Electron Physics

U. S. Department of Commerce
National Bureau of Standards
Circular 527

UNITED STATES DEPARTMENT OF COMMERCE • Sinclair Weeks, *Secretary*
NATIONAL BUREAU OF STANDARDS • A. V. Astin, *Director*

Electron Physics

Proceedings of the NBS Semicentennial Symposium on Electron Physics, held at the NBS on November 5–7, 1951

National Bureau of Standards Circular 527

Issued March 17, 1954

For sale by the Superintendent of Documents, U. S. Government Printing Office
Washington 25, D. C. - Price $2.75, Buckram

Foreword

The year 1951 marked the fiftieth anniversary of the establishment of the National Bureau of Standards. As part of the scientific program of the National Bureau of Standards for this year, twelve symposia were organized on scientific and technical topics of considerable interest. The Symposium on Electron Physics was the tenth of this series and was organized under the general chairmanship of Dr. L. Marton of the NBS Electron Physics Laboratory.

Electron physics for the purposes of this symposium was interpreted to mean the physics of the free electron, in distinction to physical electronics, which is a study of the bound electron. For many years work has been conducted at the National Bureau of Standards on the different phases of the physics of the free electron, but recent years have seen a more organized effort toward some of the objectives of research in this branch of physics. This symposium was arranged around two such topics in which NBS activities have been carried out; i. e., electron scattering and electron optics. These topics have been the subject of some earlier research conferences, but it is believed that this is the first time that a symposium has been devoted in its entirety to these subjects. That such a specialized symposium was timely is attested to by the list of distinguished scientists of nine nations who attended, and by the fact that a significant new trend in electron physics was indicated in their contributions. For the first time a large percentage of the papers in electron optics were concerned with electron-wave optics. A concerted effort seems to be under way to bridge the gap between wave and particle classification, or methods of attack, and thus to create a virtually new field for investigation.

The generous cooperation of the Office of Naval Research in making this symposium possible is gratefully acknowledged.

A. V. Astin, *Director*,
National Bureau of Standards.

Preface

The manuscripts in this volume were submitted by the authors after the symposium and hence are more or less revised accounts of the proceedings. The discussions at the end of the papers have been edited to remove questions arising from misunderstandings, and in many cases questions and answers have been paraphrased in the interest of brevity and clearness. Those involved in lengthy discussions were given the opportunity of reviewing their remarks; in the shorter exchanges, editorial liberties were taken.

The editing of this volume was done by J. Arol Simpson, aided by the staff of the Electron Physics Section of the National Bureau of Standards.

L. Marton, *Chairman,*
Committee for the Symposium on Electron Physics.

Contents

Introductory Remarks

By L. Marton [1]

The systematics of geometrical electron optics has been firmly established for the past twenty years. In a way this symposium could be regarded as a twentieth anniversary meeting if we consider the experimental and theoretical development of geometrical electron optics, or a twenty-fifth anniversary meeting if we look upon a classical paper of Busch [2] as the inception. While geometrical electron optics has been clearly defined, surprisingly enough no systematics have been established for physical electron optics in close analogy to that distinction made in the optics of light. This Symposium is devoted to electron optics in its very generalized meaning including both the geometrical and physical.

I should like to describe briefly some of the activities of the Electron Physics Section of the National Bureau of Standards. The activities of this section started with geometrical electron optics, but there has been a gradual shift of emphasis toward subjects of physical electron optics. I should like to mention first the electron optical method of field mapping that we have developed in the last few years. The technique is applicable to the mapping of magnetic or electrostatic fields that are inaccessible with other types of measurements. In this sense, inaccessible means either that the probes used in conventional field-mapping methods may be too large as compared with the physical dimensions of the field to be explored, or that these probes fail because of a possible interaction between the probe and the field. In both cases an electron beam offers a very valuable tool for field measurements. This fact in itself is not new; almost fifty years ago J. J. Thompson proposed to use the deflection of an electron beam for simple studies. The newness lies in the application of wide beams instead of narrow pencils. In its first application the method resembled very closely the Schlieren method of light optics (fig. 1). A distant electron source is imaged by an electron lens at the point C. If an obstacle is placed at this point, all rays are intercepted as long as there is no variation of the electron optical index of refraction along the trajectory. If, however, a field is placed in the plane passing through point D, some of the electrons will be deflected beyond the obstacle at C. If the plane of the image screen, S, and the plane of the field, D, are conjugate planes, a dark-field image of the field in question will appear on the screen. Figure 2 shows a typical image of this kind. The dark central strip represents the shadow of a magnetized wire. The bright spots on each side of it are dark-field images of the magnetic fields produced by a sinusoidal excitation of the moving wire.

---

[1] Chief, Electron Physics Section, National Bureau of Standards.
[2] H. Busch, Ann. Phys. **81**, 974 (1926).

This method is not completely satisfactory for a quantitative evaluation of the patterns because it involves a photometric determination of the intensity distribution across the pattern together with all the uncertainties and complications of that kind of measurement. For this reason we modified our method in a manner shown in figure 3. Again, an image, of a distant source is formed by the lens at point F, but the intercepting obstacle is placed behind that image. The image, acting as a virtual source of small dimensions, projects an enlarged shadow on the image screen as indicated by the dotted lines. In the presence of a field, the deflection of the electron beams will produce a measurable displacement of the shadow. Furthermore, if the field that we are investigating is inhomogeneous, the displacement of the virtual source will be not only lateral but axial; resulting thus not only in a displacement of the shadow but in a change of its magnification. Complications encountered in the quantitative interpretation will be dealt with by Simpson and Lachenbruch in their papers.

Figure 4 illustrates how the results look when applied to the same magnetized recording wire as shown in figure 2. For obvious reasons,

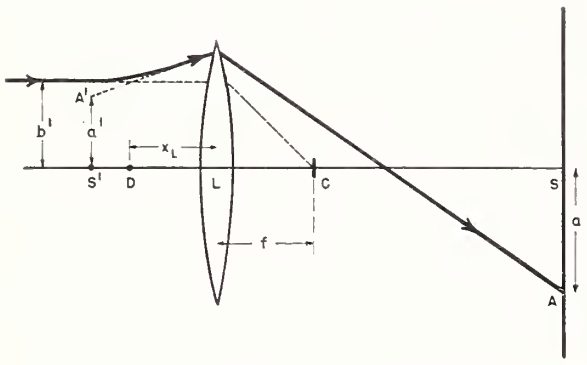


FIGURE 1. *Schematic diagram of "Schlieren" optics.*

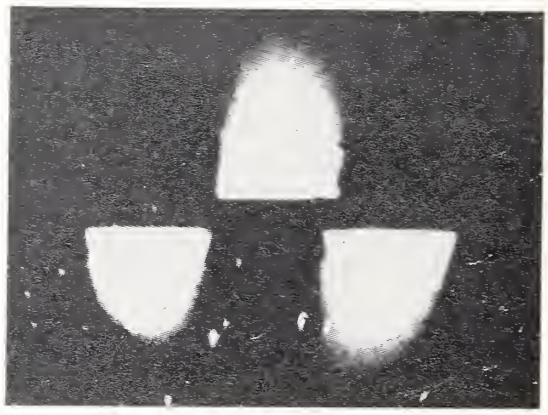


FIGURE 2. *Schlieren pattern of field about a magnetized wire.*

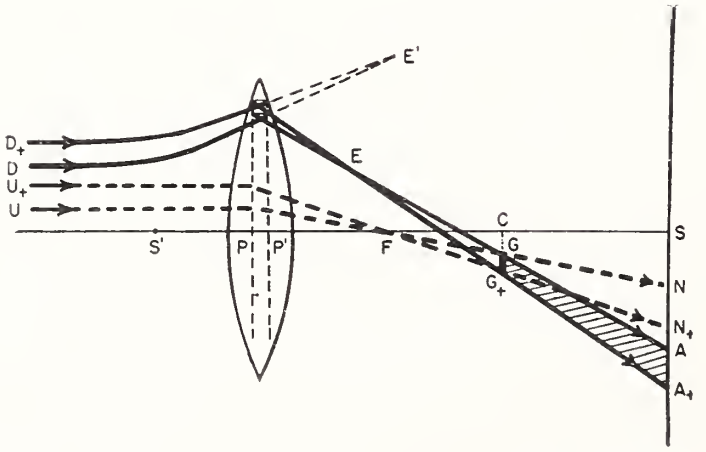


FIGURE 3. *Electron-optical diagram of shadow method.*

In the absence of a deflecting magnetic field, a parallel incident beam of electrons (dashed rays U, U+) converges to the focus F of the electron lens and then diverges past the thin obstruction GG+, which casts a shadow NN+ on the fluorescent screen S. However, when a deflecting field is placed in the midst of the incident beam, a different pair of rays (D, D+) determines the shadow boundary. The crossover point is displaced from F to some point E, and the resulting displacement of the shadow NN+ to the position AA+ may then be measured and substituted into theoretical formulas to obtain the field strength.

the single obstacle shown in figure 3 is replaced by a multiple structure of periodic nature such as, for example, a fine wire mesh. Figure 5 shows the application of the method to a small field produced by a single ferromagnetic domain. The word "small" refers to the physical dimensions of the field as indicated by the scale on the figure. Another example is the application of this method to the measurement of space-charge distribution in a d-c cut-off magnetron. A paper dealing with this subject was published by D. Reverdin,[3] formerly of this laboratory. In addition to the static fields described above, the method is applicable to time-variable fields up to relatively high frequencies (probably 300 megacycles).

The instrumentation for this type of work is relatively simple; and as an illustration, figure 6 shows an electron optical bench con-

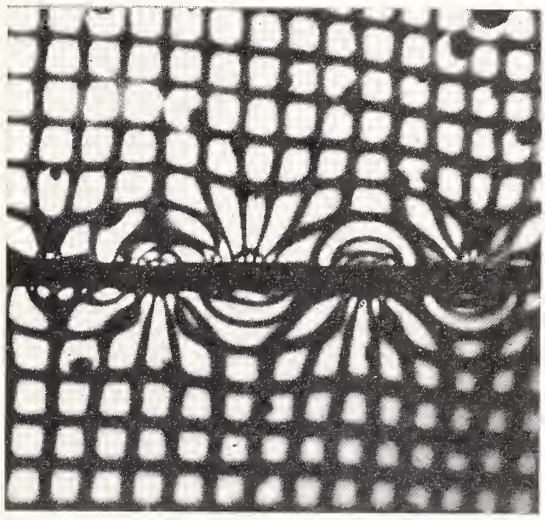


FIGURE 4. *Photograph of a typical pattern.*

Superposed on the image of a magnetic recording wire is the electron shadow of a fine wire mesh placed just beyond the back focus of an electron lens. From the displacement and reduced magnification of a selected part of the mesh, the absolute value of the magnetic field intensity at a corresponding point in the field can be accurately computed.

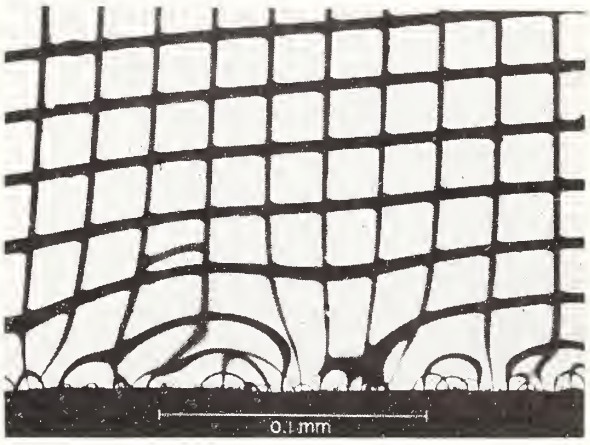


FIGURE 5. *Shadow pattern of the magnetic field at the edge of a cobalt crystal, showing large domain structure.*

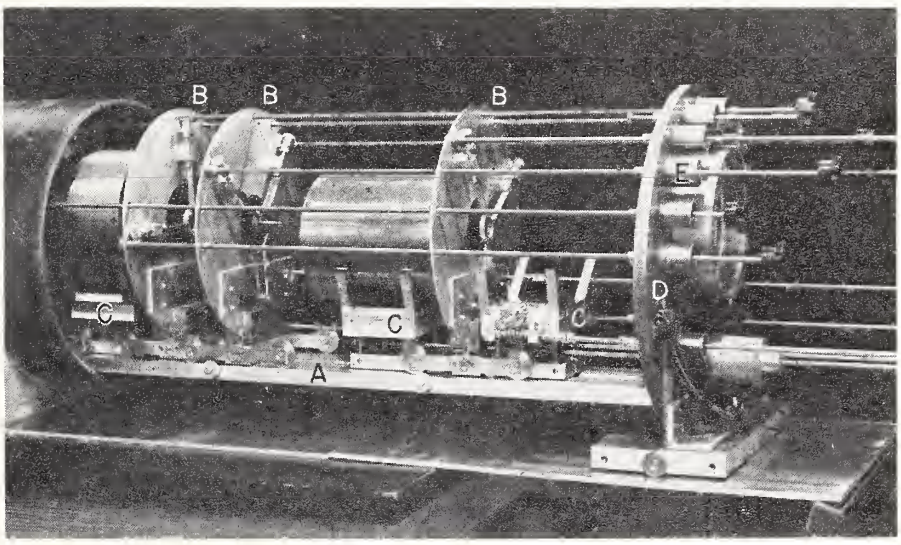


FIGURE 6. *Electron-optical bench, open.*

A, Bench; B, specimen and aperture mounts; C, lens carriages; D, electrical connections; and E, "Wilson" seals.

[3] D. L. Reverdin, J. Applied Phys. [3] **22**, 275 (1951).

structed in the Electron Physics laboratory. The original purpose in building the bench was the measurement of optical constants of electron lenses; however, due to its versatility, it has been used for carrying out field mapping and other types of experiments.

As far as future prospects of this method are concerned, we are pushing at present in three directions. One is toward its quantitative aspects, and I refer for details to papers presented later in this program. The second is the exploration of the ultimate limits of the method. These are two-fold: One consists in the measurement of the weakest possible field of large extent that can still be detected; and the other is the exploration of fields that are not necessarily very weak, but are restricted in their extent in space. We tried to detect fields around single ferromagnetic colloidal particles of about 200 A diameter, but these first attempts were not completely successful. If in repeating the experiment we do succeed, we may proceed to an electron optical study of the fields existing on surfaces of catalytic agents. The third is a continued exploration of space-charge fields in a magnetron.

Lately an increasing effort in the Electron Physics Section has been devoted to a study of electron scattering phenomena. Such investigations are limited to single scattering in medium- and low-energy ranges. Furthermore, we are more interested in inelastic scattering than in the elastic, although the experimental investigation will necessarily include both. Our first experiments in this study include measurements of energy losses in the forward direction within only a very narrow angular range. We are following the experiments originally described by Möllenstedt,[4] who used the off-axis chromatic aberration of an electrostatic lens to produce a high-precision velocity analyzer. Preliminary experiments have established that, indeed, extremely good sensitivity of such a velocity analyzer can be attained. We can distinguish without difficulty a change of 1 v in 30,000. We have found some difficulty, however, in reproducing the dark-field experiments of Möllenstedt and are continuing our investigations.

Parallel with this we are developing the instrumentation for a very comprehensive survey of single scattering in medium- and low-energy ranges for a great number of scatterers. For this survey we intend to collect data on scattered intensities by measuring both the angular and velocity distributions of the scattered electrons. One of the chief objectives of this survey is to produce reliable intercomparable data. For this purpose great care will have to be taken in the preparation of the scattering samples and careful attention paid to the conditions under which scattering occurs. To indicate the scope of the survey, I should mention that we intend to go down to very low energies and that within the framework of low-energy investigation, we plan to repeat the classical Davisson-Germer experiment. Special attention will also be given to the question of internal potential, which has caused considerable confusion over a number of years.

