A, A'_1 \) (and is usually equal to unity).
- \( C_s \) is the spherical aberration coefficient referred back to the object.

The rays with energy \( \varphi_2 \) will be well separated from those with energy \( \varphi_1 \) only if they are spread out over \( D_s \), and cover a circular area of which the radius, \( \varrho_c \), is much larger than \( \varrho_s \). If we denote \( \varphi_2 - \varphi \) by \( \delta \varphi \) and write \( \frac{\varphi_1}{A'_1 A'_2} \), the optics of the situation gives us the following expression for

\[
\frac{A'_1 A'_2}{A'_1 A'_2} = G C_c \frac{\delta \varphi}{\varphi}. \tag{25.2}
\]

If there were no spherical aberration at \( A'_2 \), the value of \( \varrho_c \) would be

\[
\varrho_c = A'_1 A'_2 \theta = G C_c \theta \frac{\delta \varphi}{\varphi}, \tag{25.3}
\]

in which \( 2 \theta \) is the angular aperture of the emergent beam at \( A'_2 \) which is given to a first approximation by

\[
\theta = \frac{\lambda}{G}, \tag{25.4}
\]

with the result that

\[
\varrho_c = C_c \lambda^3 \frac{\delta \varphi}{\varphi}. \tag{25.5}
\]

The ratio \( \varrho_c/\varrho_s \) will give us an idea of the purity of the spectrum which is obtained:

\[
\frac{\varrho_c}{\varrho_s} = \frac{1}{G} \frac{C_c}{C_s} \frac{1}{\lambda^3} \frac{\delta \varphi}{\varphi}. \tag{25.6}
\]
In reality, equation (25.5) gives too low a value for \( \varphi_c \), and we should substitute

\[
\varphi'_c = \varphi_c + \varphi_s,
\]

in order to take into account in the simplest way possible the spherical aberration which is present at \( A'_2 \) just as it is at \( A'_1 \). This leads to

\[
\frac{\varphi'_c}{\varphi_s} = 1 + \frac{1}{G} \frac{C_c}{C_s} \frac{1}{\alpha^2} \frac{\delta \varphi}{\varphi}.
\]  \hspace{1cm} (25.7)

### 25.3.2 Resolving Power

We can see, therefore, that two monochromatic lines of the same intensity and with wavelengths \( \lambda_1 \) and \( \lambda_2 \) will be separated if \( \frac{\varphi'_c}{\varphi_s} = 2 \) since the transmitted intensity, which is inversely proportional to \( \left( \frac{\varphi'_c}{\varphi_s} \right)^2 \), will fall in relative value from unity, when the spectrometer is adjusted exactly onto one or other of the lines \( \lambda_1 \) or \( \lambda_2 \), to 0.71 when it is on the average wavelength \( \frac{\lambda_1 + \lambda_2}{2} \) (see Fig. 314).

![Fig. 314. The intensity distribution between two lines at the limit of resolution.](image)

This calculation is admittedly schematic, but it is adequate to justify our selecting the value of the quantity \( \varphi/\delta \varphi \), or better (Grivet, 1950a, 1951; Persico, 1949), of

\[
P = \frac{\lambda}{\delta \lambda} = \frac{2 \varphi}{\delta \varphi},
\]
which corresponds to \( \theta_c/\theta_s \) equal to unity, to characterize the resolving power of the spectrograph. (We define \( P \) in this way in order to bring the notation into conformity with standard optical terminology.) For a point source, we find

\[
P = \frac{\lambda}{\delta \lambda} = \left( \frac{2 \gamma}{\delta \gamma} \right)_{\theta_c = \theta_s} = \frac{2C_c}{G C_s \alpha^2}.
\]

\(25.8\)

### 25.3.3 A Typical Example: the Long-coil Spectrograph

The uniform field spectrograph is of very widespread application. Its high quality makes it a very representative type. We shall use it here, therefore, as a standard for comparison. We shall first evaluate its resolving power by means of formula (25.8); formula (9.3) shows that

\[
C_s = C_c,
\]

for a uniform field, and since the magnification \( G \) is equal to unity, we obtain simply

\[
P = \frac{2}{\alpha^2},
\]

\(25.9\)

with \( \alpha \) expressed in radians.

It is usual, in \( \beta \)-ray spectrography, to relate \( P \) to the solid angle \( \Omega_s \) of the cone of useful rays. \( \Omega_s \) is expressed as a fraction of the whole sphere. The real solid angle is equal to \( 2\pi(1 - \cos \alpha) \), so that

\[
\Omega_s = \frac{2\pi(1 - \cos \alpha)}{4\pi} \approx \frac{\Delta^2}{4},
\]

\(25.10\)

and hence

\[
P = \frac{1}{2\Omega_s}.
\]

\(25.11\)

For a point source, the resolution of the long-coil spectrograph is inversely proportional to twice the brightness coefficient \( \Omega_s \).

The case of a short circular coil, of mean radius \( R_m \), is slightly less favourable, as the ratio \( C_c/C_s \) is then smaller than unity, and varies between 0.66 and 0.18 when the length, \( L \), of the instrument satisfies the relation

\[1.5 R_m < L < 5 R_m.\]

It is worth noticing that here, the resolution of an instrument appears to be independent of its size; this feature is connected with the hypothesis that a point source is being used, and the length \( L \) will be found to be shortened as soon as the radius of the source is introduced in a more realistic fashion.

25.4.1 The Two Factors Governing the Intensity: $s$ and $\alpha$

Sources of $\beta$-rays are often very weak, and the brightness is an essential quality in a spectrograph. The intensity which is received is proportional not only to the solid angle of the useful beam, but also to the surface area of the source, $S$, and is thus proportional to the product $S\Omega_s$. The source is usually in the form of a small disc of radius $s$, which lies normal to the axis at $A$, and is of surface area $S = \pi s^2$. The intensity, $I$, which falls upon the counter is therefore proportional to the product $s^2\Omega_s$, or again, to $s^2\alpha^2$:

$$I = \text{const. } s^2\Omega_s = \text{const. } s^2\alpha^2. \quad (25.12)$$

It is therefore advantageous to choose $s$ as large as possible, while retaining the same value for $\alpha$; the resolving power is not affected provided formula (25.8) remains valid, that is, provided $s$ is small. On the contrary, the intensity, by which we mean the number of useful pulses recorded at the counter, strongly increases, without the number of parasitic pulses—and it is these which now constitute background noise—being increased.

25.4.2 The Incompatibility of Resolution and Brightness

When the radius of the source becomes appreciable in comparison with the radius of the spherical aberration disc, however, the theory of § 25.2 has to be modified; equation (25.8), which was obtained for a point source, ceases to be valid, and the effective resolution drops markedly. Only rarely can the spectroscopist choose between high resolution and high brightness; a specialized instrument has to be adopted, in which one of these qualities is sacrificed to the advantage of the other. More often, he will seek a "universal" instrument, which should achieve a reasonable compromise between these mutually opposed requirements.

To attain this result, there are two parameters at the designer’s disposal: the length, $L$, and the angular aperture, $\alpha$, and thus also the diameter of the instrument. The values taken by these dimensions determine not only the quality of the spectrograph, but also its bulk, and in consequence, its price and the electrical power supply which will be necessary.

25.4.3 The Advantages of Large Spectrographs

Increasing the length offers an effective, if costly, means of resolving the intensity versus resolution dilemma. The diameter of the spherical aberration disc, $C_s\alpha^8$, is proportional to the length $L$, like $C_s$ itself, so that by
lengthening $L$, we can always arrive at a situation in which the source diameter is negligible. Large spectrographs are the best. For economic reasons, however, $L$ is never greater than two metres in modern versions, and is usually situated between $0.75$ and $1.5$ metres. It is possible, however, that designers may be led to exceed these limits in the near future, in order to unravel the complex spectra of the heavy nuclei.

After protracted experimental study, therefore, the length $L$ which has come to be adopted is of the order of a metre. We can thus certainly use a source whose diameter lies between one and a few millimetres, and this provides a suitable intensity.

These, then, are the average values of the principal data necessary for the construction of a good commercial spectrograph.

25.4.4 Source-width and Angular Aperture—the Best Compromise

One parameter is still free—the angle $\alpha$. Its choice remains at the disposal of the operator, and need not be fixed by the designer. The spectroscopist can easily change the aperture diaphragm, and choose the value of $\alpha$ as straightforwardly as that of $s$. His choice can then be made in the following way. First, the final intensity which will be necessary is fixed at an exact value and this determines the value of the product $s \times \alpha$. An optimum value of the ratio of the source radius $s$ and the angular aperture $\alpha$ then exists, for which the resolving power is a maximum. A simple calculation demonstrates this.

First, we calculate the broadening of the aberration disc which is the result of the finite size of the source. If we denote the radius of the source by $s$, then the Gaussian image, $B'$, of a point $B$ which lies on the edge of the source, will lie in the frontal plane containing $A'$, the Gaussian image of $A$, the centre of the source. The distance between $A'$ and $B'$ is given by

$$A'B' = Gs.$$ 

The rays which emerge from $B$ are very close to those which emerge from $A$, and they produce a disc in the focal plane which is very similar to that due to $A$. We shall neglect the distortions of this figure, and only take account of the displacement which occurs on transferring from $A$ to $B$. If we combine all the elementary discs which correspond to different points around the source (Fig. 315), we obtain an overall enlargement of the figure in the focal plane which is equal to twice $A'B'$. (This is because the circles which define the opposite extremities of the resultant ring correspond to points on the edge of the source taken in pairs, at the ends of the same diameter.) The total diameter of the image spot is therefore given by

$$\theta' = \theta_s + 2Gs = C_s \alpha^3 + 2Gs, \quad (25.13)$$
so that to obtain the best resolving power with a constant brightness, we must look for the minimum value of (25.13), subject to the following condition:

\[ s^2 \alpha^2 = \text{const.} = K^2. \]  

(25.14)

This ensures that the intensity at the receiver is constant when the radius of the source is varied, the brightness of the latter remaining unchanged.

\[ \text{Fig. 315. The effect of the finite size of the source on the aberration spot.} \]

(this means, from the practical standpoint, that the thickness of the radioactive deposit must remain unaltered).

The derivative \( \frac{\partial q'}{\partial \alpha} \) is zero when the value of \( \alpha \) is a root of the equation:

\[ 3C_s \alpha^2 - \frac{2G K}{\alpha^2} = 0, \]  

(25.15)

and this gives

\[ Gs = (3/2) \varrho_s \]  

(25.16)

as the optimum condition. We then find

\[ \varrho' = 4\varrho_s, \]  

(25.17)

and the resolving power, \( P_s \), falls to one-quarter of the ideal value, \( P \), which would correspond to a point source:

\[ P_s = (1/4) P; \]  

(25.18)

\( \alpha \) and \( s \) are connected by the relation

\[ s = \frac{3C_s}{2G} \alpha^3. \]

In the ordinary case for which \( G = 1 \), equation (25.19) simply expresses the fact that the optimum diameter of the source is equal to one-and-a-half times that of the spherical aberration disc.
The relations (25.9), (25.18) and (25.19) determine the dimensions of the various parts of the spectrograph completely, once the length, \( L \), and the resolving power have been settled.

Numerical calculation based upon these formulae brings out once again the difficulty of reconciling resolving power and brightness which was found in the older instruments where the simple system which we have just analysed was used—where, that is, the selective diaphragm was simply a hole. The value of Witcher’s discovery, which is described in the following sections and which avoids this problem, is obvious.

Let us consider a numerical example, in which the field is uniform, the length \( L = 1 \) m, and the resolving power with an extended source, \( P_s = 100 \).

Then

\[
P = 400, \\
\alpha = 1/14 \text{ radian} = 4^\circ 5', \\
C_s = 0.5 \text{ m},
\]

and from (25.19)

\[
s = 0.55 \text{ mm}.
\]

The brightness would then be very low since \( \Omega_s = 0.0013 \). It was for this reason that the spectrographs which were in use up to 1944 were designed to give a very modest resolution, of the order of \( P_s = 20 \), for this leads to

\[
P = 80, \quad \alpha \approx 9^\circ, \quad s = 6 \text{ mm} \quad \text{and} \quad \Omega_s = 0.006.
\]

25.5 THE ANNULAR SELECTIVE DIAPHRAGM
(WITCHER)

25.5.1 The Witcher Ring: Point Source

In all modern spectrographs, whatever type of lens be chosen, the brightness of the instrument is considerably increased by means of a procedure introduced by Witcher (1941) and developed by Frankel (1948). These two authors pointed out that the solid angle \( \Omega_s \) would be considerably increased if a hollow conical beam containing all the rays emitted at angles between \( 2(\alpha - \varepsilon) \) and \( 2(\alpha + \varepsilon) \) were used, rather than a solid cone of rays of angle \( 2\alpha \) (see Fig. 316). In the plane at which they leave the lens, these rays all touch the caustic surface \( S_t \) which has its apex at \( A_t' \); the mean rays, emitted at an angle \( \alpha \), touch the caustic along a circle \( \Gamma' \) which lies in the frontal plane \( \Pi' \); the whole beam corresponds to emission angles between \( \alpha - \varepsilon \) and \( \alpha + \varepsilon \), and thus describes in the plane \( \Pi' \) a narrow ring, the thickness \( \delta_s R \) of which is second order in \( \varepsilon \), and the outer contour of which is the circle \( \Gamma' \); the radius of the latter we shall call \( R \). We can thus place the diaphragm which performs the energy selection in the plane \( \Pi' \): we have simply to cut out an annular opening of radius \( R \) and thickness \( \delta_s R \), which will allow
only the monochromatic beam of energy \( q_1 \) to pass. So far, the theory of this system has only been worked out for the case of a uniform focusing field, but it is easy to make a qualitative estimate of the order of magnitude of the improvement which is possible. With the new value of the solid angle, we find

\[
\Omega_s = \frac{\Omega}{4\pi} = \frac{4\pi \varepsilon \sin \alpha}{4\pi},
\]

or

\[
\Omega_s = \varepsilon \sin \alpha.
\]

The brightness now depends upon the product \( \varepsilon \sin \alpha \) and as the fineness of the focusing ring is now virtually independent of \( \alpha \), and dependent only upon \( \varepsilon \), we can make \( \alpha \) as large as \( 45^\circ \). Further, the values given to \( \varepsilon \) are of the same order of magnitude as those given to \( \alpha \) in the simpler arrangement, so that by using a Witcher ring, we can hope to increase the brightness by a factor of about \( \frac{\sin \alpha}{\varepsilon} \); about seven times, for example, when the final brightness is \( 10^\circ \) \( (\varepsilon = 18^\circ, \alpha = 45^\circ) \).

Experiment confirms these predictions in the uniform field case, and shows that this device is equally effective in all the other kinds of spectrograph.

### 25.5.2 Position and Breadth of the Focusing Ring

#### (i) Methods

If the brightness is strongly increased, the energy selection mechanism remains substantially unchanged, and the arguments set out for the spectrograph with a circular opening remain valid, provided the diameter of the hole is replaced by the width of the annular slit. The calculations become far more complicated when the position of the slit makes simple Gaussian
optics inadequate. The ring is defined by three parameters, its position on
the axis, its radius and its thickness, and the determination of these requires
an exact or approximate knowledge of the form of the caustic; this in turn
implies that the calculation will be at least third order. Two theories have
been developed.

One theory is rigorous, but is restricted to the uniform homogeneous
field: it is due to Dumond (1949) and Persico (1949). The other theory is
general, but is approximate and only correct up to the third order (Grivet.
1950a, 1951). This gives good accuracy provided that the rays are not in-
clined to the axis more steeply than 30°; beyond this value, in the range which
is often employed nowadays, for which the angles lie between 30° and 45°, the
orders of magnitude which this theory yields are still useful, and above
all, it gives a clear indication of the relative importance of the various
parameters. The laws remain dimensionally correct, and the coefficients
can be obtained by measurement, so that the numerical shortcomings of
the theory can be corrected.

(ii) The theory of the uniform field; Persico (1950)

We call \( B_\theta \) the (magnetic) "rigidity" of the particle \( B_\theta = \frac{m v}{e} = \frac{p}{e} \),
where \( \theta \) is the radius of the projection of the trajectory on a frontal plane.
The equation of motion which gives the fictitious trajectory in a moving
meridian plane is

\[
r = 2q \sin \alpha \sin \left( \frac{z}{2q \cos \alpha} \right),
\]

in which \( \alpha \) is the angle between the axis and the direction of the velocity
with which the particle leaves the source. Since the ring is, geometrically,
the place at which the trajectories touch their envelope (the caustic), the
trace of the ring in the moving meridian plane will be the point of contact
of the family of trajectories (25.21) with its envelope when \( \alpha \) varies. The
second equation for the position will therefore be obtained by setting the
derivative of (25.21) with respect to \( \alpha \) equal to zero:

\[
2q \cos \alpha \sin \left( \frac{z}{q \cos \alpha} \right) + \frac{z \sin^2 \alpha}{\cos^2 \alpha} \cos \left( \frac{z}{2q \cos \alpha} \right) = 0.
\]

It is simpler to replace the pair of equations (25.21) and (25.22) by the set

\[
r = 2q \sin \alpha \sin \Theta,
\]

\[
\tan \Theta - \Theta = - \tan^2 \alpha,
\]

\[
\Theta = \frac{z}{2q \cos \alpha},
\]
in which we have introduced the parameter $\Theta$ which has a simple physical meaning: it is the angle between the moving meridian plane and the direction of the initial velocity. The function $\Theta(\alpha)$, as defined by (25.24), is plotted in Fig. 317, after Dumond (1949). On substituting for $\Theta$ in equations (25.23) and (25.25), we obtain equations for the possible positions at which

![Fig. 317. The form of the function, $\Theta(\alpha)$.](image)

the ring may be placed (Fig. 318; see Persico and Geoffrion, 1950). To each value of $\alpha$, there corresponds a value of $z$ [$z = z_s(\alpha)$] which defines the abscissa of the ring, and a value of $r$ [$r = R$] which defines its radius. This correspondence is plotted in Fig. 318.

![Fig. 318. The position ($z_s$) and radius ($R$) of the annular diaphragm, for different values of the angle $\alpha$.](image)
The width which the annulus should be given, $\delta R$, is obtained by differentiating (25.23) with respect to $\alpha$, and by taking as the value of $\delta \alpha$ the angle $2\varepsilon$, which measures the angular aperture of the useful conical region of space occupied by the beam, and is defined by

$$\alpha - \varepsilon < \alpha < \alpha + \varepsilon.$$

Let us expand (25.23) as a Taylor series; for $z = z_0$, we have $\frac{\partial r}{\partial \alpha} = 0$. If we neglect powers of $\varepsilon$ higher than 2, we obtain

$$\delta R_s = \left( \frac{\partial^2 r}{\partial \alpha^2} \right) \frac{\varepsilon^2}{2}. \quad (25.26)$$

By differentiating (25.23), and taking (25.24) into account, we obtain

$$\delta R_s = \rho \left( \frac{\sin \Theta}{\sin \alpha} \right) (3 + \cos^2 \alpha \tan^2 \Theta) \varepsilon^2. \quad (25.26')$$

The expressions which give $r$ and $z$ [namely, (25.23) and (25.25)], depend linearly upon $\rho$. A variation, $B \, d\rho$, of the momentum increases all the dimensions homocentrically about the point $S$ in the ratio $(1 + d\rho/\rho)$.

In particular, the chromatic dilatation of the radius, $\delta R_c$, is given by

$$\frac{\delta R_c}{R} = \frac{\delta \rho}{\rho}. \quad (25.27)$$

Two lines will be completely separated if the difference $\delta R_c$ is equal to $\delta R_s$. The value of the resolving power,

$$P = \frac{\rho}{\delta \rho},$$

is therefore obtained by simply replacing $\delta R_c$ by $\delta R_s$, according to (25.26'). This results in

$$P = \frac{2}{3\varepsilon^2}, \quad (25.28)$$

or, if we use (25.20) to express this as a function of the useful solid angle $\Omega_s$,

$$P = \left( \frac{2}{3} \frac{\sin \alpha}{\varepsilon} \right) \frac{1}{\Omega_s}. \quad (25.29)$$

Comparing this with (25.11), we see that the relation between $P$ and $\Omega_s$ is far more favourable when the ring is used; the coefficient $\frac{1}{2}$ is replaced by $\frac{2\sin \alpha}{3\varepsilon}$ which is of the order of 3 to 5.