The instrumentation for this survey is at present in its development stage. A circular scattering chamber, about 35 cm in diameter, has the scattering sample mounted at its center. The sample holder allows an adjustment of the sample with five degrees of freedom. An electron gun is mounted on the periphery of the chamber in such a manner that the beam generated by it enters the chamber radially.

---

[4] G. Mollenstedt, Optik. **5**, 499 (1949).

At the bottom of the chamber a turntable is mounted with its axis of rotation coinciding with the axis of the chamber. On the turntable are mounted the velocity analyzer and the beam receptor. The analyzer is an electromagnet mounted close to the edge of the turntable. The electrons, scattered in a horizontal plane, enter the analyzer and are deflected 150°. That deflection sends the selected velocities back toward the axis of the turntable where the receptor, consisting of a 12-stage electron multiplier, is located.

In addition to these experimental investigations, we are actively studying the theory of scattering, as attested by a paper to be given later by Dr. Engel, of our laboratory.

Another phase of our work is conerned with the application of electron interference phenomena for measurement purposes. For about fifteen years, I have hoped one day to be able to build an electron interferometer in close analogy to the interferometers of light optics. A simple argument shows why an electron interferometer has to be somewhat different from the usual conception of a light interferometer. Let us assume that we want to repeat Young's experiment and build an electron interferometer on the principle of the double-slit method. It is easy to calculate that we require first of all a very small source, whose dimensions probably will have to be achieved by means of an inverted electron microscope. The dimensions of the slits and their spacing is a difficult problem because a reproducible double-slit arrangement, having for example only 1,000 A separation and 100 A slit width, is rather difficult to realize. In addition, the fringe spacing in such an instrument is of the order of 100 A and therefore requires another microscope for its viewing. The total result is an instrument consisting of two electron microscopes put together end to end with something rather complicated in between.

About a year and a half ago it occurred to me that the solution lies not in using narrow-beam interferences but patterning an electron interferometer on the wide-beam interference methods used in the construction of the Michelson or the Mach-Zehnder interferometers of light optics. Fortunately, the problem of beam-splitting in such an interferometer can be very well handled by the use of diffraction on thin crystal lamellae. To explain this method let us consider figure 7. A parallel beam, limited by a convenient aperture, is directed at normal incidence onto a thin lamellar crystal. Part of the beam is transmitted and part of it is diffracted. At a certain distance a second crystal is placed parallel to the first, and the direct beam is again partially transmitted without change and partially diffracted. The diffracted part of the first beam itself is partially transmitted and partially diffracted at the second crystal. The beams meet on a third crystal where the same phenomenon is repeated and we now have two beams superimposed that have equal path lengths, having each been diffracted twice. We thus have an arrangement in all respects equivalent to the wide-beam interferometers of light optics. It is possible to produce a path difference by a superimposed electric or magnetic field gradient.

It has been demonstrated that such a scheme is a workable one. We have at present two types of confirmation. Interference fringes were observed on lamellar crystal material, first by Dr. Uyeda [5] of

[5] Mitsuishi, Nagasaki, and Uyeda, Proc. Japan Acad. 27, 86 (1951).

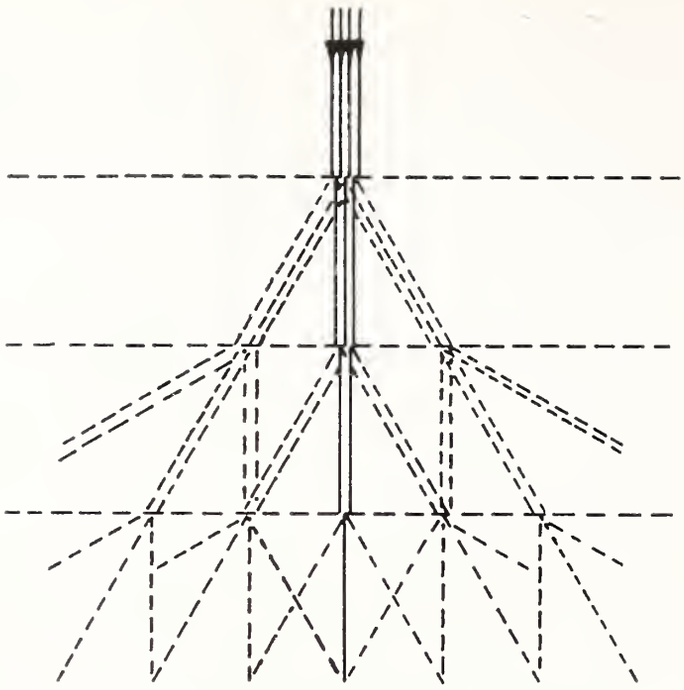


FIGURE 7. *Interferometer schema.*

Japan, and later by Dr. Rees of Australia, and Dr. Hillier of Radio Corporation of America, who will present a paper on the subject at this symposium. We obtained further evidence by setting up an analog experiment with light, in which we used transmission grating replicas and observed, as expected, a fringe structure. Although such an arrangement seems to offer a complete equivalent to the conventional type interferometers, it does not seem to be more advantageous than the instrumentation in use at present in light optics. The experiment, however, enabled us to draw conclusions about some of the requirements on accuracy of alignment of the individual gratings, and these conclusions have been confirmed and expanded by calculations. We find that the proposed system will have very reasonable tolerances and the electron beam interferometer should not be difficult to construct.

One important aspect of this type of experiment is the preparation of the necessary thin lamellar crystals. There exist natural crystals that may be useful for such experiments, but the reproduction of such crystals in case of loss or damage would be a serious problem. To have a safe reproducing method we investigated the growth of such thin crystals by means of epitaxy, that is, the oriented growth of a crystal lattice on a substratum having a different lattice. Without going into detail, I should like to say that we now have gold crystals of about 100 to 200 A in thickness and several square millimeters surface area that exhibit the diffraction pattern shown in figure 8. Until now our experiments have been limited to gold and nickel, but we hope to extend this investigation to many substances and produce good specimens not only for the interferometer study, but also for the scattering investigations that we have started.

The uses of an interferometer can be twofold. One application consists in its use as an extremely sensitive field-measuring instrument. It is easy to see that this type of instrument can detect extremely weak field gradients: and, if we succeed in developing it into as prac-

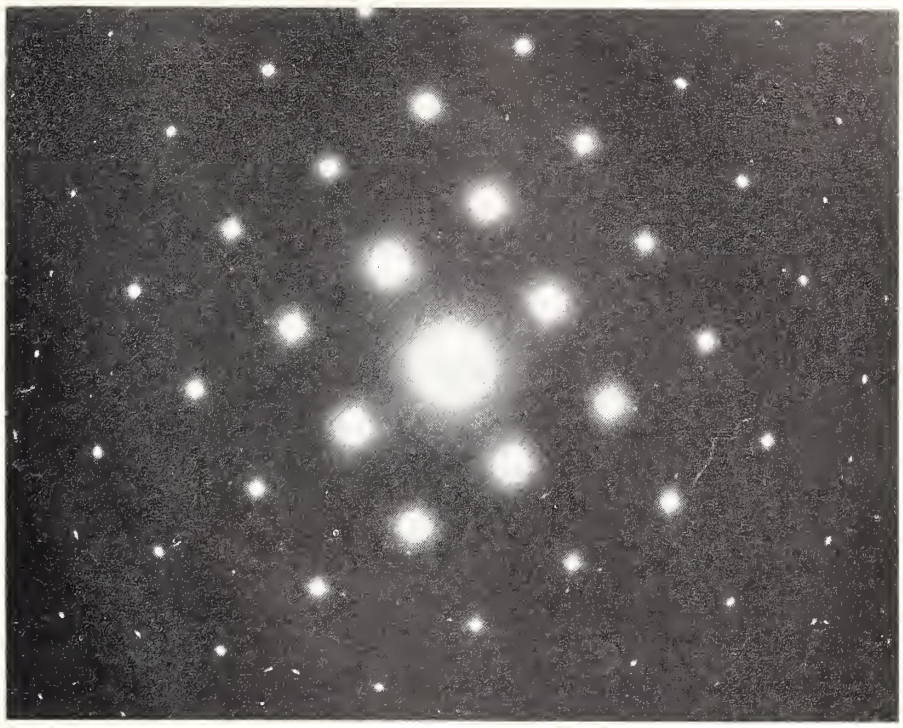


FIGURE 8. *Electron diffraction pattern of lamellar gold crystal.*

tical a device as some of the light interferometers, its usefulness may be quite extensive. In addition, it may be the only instrument that can easily answer questions of purely fundamental interest, such as the coherence of the de Broglie wave. Preliminary calculations indicate that under easily attained laboratory conditions coherence should be observable up to about 10^5 wavelengths, i. e., up to about 1 micron path difference. An experimental confirmation of these calculations should be highly interesting.

1. Diffusion of Electrons in Infinite Media [1]

By C. H. Blanchard [2]

During the past year an effort has been made to describe the distribution of electrons as they penetrate, diffuse, and slow down in an infinite medium. The points of departure are the formal theory of Lewis [3] and the parallel study of X-ray distributions that has been completed by Spencer and Fano.[4] Although the study of electron distributions is far from complete, enough has been learned, especially in the way of approximate schematizations, to obtain a fairly good qualitative understanding of the phenomenon. Accordingly, we present here a summary of initial results.

We consider the distribution of electrons from a given source under multiple elastic scattering and continuous energy loss. We wish to describe this distribution in space, energy, and direction of travel. We use Lewis' solution of the transport equation, in which a formal expression is given, in terms of the scattering cross section, for the angular distribution; and a procedure for deriving the moments of the spatial distribution from this angular distribution is indicated. The purpose of this paper is to point out a simple schematization of the cross section that should give quite accurately the main features of the distribution; in particular, it exhibits the tendency for the distribution to depend only on the fractional energy loss. The discussion follows the actual method of solution in that we consider first the angular distribution, and then the derivation of the spatial distribution from the angular distribution. Finally, some preliminary numerical results on the spatial distribution are discussed briefly.

The energies considered here are low enough for radiation energy loss to be neglected, and well above atomic binding energies. We use for the rate of the continuous energy loss the average excitation-ionization stopping power. (This formulation implies a unique correspondence between the kinetic energy, T, of the electron and the *path-length* distance, s, which the electron has traveled since it had the source energy, T_0.)

If θ is the angle between the directions of travel of an electron and its initial direction (at the source), and if P_l is the usual Legendre polynomial, then it is easily shown [5] that

$$\langle P_l(\cos\theta)\rangle_s = e^{-\int_0^s K'(s')ds'}, \tag{1}$$

with

$$K_l(s) = N\int_{-1}^{1} \sigma_{\text{elastic}}(\cos\theta)\,[1-P_l(\cos\theta)]2\pi d\cos\theta, \tag{2}$$

---

[1] This work was supported by the Office of Naval Research.
[2] National Bureau of Standards, Washington, D. C.
[3] H. W. Lewis, Phys. Rev. **78**, 526 (1950).
[4] L. V. Spencer and U. Fano, J. Research NBS **46**, 446 (1951) RP 2213.
[5] See footnote 3.

where $\langle \quad \rangle_s$ denotes the average over all electrons that have traveled a pathlength distance s (regardless of position in the medium), and N is the number of scattering atoms per unit volume. $K_l(s)$ is a function of s because the cross section depends on the energy. Using the stopping power formula, $dT/ds = -2\pi NZe^4B/mv^2$, to relate T and s, we find

$$\int_0^s K_l(s')ds' = Z\int_T^{T_0} \frac{C_l}{B}\frac{mc^2dT}{T(T+2mc^2)}. \tag{3}$$

We have introduced the quantity

$$C_l = \int_{-1}^{1} \frac{\sigma_{\text{elastic}}(\cos\theta)}{\sigma_{\text{Rutherford}}(\cos\theta)}\frac{1-P_l(\cos\theta)}{(1-\cos\theta)^2}\, d\cos\theta. \tag{4}$$

By taking the ratio of σ_{elastic} to the Rutherford scattering cross section $(Z^2e^4/p^2v^2(1-\cos\theta)^2)$, the most violent T and Z dependences of σ_{elastic} fail to appear in C_l, and leave the C_l as slowly varying parameters characteristic of the shape of the elastic cross section. We call C_l the lth "scattering number" because it contributes to the decrement of $\langle P_l(\cos\theta)\rangle_s$ with s in much the same way as the stopping number B contributes to the decrement of T with s. The relationship between the two quantities is stronger than this formality, since both C_l and B are integrals over Coulomb cross sections, and, as a matter of fact, the ratio C_l/B tends to be much more constant than either C_l or B alone (both logarithmically energy dependent). Therefore, for an initial approximation, we remove the factor C_l/B from the integration, which is then elementary, giving the result [6]

$$\langle P_l(\cos\theta)\rangle_T = \left[\frac{T}{T_0}\frac{T_0+2mc^2}{T+2mc^2}\right]^{\frac{ZC_l}{2B}} \tag{5}$$

from which the angular distribution at the energy T may be constructed. $C_l/B \approx 0.6$ for all Z, so, with energies in Mev, we have the useful rule of thumb [7]

$$\langle\cos\theta\rangle_T = \left[\frac{T}{T_0}\frac{T_0+1}{T+1}\right]^{0.3Z}. \tag{6}$$

Note that for energies much below 1 Mev, this angular distribution depends only on T/T_0 and not on T and T_0 separately.

Lewis [8] has given a formal solution of the problem of finding the distribution of electrons in space. He presents a procedure for generating successive moments of the spatial distribution, starting from the angular distribution. One can calculate the average over all space of successive powers of the penetration variable for all electrons with a given s, regardless of direction. In order to use this formal solution, the angular distribution must be written as a function of s (rather than T, as above). It can be shown that the essential features of the scattering and stopping processes should be given by

$$K_1(s) = \frac{Z}{4}\frac{C_l}{B}\frac{1}{R-s}, \tag{7}$$

[6] Similar considerations have been published by Danzer, Z. Physik **128**, 79 (1950).
[7] This result was given by C. H. Blanchard and U. Fano, Phys. Rev. **82**, 767 (abs. G13).
[8] See footnote 3.

where R is the average total pathlength executed by an electron in losing all its initial kinetic energy ("true range"); and where, as above, we will treat C_l/B as energy independent. This form is just what would be expected in a very simple nonrelativistic formulation with $\sigma \sim T^{-2}$ and $dT/ds \sim T^{-1}$. It is to be emphasized, however, that the validity of the schematization (7) can be shown, from the detailed form of the energy dependences of the scattering and stopping, to extend well into the relativistic range of energies. Equation (7) implies that the "mean free path for l scattering" is proportional to the residual range. This proportionality is sufficient to insure that the solutions of the transport equation will have the property that the *shape* of the distributions, for a given s, does not depend on R (i. e., the initial energy) and s separately, but only on s/R (i. e., on the fraction of the energy that has been lost); and, of course, on Z. Thus we have a formulation that embodies the *scaling* property of electron distributions, well known from experiment, especially from the observation that fractional backscattering is independent of incident energy.[9]

We have derived depth moments of the distribution of electrons from a plane, monodirectional, monoenergetic source. Algebraic complication precludes in practice the use of more than the first few moments. These moments can be exhibited as functions of s analytically. From these functions, the corresponding first few moments of the distribution of the ionization in a medium can be derived. For this type of distribution, the first few moments determine only the gross features and do not specify very closely a detailed form. Initial numerical results, on comparison with experimental "depth-dose" curves [10] (which are of necessity for semi-infinite media), give good agreement for the extrapolated range, but both the depth of the maximum dose and the ratio of maximum dose to surface dose are predicted substantially higher than experiment, for low Z elements. Refinements and improvements of the theory are in progress.