The progress which we have made by introducing the ring is thus expressed by $K$ in

$$P = \frac{K}{\Omega_s};$$

there is a marked advantage, and $K$ may increase from 0.5 to 5.
(iii) The arbitrary lens. The straight ray approximation

(A) The method. The same advantage is obtained whatever type of magnetic lens be employed, as experiment has shown in every recent design. Theory, however, can no longer give the position of the ring, its breadth, or its chromatic displacement, with the excellent accuracy which characterized the formulae which have been obtained for the uniform field. Now, we have to be satisfied with a third order theory, the predictions of which become only qualitative for $\alpha > 30^\circ$; and in practice $\alpha$ is often given a value greater than $30^\circ$ in order to improve the clarity. It is for this reason that only the principle underlying the calculations will be given.

We need to find an expression describing the ends of the trajectories, which is valid at least to the third order approximation; we obtain such an expression by representing the stretch of trajectory between the plane in which the focusing ring lies and the plane of the Gaussian image $A'$ of the source by a straight line, and by expressing the equation for this straight line as a series expansion in $\alpha$ (the initial emission angle) up to $\alpha^3$ inclusive. It is legitimate to represent the emergent ray by a straight line in the case of the short-coil spectrograph since the magnetic field is completely negligible in the region between the ring and the Gaussian image $A'$. In the other models, this supposition is only approximate, and can only be justified by considering the form of the trajectories, which usually pass through a point of inflexion at $A'$; this is true, for example, of the sinusoidal path in the uniform field spectrograph.

![Diagram](image)

Fig. 319. The intersection of the straight emergent ray with the Gaussian image plane.

As the emergent ray $\Delta$ is assumed to be straight, it will be completely defined (Grivet and Paquien, 1950) by the point $P$ at which it intersects the Gaussian image plane $A'$ and by its gradient, $\tan \theta$, at this point (see Fig. 319); the distance $A'P$, of $P$ from the axis is given by the spherical aberration coefficient

$$A'P = G C_s \alpha^3,$$

and the slope is given by a third order expansion of the following kind:

$$\tan \theta = \frac{\alpha}{G} - \mu \alpha^3.$$
With the aid of the theory of third order aberrations (Grivet, 1950e), it is possible to calculate $\mu$ as a function of the coma coefficient $F$, the spherical aberration coefficient $C_s$, the magnification $G$ and the object focal length, $f_0 = \overline{FH_0}$ by means of the general expression for $\mu$:

$$\mu = \frac{1}{G} \left( \frac{1}{6} - F \right) + \frac{C_s}{g M_0}. \quad (25.32)$$

This expression contains both the algebraic distance $M_0$ between the source and the aperture diaphragm, and the magnification $g$ relative to this plane, but only two cases are of practical interest:
(a) spectrographs with unit magnification $G$, for which $\mu$ vanishes from the results;
(b) spectrographs with a high magnification $G$; the role of the constant $\mu$ is then an important one, but in instruments of this kind, the source is always placed at, or very close to, the focus of the lens (large $G$) and equation (25.32) can then be simplified. We find

$$\mu = \frac{1}{G} \left( \frac{1}{6} - F \right) + \frac{C_s}{f_0}. \quad (25.33)$$

Details of the proofs of these formulae are to be found in an article by Grivet (1951). To lighten the labour involved in writing down the full expression, we shall continue to use "$\mu$" in the equations, and only substitute the value given by equation (25.33) when we arrive at the final results.

**Fig. 320.** Trajectories in a rotating meridian plane.

For simplicity, we shall always study the trajectory which lies in the rotating meridional plane which contains at every instant the electron which is describing the ray in question; $Oz$ will be the axis of revolution, and the axis normal to $Oz$ at $A'$ we call $Oy$. The origin of coordinates is thus at $A'$, the Gaussian image of $A$ (see Fig. 320).

In this system, the equation of the emergent ray is of the form

$$y = G C_s \alpha^3 + z \tan \theta, \quad (25.34)$$
in which the algebraic values of $G$, $C_s$, $\alpha$ and $\theta$ are defined according to the usual optical conventions. The angle $\theta$ is related to the principal parameter, $\alpha$, by relation (25.31).
(B) The focusing ring

(a) The position and radius of the ring. The point of contact of the straight line and the meridian of the caustic surface is obtained from equations (25.31) and (25.34) together with the derivative of (25.34) with respect to \( \alpha \), namely:

\[
0 = 3G C_s \alpha^2 + z \frac{d(tan \theta)}{d \alpha}.
\]  

(25.35)

The ordinate of the point of contact is the radius \( R \) of the focusing ring \( I' \), and its abscissa \( z_R \) fixes the position of the circle \( A \) with respect to the origin, that is, to the Gaussian image \( A' \). Equations (25.31), (25.34) and (25.35) lead to the following representation of the various quantities with \( \alpha \) as parameter:

\[
z_R = -\frac{3G^2 C_s \alpha^2}{1 - 3 \mu G \alpha^2}.
\]  

(25.36)

\[
R = \frac{2G C_s \alpha^3}{1 - 3 \mu G \alpha^2}.
\]  

(25.37)

(b) The width of the ring. We shall obtain the width of the focusing ring, \( \delta_s R \), when the emission angle varies about a mean value \( \alpha_0 \) between the limits \( \alpha_0 + \varepsilon \) and \( \alpha_0 - \varepsilon \) (see Fig. 321). The equation for the emergent

![Fig. 321. The breadth of the ring.](image)

ray which corresponds to the value \( \alpha = \alpha_0 + \varepsilon \) is obtained by differentiating equation (25.34) with respect to \( \alpha \). We write \( \delta \alpha = \varepsilon \), and for \( \delta Y \), which, for geometrical reasons, is second order in \( \varepsilon \) (it is the distance of a point on a curve to the tangent at a neighbouring point), we obtain the following expression:

\[
\delta Y = G C_s [3 \alpha_0^3 \varepsilon + 3 \alpha_0 \varepsilon^2 + z \delta(tan \theta)],
\]  

(25.38)

\[
\delta(tan \theta) = \frac{\varepsilon}{G} - 3 \mu \alpha_0 \varepsilon^2.
\]  

(25.39)

We write \( \alpha \) simply, instead of \( \alpha_0 \), and we consider the plane \( z = z_R \), and finally obtain:

\[
\delta_s R = \frac{3G C_s \alpha}{1 - 3 \mu G \alpha^2},
\]  

(25.40)

for the width which the annulus in the diaphragm ought to be given.
(c) Some special ring positions

(1) The normal position—the unit magnification spectrograph. In the majority of spectrographs which have been used so far, which have either a uniform field, a short coil, or a Siegbahn field (Siegbahn, 1944, 1946; Siegbahn and Svartholm, 1946), the magnification \( G \) is unity \( (G = -1) \), the angle \( \alpha \) is about twenty degrees, and the ring remains close to the Gaussian image \( A' \). These are cases in which we may neglect the terms \( \mu G \alpha^2 = \mu \alpha^2 \) in comparison with unity, and the formulae reduce to:

\[
\begin{align*}
z_R &= 3 C_3 \alpha^2, \\
R &= 2 C_3 \alpha^3, \\
\delta'_3 R &= 3 C_3 \alpha e^2.
\end{align*}
\]

(2) The Siegbahn and Slättis position: the high magnification spectrograph. The arrangement which Siegbahn and Slättis suggested (1949) is characterized by the fact that the mean ray of the beam emerges from the first lens parallel to the axis of the instrument (see Fig. 322). It is therefore defined by \( \theta = 0 \), or, using equation (25.31), by

\[
\frac{1}{G} - \mu \alpha^2 = 0,
\]

so that \( G \) is very large and remains considerably greater than one even when \( \alpha \) is itself large. In practice, we go up to \( \alpha = 45^\circ \). The lens characteristics then give values of \( G \) of the order of 5, as is shown in the references quoted above. In this extreme case, the theory can no longer hope to be accurate, but it still provides a useful model which takes the peculiarities of the
arrangement satisfactorily into account. In every case, the term in $\alpha$ in the
denominators of formulae (25.36), (25.37) and (25.40) becomes first order;
if we take equation (25.25) into account and leave $G$ and $\alpha$ as parameters
(linked as they are to the value of $\mu$ which is characteristic of the lens),
we find:

$$z_R = \frac{3}{2} G^2 C_s \alpha^2,$$  \hspace{1cm} (25.45)

$$R = G C_s \alpha^3,$$  \hspace{1cm} (25.46)

$$\delta s R = \frac{3}{2} G C_s \alpha \varepsilon^2.$$  \hspace{1cm} (25.47)

Equation (25.45) is not very suggestive here, as the Gaussian image $A'$
is very distant on the left-hand side, but it is possible to show that the ring
remains at a reasonable distance from the lens, and that this type of spectro-
graph will, in consequence, be of about the same length as the others.

The reader will find numerous references in the Bibliography which
complement those already cited; for an overall survey of all types of $\beta$-ray
spectrograph, the reader should refer to the article by Persico and Geoffrion
(1950), which we have already mentioned, and to a long article by Gerholm
(1956) which includes a very extensive bibliography.

25.6 MODERN TRENDS

25.6.1 Choice of Diaphragms, and Line-shape

The resolving power which we have calculated for various situations in
the preceding paragraphs gives a fair estimate of the quality of the device
for general comparison, but nowadays, the spectroscopist requires a more
refined measure of the performance of his spectrograph. A very satisfactory
answer to this demand is provided by a detailed study of the line-shape
produced by the instrument for a hypothetical monochromatic source. This
way of characterizing the "quality" was introduced in 1949 for the solenoid
spectrograph by Dumond, who later considerably developed it (1952);
Hubert (1953) went into it still more deeply, and showed that the large pe-
destal of the line which is obtained with a simple stop could be largely
suppressed by using instead a set of diaphragms, placed at rather widely
separated points along the axis. With this arrangement, the ordinary stop
is still used, but its radius (or its width, in the case of a ring focus) must
be adjusted, and is, in fact, a new parameter; the width which was calcu-
lated earlier and found to be equal to the spherical aberration spot is no
longer the best choice. Hubert's theory has been substantiated by all recent
theoretical work (cf. Beard, 1957) and by experiment: his baffles are in-
corporated in all modern instruments (Dumond, 1957; Jungermann et al.,
1962). The action of this system of diaphragms is very different from that of the ordinary "helical baffle" commonly used in nearly all magnetic devices, which cuts down the parasitic background by distinguishing between the useful rays—helical in shape—and the parasites, which are more randomly curved.

The calculations of Dumond and Hubert apply only to the uniform field, but it is also possible to develop a line-shape theory for other kinds of lens, starting from Grivet’s theory; this development was successfully undertaken by Schneider (1957), who was able to calculate the optimum choice of ring diaphragm in terms of the line-width in a Siegbahn–Slätis spectrograph.

25.6.2 Correction of the Spherical Aberration

It was early discovered by Kervin (1949) and Persico (1950) that spherical aberration could be corrected in a simple way for the hollow conical bundle in the ring spectrograph. In principle, this may be achieved by placing a magnetic dipole of suitable strength at the centre of the spectrograph, inside the basic stop. This dipole can consist of a fairly small coil. In fact, the early attempts to correct a particular spectrograph by this means met with very little success (see Grivet, Hubert and Seiden, 1952), probably because the correction requires a far higher degree of mechanical and magnetic accuracy than does the standard apparatus. Recently, however, a Russian team has constructed an instrument which is appreciably improved by this method, after a careful sequence of preliminary investigations (Dolmatova and Kel'man, 1957, 1959).
TABLE 1
SOME USEFUL CONSTANTS
(in R.M.K.S. units)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dielectric constant of free space:</td>
<td>( \varepsilon_0 = \frac{10^{-9}}{36\pi} \times 8.855 \times 10^{-12} \text{ F m}^{-1} )</td>
</tr>
<tr>
<td>Permeability of free space:</td>
<td>( \mu_0 = \frac{4\pi}{10^7} = 1.257 \times 10^{-6} \text{ henry m}^{-1} )</td>
</tr>
<tr>
<td>Velocity of light in free space:</td>
<td>( c = 2.998 \times 10^8 = 3 \times 10^8 \text{ m sec}^{-1} )</td>
</tr>
<tr>
<td>Charge of the electron or proton:</td>
<td>( e = 1.602 \times 10^{-19} \text{ coulomb (C)} )</td>
</tr>
<tr>
<td>Rest mass of the electron:</td>
<td>( m_0 = 9.107 \times 10^{-31} \text{ kg} )</td>
</tr>
<tr>
<td>Rest mass of the proton:</td>
<td>( M_0 = 1.672 \times 10^{-27} \text{ kg} )</td>
</tr>
<tr>
<td>Ratio ( e/m_0 ) (electron):</td>
<td>( \frac{e}{m_0} = 1.759 \times 10^{11} \text{ C kg}^{-1} )</td>
</tr>
<tr>
<td>Ratio ( e/M_0 ) (proton):</td>
<td>( \frac{e}{M_0} = 0.958 \times 10^8 \text{ C kg}^{-1} )</td>
</tr>
<tr>
<td>Ratio ( m_0/M_0 ):</td>
<td>( \frac{m_0}{M_0} = 1836 )</td>
</tr>
<tr>
<td>Relativistic correction factors; electron:</td>
<td>( \frac{e}{m_0 c^2} = 1.957 \times 10^{-4} \text{ V}^{-1} )</td>
</tr>
<tr>
<td>proton:</td>
<td>( \frac{e}{M_0 c^2} = 1.066 \times 10^{-9} \text{ V}^{-1} )</td>
</tr>
<tr>
<td>Boltzmann’s constant:</td>
<td>( k = 1.380 \times 10^{-23} \text{ J deg}^{-1} )</td>
</tr>
<tr>
<td>Reciprocal ( 1/k ):</td>
<td>( \frac{1}{k} = 11,600 \text{ deg eV}^{-1} )</td>
</tr>
<tr>
<td>Planck’s constant:</td>
<td>( h = 6.624 \times 10^{-34} \text{ J sec}^{-1} )</td>
</tr>
<tr>
<td>Electron volt:</td>
<td>( 1 \text{ eV} = 1.602 \times 10^{-19} \text{ J} )</td>
</tr>
<tr>
<td>Velocity of a 1 eV electron:</td>
<td>( v_e = \sqrt{2e/m_0} = 5.932 \times 10^8 \text{ m sec}^{-1} )</td>
</tr>
<tr>
<td>Velocity of a 1 eV proton:</td>
<td>( v_p = 1.384 \times 10^4 \text{ m sec}^{-1} )</td>
</tr>
</tbody>
</table>
TABLE 2

SOME USEFUL FORMULAE

(The numerical values are given in R.M.K.S. units for electrons)

(1) Low velocities \((v \ll c)\)

\[
v = \sqrt{\frac{2e}{m} \cdot \sqrt{V}} = 0.5932 \times 10^6 \sqrt{V}
\]
\[
p = m v = 5.425 \times 10^{-20} \sqrt{V}
\]
\[
B(q) = \frac{mv}{e} = 5.685 \times 10^{-12} v
\]
\[
B(q) = \sqrt{\frac{2mV}{e}} = 3.372 \times 10^{-6} \sqrt{V}
\]
\[
v = \frac{e}{m} B(q) = 1.759 \times 10^{11} (B q)
\]
\[
\lambda = \frac{h}{p} = \frac{12.26}{\sqrt{V}} \times 10^{-10}
\]
\[
or\]
\[
\lambda = \frac{12.26}{\sqrt{V}} \AA
\]

\(v\) in m sec\(^{-1}\); \(B\) in teslas; \(p\) in metres; \(V\) in volts; \(m\) in kg; \(e\) in coulombs.

(2) Relativistic velocities

\[
\varepsilon = \frac{e}{2m_0 c^2} = 0.978 \times 10^{-6}
\]
\[
V^* = V(1 + \varepsilon V)
\]
\[
\frac{m}{m_0} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} = 1 + 2\varepsilon V
\]
\[
p = m v = 2m_0 c \sqrt{\varepsilon V(1 + \varepsilon V)}
\]
\[
\beta = \frac{v}{c} = 2 \frac{\sqrt{\varepsilon V(1 + \varepsilon V)}}{1 + 2\varepsilon V}
\]
\[
B(q) = \frac{p}{e} = 2 \frac{m_0 c}{e} \frac{\sqrt{\varepsilon(1 + \varepsilon V)}}{\varepsilon(1 + \varepsilon V)}
\]
\[
\lambda = \frac{h}{p} = \frac{h}{2m_0 c} \frac{1}{\varepsilon(1 + \varepsilon V)}
\]
### TABLE 3

**FUNDAMENTAL CHARACTERISTICS OF FREE ELECTRONS IN MOTION**

In the following table, the following quantities are tabulated to four significant figures:

- $V$ — The potential necessary to raise the kinetic energy of an electron initially at rest to a value $E$. $V$ is expressed in volts.

- $V^*$ — The potential, relativistically corrected in such a way that the relativistic value of the momentum is obtained by simply replacing $V$ by $V^*$ in the non-relativistic expression. $V$ and $V^*$ are related by the equation:
  
  $$V^* = V + \frac{e}{2m_0 c^2} V^2,$$

  and the relativistic momentum $p$ has the form:

  $$p = m_0 c \sqrt{\frac{2E}{m_0 c^2} + \left(\frac{E}{m_0 c^2}\right)^2} = (2m_0^2 V^*)^{1/2}.$$

- $B\rho$ — The product of the magnetic induction $B$, measured in teslas ($1 \text{ tesla} = 10^4 \text{ gauss}$) and the radius of curvature of the electron trajectory $\rho$, measured in meters, in a plane perpendicular to the field ($1 \text{ tesla} \cdot \text{m} = 10^6 \text{ gauss} \cdot \text{cm}$).

- $\lambda$ — The de Broglie wavelength of the electron in the absence of a magnetic field; $\lambda$ is expressed in metres ($1 \text{ m} = 10^{10} \text{ Å}$).
### (a) Thermal and photoelectric velocities

(0·1 to 10 eV)

<table>
<thead>
<tr>
<th>(V) volts</th>
<th>(V^*) volts</th>
<th>(B\rho) tesla ⋅ m</th>
<th>(\lambda) m</th>
</tr>
</thead>
<tbody>
<tr>
<td>0·1</td>
<td>0·1000</td>
<td>(1·846 \times 10^{-6})</td>
<td>2·240 \times 10^{-9}</td>
</tr>
<tr>
<td>0·2</td>
<td>0·2000</td>
<td>2·132</td>
<td>1·940</td>
</tr>
<tr>
<td>0·3</td>
<td>0·3000</td>
<td>2·385</td>
<td>1·734</td>
</tr>
<tr>
<td>0·4</td>
<td>0·4000</td>
<td>2·616</td>
<td>1·583</td>
</tr>
<tr>
<td>0·5</td>
<td>0·5000</td>
<td>2·821</td>
<td>1·469</td>
</tr>
<tr>
<td>0·6</td>
<td>0·6000</td>
<td>3·024</td>
<td>1·371</td>
</tr>
<tr>
<td>0·7</td>
<td>0·7000</td>
<td>3·184</td>
<td>1·293</td>
</tr>
<tr>
<td>0·8</td>
<td>0·8000</td>
<td>3·372</td>
<td>1·226</td>
</tr>
<tr>
<td>0·9</td>
<td>0·9000</td>
<td>4·769</td>
<td>8·671 \times 10^{-10}</td>
</tr>
<tr>
<td>1</td>
<td>1·0000</td>
<td>5·840</td>
<td>7·080</td>
</tr>
<tr>
<td>2</td>
<td>2·0000</td>
<td>6·744</td>
<td>6·132</td>
</tr>
<tr>
<td>3</td>
<td>3·0000</td>
<td>7·540</td>
<td>5·484</td>
</tr>
<tr>
<td>4</td>
<td>4·0000</td>
<td>8·260</td>
<td>5·006</td>
</tr>
<tr>
<td>5</td>
<td>5·0000</td>
<td>8·922</td>
<td>4·635</td>
</tr>
<tr>
<td>6</td>
<td>6·0000</td>
<td>9·538</td>
<td>4·336</td>
</tr>
<tr>
<td>7</td>
<td>7·0000</td>
<td>10·12</td>
<td>4·088</td>
</tr>
<tr>
<td>8</td>
<td>8·0000</td>
<td>10·66</td>
<td>3·878</td>
</tr>
</tbody>
</table>