[9] W. Bothe, Z. Naturforsch. **4a**, 542 (1949); H. H. Seliger, Paper No. 9 in this symposium and Phys. Rev. (paper to be published).
[10] J. Fleeman, Paper No. 10 in this symposium, and Frantz and Fleeman (paper to be published).

2. Scattering of Electrons by Nuclei [1]

By Herman Feshbach [2]

Introduction

In this paper we shall review the present status of the theory of the scattering of electrons by nuclei. This phenomena is of interest because of the light it throws upon (1) electrodynamics—the interaction of the electron with the electromagnetic field, and the (2) properties of the target nucleus.

Let us elaborate on (2). For electron energies less than about 10 Mev (the actual figure depends upon the nuclear radius) the elastic scattering of electrons may be treated accurately by considering the nucleus as a point charge. The other electric and magnetic properties such as the magnetic dipole and electric quadrupole moments affect the scattering in a very minor way and thus are unobservable with present techniques. Nuclear information is provided, however, by the inelastic scattering of electrons in which the target nucleus is left in an excited state.

Above 10 Mev, the elastic scattering becomes sensitive to the finite size of the nucleus since for these energies the electron wavelengths become comparable with the nuclear radius. This generally results in a decrease in the scattering as well as the appearance of typical diffraction phenomena analogous to the Ramsauer effect in atomic scattering. As the electron energy increases, the electron wavelength becomes comparable with the interparticle distance in the nucleus so that the model picturing the latter as a smeared-out current and charge distribution is no longer valid. The passage from the smeared-out picture to one in which the total scattering is simply the sum of the intensities of scattering from each nucleon depends very naturally upon the correlation in position which exists along the protons of the nucleus, very similar once more to the atomic case. We may expect these effects to become important for $E \gtrsim 50$ Mev.

By roughly 300 Mev the coherent scattering becomes small except for small angle scattering. Elsewhere the observed scattering just equals Z (the nuclear charge) times the scattering from one proton. In this energy range, we therefore begin to investigate the structure of the proton. Assuming that this fundamental interaction is known (by measurement of electron-proton scattering) further nuclear information may be obtained by examining the electron energy spectrum, at a given scattering angle. The electron energy, if the protons in the nucleus were at rest, would be unique. The effect of nuclear motion is to spread this "line" out into a broadened "Compton line" whose shape and width would indicate the nature of the velocity dis-

[1] This work was supported in part by the Office of Naval Research and the Atomic Energy Commission.

[2] Laboratory for Nuclear Science and Engineering, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Mass.

tribution of particles within the nucleus. One sould be careful to remember that at these energies both meson production and bremsstrahlung have cross sections comparable with elastic scattering cross sections. The large energy loss in the first of these and the continuous energy spectrum of the second may permit the experimental separation of this background if a detector with good energy resolution is employed. One should also remember that our understanding of nuclear structure is incomplete, that this last picture in which scattering from the separate nucleons is considered may be too naive, for it may be the case, as, for example, the esistence of exchange currents indicates, that the nucleons are too strongly coupled to be considered as separate entities. In this listing we have also omitted special effects, such as a large polarizability, which the Goldhaber-Teller [1][3] model would give.

To obtain quantitative results for these nuclear effects, it is essential that the dynamics of the electron be completely understood. Fortunately, an energy region ($E<10$ Mev) exists for which the deviations from point-charge scattering is negligible. However, the motion of an electron in the field of a point charge is not exactly given by the Dirac equation because of the modifications introduced by "radiative" forces that arise from the vacuum fluctuations of the electromagnetic and electron-positron fields. An example of the effects of the first of these is represented by the figure 2.1, a, in which an electron radiates a virtual photon represented by the wavy line, is scattered by the nuclear potential, V, and reabsorbs the photon. These virtual emissions and absorptions result in fluctuations in position giving rise to a "smearing out" of the Coulomb potential, thus reducing the scattering. The effect of the electron-positron field is contained in part in the modification of the Coulomb law of force arising from the polarization of the vacuum. This type of correction is shown in figure 2.1, b, in which the nuclear potential produces a virtual pair, which annihilates, the virtual photon thus produced being absorbed by the incident electron. It is important to note that these very same processes are responsible for the Lamb shift in the 2S level in hydrogen where quantum electrodynamics has been found to hold very closely. Verification by scattering experiments would extend confirmation of the theory to the relativistic domain.

In addition to the above virtual effects, it is necessary to take into account the real emission of photons, as all electrons radiate upon scattering. Hence, the measured cross section will depend upon the energy resolution of the detection apparatus. This last effect together with the virtual effects form the radiative "correction" to the scattering of electrons.

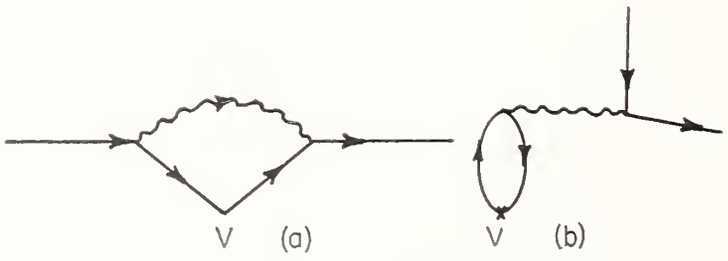


FIGURE 2.1. *Radiative effects in electron scattering.*

[3] Figures in brackets indicate the literature references on p. 23.

We have outlined the phenomena that are important for electron scattering and shall now proceed to summarize the present state of knowledge. Consider first the simplest problem, the scattering of electrons by point nuclei, omitting any radiative effects.

Scattering by a Point Nucleus

The exact solution of the Dirac equation for this problem was given by Mott [2], the differential cross section being expressed in the form of a series, which is, however, poorly convergent. The series has been summed by Bartlett and Watson [3] for $Z=80$ and by Massey [4] for positron scattering by the same element. A simple formula [5] exists for low Z,

$$\sigma_c=\sigma_R[1-\beta^2 \sin^2 \vartheta/2+\pi\alpha Z\beta(\sin \vartheta/2)(1-\sin \vartheta/2)], \qquad (1)$$

where σ_R=Rutherford scattering, $\beta=v/c$, α=fine structure constant. This is a first term in an expansion in powers of αZ and $\alpha Z/\beta$ that is valid for medium Z. Recently we [6] have summed the Mott series directly for $v/c\simeq 1$. An indication of the accuracy of our method is given by the comparison of our results with those of Bartlett and Watson in table 2.1. The notation is that of Mott. Generally, the ratio of Mott scattering to Rutherford is greater than unity for small angles, but becomes less than unity as the angle increases.

TABLE 2.1. *Comparison of values of G_1 for H_g with those obtained by Bartlett and Watson* [1]

θ	R_eG_1		I_mG_1	
	Bartlett and Watson	Computed by Feshbach	Bartlett and Watson	Computed by Feshbach
deg.				
30	1.349	1.349	0.026	0.027
45	.851	.851	.292	.290
60	.538	.538	.331	.330
90	.213	.213	.227	.228
120	.0734	.0735	.107	.107
150	.0159	.0158	.0272	.0272

[1] See reference [3].

These results have not been tested experimentally in a very complete manner until recently [7]. For small angles $\vartheta \leq 50°$ and for energies between 1.3 and 2.3 Mev the agreement obtained by Van de Graaff's group is indicated in figure 2.2. Agreement has also been found by

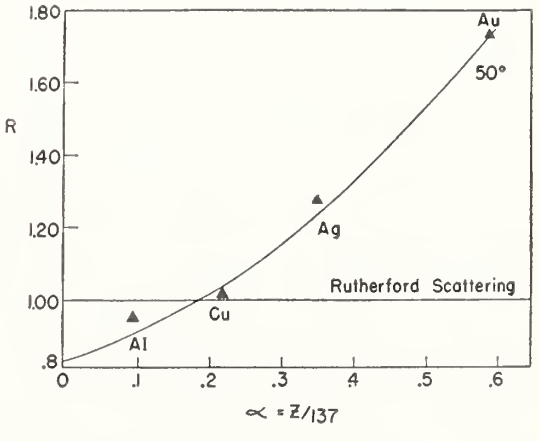


FIGURE 2.2. *Comparison of theory and experiment at electron energies of 2 Mev.*

The solid line is given by theory. The triangles give the experimental points as obtained by Van de Graaff, Buechner, et al.

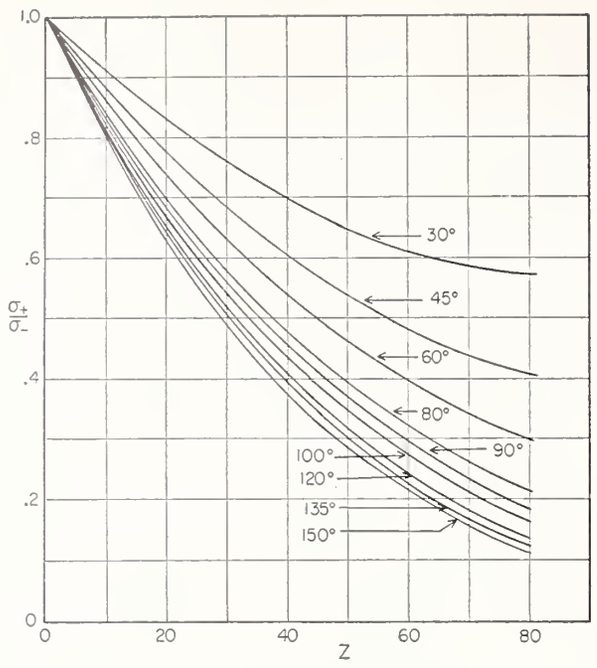


FIGURE 2.3. *Ratio of the scattering of positrons to the scattering by electrons.*

the same group for Be and Al for $\vartheta \leq 70°$. Recently Paul [8] with electrons of 2-Mev energy and Bothe [8] with electrons of 0.245-Mev energy have extended the measurements to larger angles. They found agreement for $\vartheta \leq 60°$ for all elements. At larger angles there is a gradual departure from theory as Z increases to 79, the experimental values falling below the theory. At 120°, $Z=79$, Paul reports a deviation of about 13 percent. The effect seems to be energy independent and quite similar for angles 90°, 120°, and 150°. Such deviations are very small compared to those reported in prewar experiments. No explanation of this effect has as yet been given, although a deviation in the same direction is to be expected to arise from radiative effects.

The scattering of positrons is another prediction of the theory. The Lipkin and White experiments [9], which are difficult, have yielded only an approximate agreement. The ratio of the cross section for positron scattering to Rutherford is always smaller than unity and decreases monotonically with increasing angles. The variation of (σ_+/σ_-) for several elements is shown in figure 2.3. Calculations for $Z=80$ have been performed by Massey [4].

Radiative Corrections to Electron Scattering

The most complete discussion has been given by Schwinger [10]. These calculations evaluate the correction to the Born approximation for point coulomb scattering. The correction is applied as a multiplicative factor $e^{-\delta}$. For a slowly moving particle,

$$\delta = \frac{8\alpha}{3\pi}\beta^2 \sin^2\frac{\vartheta}{2}\left[\ln\frac{mc^2}{2\Delta E}+\frac{19}{30}\right]. \tag{2}$$

ΔE is maximum energy loss permitted by the detection apparatus. This formula can be readily obtained from the Lamb shift calcula-

tions, wherein an approximate potential is derived [11]:

$$V=\frac{\alpha}{3\pi}\left(\frac{\hbar}{mc}\right)^2[\nabla^2 V]\left[\int_{k_0}^{vmc^2}\frac{dk}{k}-\ln 2+\frac{19}{30}\right]. \tag{3}$$

The symbol k_0, the Bethe lower limit, is in the present connection replaced by ΔE, the energy resolution. Calculation of the additional scattering arising from this term gives Schwinger's result, indicating the close connection of the radiative corrections to the Lamb shift. For relativistic energies ($2E^2/mc^2 \sin \frac{1}{2}\vartheta \gg 1$) Schwinger gives:

$$\delta=\frac{4\alpha}{\pi}\left[\left(\ln\frac{E}{\Delta E}-\frac{13}{12}\right)\left(\ln\frac{2E}{mc^2}\sin\frac{\vartheta}{2}-\frac{1}{2}\right)+\frac{17}{72}+\frac{1}{2}\sin^2\vartheta/2 f(\vartheta)\right], \tag{4}$$

where $f(\vartheta)$ is a rather complicated function. For $\vartheta=\pi/2$, $f=1.167$, for $\vartheta=\pi$, $f=\pi^2/12$. Some representative numerical values are given in table 2.2. The logarithmic terms are dominant so that $\delta\sim(\ln E)^2$.

TABLE 2.2. *Radiative correction to electron scattering*

$\frac{\Delta E}{E}$=40 kev; E=3.1 mev.

ϑ	δ
$\pi/4$	0.042
$\pi/2$	.063
$3\pi/4$	.072

It is clear that these radiative corrections reduce the discrepancy between theory and the experiments of Bothe and Paul, but that a sizeable gap still remains. An additional Z dependence may arise if the distortion in the wave functions coming from the coulomb field is taken into account. These calculations are still in progress.

We turn now to the nuclear properties that may be investigated by means of electron scattering.

Inelastic Scattering

The prediction of the cross sections for the excitation of the nucleus for dipole (magnetic and electric) and for electric quadrupole transitions was first given by Wick [12] who employed the Born approximation expressing this result in terms of the transition electric and magnetic moments. More accurate calculations taking coulomb effects into account have been made by Guth, Mullin, and Thie [13]. Precise experimental data is as yet not available.

Effects of the Finite Size of the Nucleus

At 15 Mev the electron wavelength is about 2.5×10^{-12} cm, which is between 2 to 10 times larger than the radii of most nuclei. Hence, it should be possible to detect the effect of nuclear size. This has been done by Lyman, Hanson, and Scott [14]. The results are shown in figure 2.4, where σ_c is the scattering from point nuclei. The cross

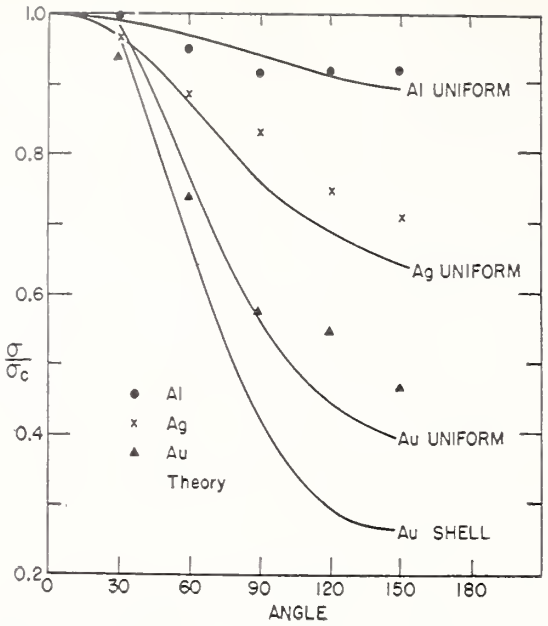


FIGURE 2.4. *Experimental results shown as ratios to Coulomb scattering.*

The solid curves represent the reduction in scattering arising for a uniform-distribution charge except for gold where the curve for a surface-charge distribution is also shown.

section is of course reduced from that of a point nucleus, the decrease being greatest for the heavier nuclei. Theoretical calculations employing the Born approximation and assuming that the nucleus can be represented by a smeared-out charge distribution have been performed by Guth and Rose [15]. Precise calculations have been made by Elton [16] for Au and $E=40$ mc^2, by Parzen [17] for Pb at 100 Mev, and by Acheson [18] for Al, Au, Ag, and Au for 15 Mev $\leq E \leq 30$ Mev. The theory as well as the analysis of the experimental results may be simplified considerably by taking advantage of the fact that the electron energy is much greater than its rest mass and that on the other hand the electron wavelength is large compared to the nuclear radius. In symbols $\epsilon \gg 1$, $\epsilon R \ll 1$. We are using units of energy mc^2 and length $(\hbar/mc)$. From the first of the above inequalities, it immediately follows that the phase shift for a given j, total angular momentum of the electron, is independent of l, the orbital angular momentum which may be associated with this j. There are, of course, two such values of l, $j \pm \frac{1}{2}$. The corresponding phase shifts are equal.

$$\eta_l = \eta_{-(l+2)}. \tag{5}$$

The notation here is that of Mott. This arises formally from the fact that in the high-energy domain the "small" component and "large" component Dirac wave functions become equal in magnitude. This relation may be verified for point coulomb scattering where

$$\exp i[\eta^{(c)}_{-(l+2)} - \eta_l^{(c)}] \underset{\epsilon \gg 1}{\rightarrow} 1 - i\alpha/\epsilon(l+1). \tag{6}$$

If we write

$$\eta_l = \eta_l^{(c)} + \eta_l', \tag{7}$$

where η_l' then measures the charge in phase shift arising from the

finite size of the nuclear radius, then

$$\eta_l' = \eta'_{-(l+2)} . \tag{8}$$

The accuracy with which this relation holds may be seen from table 2.3.