### (b) Vacuum tube region (10 to 10⁴ eV)

<table>
<thead>
<tr>
<th>(V) volts</th>
<th>(V^*) volts</th>
<th>(B\rho) tesla ⋅ m</th>
<th>(\lambda) m</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10·00</td>
<td>10·66</td>
<td>3·878 \times 10^{-10}</td>
</tr>
<tr>
<td>20</td>
<td>20·00</td>
<td>15·08</td>
<td>2·742</td>
</tr>
<tr>
<td>30</td>
<td>30·00</td>
<td>18·47</td>
<td>2·239</td>
</tr>
<tr>
<td>40</td>
<td>40·00</td>
<td>21·32</td>
<td>1·939</td>
</tr>
<tr>
<td>50</td>
<td>50·00</td>
<td>23·84</td>
<td>1·734</td>
</tr>
<tr>
<td>100</td>
<td>100·0</td>
<td>33·72</td>
<td>1·226</td>
</tr>
<tr>
<td>150</td>
<td>150·0</td>
<td>41·30</td>
<td>1·001</td>
</tr>
<tr>
<td>200</td>
<td>200·0</td>
<td>47·69</td>
<td>8·670 \times 10^{-11}</td>
</tr>
<tr>
<td>250</td>
<td>250·0</td>
<td>53·32</td>
<td>7·755</td>
</tr>
<tr>
<td>300</td>
<td>300·0</td>
<td>58·41</td>
<td>7·079</td>
</tr>
<tr>
<td>500</td>
<td>500·2</td>
<td>75·42</td>
<td>5·483</td>
</tr>
<tr>
<td>1000</td>
<td>1001</td>
<td>106·7</td>
<td>3·876</td>
</tr>
<tr>
<td>2500</td>
<td>2506</td>
<td>168·8</td>
<td>2·449</td>
</tr>
<tr>
<td>5000</td>
<td>5024</td>
<td>239·0</td>
<td>1·730</td>
</tr>
<tr>
<td>7500</td>
<td>7555</td>
<td>293·1</td>
<td>1·410</td>
</tr>
<tr>
<td>10000</td>
<td>10097</td>
<td>338·8</td>
<td>1·220</td>
</tr>
</tbody>
</table>
(c) X-ray tubes, microscopes, diffraction cameras
(10^4 to 2 \cdot 10^6 eV)

<table>
<thead>
<tr>
<th>$V$ volts</th>
<th>$V^*$ volts</th>
<th>$B \rho$ tesla \cdot m</th>
<th>$\lambda$ m</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>10097</td>
<td>338.8</td>
<td>$1.220 \times 10^{-11}$</td>
</tr>
<tr>
<td>20000</td>
<td>20391</td>
<td>481.5</td>
<td>$8.588 \times 10^{-12}$</td>
</tr>
<tr>
<td>30000</td>
<td>30880</td>
<td>592.5</td>
<td>6.979</td>
</tr>
<tr>
<td>40000</td>
<td>41560</td>
<td>687.5</td>
<td>6.015</td>
</tr>
<tr>
<td>50000</td>
<td>52440</td>
<td>772.2</td>
<td>5.355</td>
</tr>
<tr>
<td>60000</td>
<td>63520</td>
<td>849.9</td>
<td>4.866</td>
</tr>
<tr>
<td>70000</td>
<td>74790</td>
<td>922.2</td>
<td>4.484</td>
</tr>
<tr>
<td>80000</td>
<td>86260</td>
<td>990.4</td>
<td>4.175</td>
</tr>
<tr>
<td>90000</td>
<td>97920</td>
<td>1055</td>
<td>3.918</td>
</tr>
<tr>
<td>$100 \times 10^3$</td>
<td>$109.8 \times 10^3$</td>
<td>1117</td>
<td>3.701</td>
</tr>
<tr>
<td>120</td>
<td>134-0</td>
<td>1235</td>
<td>3.349</td>
</tr>
<tr>
<td>140</td>
<td>159-1</td>
<td>1345</td>
<td>3.073</td>
</tr>
<tr>
<td>160</td>
<td>185-0</td>
<td>1450</td>
<td>2.850</td>
</tr>
<tr>
<td>180</td>
<td>211-7</td>
<td>1551</td>
<td>2.665</td>
</tr>
<tr>
<td>200</td>
<td>239-1</td>
<td>1649</td>
<td>2.507</td>
</tr>
</tbody>
</table>

(d) $\beta$-ray devices and accelerators
($2 \times 10^3$ to $20 \times 10^6$ eV)

<table>
<thead>
<tr>
<th>$V$ volts</th>
<th>$V^*$ volts</th>
<th>$B \rho$ tesla \cdot m</th>
<th>$\lambda$ m</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2 \times 10^5$</td>
<td>$2.391 \times 10^4$</td>
<td>1649</td>
<td>$2.507 \times 10^{-12}$</td>
</tr>
<tr>
<td>3</td>
<td>3.880</td>
<td>2100</td>
<td>1.968</td>
</tr>
<tr>
<td>4</td>
<td>5.565</td>
<td>2515</td>
<td>1.644</td>
</tr>
<tr>
<td>5</td>
<td>7.446</td>
<td>2909</td>
<td>1.421</td>
</tr>
<tr>
<td>6</td>
<td>9.522</td>
<td>3290</td>
<td>1.257</td>
</tr>
<tr>
<td>$8 \times 10^5$</td>
<td>$1.426 \times 10^6$</td>
<td>4027</td>
<td>1.026</td>
</tr>
<tr>
<td>$1 \times 10^6$</td>
<td>1.978</td>
<td>4743</td>
<td>$8.718 \times 10^{-13}$</td>
</tr>
<tr>
<td>1.5</td>
<td>3.701</td>
<td>6487</td>
<td>6.374</td>
</tr>
<tr>
<td>2</td>
<td>5.914</td>
<td>8200</td>
<td>5.042</td>
</tr>
<tr>
<td>10</td>
<td>107-8</td>
<td>35020</td>
<td>1.180</td>
</tr>
<tr>
<td>20</td>
<td>411-4</td>
<td>68390</td>
<td>$6.046 \times 10^{-14}$</td>
</tr>
<tr>
<td>θ</td>
<td>n = 0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>5°</td>
<td>0.365201</td>
<td>-0.254581</td>
<td>-0.813813</td>
</tr>
<tr>
<td>10°</td>
<td>0.501717</td>
<td>0.005894</td>
<td>-0.453932</td>
</tr>
<tr>
<td>20°</td>
<td>0.638358</td>
<td>0.268268</td>
<td>-0.088558</td>
</tr>
<tr>
<td>30°</td>
<td>0.718190</td>
<td>0.423320</td>
<td>0.130467</td>
</tr>
<tr>
<td>40°</td>
<td>0.774511</td>
<td>0.534092</td>
<td>0.289416</td>
</tr>
<tr>
<td>50°</td>
<td>0.817704</td>
<td>0.620144</td>
<td>0.414869</td>
</tr>
<tr>
<td>60°</td>
<td>0.852374</td>
<td>0.690081</td>
<td>0.518406</td>
</tr>
<tr>
<td>70°</td>
<td>0.880955</td>
<td>0.748422</td>
<td>0.606031</td>
</tr>
<tr>
<td>80°</td>
<td>0.904886</td>
<td>0.797813</td>
<td>0.681210</td>
</tr>
<tr>
<td>90°</td>
<td>0.925086</td>
<td>0.839927</td>
<td>0.746089</td>
</tr>
<tr>
<td>100°</td>
<td>0.942171</td>
<td>0.875872</td>
<td>0.802069</td>
</tr>
<tr>
<td>110°</td>
<td>0.956571</td>
<td>0.906416</td>
<td>0.850092</td>
</tr>
<tr>
<td>120°</td>
<td>0.968597</td>
<td>0.932102</td>
<td>0.890814</td>
</tr>
<tr>
<td>130°</td>
<td>0.978471</td>
<td>0.953322</td>
<td>0.924694</td>
</tr>
<tr>
<td>140°</td>
<td>0.983632</td>
<td>0.970362</td>
<td>0.952059</td>
</tr>
<tr>
<td>150°</td>
<td>0.992387</td>
<td>0.983428</td>
<td>0.973140</td>
</tr>
<tr>
<td>160°</td>
<td>0.996635</td>
<td>0.992665</td>
<td>0.988095</td>
</tr>
<tr>
<td>170°</td>
<td>0.999161</td>
<td>0.998171</td>
<td>0.997028</td>
</tr>
<tr>
<td>180°</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
</tbody>
</table>
Table 4 (continued)