TABLE 2.3: *Comparison of* η_l' *and* $\eta'_{-(l+2)}$

l	η_l'	$\eta'_{-(l+2)}$	Author
0	−0.196	−0.195	Elton.
1	−.002	−.001	Do.
0	−.271	−.268	Do.
1	−.004	−.004	Do.
0	−.5339	−.5307	Parzen.
1	.2064 − π	.2991 − π	Do.
2	.2507 − π	.2572 − π	Do.
3	.9678 − π	.0711 − π	Do.
4	.0115 − π	.0124 − π	Do.
5	.0013 − π	.0015 − π	Do.
6	.0001 − π	.0001 − π	Do.

The second simplication, which follows from $\epsilon \gg 1$, has to do with the potential associated with the finite charge distribution.

Let

$$U(\rho) = -(Z\alpha/\rho)[1 - q(\rho/R)], \tag{9}$$

where R is a parameter measuring the nuclear radius, and $q(x)$ is a function that goes to zero rapidly once $x > 1$. It then follows that

$$\eta_l = \eta_l(\epsilon R, Z\alpha). \tag{10}$$

In words, the phase shift depends upon energy and nuclear radius through the dimensionless combination ϵR. This relation shows that a change in radius with energy constant is equivalent to a change in energy with radius constant.

The next result requires not only that $\epsilon \gg 1$ but also $\epsilon R \ll 1$. We ask under what conditions are two models for the nuclear charge distribution equivalent. It may be shown that the equality of the phases η_l' for two models requires that

$$\int_0^\infty q_1(\rho)[f_l^2 + G_l^2]\, d\rho/\rho = \int_0^\infty q_2(\rho)[f_l^2 + G_l^2]\, d\rho/\rho. \tag{11}$$

The requirement that the energy derivative of η_l' be model independent is given by

$$\int_0 q_1 \frac{d}{d\rho}(f_l^2 + G_l^2) d\rho = \int_0^\infty q_2 \frac{d}{d\rho}(f_l^2 + G_l^2) d\rho. \tag{12}$$

We see that these conditions can be met if, in the region occupied by the nucleus,

$$(f_l^2 + G_l^2) \sim \rho^t. \tag{13}$$

For $\epsilon R \ll 1$, only η_0 is important, and for that case for nonsingular potentials it is a good approximation to place $t=2$; i. e., the potential is not strong enough to make the wave function have any appreciable curvature over the volume of the nucleus. Going back to the potential U, we see that if R_1 and R_2 are chosen so that

$$\int [U_1(\rho)+(\alpha/\rho)]\rho^2 d\rho = \int [U_2(\rho)+(\alpha/\rho)]\rho^2 d\rho, \tag{14}$$

then η_l' is the same for both U's. This result shows that the scattering essentially measures the volume integral of the perturbing potential. These results have been checked by comparing Achesons' calculation for the homogeneous and shell models.

For this case $R_s=(6/10)^{\frac{1}{2}}R_h$. Figure 2.5 demonstrates the great accuracy with which this relation holds.

We learn from this that in the energy range $\epsilon R \ll 1$, that elastic scattering experiments determine only the value of one parameter; that for any charge distribution a radius may be chosen which permits reproduction of the experimental data. In this energy range electron scattering cannot choose between models, it can only tell us what the appropriate radius for a given model is. However, it would obviously be of great interest to compare the scattering of isotopes.

We have started an analysis of the experimental data and we can only quote first results now. These indicate a homogeneous distribution with radius somewhat smaller than the usual $(1.37\times10^{-13}$ cm$)A^{\frac{1}{3}}$. This can only be a rough statement, because the error on the phase shift determined in this fashion is quite large.

In concluding this part of the discussion, it should be noted that only the effect of a smearing out the charge distribution has been investigated. In addition, there may be effects due to current distributions and polarizability. That part of the current distribution that gives rise to the magnetic moment has very little effect on the scattering. However, in electron scattering we are not dealing with the expectation value of the magnetization alone, so additional effects might occur.

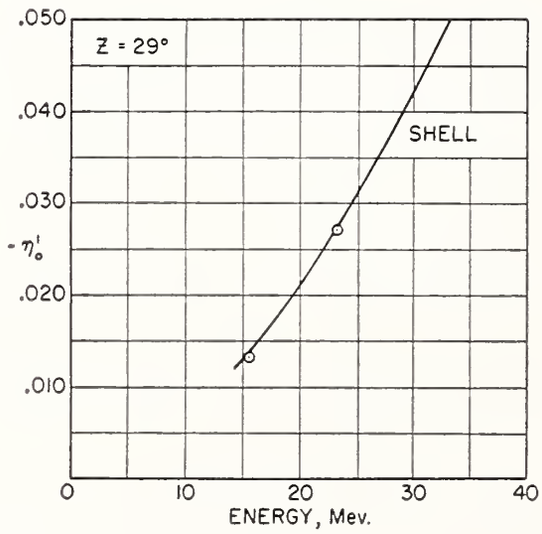


FIGURE 2.5. *Comparison of phase shifts η_0' for the shell and homogeneous models.*

The solid line gives Acheson's results for the shell model, the radius R_s is taken from $(1.5\times10^{-13}\text{cm})A^{\frac{1}{3}}$. The circles are computed from Acheson's curves for the homogeneous model employing, however, a radius R_h equal to $(10/6)^{\frac{1}{2}}R_s$.

Approach to Incoherent Scattering

As the energy increases, we approach an electron wavelength of the order of the distance between particles in the nucleus. The nucleons then scatter separately so that the net scattering from a nucleus will be sensitive to the correlation in the motion of the individual nucleons. The approach to incoherent scattering may be discussed in much the same way as the scattering of electrons or X-rays from atoms. If we make the approximation that all final states are permitted, one may use the sum rule (closure) and obtain in the Born approximation

$$\sigma=|f_B|^2[Z+Z(Z-1)]\int \rho(\mathbf{r}_i,\mathbf{r}_j)e^{\frac{\Delta\mathbf{p}\cdot(\mathbf{r}_i-\mathbf{r}_j)}{\hbar}}\,dV_i\,dV_j.$$

Here ρ is the probability that particles i and j will be at $\mathbf{r}_i$ and $\mathbf{r}_j$; $|\Delta p|=2p\sin\frac{1}{2}\vartheta$, while f_B is the Born approximation amplitude. The correlation integral goes to one if all the particles are at one point, $\rho=\delta(\mathbf{r}_i-\mathbf{r}_j)$ or if $(\Delta p/\hbar)R\ll 1$, where R is the nuclear radius so that it is always unity for zero angle scattering. The angular region in which it is appreciable narrows as p increases. A rough calculation indicates that the correlation integral is approximately given by

$$\left[\frac{1}{1+\left(\frac{(\Delta p)r_0}{\hbar}\right)^2}\right]^2.$$

From this factor we see that, as the energy increases, the coherent cross section will decrease while the incoherent will increase relatively. However, since the latter is proportional to Z, it will be some time before the incoherent term will dominate. This will occur at about 300 Mev, as is indicated by table 2.4, which is based on the approximate formula given above.

Amaldi [19] and his coworkers have carried out a more elaborate calculation in which a rather definite model for the nucleus is employed. His results for lithium are shown in figure 2.6. We note that the cross-section is completely incoherent for $\vartheta\geq 50°$.

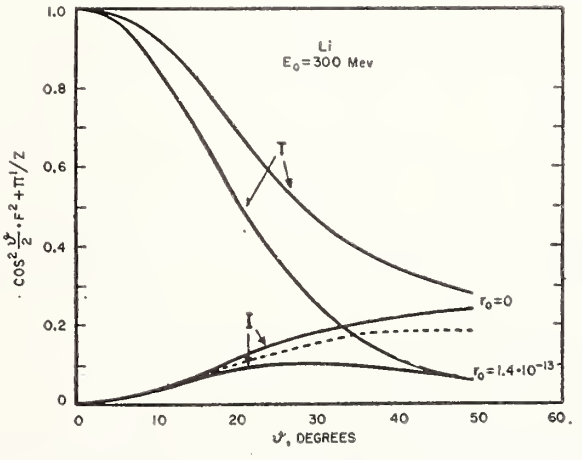


FIGURE 2.6. *Scattering of 300 Mev electrons from lithium.*

T designates the total cross section, I the incoherent cross section while r_0 is the assumed proton radius. The ordinate is the ratio to the Rutherford scattering.

TABLE 2.4. *Comparison of coherent and incoherent electron scattering*

Energy	$(\sigma_{inc}/\sigma_{co})_{90°}$	$(\sigma_{co}/Z^2)_{90°}$	$(\sigma_{inc}/\sigma_{co})_{Au\ at\ 90°}$	$(\sigma_{inc}/\sigma_{co})_{Al\ at\ 90°}$
Mev				
10	----------	0.98	----------	----------
20	----------	.93	----------	----------
50	----------	.64	----------	----------
100	(3/Z)	.25	0.038	0.23
200	(24/Z)	----------	.30	1.8
300	(99/Z)	----------	1.24	6

Incoherent Scattering

Above about 300 Mev, the scattering is incoherent and is thus equal to the sum of the scattered intensities from each nucleon in the nucleus. The fundamental electron-proton cross section is most clearly determined in scattering from hydrogen. As the wavelengths here are only several times the Compton wavelength of the proton, such experiments will be probing the internal proton charge and current distributions. These presumably arise from the interaction of the proton and charged meson fields. Unfortunately, the description of this interaction is not yet known quantitatively. However, the qualitative features may be seen in the calculations by Rosenbluth [20], in which perturbation theory was employed for theories in which the proton is coupled with either scalar or pseudoscalar mesons. The coupling constant is adjusted so as to give the experiment magnetic moment. His results are given in figure 2.7 and figure 2.8, in which the effective magnetic moment and the effective charge are given in terms of a quantity that is closely related to the impact parameter. We see that both of these quantities decrease with energy or increasing angle of scattering. This may be interpreted as arising from the finite proton radius. In the long wavelength range, the major effect arises from the spreading

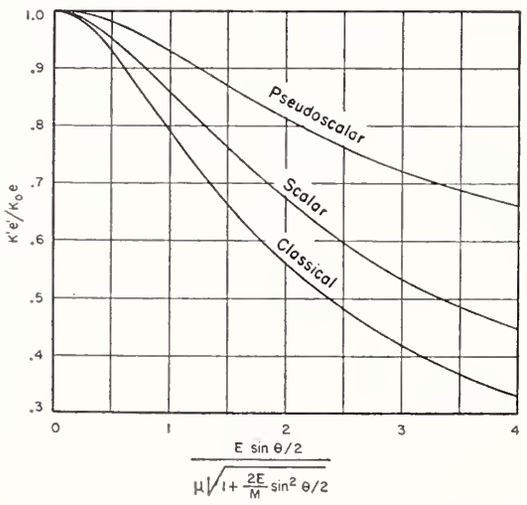


FIGURE 2.7. *Ratio of the effective anomalous proton magnetic moment to its zero-energy value for charged-meson theories.*

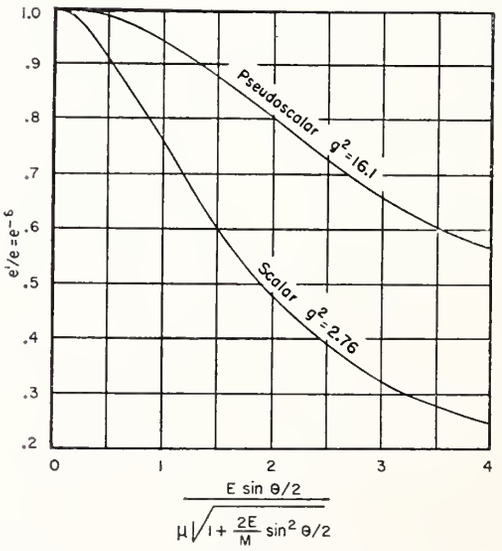


FIGURE 2.8. *Effective proton charge for charged-meson theories with coupling constant chosen to fit the magnitude of the observed proton anomalous magnetic moment.*

of the proton electric charge. We may there employ the theory described earlier, for the effects arise from finite nuclear size, and obtain a model independent theory. We may, moreover, employ Rosenbluth's results to obtain a number for the proton radius. If we assume a homogeneous charge distribution, we obtain (¼) the meson Compton wavelength, or about 3.5×10^{-14} cm.

The scattering in an actual nucleus will be modified by the motion of the scattering nucleons giving rise to a "Compton line" at a given angle of scattering, instead of a unique electron energy. We may expect that the experimental energy distribution may be analyzed to give the nucleon momentum distribution.

Experiments on this energy require that sufficiently fine energy analysis of the scattered electrons be made so that the effects of inelastic processes such as meson production, electron-electron scattering and bremsstrahlung of hard X-rays be accounted for.

References

[1] M. Goldhaber and E. Teller, Phys. Rev. (in press).
[2] N. F. Mott, Proc. Roy. Soc. [A] **124,** 426 (1929; [A] **135,** 429 (1932).
[3] J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74,** 53 (1940).
[4] H. S. W. Massey, Proc. Roy. Soc. [A] **181,** 14 (1942).
[5] W. A. McKinley, Jr., and H. Feshbach, Phys. Rev. **74,** 1759 (1948).
[6] H. Feshbach, Phys. Rev. (to be published).
[7] R. J. Van de Graaff, W. W. Buechner, and H. Feshbach, Phys. Rev. **69,** 452 (1946); W. W. Buechner, R. J. Van de Graaff, A. Sperduto, A. Burrell, Jr., and H. Feshbach, Phys. Rev. **72,** 678 (1947).
[8] W. Paul and H. Reich, Z Physik (to be published); W. Bothe and P. Kinzinger (unpublished).
[9] H. J. Lipkin and M. G. White, Phys. Rev. **79,** 892 (1950).
[10] J. Schwinger, Phys. Rev. **76,** 790 (1949). See also, R. P. Feynman, Phys. Rev. **76,** 769 (1949); H. W. Lewis, Phys. Rev. **73,** 173 (1948); Z. Koba and S. Tomanaga, Prog. Theoret. Phys. **3,** 290 (1948).
[11] See, for example, B. French and V. F. Weisskopf, Phys. Rev. **75,** 1240–1248 (1949).
[12] G. Wick, Ricerca Sci. **11,** 49 (1940).
[13] E. Guth, W. Mullin, and J. Thie, Bul. Am. Phys. Soc., annual New York meeting (1952).
[14] E. M. Lyman, A. O. Hanson, and W. T. Scott, Phys. Rev. **84,** 626 (1951).
[15] E. Guth, Setzingber. Akad. Wiss. in Wien (Nov. 22, 1934); M. E. Rose, Phys. Rev. **73,** 279 (1948).
[16] L. R. B. Elton, Proc. Phys. Soc. (London) [A] **63,** 1115 (1950).
[17] G. Parzen, Phys. Rev. **82,** 355 (1950).
[18] L. K. Acheson, Jr., Phys. Rev. **82,** 488 (1951).
[19] E. Amaldi, G. Dedecaro, and F. Mariani, Nuovo Cimento **7,** 757 (1950).
[20] M. N. Rosenbluth, Phys. Rev. **79,** 615 (1950).

Discussion

Dr. W. Bothe, University of Heidelberg, Heidelberg, Germany: In connection with the very interesting nuclear interferences, can one be quite sure that, with a wavelength comparable to this so-called length, one can still apply motion of charge distribution and motion of charges within the border?

Dr. Feshbach: There are some possibilities for deviation. For example, although the magnetic moment effect is small, the magnetic moment is the expectation of the current in the ground state. In electron scattering we are interested in high components of these, and it is not altogether clear that these will be small.

Dr. Bothe: If the basis of all these calculations is the Mott theory for point of charge, I am not sure it is applicable, because there may be a smallest length beyond which the Mott theory does not apply.

Dr. Feshbach: That is what the radiative corrections do. They take into account the fact that there is a change when the distances get small and that is what the present point of electrodynamics would predict. If the present point of electrodynamics is wrong, then this is, of course, extremely interesting.

3. Low Electron-Energy Losses in Zinc Oxide

By Olive G. Engel [1]

We have been interested in the energy losses [2] sustained by bombarding electrons as an electron beam passes theough a section of zinc oxide that is thin enough so that plural and multiple inelastic scattering can be neglected. When the accelerating potential of the beam is sufficiently high, these losses correspond to the excitation and ionization energies of each of the electron shells of the atom. We have restricted the present approximate treatment to losses of 20 ev or less and have made the simplifying assumption that scattering takes place from atoms whose wavefunctions have not been modified by the presence of neighbor atoms. We have some vindication for this in the conclusion of Coster and Kiestra [1] [3] that atomic wave functions are much more maintained in the oxide than in the solid metal, but we are aware that, while this assumption is good for ionization from inner shells, it is poor for energy losses resulting from excitation of the outer electrons. Nevertheless, as is pointed out by Massey [2], it is usually only possible to work to this approximation in making a theoretical estimate of the probability of inelastic scattering.

The divalent zinc ion has completed shells through the $3d$ shell. The two $4s$ electrons of the neutral zinc atom were lost in the ionization process. Excitation and ionization of the $3d$ electrons should take place as is indicated schematically in figure 3.1. Assuming that atomic selection rules apply, the $3d \rightarrow 4p$ excitation is allowed, but the $3d \rightarrow 4s$ excitation is forbidden. This allowed transition involves losses of 18 electron volts [4] to the bombarding electrons. Ionization of a $3d$ electron would involve a loss of 40 ev. This, however, is in excess of the arbitrary limit of 20 ev that we have set and the cross section for it will not be calculated.

As will be shown in a moment, the presence of monovalent zinc ion is also to be anticipated. It has one $4s$ electron that may be excited to the $4p$ state, or to the $5p$ state if this state exists as such in the solid. These excitations, also shown schematically in figure 3.1, would involve losses to the primary electrons of 6 ev and 12.6 ev, respectively. Furthermore, ionization of the $4s$ electron may occur involving a loss of 17.89 ev.

The state of the divalent oxide ion is not clear. It is usually conceded that a gain of two electrons gives the neutral oxygen atom the stable neon structure in which the second shell is complete. Yet O'Bryan and Skinner [4] state that the oxide ion is unstable in the

[1] National Bureau of Standards. This work was carried out under contract with the Office of Naval Research.

[2] Energy losses of the bombarding electrons correspond to energy gains of atomic electrons. We shall look at the scattering problem from the viewpoint of the bombarding electrons because it is the energy losses of these electrons that we shall observe experimentally.

[3] Figures in brackets indicate the literature references on p. 30.

[4] With the exception of the transition from the oxygen full band to the empty band of zinc ion, all excitation and ionization energies are taken from Bacher and Goudsmit [3].

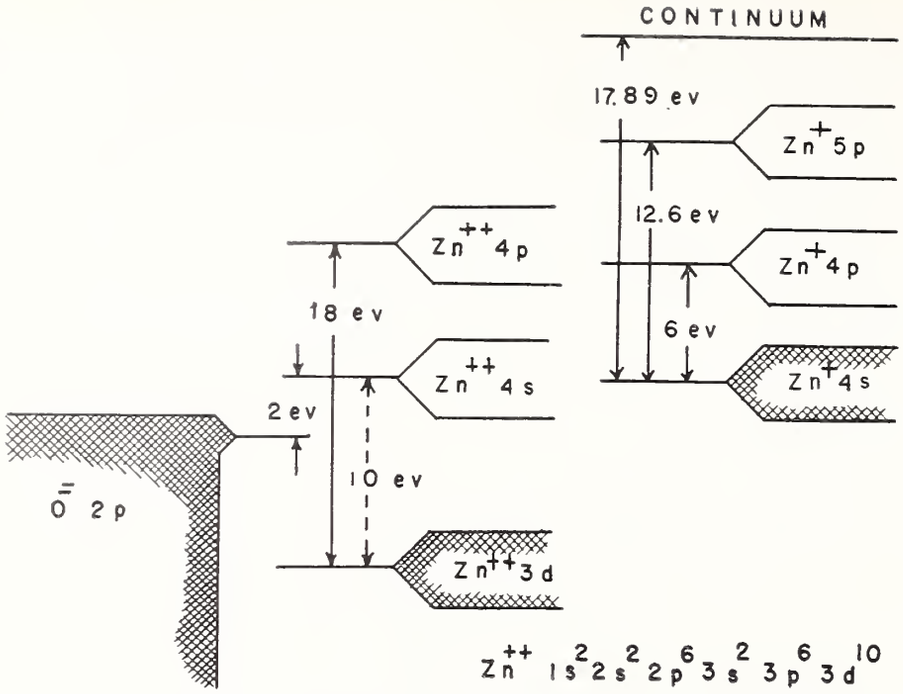


FIGURE 3.1.

free state and question its existence in the crystal lattice. In addition to this, Hartree, Hartree, and Swirles [5] have abandoned their attempt to obtain the Hartree field for the divalent oxide ion and have concluded that, to the accuracy of this approximation to the structure of a many electron atom, the $O^{=}$ ion cannot exist in the free state. We may assume that it is stabilized by the lattice and that it exists in the lattice at least initially. Then one of its $2p$ electrons may be excited to the empty $4s$ state of the divalent zinc ion producing monovalent zinc ion in the lattice.[5] The energy required for this transition was estimated first by Seitz [6] and later by Hahn [7]. Seitz neglected the width of the bands and the polarization energies. Hahn estimated the band widths and included polarization energies found by Wright [8]. Both Seitz and Hahn neglected the repulsive potential of the lattice energy. Now the formation of bands in a solid is a direct consequence of the overlapping of the outer shells of the ions and so is related to the repulsive potential. We shall therefore use a modification of Seitz's procedure and take the repulsive potential into account in determining the energy required for this transition.

First, the energy required to remove an electron from a free oxide ion and place it on a free zinc ion is considered. This is shown diagrammatically in figure 3.2. Seitz, using the total electron affinity of oxygen found from the Born-Haber cycle, and the affinity of neutral oxygen for one electron found experimentally by Lozier, has estimated the energy needed to remove one electron from the divalent oxide ion at about -9 ev. From spectroscopic data, the first ionization potential of the monovalent zinc ion is 17.89 ev. The energy to remove an electron from the free oxide ion and place it on a free zinc ion is the total energy separation or approximately 27 ev where the zero of energy is taken to be $Zn^{+}O^{-}$.

The lattice energy $\Phi(r)$ consists of two parts, namely, the attractive electrostatic potential and the repulsive potential that arises in the overlap of the outer shells.

$$\Phi(r) = -\frac{z^2 e^2 a_r}{r_0} + b r_0^{-n}. \tag{1}$$

[5] This may be regarded as a first step in reduction. The second step would produce the free metal and in the case of thin films this would be favored by the escape of free oxygen. The extent to which this takes place under electron bombardment has not been ascertained. O'Bryan and Skinner found that most oxides were stable to X-rays although aluminum oxide decomposed slowly to give the metal.

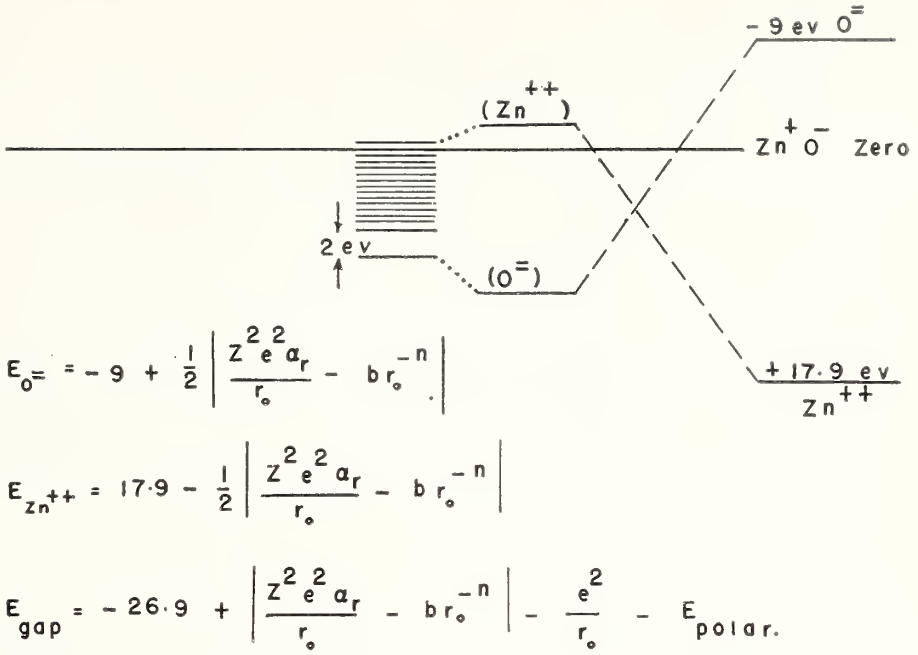


FIGURE 3.2.

In the attractive potential z is the charge of the ions, e is the electronic charge, a_r is the Madelung constant, and r_0 is the equilibrium cation-anion distance. Born [9] has stated that the repulsive potential can be developed in power series in powers of $1/r$. Assuming that only one term in the series gives an important contribution, he has written the repulsive energy in the form br^{-n} where b and n are constants. When the ions are brought into their characteristic lattice configuration, they approach until the repulsive term of the lattice energy just balances the attractive electrostatic potential. The amount by which the negative ion energy is lowered and that by which the positive ion energy is raised is less than the Madelung potential to the extent of the repulsive energy. This lowering and raising is shown by the dashed lines in figure 3.2.

In the lattice the energy of the oxide ion and the energy of the zinc ion are given by the expressions

$$E_{0=} = -9 + \tfrac{1}{2}\left|\frac{z^2e^2a_r}{r_0} - br_0^{-n}\right|, \tag{2}$$

$$E_{Zn^{++}} = 17.9 - \tfrac{1}{2}\left|\frac{z^2e^2a_r}{r_0} - br_0^{-n}\right|. \tag{3}$$

The energy to take an electron from the oxide ion in the lattice to a *very distant* zinc ion in the lattice is the difference in these two equations plus an additional term. As this transition occurs, the oxide ion and the zinc ion involved are each left with unit charge. An additional energy is involved when the process is carried out in the lattice because of the polarization of the medium about a lattice site to which an electronic charge has been added or from which an electronic charge has been subtracted. This is indicated by the dotted lines in figure 3.2. The energy to take an electron from the oxide ion to a less distant zinc ion is less than this result by e^2/r, where r is the distance between the ions. This has been shown by Seitz [10] who points out that the Madelung potential at the cation is decreased by e^2/R if an electron is removed from an anion at distance R. It is indicated by the series of horizontal lines in figure 3.2. To obtain the energy to take an electron from an oxide ion to its very nearest neighbor zinc ion necessitates the further term of e^2/r_0, where r_0 is the equilibrium

cation-anion distance. The energy of the gap between the oxide ion full band and the zinc ion empty band is then given by the following equation, where $E_{\text{polar.}}$ represents the sum of the absolute values of the anion and cation polarization energies which in the case of zinc oxide is estimated to be about 4.5 ev [8].

$$E_{\text{gap}} = -26.9 + \left| \frac{z^2 e^2 a_r}{r_0} - b r_0^{-n} \right| - \frac{e^2}{r_0} - E_{\text{polar.}} \tag{4}$$

All quantities are known in eq (4) except r_0, b, and n. The constant b is eliminated by use of the equilibrium condition that the derivative of the lattice energy per ion pair with respect to displacement is zero. We proceed to evaluate r_0.

Zinc oxide has the wurtzite structure [11]. Each zinc ion has about it a tetrahedron of oxide ions but is nearer by 0.1 A to one of the oxygens than to the others [12]. This has been construed as indicating homopolar character. Similarly, each oxide ion has about it a tetrahedron of zinc ions. Ions of the same kind form a hexagonal close packed system. Consequently, we may take the equilibrium distance r_0 to be the sum of the Goldschmidt radii decreased by the factor 0.93 because of the coordination number of four [13]. In this way the equilibrium distance is found to be 2 A.

Born's [14] method can be used to find the value of the constant n, and the lattice energy can then be calculated by making allowance for the homopolar character of zinc oxide. On the other hand, the value found by Sherman [15] by means of the Born-Haber cycle may be used as well. Sherman's value of 42.1 ev includes the repulsive potential that was neglected by Seitz and Hahn. On substituting this value for the lattice energy in eq (4) the energy of the gap between the oxide ion full band, and the zinc ion empty band is found to be approximately 2 ev.

This is a reasonable value. It compares with the cathodoluminescent peak of zinc oxide [16], with the excitation energy of ZnO:(Zn) phosphor [16], and with the conductivity measurements of Hahn [7]. We may, however, question whether or not a loss of 2 ev will be observed for zinc oxide under our experimental setup. The sensitivity is such that it certainly could be observed. On the other hand, the crystal system may be energetically more stable after the transition has occurred than before, and the return to the initial state may be slow. The sample is exposed to the beam over a considerable period of time before a photographic observation is made so that possibly the observed losses should be interpreted in terms of a Zn^+O^- lattice.