<table>
<thead>
<tr>
<th>θ</th>
<th>n = 1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
<th>1.6</th>
<th>1.7</th>
<th>1.8</th>
<th>1.9</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>5°</td>
<td>-0.542651</td>
<td>-0.058480</td>
<td>+0.413527</td>
<td>+0.833237</td>
<td>1.165932</td>
<td>1.385435</td>
<td>1.476543</td>
<td>1.433916</td>
<td>1.265476</td>
<td>0.989351</td>
</tr>
<tr>
<td>10°</td>
<td>-0.672324</td>
<td>-0.318188</td>
<td>+0.047601</td>
<td>+0.395130</td>
<td>0.696181</td>
<td>0.930007</td>
<td>1.077683</td>
<td>1.130627</td>
<td>1.087486</td>
<td>0.954769</td>
</tr>
<tr>
<td>20°</td>
<td>-0.775450</td>
<td>-0.562976</td>
<td>-0.319350</td>
<td>-0.063159</td>
<td>-0.186880</td>
<td>0.412909</td>
<td>0.599730</td>
<td>0.735204</td>
<td>0.811308</td>
<td>0.824533</td>
</tr>
<tr>
<td>30°</td>
<td>-0.792525</td>
<td>-0.672251</td>
<td>-0.514788</td>
<td>-0.331564</td>
<td>-0.135126</td>
<td>0.061648</td>
<td>+0.246321</td>
<td>0.407651</td>
<td>0.536224</td>
<td>0.625000</td>
</tr>
<tr>
<td>40°</td>
<td>-0.756377</td>
<td>-0.705113</td>
<td>-0.617004</td>
<td>-0.498494</td>
<td>-0.357352</td>
<td>-0.202235</td>
<td>-0.042220</td>
<td>-0.113682</td>
<td>+0.260689</td>
<td>0.380236</td>
</tr>
<tr>
<td>50°</td>
<td>-0.679240</td>
<td>-0.680298</td>
<td>-0.648801</td>
<td>-0.587376</td>
<td>-0.500235</td>
<td>-0.392577</td>
<td>-0.270327</td>
<td>-0.139860</td>
<td>-0.007413</td>
<td>-0.119764</td>
</tr>
<tr>
<td>60°</td>
<td>-0.567487</td>
<td>-0.607957</td>
<td>-0.621232</td>
<td>-0.608117</td>
<td>-0.570350</td>
<td>-0.510520</td>
<td>-0.431952</td>
<td>-0.338567</td>
<td>-0.234713</td>
<td>-0.125000</td>
</tr>
<tr>
<td>70°</td>
<td>-0.428902</td>
<td>-0.496128</td>
<td>-0.542575</td>
<td>-0.576790</td>
<td>-0.571540</td>
<td>-0.555254</td>
<td>-0.520788</td>
<td>-0.468913</td>
<td>-0.402560</td>
<td>-0.324533</td>
</tr>
<tr>
<td>80°</td>
<td>-0.269688</td>
<td>-0.352628</td>
<td>-0.421019</td>
<td>-0.473792</td>
<td>-0.510309</td>
<td>-0.530315</td>
<td>-0.533992</td>
<td>-0.521924</td>
<td>-0.495078</td>
<td>-0.454769</td>
</tr>
<tr>
<td>90°</td>
<td>-0.096662</td>
<td>-0.185629</td>
<td>-0.265506</td>
<td>-0.335101</td>
<td>-0.393447</td>
<td>-0.439820</td>
<td>-0.473745</td>
<td>-0.495011</td>
<td>-0.503662</td>
<td>-0.500000</td>
</tr>
<tr>
<td>100°</td>
<td>+0.083160</td>
<td>-0.003726</td>
<td>-0.085870</td>
<td>-0.162209</td>
<td>-0.231828</td>
<td>-0.293895</td>
<td>-0.347694</td>
<td>-0.392684</td>
<td>-0.428461</td>
<td>-0.454769</td>
</tr>
<tr>
<td>110°</td>
<td>0.262707</td>
<td>+0.184212</td>
<td>+0.107352</td>
<td>+0.032911</td>
<td>-0.038356</td>
<td>-0.105747</td>
<td>-0.168608</td>
<td>-0.226363</td>
<td>-0.278485</td>
<td>-0.323533</td>
</tr>
<tr>
<td>120°</td>
<td>0.435048</td>
<td>0.369302</td>
<td>0.303277</td>
<td>0.237487</td>
<td>+0.172439</td>
<td>+0.108625</td>
<td>+0.046528</td>
<td>+0.013395</td>
<td>+0.070707</td>
<td>+0.125000</td>
</tr>
<tr>
<td>130°</td>
<td>0.593260</td>
<td>0.542960</td>
<td>0.491146</td>
<td>0.438457</td>
<td>0.385144</td>
<td>0.331534</td>
<td>0.277906</td>
<td>0.224546</td>
<td>0.171992</td>
<td>+0.119764</td>
</tr>
<tr>
<td>140°</td>
<td>0.732358</td>
<td>0.697230</td>
<td>0.660787</td>
<td>0.623160</td>
<td>0.584843</td>
<td>0.544892</td>
<td>0.504529</td>
<td>0.463536</td>
<td>0.422057</td>
<td>0.380236</td>
</tr>
<tr>
<td>150°</td>
<td>0.846085</td>
<td>0.825099</td>
<td>0.803106</td>
<td>0.780153</td>
<td>0.756288</td>
<td>0.731556</td>
<td>0.706009</td>
<td>0.679698</td>
<td>0.652678</td>
<td>0.625000</td>
</tr>
<tr>
<td>160°</td>
<td>0.930510</td>
<td>0.920782</td>
<td>0.910521</td>
<td>0.899731</td>
<td>0.888247</td>
<td>0.876616</td>
<td>0.864311</td>
<td>0.851520</td>
<td>0.838256</td>
<td>0.824533</td>
</tr>
<tr>
<td>170°</td>
<td>0.982463</td>
<td>0.979971</td>
<td>0.977330</td>
<td>0.974544</td>
<td>0.971609</td>
<td>0.968530</td>
<td>0.965306</td>
<td>0.961937</td>
<td>0.958424</td>
<td>0.954769</td>
</tr>
<tr>
<td>180°</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

This table has been kindly provided by M. Schelkunoff.
BIBLIOGRAPHY

Books of a more general nature are to be found listed in the Bibliography for Part 1. More specialized works, devoted to a single electron optical instrument, are mentioned in the appropriate section of Part 2.

BIBLIOGRAPHY FOR PART 1

I. Basic Works
   (A) on electron optics
   (B) on glass optics
   (C) on field calculations and basic techniques
   (D) on luminescent screens

II. Articles

BIBLIOGRAPHY FOR PART 2

I. Cathode Ray Tubes
   (A) books
   (B) articles

II. The Emission Microscope

III. Image Converters and Intensifiers
   (A) books and surveys on photo-electronic image devices
   (B) articles

IV. Electron Microscopy
   (A) books
   (B) conferences on electron microscopy
   (C) articles

V. Specimen Preparation
   (A) basic works
   (B) articles

VI. The Diffraction Camera and Electron Probe Devices
   (A) books
   (B) articles

VII. The Mass Spectroscope
   (A) books
   (B) articles

VIII. The β-ray Spectrograph
   (A) general references
   (B) articles

xxiii
BIBLIOGRAPHY FOR PART 1†

I. BASIC WORKS

(A) On Electron Optics

BUSCH, H. and E. BRÜCHE (1937) *Beiträge zur Elektronenoptik*, Barth, Leipzig.
(First edition distributed by Edwards, Ann Arbor.)

(B) On Glass Optics


† A number of the references in the bibliography are to communications to scientific congresses. The text in each case is to be found in the Proceedings, which used always to be published with an appreciable delay. A list of these congresses, with their full titles, is to be found in Section IVB. In the text, only the date and the place where the congress was held, and the number or page number of the communication are given.
XXVI

BIBLIOGRAPHY


(C) On Field Calculations and Basic Techniques

SOUTHWELL, R. V. (1946) Relaxation Methods, Oxford University Press.

(D) On Luminescent Screens


II. ARTICLES

ALBRECHT, R. (1956) Z. Naturforsch. 11a, 156.
BECKER, H., A. WALLRAPP (1940) Arch. Elektrotechn. 34, 115.
BIBLIOGRAPHY

Borries, B. von (1942a) Z. Physik 119, 498; (1942b) Z. angew. Photograph. 4, 42; (1944)
Z. Physik 122, 539 (1948); Optik 3, 321 and 389.
BRÜCHE, E. (1942) Kolloid Z. 100, 192.
BRÜCHE, E., and H. JOHANNSON (1932) Naturwiss. 20, 49 and 353.
BIBLIOGRAPHY

COURANT, E. D., M. S. LIVINGSTON and H. S. SNYDER (1952) Phys. Rev. 88, 1168;
(1953) 91, 202.
DOSSE, J. (1940) Z. Physik 115, 530; (1941a) 117, 437; (1941b) 117, 316; (1941c) 117,
722; (1941d) 118, 375.
DYKIMKOV, A. D., T. YA. FISHKOVA and S. YA. YAVOR (1964a) Zh. Tekh. Fiz. 35, 759
Phys. 8, 639).


GLASER, W. (1933a) Z. Physik 81, 647; (1933b) 83, 104; (1940a) 116, 19; (1940b) 116, 56; (1941a) 117, 285; (1941b) 118, 264; (1943a) 120, 1; (1943b) 121, 647; (1948) Nature 162, 455; (1950a) Proc. Phys. Soc. B64, 114; (1950b) ZAMP 1, 363; (1950c) Paris, communication No. 24; (1950d) Ann. der Phys. 7, 213; (1956) "Elektronen und Ionenoptik" in Handbuch der Physik (Encyclopedia of Physics) 33, 123-395.


GLASER, W., and F. LENZ (1951) Ann. der Phys. 9, 19.

GLASER, W., and ROBL (1951) ZAMP 2, 444.


GOBRECHT, R. (1941) Arch. Elektrotechn. 35, 672; (1942) 36, 484.


HEISE, F. (1949) Optik 5, 479.

HEISE, F., and O. RANG (1949) Optik 5, 201.

BIBLIOGRAPHY

HIMPAK, J. (1939) Telefunkenzöhe 16, 198; (1942) Elektrotechn. Z. 63, 349.
Lond. B 63, 75; (1950b) J. Appl. Phys. 21, 966.
USSR. Phys. Ser. 27, 1116.
LE POOLE, J. B., and A. C. VAN DORSTEN (1951) Communication to the St Andrews Congress of the British Microscopical Society.
LIPPERT, W., and W. PohlIT (1952) Optik 9, 456.
LOEB, J. (1947) Onde electr. 27, 27.
MAHL, H., and A. RECKMAG (1944) Z. Physik 122, 60.
PAUL, W., H. P. REINHARD, and V. VON ZAHN (1958) Z. Physik 152, 143.


POCKE, M. (1951) ZAMP 3, 441; (1952) 4, 1.


RANG, O. (1949) Optik 5, 518.


RECKNAGEL, A. (1937) Z. Physik 104, 381; (1940) 117, 67; (1941) 117, 689; (1943) 120, 313.


Schiessel, M. (1952) Optik 9, 145.


Seeliger, R. (1949) Optik 5, 490; (1951) 8, 311; (1953) 10, 29.


SIEGBAHL, K. (1946) Phil. Mag. 37, 162.