The experimental observations are being made by Harold Mendlowitz of the Electron Physics Section with 30,000-v electrons and by use of dark-field illumination. Use of dark-field illumination limits the scattering angle to the aperture of the microscope, which is 0.001 radian. It is necessary then to have an equation which will give the cross section for excitation of zinc to a state n, the scattering angle being 0.001 radian, and the accelerating voltage 30,000 v. The single inelastic scattering cross section of Mott and Massey [17] is

$$Q_n = 2\pi \int \frac{k_n}{k} [f_n(\theta, \varphi)]^2 \sin\theta d\theta, \tag{5}$$

where the energy of the colliding electron is $k^2\hbar^2/2m$ and its energy after the scattering incident is $k_n^2\hbar^2/2m$. Massey [20] has derived the

expression for the scattered amplitude $f_n(\theta,\phi)$ by writing the Hamiltonian for the system atom plus colliding electron. This is

$$f_n(\theta,\varphi)=-\frac{8\pi^2me^2}{k^2K^2}\int\phi_0\phi_n^*\left[\sum_{s=1}^{N}e^{iK\vec{n}\cdot\vec{r}_s}\right]d\tau_1.\ldots.d\tau_n, \tag{6}$$

where $\vec{n_0}$ and $\vec{n_1}$ are unit vectors at angle θ, the change in momentum is $K\vec{n}=k\vec{n_0}-k\vec{n_1}$, $\phi_0(\vec{r_1}\ldots\vec{r_N})$ is the wave function of the atom in its ground state, and $\phi_n(\vec{r_1}\ldots\vec{r_N})$ is the wave function for the atom in the state n. Substituting this expression for the scattered amplitude into the equation for the cross section, Q_n, expanding the exponential, using the orthogonality condition, and the expansion of the function of θ results in the equation

$$Q_n=\int_{\theta=\frac{E_n-E_0}{2E}}^{\theta=0.001}\frac{32\pi^3e^4}{h^2v^2}\left|\left(\sum_s Z_s\right)_{0n}\right|^2\frac{d\theta}{\theta}, \tag{7}$$

where E_n-E_0 is the binding energy of the electron to be excited and E is the accelerating potential. On integrating, we have for 30,000-volt electrons and a scattering angle of 0.001 radian,

$$Q_n=0.0261\left|\left(\sum_s Z_s\right)_{0n}\right|^2\log\frac{60}{(E_n-E_0)}. \tag{8}$$

When only a single electron takes part in the transition, only the off-diagonal matrix element associated with this one electron need be computed. The other electrons are left unchanged and so do not contribute anything to the cross section. Consequently, for this case we have,

$$Q_n=0.0261|Z_{0n}|^2\log\frac{60}{(E_n-E_0)}, \tag{9}$$

where $|Z_{0n}|^2$ is the square of the dipole moment.

The values of the square of the dipole moment for discrete transitions were computed by using the Coulomb approximation of Bates and Damgaard [18]. In the case of ionization, a plane wave was used for the final state as was done by Rudberg and Slater [19], and integration was carried over states in the continuum to the extent of an ionization energy. Lorenzo P. Greene and Morton Lutzky, both of the Electron Physics Section, assisted with the computational work involved in evaluating the integral. In the case of the transition of an electron from the full band of the oxide ion to the empty 4s band of the divalent zinc ion, the square of the equilibrium cation-anion distance was used for $|Z_{0n}|^2$. The results are given in table 3.1.

Our calculations have indicated that the most probable transition is that of an electron from the oxide ion full band to the 4s band of divalent zinc ion. After this, the most probable transitions are excitation of the 4s electron of monovalent zinc ion to the 4p state and the complete loss of this electron by ionization. About 10 times less probable than these is the excitation of the 3d electrons. Excitation of the 4s electron of monovalent zinc ion to the 5p state, if, indeed, this state exists as such in the solid, is a thousand times less probable

Table 3.1

Transition	Energy loss	Cross section
	ev	cm^2
$Zn^{++}\ 3d^{10} \rightarrow Zn^{++}\ 3d^9\ 4p$	18	4.4×10^{-19}
$Zn^{+}\ 3d^{10}\ 4s \rightarrow Zn^{+}\ 3d^{10}\ 4p$	6	1.4×10^{-18}
$Zn^{+}\ 3d^{10}\ 4s \rightarrow Zn^{+}\ 3d^{10}\ 5p$	13	6.9×10^{-22}
$Zn^{+}\ 3d^{10}\ 4s \rightarrow Zn^{++}\ 3d^{10}$	18	1.2×10^{-18}
$O^{-}\ 2p^6 + Zn^{++}\ 3d^{10} \rightarrow O^{-}\ 2p^5 + Zn^{+}\ 3d^{10}\ 4s$	2	1.5×10^{-17}

than the excitation of the $3d$ electrons. It will be interesting to see what justification experimental results may give to the assumption that atomic states continue to exist in the solid oxide and to the use of this simple model.

I acknowledge the suggestions and assistance of my associates at the National Bureau of Standards, and several very helpful conversations with Dr. K. F. Herzfeld of Catholic University and with Dr. R. W. Gurney of the Institute of Fluid Dynamics of the University of Maryland.

---

[1] D. Coster and S. Kiestra, Empty electron bands of lowest energy of the transition metals, Mn and Fe, and their oxides, Physica **14,** 175 (1948).
[2] H. S. W. Massey, Electron scattering in solids, Advances in Electronics, to be published.
[3] Robert F. Bacher and Samuel Goudsmit, Atomic energy states (McGraw-Hill Book Co., Inc., New York, N. Y., 1932).
[4] H. M. O'Bryan and H. W. B. Skinner, The soft X-ray spectroscopy of solids II. Emission spectra from simple chemical compounds, Proc. Roy. Soc. [A] **176,** 229 (1940).
[5] D. R. Hartree, W. Hartree, and Bertha Swirles, Self-consistent field, including exchange and superposition of configurations with some results for oxygen, Phil. Trans. Roy. Soc. [A] **238,** 229 (1939–40).
[6] F. Seitz, The modern theory of solids, p. 446 (McGraw-Hill Book Co., Inc., New York, N. Y., 1940).
[7] E. E. Hahn, Some electrical properties of zinc oxide semi-conductor (University of Pennsylvania Technical Report No. 17, 1949).
[8] D. A. Wright, Work function and energy levels in insulatois, Proc. Phys. Soc. London **60,** 13 (1948).
[9] M. Born and M. Mayer, Handbuch der Physik **24/2** (Berlin, 1933).
[10] F. Seitz, The modern theory of solids, p. 409–410 (McGraw-Hill Book Co., Inc., New York, N. Y., 1940).
[11] P. P. Ewald and C. Hermann, Strukturbericht, 1913–1928, pp. 78 and 119 (Leipsig, 1931).
[12] H. M. James and Vivian A. Johnson, Electron distribution in ZnO crystals, Phys. Rev. **56,** 119 (1939).
[13] Ralph W. G. Wyckoff, Crystal Structures, chap. III, p. 21 (Interscience Publishers, Inc., New York, 1948).
[14] M. Born, Über die elektrische Natur der Kohäsionskräfte fester Körper, Ann. Physik **61,** 87 (1920).
[15] J. Sherman, Chem. Rev. **11,** 93 (1932).
[16] H. W. Leverenz, Luminescence and tenebrescence as applied in radar, R. C. A. Review **7,** 199 (1946).
[17] N. F. Mott and H. S. W. Massey, The theory of atomic collisions, 2d ed., p. 137–137 (Oxford, 1949).
[18] D. R. Bates and Agnete Damgaard, The calculation of the absolute strengths of spectral lines, Phil. Trans. Roy. Soc. [A] **242,** 101 (1949–50).
[19] E. Rudberg and J. C. Slater, Theory of inelastic scattering of electrons from solids, Phys. Rev. **50,** 150 (1936). See pages 155 and 156.

4. Summary Calculations of Multiple Scattering

By W. T. Scott [1]

The recent and successful theories of Molière [1, 2] [2] and Snyder-Scott [3, 4] for the small-angle multiple scattering of fast charged particles have not been presented in a form as convenient for comparison with experiment as the approximate Gaussian form of the older theory. It is the purpose of this paper to present calculations that summarize the newer theories in a convenient form.

Any theory of multiple scattering involves, first, a satisfactory single-scattering law, second, a satisfactory statistical study of the result of a random multiplicity of single events, and third, an adequate numerical computation, including some types of mean-value quantities, that can be directly compared with experiment or used to evaluate experimental errors. We shall consider here the single-scattering law as calculated by Molière, the statistical theory common to both treatments along with the Snyder-Scott extension to cover the lateral distribution, the two differing methods of numerical computation, and some new and convenient mean-value results. In addition, we shall give an indication of the degree of departure from the Gaussian approximation, a discussion of other methods of measuring scattering than use of tangents and chords, a brief comparison with E. J. William's [5] scattering, and finally some values of the so-called scattering constants for photographic emulsions.

The statistical theory depends on the use of a single scattering law, which has a reasonably simple Fourier-Bessel transform. Molière has used the same form of scattering law as have Snyder and Scott and a number of other authors, namely, that the differential cross section per unit solid angle is

$$Q(\chi)=\left(\frac{2Z'Ze^2}{pv}\right)^2 \frac{1}{(\chi^2+\chi_\gamma^2)^2}. \tag{1}$$

The parameter χ_γ depends on the screening of the atomic nucleus by its electrons, and Molière has evaluated it by fitting eq (1) to an accurate calculation based on two improvements over the usual Born approximation. For small angles, he used a combination of WKB ray optics near the atom and strict wave mechanics for the scattered waves beyond the atom. For angles up to 90°, a partial wave analysis, using Legendre functions was employed. Both methods were evaluated for Thomas-Fermi atomic wave functions, using a sufficiently accurate analytic approximation to the latter. Unlike the Born approximation, which is valid only when the parameter

$$\gamma=\frac{ZZ'}{137\beta} \tag{2}$$

[1] Smith College, Northampton, Mass.
[2] Figures in brackets indicate the literature references on p. 37.

is much less than 1 (Z, Z' are atomic numbers of scattering atom and scattered particle; $\beta=v/c$ for the particle), these calculations of Molière are valid for all γ.

To evaluate χ_γ, he calculated a root-mean-square angle, with an arbitrary cut-off, using the accurate theory, and fitted the same value from (1) for each of several values of γ. A simple analytic interpolation formula yielded finally

$$\chi_\gamma=1.13\,\frac{me^2Z^{1/3}}{\hbar p}\,(1.13+3.76\gamma^2)^{1/2}, \tag{3}$$

where m and e refer to the electrons and p is the momentum of the scattered particles.

The Snyder-Scott η_0 is $me^2Z^{1/3}\hbar p$. The 1.13 factor in front of this formula arises from the use of the Thomas-Fermi atom radius instead of the Bohr radius multiplied by $Z^{1/3}$. The radical is Molière's interpolation formula. The results are certainly accurate enough for our purposes since the multiple-scattering distribution is relatively insensitive to χ_γ. We shall therefore use eq (3) in what follows.

The statistical methods used by both Molière and Snyder and Scott are mathematically equivalent, and were derived in a straightforward way by Bothe [6] in 1921. One calculates, in fact, the probability of a given angle θ occurring as a result of exactly one, exactly two, exactly three, etc., scatterings, and performs an infinite sum. Each term has the form of an n-fold faltungs-integral over the single-scattering curve, so that the use of Fourier-Bessel transforms reduces each term to a power of the transform of the single-scattering function. An exponential series results, and the summation is immediate. Bothe used this method to show the limitations of the Gauss error law, and, of course, the deviations from Gaussian behavior are clearly discernible.

Molière used the Bothe method directly. Snyder and Scott used a diffusion equation whose solution is identical. However, the Bothe method is simple only for the problem of the tangent angle distribution at the end of the track, whereas the Snyder-Scott method is equally good for the laterial distribution of the end of the track, or, what is the same thing, the distribution of chord angles with respect to the initial direction.

Figure 4.1 shows the two angles in question. It is proved in the paper of Scott and Snyder [4] that the chord angle distribution for two chords in a given cell or length of track is independent of the division point of the cell, so the lateral distribution yields also the distribution between adjacent chord angles.

These results, of course, apply only to the small-angle approximation. Lewis [7], and Wong and Guth [8], have shown how our theory can be derived from a calculation valid for all angles.

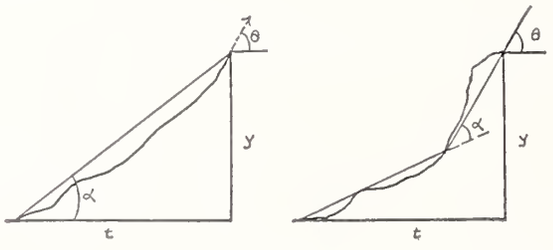


FIGURE 4.1.

Molière's numerical calculation was based on a clever choice of an expansion parameter. Now, the simplest physically significant parameter for the scattering in a given thickness of material is the mean number of scatterings that occur. We shall call this number Ω, omitting Molière's subscript b, and distinguish Ω from the Snyder-Scott z, which has the same significance but was not accurately calculated. Using eq (1) and (3), we obtain

$$\Omega=\frac{4\pi Nte^4(Z')^2Z(Z+1)}{(pv)^2\chi_\gamma^2}. \tag{4}$$

N is the number of atoms per cubic centimeter, and t is the thickness of the scatterer in centimeters. Z^2 has been replaced by $Z(Z+1)$ to account for the electron-electron scattering, in accord with the recent work of Hanson, Lanzl, Lyman, and Scott [14].

Molière's expansion parameter is a quantity B, related to Ω by

$$1.167\frac{e^B}{B}=\Omega. \tag{5}$$

It turns out that if the Fourier-Bessel transform for the multiple-scattering distribution is expanded in inverse powers of B, the separate terms can be evaluated by rather simple calculations. Molière's second paper gives numerical results to three-figure accuracy valid for a fairly wide range of angles and a wide range of values of Ω. Results are given for both the actual spatial angle and the angle projected on a plane of observation.

The Snyder-Scott calculations were performed by direct, numerical integration involving considerable effort, including the use of ten-place cosine tables, and are somewhat more accurate than Molière's results. Tables are available for both tangent and chord angle distributions for the projected case, and are nearly complete for the three dimensional case. These tables were, of course, based on a different value of χ_γ than that given here, and a different value of Ω. However, by suitably defining the dimensionless variables of the tables, they become a numerical evaluation of Molière's results. In fact, we simply take the dimensionless angle η to be the actual projected angle Φ divided by χ_γ, and we use Ω for z.

We have compared our tables with Molière's results, and find excellent numerical agreement in the projected tangent angle case for six values or Ω from 100 to 84,000—in most cases better than 1 percent and always within about 2 percent. Molière's functions $f^{(1)}$ and $f^{(2)}$ are not tabulated to much greater accuracy, and one or two of his numbers may be in error. We have also extended his tables of these functions for his dimensionless angle $\varphi(\varphi=\Phi/\chi_\gamma\ (B\Omega)^{1/2})$ from 4 to 13.5 in order to compare the tails of our curves, and again find excellent agreement. His asymptotic formulae agree with ours, but the terms are rearranged because of his expansion in inverse powers of B as well as of φ.

Butler [9] has an asymptotic expansion which is almost identical to Molière's $f^{(1)}$ term. The $f^{(2)}$ term, which Butler does not have, contributes, at $\Omega=100$, 3 percent for $\varphi=4$ and 1 percent for $\varphi=13$; and one-half as much at $\Omega=24{,}000$.

Miss Jean Snover of the computing staff at the Brookhaven Laboratory has recently reported to me that similar close agreement exists for the three-dimensional case.

Now, while there are some experiments that yield distribution curves directly comparable with the tables, in most cases some mean-value quantity is obtained, and compared. The root-mean-square has frequently been used, largely because of the well-known theorem that the mean square angle for multiple-scattering is just Ω times that for single scattering. However, this is a bad choice experimentally, since it overweights the large angles for which the statistics are poor. Furthermore, the rms value for single scattering can only be found by cutting off the single scattering law in a somewhat arbitrary and ambiguous way, leading to a minor uncertainty in the theoretical results.

If, however, a mean absolute value is calculated, the effect of any such cut-off (such as is produced by the finite size of the nucleus) is completely negligible. The use of a median absolute value in cases where it is practical allows even less influence of the poor statistics at large angles. If background counts, as in some experiments, prevent counting the numbers of large angles, a measure of half-width would be more practical. If the experiments allow it, the $1/e$ width might be even more useful.