BIBLIOGRAPHY FOR PART 2

I. CATHODE RAY TUBES

(Chapter 12)

(A) Books


XXXIV  BIBLIOGRAPHY


(B) Articles


DESENO, P. (1953) Arch. Elektrotechn. 29, 139.

DOSE, J. (1940) Z. Physik 115, 530.


EPSTEIN, D., and L. PENSAK (1946) RCA Rev. 7, 55.


NGUYEN, TIEN CHI (1952) Techniques de l’Ingénieur “Electronique” I.


ROGOWSKI, W., and H. THIILEN (1939) Arch. Elektrotechn. 33, 411.


TERRILL, H. M., and L. SWEENEY (1944a) J. Frank Inst. 237, 495; (1944b) 238, 220.
BIBLIOGRAPHY


II. THE EMISSION MICROSCOPE

(Chapter 13)

BARKER, A. N., H. O. W. RICHARDSON and N. FEATHER (1950) Research Correspondence, suppl. 3–9, 431.


BURGERS, W. G., and J. J. A. PLOOS VAN AMSTEL (1937a) Physica 4, 5; (1937b) 4, 15;
(1938a) 5, 305; (1938b) 5, 313.


GAUZIT, M., and A. SEPTIER (1951) Microscopie, (2) 1, 109.


BIBLIOGRAPHY

RECKNAGEL, A. (1943) Z. Physik 120, 331.

III. IMAGE-CONVERTERS AND INTENSIFIERS
(Chapter 14)

(A) Books and Surveys on Photo-electronic Image Devices


(B) Articles

Möltenstedt, G., Innsbruck, 1953.
Sternglass, E. J. (1954) Research Memo. 60.8.10.43, ML.
Timofeyev, P. V., and V. V. Sorokina (1960) Radio Eng. and Electronics 5, 201.

IV. ELECTRON MICROSCOPY
(Chapters 15 to 19)

(A) Books

XXXVIII

BIBLIOGRAPHY


(B) Conferences on Electron Microscopy (some of the communications to these conferences are included in the bibliography)


1951 (WASHINGTON), Proceedings, published in 1954 under the auspices of the N.B.S., with the title: Electron Physics, circular 527.

1953 (INNSBRUCK), Verhandlungen, unpublished.

1954 (GHENT), Comptes rendus, published in 1956 under the auspices of Ghent University.


1966 (KYOTO) Abstracts of the Sixth International Conference, published by the Maruzen Co. of Tokyo.


(C) Articles


BORRIES, B. VON, and E. RUSKA (1940a) Naturwiss. 28, 366; (1940b) Z. Physik 116, 249.
DIETRICH, W. (1958a) Z. Physik 151, 519; (1958b) 152, 87; (1958c) 152, 306.
DYSON, J. (1950) 127.
EITEL, W., and E. GOTTHARDT (1940) Naturwiss. 28, 367; (1942) Z. Physik 118, 714.


HALL, C. G. (1947) J. Appl. Phys. 18, 588; (1948) 19, 198 and 271; (1949) 20, 631;
(1951) 22, 655.


HANAK, J. J. (1964) RCA Rev. 25, 551.


HEARD, H. G., et al. (1952) Amer. UCRL Report, 1622; (1953) 2252; (1953a) 2321.


HEISE, F. (1949) Optik 5, 479.


HERCENBACH, W., and H. DUCKER (1956) Optik 13, 375.


KINDER, E. (1943) Naturwiss. 31, 149; (1946) 33, 367.
KOPP, C., and G. MÖLLENSTEDT (1946) Optik 1, 327; (1947) 2, 283.
STEIGERWALD, R. (1949) Optik 5, 469.
TRIMMER, R., and G. PEARLMAN (1951) Amer. TID Report, 52 N.
VANCE, A. W. (1941) RCA Rev. 5, 293.
VASTEL, J. (1951) Ann. Radioïél. 6, 84.
WALCHER, P. (1943) Z. Physik 121, 604.
WEICHAN (1962): see ENGEL et al. (1962).
V. SPECIMEN PREPARATION
(Chapter 20)

(A) Basic Works


(B) Articles

ARDENNE, M. VON (1948) Kolloid Z. 111, 22.
BORRIES, B. VON, and E. RUSKA (1939) Naturwiss. 27, 281.
CASTAING, R., and A. GUINIER (1949a) C. R. Acad. Sci. Paris 228, 2033; (1949b) 229,
1146.
RANG, O. (1952) Le Vide 7, 1211.
STOJANOVA, I. C. (1958) Berlin, 82.
New York.
VI. THE DIFFRACTION CAMERA
AND ELECTRON PROBE DEVICES
(Chapters 21 and 22)

(A) Books


Proceedings of the 2nd, 3rd, 4th, and 5th Symposia on Electron Beam Technology;

(B) Articles

and X-ray Microanalysis, Stockholm, No. 52.

Laboratoires 17, 7; (1960) Adv. in Electronics 13, 317.

nautique 63, 41.

CORRET, H. C., G. M. VAN KOPPEN, J. KRÄMER, M. VAN MENTS, J. B. LE POOLE and I. L.


DUNCUMB, P., and D. A. MELFORD (1959) Proceedings of an International Symposium on
X-ray Microscopy and X-ray Microanalysis. Stockholm, No. 82.


BIBLIOGRAPHY

SCHLEICH, F. (1958) Berlin, 278.
VERDIER, P. (1960) Thèse, Toulouse; (see DUPOUY, PERRIER and VERDIER, above).
BIBLIOGRAPHY

VII. THE MASS SPECTROSCOPE
(Chapters 23 and 24)

(A) Books


(B) Articles

EWALD, H., (1946) Z. Naturforsch. 1, 131; (1950) 5a, 229; (1951) 6a, 293.
HEIL, H. (1943) Z. Physik 120, 212.
KOCH, J. (1936) Z. Physik 100, 669.
MATTACH, J., and R. HERZOG (1934) Z. Physik 89, 786.
PAUL, W. (1948) Z. Physik 124, 244.
PAUL, W., and M. RAETHER (1953) Z. Physik 140, 262.
BIBLIOGRAPHY

Roth, E. (1949) J. Phys. 10, 17D.
Tate, J., and P. Smith (1934) Phys. Rev. 46, 773.
Thomson, J. J. (1910) Phil. Mag. 20, 752; (1911) 21, 225.
Transactions of the 1954 Vacuum Symposium, Boston, 1955, published by the Committee on Vacuum Techniques, Inc.

VIII. THE β-RAY SPECTROGRAPH
(Chapter 25)

(A) General References


(B) Articles

Nucl. Instrum. and Methods 5, 269.
Glaser, W., and E. Lammel (1943) Arch. Elektrotechn. 37, 347.
(1950c) 230, 1048; (1950d) 230, 1652; (1950e) 230, 1152.
Grivet, P., P. Hubert, and J. Seiden (1952) Physica 18, 1136.
and Methods 15, 1.
Kapitsa, P. (1924) see Tricker.
Schneider, W. (1957) Z. Instrumentent. 65, 103 and 126.
Siegbahn, K. (1942) Ark Math. Astr. Fysik 28, 1; (1944) 30A, 1; (1946) Phil. Mag. 37,
162.
Witcher, C. M. (1941) Phys. Rev. 60, 32.
INDEX

Abbe 529
Aberrations
cromatic 176, 179, 540
correction of 201
correction of 198
diffraction 533
general 155–206
geometrical 156, 164, 168
mechanical 182, 192, 220, 544
correction of 202
of quadrupole lenses 313
space charge 391
spherical 532
coefficient of 159, 194
correction of 199, 203
measurement of 186
Adjustment of microscopes 633
Alignment of microscopes 634
Analogue methods of determining
electric fields 45 ff
magnetic fields 77 ff
trajectories
in electrostatic lenses 118
in magnetic lenses 133
Angular focusing 798
Astigmatism 161
cathode ray tubes 411
geometrical in origin 161, 169, 176
mechanical in origin 183, 192, 202
of microscopes 545 ff
Atomic masses, determination of 837
Autobias for electron guns 568

Brightness
of an electron spot 107
of fluorescent screens 11

Cardinal elements
classical 103, 119, 215
immersion 105, 131
Cathode ray oscillograph 377
Cathode ray tube 377, 380–426
Chemical analysis by means of spectroscopes 840
Circuitry in spectroscopes 828
Coils for field measurements 73 ff
Cold-cathode gun 566
Collective theory 644
Collector for spectrometers 827
Collision cross-section 642
Coma 159
Condenser lens 571
Condenser-objective 594
Conjugate points 102
Contamination 621
Contrast 638–90
defocusing 669
improvement of 652, 673
in crystals 662
in self-emitting objects 436
in transparent objects 648 ff, 709
phase 655
photo-emission 447
secondary emission 441 ff, 764
thermal emission 436
Crossover 384, 428
Current density in the spot 389
Current regulation in magnetic lenses 608
Cut-off 398
Cylindrical symmetry
fields 48, 80
lenses 202, 203 ff, 289
INDEX

Dark-field microscopy 652
Deflexion of the spot in cathode ray tubes 407
Dempster spectrometer 790
Depth
of field 505
of focus 508
Deviation
electrostatic 85
magnetic 123
Diaphragms
as stops 107 ff
contrast 7, 456, 545, 649
field 522, 748 ff
influence on aberrations 198
Diffraction camera 716–51
accuracy 729
fidelity 728
high resolution 734
resolving power 721
Diffraction devices 716, 744
applications of 749
Diffraction of electrons 531
Dispersion
of mass and velocity in spectroscopes 803
of velocity 178
initial velocities 541
interaction with the object 542
Distortion 160, 192
Double-focusing in spectrometers 803

Edge
effects in spectroscopes 810
field of an electron prism 810
Eikonal function 170
Einzel lenses (unipotential lenses)
symmetrical 211 ff, 573
unsymmetrical 578
Electrical analogues 118
Electrical supplies
electron guns 568
mass spectroscopes 828
microscopes 602
Electrolytic tank
conjugate analogy 78
direct analogy 45
Electron
microscope 691–715

mirror 91, 100, 195
optical bench 15
probes 691–715, 752
high intensity 779
sensitometry 512
telescopes 378
Electrostatic velocity analysis
with a decelerating field 678
with a lens 681
Ellipticity see Aberrations, mechanical
Energy losses when electrons traverse the
object 644
Equation
for an electric field 20–57
for a magnetic field 58–84
of motion
in electrostatic lenses 95 ff, 102
in magnetic lenses 124
reduced, for trajectories 97, 135
Equivalent length of a quadrupole 296
Evaporation 694

Fermat's principle 142
Fernfokus 247
Field
curvature 161, 194
electrostatic, analogue methods 21, 54
electrostatic, direct calculation 20–57
electrostatic, in the presence of space
charge 51
electrostatic, measurement 48
emission 463 ff
magnetic, analogue methods 77
magnetic, direct calculation 58
magnetic, measurement 72
models
for electrostatic lenses 207
for magnetic lenses 259–86
Filter lens 688
Fluorescent screen 11, 401, 509
Focal length
of electrostatic lenses and mirrors 86,
91, 96 ff, 104, 119 ff
of magnetic lenses 130, 138
Focus
classical 102
immersion 105
Focusing a microscope 637
Foucault contrast 652
INDEX

Gans’ method 115
Gaussian approximation
  in electrostatic lenses 85–122
  in magnetic lenses 123–38
  limit of 155
Glaser’s bell-shaped model 260
Glass optics and electron optics 139–54
  electrostatic lenses 139
  magnetic lenses 144
Grid (or Wehnelt, or control electrode)
  231 ff, 380, 383 ff, 431
  as a specimen-holder 692
Grid lenses 224
  as velocity analysers 686
Guns 207–58
  cathode ray tube 383 ff
  electron microscope 241, 563
  triode 240

Hall effect 76
Halo test 191
Hartmann test 186
High voltage generator 602
Hodoscope 133

Iconoscope 382
Image 378
  converter 474–97
  formation
    by electrostatic lenses 101
    by magnetic lenses 125, 130
    immersion 105
    intensifiers 474–97
    real and virtual 102
    rotation in magnetic lenses 129
Immersion elements 105
Integration
  of Laplace’s equation 20–57
  of the trajectory equation
    in electrostatic lenses 111 ff
    in magnetic lenses 135 ff
Interference and electron beams 658
Ion collector 827
Ion emission 456
Ion optics in fields with a symmetry plane
  793
Ion trap 406
Ions, harmful effect of 12, 93
Isotope abundance 839

Lagrange–Helmholtz relation 104
Laplace’s equation 20
Larmor rotation 127
Lattice resolution 537
Lens
  spectrographs
    principle 853
    resolution 855
  supplies 602 ff
Lenses
  electrostatic 85–122, 207–58, 573 ff
  aberrations 220
  Gaussian optics 94
  grid lenses 224
  immersion lenses 195
  three-electrode lenses 195, 211
  two-cylinder lenses 194, 207
  magnetic 123–38, 579 ff
  aberrations 195
  Gaussian optics 130, 144
  long coil 259
  models 259–86
  shielded lens 276
  short coil 260, 273
  with permanent magnets 595
Liebmünn’s method 114
Line focus 161, 183, 192, 299 ff
Long coil spectrograph 856
Low voltage microscopy 673

Machining by electron bombardment 775
Magnetic fields, calculation of 58, 273 ff
Magnetic filter (Castaing and Henry) 690
Magnetic prism analyser 674
Magnets for microscope lenses 579
Magnification of a microscope 498, 514, 522
Mass
  atomic and molecular 837 ff
  spectrocope, spectrograph, spectrometer
    375, 787
  high frequency 848
  survey of 842
  thickness 638
Maxwell’s “fish-eye” 199
Microanalyser 752
  applications 760
Microscope
  β-ray 450
  CSF M1X 622
INDEX

Microscope (cont.)
  emission 427–73
  filament 460
  Gabor 559
  high voltage 629
  ion 456, 557
  lenses
    electrostatic 573
    magnetic 279, 579
  O.P.L. 625
  photoelectric emission 447, 458
  point-projection 460, 463, 467
  recoil nucleus 450
  reflexion 702
  scanning 762
  secondary emission 441, 455
  shadow 559
  “sphere” 460
  thermal emission 435, 452
  transmission 498–528
  wire 462
  X-ray 501, 530, 766
Microtome 697
Modulation of the spot intensity 397

Numerical integration 112

Objective
  immersion 98, 228, 231, 430
  of transmission microscope 223
Observation
  by reflexion 702
  indirect 705
  of living specimens 703
Omegatron 848

Parasitic fields 554
Permanent magnets 595
Pervance of a gun 248
Photo-emission 447, 448
Photographic emulsions 511 ff
Photography 14, 511, 827
Photometry 107
Plasma 645
Post-acceleration 423
Potential
  electrostatic 48

scalar magnetic 59
vector magnetic 79
Principal planes
  of a lens 103, 130
  of a prism 410
Principle of least action 142
Prisms 786–814
  aberrations 348, 367
  cardinal elements 327, 361
  electrostatic 354, 413
  fringing fields 340, 369
  magnetic 324, 418
  transfer matrices 329, 361
Projective lens
  electrostatic 223
  magnetic 51 ff
  role of 517
Pseudo-stigmatic image formation 303

Quadrupole
  doublet 301
  lenses 289
  aberrations of 313
  cardinal elements 299
  chain of 307
  equations of motion 296
  matrix representation 310
  potential in 291
  quadruplet 304
  triplet 304

Radioscope images 496
Ray
  equation in electrostatic fields 95
    integration of 111
  equation in magnetic fields 129, 133
  tracing 116 ff
    analogue methods 118
Refraction 139
Refractive index
  in electrostatic systems 140 ff
  in magnetic systems 144
  in Störmer's rotating system 145
Relativistic effects 110, 132
Replicas 705
Resistance network 54, 83
Resolving power
  of diffraction devices 722
  of emission microscopes 428
INDEX

Resolving power (cont.)
of image-converters and image-intensifiers 487
of point-projection microscopes 469
of spectrometers and spectrographs 812
of transmission microscopes 498, 524, 529, 531
of X-ray microscopes 766
Runge–Kutta method 113

Spherical aberration see also Aberrations
  in spectrometer 808
Stereographic reconstruction 506
Stigmators 203, 551
Störmer–Adams method 112
Strong focusing 287–322
Superconducting coils 283
Superconducting lenses 599
Supports for microscope objects 693
Symmetric and unsymmetric arrangements
  in cathode ray tubes 407

Saturation 280, 579 ff
Scanning
electron diffraction 741
electron microscope 762
in cathode ray tubes 418
Scattering of electrons 639
  inelastic 641
  multiple 639
  plural 639
  single 639
Secondary emission 441, 455
Sector fields 798, 803
Sensitivity
  of cathode ray tubes 418, 423
  of image-converters and intensifiers 483, 489
Sensitometry of photographic plates used
  with electrons 512
Shadowing technique 652
Similitude
  in cathode ray tubes 381
  in electrostatic lenses 92
  in magnetic lenses 131
Snell's law 139
Sources
  electron 565
  ion
    for microscopes 558
    for spectroscopes 816
Space charge see also Aberrations
  in spectrometers 811
Specimen-holder 625
Specimen preparation 691–715

Temperature changes in the object 712
Thermal emission 435, 452
Thin sections 697
Trajectory tracing 116
Transgaussian operation 681
Transit time of electrons 421

Ultra-high vacuum 620

Vacuum
  in a microscope 619
  in an emission microscope 450, 467
  in a spectroscope 833
Velocity
  analysis, analysers 376, 674
  phase and group 142
Velocity filters 685

Wavelength, de Broglie 501, 717
Wehnelt see Grid
Wien's velocity filter 791
Witcher velocity ring 860
Writing speed 418

X-ray diffraction 719, 748
X-ray microscope 501, 530, 766
X-ray projection microscope 766
PRINCIPLES OF OPTICS
Electromagnetic Theory of Propagation, Interference and Diffraction of Light
Fourth Edition
Max Born & Emil Wolf

This is the latest edition of the authors' classical work in the field of physical optics, and incorporates corrections to the text and illustrations, as well as references to recent contributions to the literature on the subject. The text deals with optical phenomena that can be described with the help of a continuous distribution of matter, and within this framework provides a complete picture of the present knowledge of optics. The book has been prepared with the authors' usual attention to accuracy, and its faithful representation of the ever growing field of optics marks this book as a major work of reference, ideal for all scientific, technical and educational libraries, as well as a valuable addition to the library of any scientist.

PERGAMON PRESS
Headington Hill Hall, Oxford OX3 0BW
Maxwell House, Fairview Park, Elmsford, New York 10523
207 Queen's Quay West, Toronto 1
19a Boundary Street, Rushcutters Bay, NSW 2011, Australia
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig

Printed in Great Britain/Bradley
some authoritative and independent reviews of the first edition:

"... this excellent book is abreast of our time. It is good that Professor Grivet never places the technological aspects of equipment in the foreground; he always endeavours to pass from the physical principles of equipment and methods of application to the latest developments in scientific research. Thus he is in a position to make remarks on development trends in various places. His book, as a consequence, is particularly 'close to industry'. The entire manner of the subject and treatment, as well as the textual part, the excellent diagrams and outstanding photographs show that the French are not only masters of conventional photo-optics, but have a mastery born from tradition."

JENA REVIEW

"A feature of the book is the remarkable uniformity of treatment, a tribute to Septier's editing and Hawkes' superb translation. The book will be especially useful to postgraduate students and research workers entering the field."

PHILOSOPHICAL MAGAZINE

"... the most complete account of electron optics available in any language. It will be invaluable as a text for introductory courses in the theory of the subject, previous treatments being either too cursory or too detailed. ... As for the translation, it is rarely possible to detect that it is such. Dr. Hawkes ... is to be congratulated on the easy flow and correctness of his English rendering of a long and, in parts, mathematically intricate text."

NATURE

"The wealth of material presented, its concise formulation, and the ample bibliography should make it a very valuable addition to the bookshelves of anyone who is either interested or engaged in electron optical design or study."

IEEE SPECTRUM

"The English translation of Grivet's book constitutes a valuable assistance to experts working in any sphere of electron optics."

ACTA PHYSICA ACADEMIAE SCIENTIARUM HUNGARICAE