The four above-mentioned quantities have been calculated for both tangent and chord projected distributions, and in addition we have included two more. The fifth is the angle measured by the ratio of intensity at zero angle to the total number of observations, an angle that could be defined as the $1/e$ width of a Gaussian curve of the same total area and same maximum height. This angle is equal to $1/P_0\pi^{1/2}$, where P_0 is the normalized intensity per unit angle at zero angle. Lastly, we give the somewhat arbitrary, but very practical, mean absolute angle taken when the distribution is cut off at 4 times this mean angle, by a relaxation calculation.

The mean absolute values were calculated by numerical integration, and the results checked closely with an analytic formula for the mean value obtained by Molière. The $\frac{1}{2}$ and $1/e$ widths were found by using a linear interpolation between Φ and $\log P$, an accurate procedure because of the approximate Gaussian behavior of the curves. The medians were found by integrating a not-quite-normalized Gauss curve fitted to the central maximum and to the $\frac{1}{2}$ width. The error in this procedure was shown to be quite small. The other two calculations were straightforward.

All of these twelve quantities, six for each distribution, have remarkably simple dependencies on the mean number of scatterings Ω. These were discovered by considering that if the distribution were truly Gaussian, any one of the quantities—call it ψ—would vary strictly as $\Omega^{\frac{1}{2}}$. A plot of some of the quantities $\psi/\Omega^{1/2}$ against $\log \Omega$ yielded curves that were nearly straight, and just as a guess ψ^2/Ω was plotted against $\log \Omega$. For the range of Ω from 10^2 to 10^5, lines straight to 1 percent resulted, and the coefficients are listed in table 4.1.

We call the various (ψ^2/Ω)'s for the tangent distribution, T's, and the quantities for the chord angle distribution, C's. Figures 4.2 and 4.3 show the calculated points and the lines fitted to them by a least-square-relative-error method. Only three points in all these curves deviate by more than 1 percent, and they by less than 2 percent.

TABLE 4.1. *Coefficients for the relations* T *or* $C=\psi^2\Omega=A+A'\log_{10}\Omega$.

Mean-value angle, ψ	Tangent values		Chord values	
	A	A′	A	A′
Arithmetic mean	1.044	0.809	0.299	0.269
Arithmetic mean cut-off at 4x	.418	.818	.090	.272
Median absolute angle	.222	.596	.033	.200
½ width	.035	1.831	−.123	.613
1/e width	.253	2.636	−.105	.879
$[\sqrt{\pi}P(0)]^{-1}$	.806	2.656	.088	.885

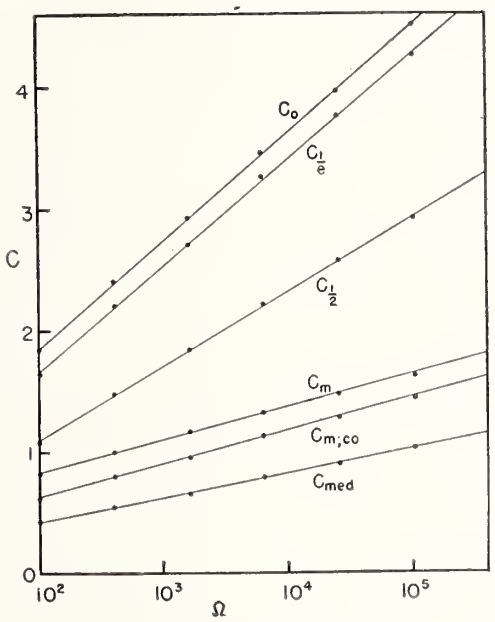


FIGURE 4.2.

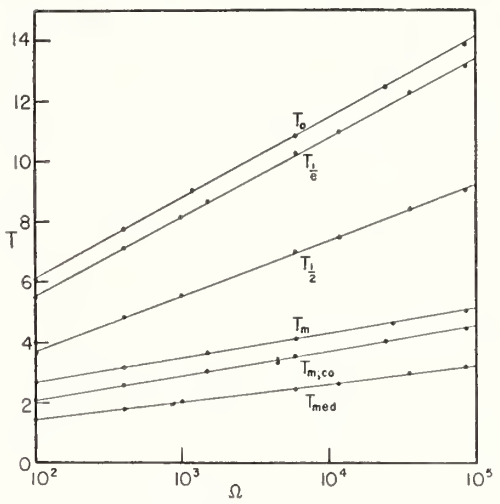


FIGURE 4.3.

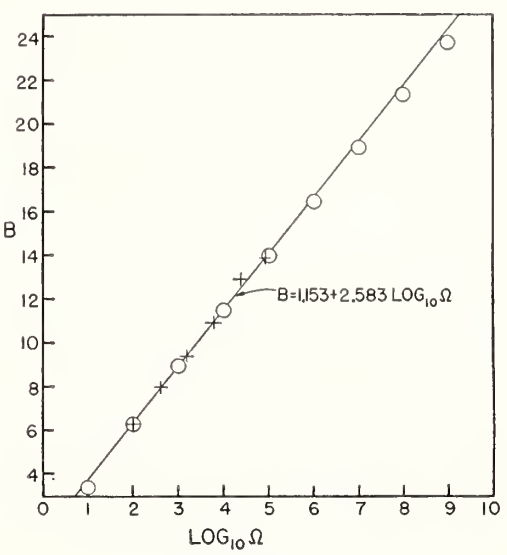


FIGURE 4.4.

The results are not very good for Ω very much below 100. However, for Ω from 10^5 to 10^9 we believe they are good. The mean absolute angle for the tangent case fits Molière's analytic formula to 1 percent out to Ω equal to 10^7, and to 3 percent out to 10^9. One expansion of Molière's, good for small angles, has a Gaussian term plus corrections, and the 1/e width of this Gaussian is just B. B is thus closely related to $T_{1/2}$, $T_{1/e}$, and T_0, and a linear relation for it is good to 3 percent for Ω out to 10^9. Figure 4.4 shows Molière's values of B, marked by circles, and more accurate calculations, made by Miss Snover, marked by crosses. The linear relationship

$$B=1.153+2.583\ \log_{10}\Omega, \tag{6}$$

which was fitted to the crosses is seen to fit fairly well over a wide range. The deviation for $\Omega=10$ is, however, about 10 percent and will get rapidly worse above 10.

The extent to which the Gaussian approximation is valid is partly shown by the slow logarithmic variation of the T's and C's. A Gaussian expression in which the various ψ's propagate like the root-mean-square value would show no such variation. However, a Gaussian expression with a different law of propagation should still have certain particular relations between the various mean value quantities. Let us characterize a Gauss curve by its $1/e$ width w, so $P(\Phi)=P_0e^{-\Phi^2/w^2}$. Then $|\Phi|_{\text{avg}}=w/\pi^{1/2}$; $|\Phi|\,\text{med}=0.477w$; $\Phi_{1/2}=w\sqrt{\log_e 2}$; $\Phi_0=w$ and $\Phi_{\text{avg cut-off}}=0.561w$. If we calculate w for large Ω from the formulas given for the T's, using the coefficients of log Ω, we get, in order of the table in the abstract, 2.593, 2.595, 2.637, 2.642, 2.636, 2.656. The last four values in this list are pretty much independent of the tail of the curve, and show a good agreement. The same thing is true of the C's. However, no simple relationships hold for the constant terms, i. e., for small Ω. We see that for large Ω the distribution becomes Gaussian in form, but with a different law of variation for w than would be calculated from an rms value.

In the Gaussian approximation, it is easy to derive a relationship between the T's and C's. The mean angle between tangents is $\sqrt{3/2}$ times the mean angle between successive chords for the same cell length. We have here the case of half cell lengths for the chord angles, which reduces the mean chord angles by $\sqrt{2}$, and the T's and C's involve the squares of angles. Thus the T's should be three times the C's. The coefficients of the logarithms obey this rule within 0.4 percent, but the other terms do not, showing that this frequently quoted relationship is only accurate for large Ω.

Although as stated before, a root-mean-square angle is not very useful, the estimations of errors in a sample mean is most easily made by use of a mean-square error. Therefore, we calculated the mean-square angle with the same cut-off that was used for the cut-off mean absolute angle. Similar results occur when the same plots are made, although the deviation from linearity is greater. If we calculate the rms deviation from the mean absolute angle, we find values in both tangent and chord case that vary from 0.77 to 0.82 times the mean value. If several successive chords or tangents are measured on a track, one can estimate the error in the mean arithmetic angle by use of the well-known theorem of the sample mean; namely, that the rms deviation of the sample mean is $n^{-1/2}$ times the expected rms deviation for a single observation. The result is not strictly true in this case, for the successive angles are correlated, but an earlier calculation [12] on 8 chords showed that the effect of correlation is to increase the standard deviation by about 5 percent. Hence we feel justified in quoting $0.85\ n^{-1/2}$ for the relative rms deviation when n chord angles are measured on a track.

A comparison of our results with Williams' [5] theory can be made by fitting formulas of our type to his results. We find

$$\left.\begin{aligned} T^w_m &= 1.63+0.79\ \log_{10}\Omega \\ T^w_{mco} &= 1.18+0.79\ \log_{10}\Omega \end{aligned}\right\}, \tag{7}$$

which agree fairly well with our results for large Ω but not for small Ω.

Lastly we shall discuss the scattering constant K. In photographic emulsions we have a mixture of nuclear species, and must employ the appropriate average values for Ω and ψ_γ. The work of Molière [2] indicates that the appropriate quantities to average are $\chi_\gamma^2\Omega$ and $(Z^2 \log \chi^2)\langle Z^2\rangle_{av}$ rather than $\chi_\gamma^2\Omega$ and Ω/t as indicated by Scott and Snyder [4] and used recently by Snyder [13].

The scattering constant K has usually been defined for mean chord angles α, and in our case can be expressed in terms of C_m. However, other C's can be used. The relation, for a mean-value type ψ is

$$K_\psi = 2p^2v^2\chi_\gamma^2\Omega C_\psi/t. \tag{8}$$

The 2 results because the cell length $s=t/2$.

Using eq (3) and (4) and the results in the last paragraph, we have then

$$K_\psi = [8\pi e^4(Z')^2\textstyle\sum_i N_i Z_i(Z_i+1)]C_\psi. \tag{9}$$

Using the manufacturer's data for Ilford G–5 emulsion, we find the bracket in (9) to have a value 25.25. We also find for this emulsion $\Omega = 631 s_0$, where s_0 is the cell length in microns. (This value contrasts with $\Omega = 1248 s_0$ given by Snyder [13]). Several results are obtainable, according as we calculate the C from chord angles, from tangent angles with a factor 3, from smoothed-out successive chords, or from alternate smoothed chords, and are quoted in the accompanying table:

Scattering constant K for 100-micron tracks				
	From C	From $T/3$	Smoothed-out chords	Alternate smoothed chords
Complete mean value	25. 9	26. 5	27. 3	30. 3
Cut-off mean value	23. 3	24. 0	24. 6	27. 5

The author acknowledges the support of the AEC at Brookhaven, and the assistance of the Misses Theresa Danielson (Mrs. William Ames), Jean Snover, Rika Sarfaty, and Gertrude Nissenbaum in performing the calculations.

[1] G. Molière, Z. Naturforsch. **2a**, 133 (1947).
[2] G. Molière, Z. Naturforsch. **3a**, 78 (1948).
[3] H. S. Snyder and W. T. Scott, Phys. Rev. **76**, 220 (1949).
[4] W. T. Scott and H. S. Snyder, Phys. Rev. **78**, 223 (1950).
[5] E. J. Williams, Phys. Rev. **58**, 292 (1940); L. Voyvodic and E. Pickup, Phys. Rev. **85**, 91 (1952).
[6] W. Bothe, Z. Phys. **4**, 161, 300 (1921); **5**, 63 (1921).
[7] H. W. Lewis, Phys. Rev. **78**, 526 (1950).
[8] M. C. Wang and E. Guth, Phys. Rev. **84**, 1092 (1951).
[9] S. T. Butler, Proc. Phys. Soc. (London) **63A**, 599 (1950).

[10] See, e. g., W. T. Scott, Phys. Rev. **76,** 212 (1949).
[11] Statistical factors for smoothing are given by G. Molière in an unpublished manuscript. I am indebted to Dr. Molière for a copy of this work.
[12] W. T. Scott, Phys. Rev. **75,** 1763 (1949).
[13] H. S. Snyder, Phys. Rev. **83,** 1068 (1951).
[14] A. O. Hanson, L. H. Lanzl, E. M. Lyman, and M. B. Scott, Phys. Rev. **84,** 639 (1951).

Discussion

Dr. W. Bothe, University of Heidelberg, Heidelberg, Germany: Have you made comparison between your results and Williams' formula?

Dr. Scott: Yes. What I did was to take mean values and calculate similar formulae. The Williams' theory was recently summarized in a paper submitted by Voyvodic and Pickup to the Physical Review. Using that theory one can get the quantity I call T for the mean value: $T_{\text{mean}}=1.63+0.79$ log Ω; and the cut-off value is $T_{\text{cut-off}}=1.18+0.79$ log Ω. The numbers in table 4.1 indicate that in our case the coefficients of log Ω are nearly the same, but the first numbers are smaller, so for large values of Ω the expressions agree.

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: You mentioned that you don't have an explanation for the results in that relation. Do you have something physical to offer for it?

Dr. Scott: As to why it is a straight line? Of course, why it is slowly varying is quite clear. It ought not to have much variation. Snyder and I tried to see from our theory how one would get this linear relation, but we could not. It might be that by studying the way in which this quantity B of Molière leads to a simple expansion, one could see that the quantity ought to vary in this way. For instance, Molière would give for the mean value a formula of this sort: A_0+A_1/B. I don't really see why this comes out a straight line, but the fact that it does is very helpful.

Dr. E. Guth, University of Notre Dame, Notre Dame, Ind.: Are your computations more complete than Molière's?

Dr. Scott: Yes. Molière gives a brief table from which you can cover nearly all of our tables but not very far out. We extended his tables for his ϕ, which he takes up to 4. We extended them from 4 to 13½ and calculated to another decimal place, and his tables would be easier to carry in your pocket than ours and give actually the same results. Although ours are more complete, I am not sure there is a great advantage to them.

5. On the General Theory of Multiple Scattering, Particularly of Charged Particles[1]

By Ming Chen Wang [2] and Eugene Guth [2]

The general theory of the elastic multiple scattering of particles with a strongly anisotropic scattering function is investigated without making the small-angle approximation. The rigorous transport equation is used and approximations are introduced at a later stage. First the general formulation of the problem is given. Some of the approximations involved in the existing theories of small-angle forward scattering are discussed. The solution for the problem of anisotropic scattering is derived by means of the spherical harmonic method. The general nth approximation is given, and is shown to be equivalent in all details with the nth approximation obtained by means of the gaussian quadrature method. Finally, the problem of anisotropic multiple scattering is reduced to a quasi-isotropic one by using the Goudsmit-Saunderson distribution function (defined for all angles) as a first approximation. Three different methods are given for forward scattering (including large angles). The first method is a perturbation treatment. The second method is based on the approximate delta-function character of the single scattering function, and employs a Fokker-Planck-type development for the peaked part of the scattering function. The third method is a Liouville-Neumann type of iteration applied directly to the transport equation. For back scattering the second and third of these methods also apply. In addition a special method is developed, based on the smallness of the back single scattering cross section. The Goudsmit-Saunderson distribution function is developed in powers of the thickness of the scatterer, and it is shown that all three methods lead to the same single scattering tail.

1. Introduction

The theory of multiple scattering has been developed in the past primarily for three groups of problems, namely, (1) radiative transfer, (2) multiple scattering of neutrons and (3) multiple scattering of charged particles, in particular of electrons. In any problem of multiple scattering we have two steps to consider: (a) the law of single scattering, and (b) the statistical problem of obtaining the spatial and angular distribution of the multiply scattered light or particles, which is properly governed by a Boltzmann integro-differential equation. We shall be concerned with the case of axial symmetry in the single scattering law, and multiple scattering in a plane-parallel stratified medium. The distribution function will then depend only on two variables: a cartesian and a polar coordinate describing the spatial and angular behavior respectively.

A method of expanding the distribution function into Legendre polynomials (spherical harmonic method, called SH in the sequel) has been applied by Gratton [1]* and Chandrasekhar [2] to the problem of isotropic scattering. Unfortunately these authors did not go far enough to get the general nth approximation so that, at that time, this method seemed to be inferior to the method of gaussian quadrature

[1] Supported in part by ONR. This is an abbreviated version of a paper which appeared in Dec. 15, 1951, Phys. Rev.

[2] University of Notre Dame, Notre Dame, Ind.

*Figures in brackets indicate the literature references on p. 59.

(called GQ in the sequel). We shall derive the nth approximation for anisotropic scattering by means of the SH method and show the relation between the SH method and the GQ method.

In both the SH and the GQ method, there is an ambiguity in the way of defining (a) successive approximations and (b) approximate boundary conditions. This ambiguity does not seem to have been recognized in the literature. The usual procedure for (a) and (b) in the GQ method was chosen for its analytical simplicity. We shall show that for the SH method (a) can be chosen in a natural way so that the GQ and the SH methods give exactly the same solution. Our (b) was chosen among other reasonable choices to give the best approximation to the exact solution for the isotropic case.

The simpler features of the multiple scattering of charged particles have been described in a very instructive manner in a recent paper by Bohr [3]. For this reason, we shall restrict ourselves to a brief summary of the attempts toward a more rigorous theory Bothe [4] started with the correct Boltzmann integro-differential equation, but did not state the exact boundary conditions. From the Boltzmann equation he derived a Fokker-Planck type of differential equation. Though the transition "from Boltzmann to Fokker-Planck" assumes the small angle approximation, Bothe retained in the latter a factor ($\cos\theta$) which goes to unity in that approximation. He then tried to solve this Fokker-Plancktype equation with an inexact boundary condition. Recently he [5] has given elementary considerations on back scattering. Bethe, Rose, and Smith [6] have tried to obtain a solution of the same differential equation with the exact boundary conditions. However, it seems doubtful whether such a solution does exist. They also derived a diffusion equation for back scattering from thick foils by a procedure similar to the "age" theory of neutron diffusion. Williams [7] developed a consistent theory in the small angle approximation based on the Fokker-Planck equation, putting the factor mentioned ($\cos\theta$) equal to unity and using an approximate boundary condition neglecting back scattering. Goudsmit and Saunderson [8] developed a more accurate theory, which can be formulated as follows: One applies the small angle approximation (putting $\cos\theta=1$) to the Boltzmann equation (instead of the Fokker-Planck equation of Williams), and uses the same approximate boundary condition as Williams does, neglecting back scattering.

Goudsmit and Saunderson pointed out that a parameter in their theory can be considered either as the thickness of the medium or as a path length, and that their solution is an exact one for the latter case. Molière [9] used an old theory of Wentzel [10] to derive an expression which is just that of Goudsmit and Saunderson if one replaces a series by an integral, and evaluates the integral. Snyder and Scott [11] derived essentially the same integral from equations equivalent to the approximate Boltzmann equation which leads to the Goudsmit-Saunderson theory. These authors give the most extensive numerical evaluation of their formulas. For thin scatterers both Molière and Snyder-Scott showed how the multiple scattering approaches the single scattering tail. A different approach by Butler [12] leads, for thin scatterers, also to a separation of the gaussian multiple scattering and the single scattering tail in the small angle approximation. Butler's procedure was generalized to larger angles, still neglecting back scattering, by Teichmann [13]. Lewis [14] treated multiple scattering in an infinite medium using the path-length as a variable.

He rederived the Goudsmit-Saunderson solution and showed the transition to the Snyder-Scott integral form. Weymouth [15] obtained some approximate formulas for the spatial, angular and path-length distribution, in the form of integrals, in a semi-infinite medium with isotropic sources uniformily distributed in the medium.

One may solve the problem by getting a joint distribution of path length and actual thickness of the foil, since we already have the distribution of path length and angle, the Goudsmit-Saunderson distribution. This is, of course, just as complicated as the original problem. Yang [16] has derived such a joint distribution in the small angle approximation. The average value of path length obtained by him is the same as that obtained previously by Rose [17].

2. General Formulation

2.1. Law of Elastic Single Scattering for Axially Symmetric Scattering Potential

Let α be the angle between the incident velocity and the scattered velocity, and $\Phi(\cos\alpha)$ be the single scattering law. The total scattering cross section σ is given by

$$\sigma=\int_0^{2\pi} d\phi \int_0^{\pi} \Phi(\cos\alpha)\sin\alpha d\alpha,$$

where ϕ is the azimuth angle. Then we define our scattering function by

$$p(\cos\alpha)=\frac{4\pi}{\sigma}\Phi(\cos\alpha), \tag{1}$$

so that, when we develop $p(\cos\alpha)$ in Legendre polynomials

$$p(\cos\alpha)=\sum_{l=0}^{\infty}\omega_l P_l(\cos\alpha), \tag{2}$$

we have $\omega_0=1$.

Now with reference to an arbitrary spherical coordinate system, the directions of the incident velocity and the scattered velocity can be specified by the angles θ, ϕ and θ', ϕ', respectively. These angles are related to the angle α by the equation

$$\cos\alpha=\mu\mu'+[(1-\mu^2)(1-(\mu')^2)]^{\frac{1}{2}}\cos(\phi-\phi'), \tag{3}$$

where $\mu=\cos\theta$ and $\mu'=\cos\theta'$. Equation (2) then becomes

$$p(\mu,\phi,\mu',\phi')=\sum_{l=0}^{\infty}\omega_l P_l\{\mu\mu'+[(1-\mu^2)(1-\mu'^2)]^{\frac{1}{2}}\cos(\phi-\phi')\}.$$

Expanding the Legendre polynomials by the addition theorem and integrating over the variable ϕ, we get

$$p(\mu,\mu')=\frac{1}{2\pi}\int_0^{2\pi} p(\mu,\phi,\mu',\phi')d\phi=\sum_{l=0}^{\infty}\omega_l P_l(\mu)P_l(\mu'). \tag{4}$$

2.2. Boltzmann Equation

An axially symmetric beam of particles is incident on a plane slab of scattering material with two of its dimensions infinite. Let the axis parallel to the finite dimension be the x-axis, which is also the axis of symmetry of the incident beam. The beam is incident on the surface $x=0$, and the other surface is $x=\alpha>0$. The steady state distribution function of the scattered particles obeys the well-known Boltzmann equation

$$\mu \frac{\partial f}{\partial \tau}+f=\frac{1}{2}\int_{-1}^{1} f(\tau,\mu')p(\mu,\mu')d\mu', \tag{5}$$

where μ is the cosine of the angle between the velocity of the particle and the x-axis, $\tau=N\sigma x$, and N is the number of scattering centers per unit volume. σ and $p(\mu,\mu')$ are defined above. The problem is to solve eq (5) with the boundary conditions

$$f(0,\mu)=\pi F\delta(\mu-\mu_0) \qquad \mu>0 \tag{6a}$$

$$f(t,\mu)=0 \qquad \mu<0, \tag{6b}$$

where $t=N\sigma a$ and $\mu_0>0$. For simplicity, we have used a δ-function as the incident beam. As the integral equation is linear, the solution for any arbitrary axially symmetric incident beam is just a superposition of such fundamental solutions. The integral of the δ-function over the whole solid angle is normalized to unity, while the strength of the incident beam is governed by the constant F.

If we define the integral in eq (5) by the function $J(\tau,\mu)$, i. e.,

$$J(\tau,\mu)=\frac{1}{2}\int_{-1}^{1} f(\tau,\mu')p(\mu,\mu'd\mu', \tag{7}$$

eq (5) becomes

$$\mu \frac{\partial f}{\partial \tau}+f=J(\tau,\mu). \tag{8}$$

The formal solution of eq (8), satisfying the boundary conditions, eq (6), is clearly

$$f(\tau,+\mu)=e^{-\tau/\mu}\int_0^{\tau} J(\tau,\mu)e^{\tau/\mu}\frac{d\tau}{\mu}+\pi Fe^{-\tau/\mu}\delta(\mu-\mu_0) \tag{9a}$$

$$f(\tau,-\mu)=e^{\tau/\mu}\int_{\tau}^{t} J(\tau,-\mu)e^{-\tau/\mu}\frac{d\tau}{\mu}. \tag{9b}$$

Here μ is an absolute value. The expressions (9a) and (9b) correspond to the range $\mu>0$ and $\mu<0$, respectively.[3]

2.3. Approximations for Forward Scattering

If the single scattering is mainly in the forward direction, a reasonable approximation is to replace the factor μ in the first term of eq (5)

[3] The notations $f(\tau,+\mu)$ and $f(\tau,-\mu)$ were used by Chandrasekhar. We use them in some of the equations in this paper so that our results can be readily compared with his.

by μ_0 and neglect back scattering in the boundary condition. If we take $\mu_0=1$ for simplicity, the problem now becomes

$$\frac{\partial f}{\partial \tau}+f=\frac{1}{2}\int_{-1}^{1} f(\tau,\mu')p(\mu,\mu')d\mu', \tag{10}$$

with the boundary condition

$$f(0,\mu)=\pi F\delta(\mu-1), \qquad -1\leq\mu\leq 1. \tag{11}$$

The exact solution of this approximate problem is

$$f(\tau,\mu)=\frac{F}{4}\sum_{l=0}^{\alpha}(2l+1)P_l(\mu)e^{-k_l\tau}, \tag{12}$$

where $k_l=[1-\omega_l/(2l+1)]$ with ω_l given by eq (2). This series was first derived by Goudsmit and Saunderson [8] without explicit use of the Boltzmann equation. Their assumptions are thus equivalent to eqs (10) and (11).

Another kind of small-angle approximation is to replace the exact integral equation, eq (5), by a differential equation of the Fokker-Planck type. Using the definition of the function $p(\mu,\mu')$, we can rewrite eq (5) into the form

$$\mu\frac{\partial f}{\partial \tau}=\frac{1}{4\pi}\int_{-1}^{1}d\mu'\int_{0}^{2\pi}d\phi p(\cos\alpha)[f(\tau,\mu')-f(\tau,\mu)]. \tag{13}$$

The three angles α, $\theta=\cos^{-1}\mu$ and $\theta'=\cos^{-1}\mu'$ form a spherical triangle as shown by eq (3). Now let us use the edge opposite the angle θ' as the polar axis. We then have a relation similar to (3)

$$\cos\theta'=\cos\theta\cos\alpha+\sin\theta\sin\alpha\cos\beta, \tag{14}$$

where β is the azimuth angle in this case. Then one develops $f(\tau,\mu')$ into a Taylor series around μ, keeping terms up to α^2, and gets

$$f(\tau,\mu')=f(\tau,\mu)+\frac{\partial f}{\partial\mu}[(1-\mu^2)^{1/2}\alpha\cos\beta-\mu\alpha^2/2]+\frac{1}{2}\frac{\partial^2 f}{\partial\mu^2}(1-\mu^2)\alpha^2\cos^2\beta\ldots \tag{15}$$

By putting the last expression for $f(\tau,\mu')$into eq (13) and integrating over the angles α and β instead of $\mu'$and φ, we get a constant,

$$\mu\frac{\partial f}{\partial\tau}=\frac{1}{\lambda}\frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial f}{\partial\mu}, \tag{16}$$

where $1/\lambda=(1/8)\int_0^{\pi}p(\cos\alpha)\,\alpha^2\sin\alpha\,d\alpha$. This is the differential equation that Bethe, Rose, and Smith have tried to solve with the exact boundary condition, eq (6), but did not succeed, since eq (16) is incompatible with eq (6).

If one applies the same "forward" approximations to the Fokker-Planck equation, eq (16), as we did to the original integral equation,

i. e., putting $\mu=1$ in the right-hand member of eq (16) and using the approximate boundary condition eq (11), one gets the solution

$$f(\tau,\mu)=\frac{F}{4}\sum_{l=0}^{\alpha}(2l+1)P_l(\mu)e^{-l(l+1)\tau/\lambda},$$

which is, of course, an approximate form of the Goudsmit-Saunderson solution, eq (12). This type of small-angle approximation leads to Williams' theory of multiple scattering.

3. Anisotropic Scattering by Spherical Harmonic Method

It is convenient to reformulate the boundary value problem defined by eq (5) and (6). Let

$$f(\tau,\mu)=T(\tau,\mu)+\pi F\delta(\mu-\mu_0)e^{-\tau/\mu_0}. \tag{17}$$

Putting eq (17) into eq (5), we get the integral equation for $T(\tau,\mu)$

$$\mu\frac{\partial T}{\partial\tau}+T\frac{1}{2}=\int_{-1}^{1}T(\tau,\mu')p(\mu,\mu')d\mu'+\frac{F}{4}p(\mu,\mu_0)e^{-\tau/\mu_0}. \tag{18}$$

The boundary conditions, eq (6), become

$$T(0,\mu)=0 \qquad \mu>0 \tag{19a}$$

$$T(t,\mu)=0 \qquad \mu<0. \tag{19b}$$

Here we define $J(\tau,\mu)$ as

$$J(\tau,\mu)=\frac{1}{2}\int_{-1}^{1}T(\tau,\mu')p(\mu,\mu')d\mu'+\frac{F}{4}p(\mu,\mu_0)e^{-\tau/\mu_0}. \tag{20}$$

Then the formal solution, similar to eq (9), is

$$T(\tau,+\mu)=e^{-\tau/\mu}\int_0^{\tau}J(\tau,\mu)e^{\tau/\mu}\frac{d\tau}{\mu} \tag{21a}$$

$$T(\tau,-\mu)=e^{\tau/\mu}\int_{\tau}^{t}J(\tau,-\mu)e^{-\tau/\mu}\frac{d\tau}{\mu}. \tag{21b}$$

3.1. Approximate Boltzmann Equation

We have made a small-angle approximation for eq (5) and (6) to get eq (10) and (11). Now if we make the same approximation for eq (18) and (19), we get

$$\frac{\partial T}{\partial\tau}+T=\frac{1}{2}\int_{-1}^{1}T(\tau,\mu')p(\mu,\mu')d\mu'+\frac{F}{4}p(\mu,1)e^{-\tau} \tag{22}$$

and

$$T(0,\mu)=0 \qquad -1\leq\mu\leq 1. \tag{23}$$

Although we know the exact solution of this approximate boundary value problem, yet we shall work out the approximations of this problem by the SH method, and compare them with the corresponding approximations of the exact problem later.

First we develop both functions $T(\tau,\mu)$ and $p(\mu,\mu')$ in eq (22) in series of Legendre polynomials. The development for $T(\tau,\mu)$ is

$$T(\tau,\mu)=\frac{1}{2}\sum_{l=0}^{\infty}(2l+1)T_l(\tau)P_l(\mu) \tag{24}$$

and that for $p(\mu,\mu')$ is given by eq (4). Substituting both developments into eq (22), and equating coefficients of $P_l(\mu)$, we get the infinite system

$$\frac{dT_l}{d\tau}=-k_lT_l+\frac{F}{2}(1-k_l)e^{-\tau}, \qquad l=0,1,2\ \ldots \tag{25}$$

where $k_l=1-\dfrac{\omega_l}{2l+1}$. The system eq (25) can be immediately integrated, and we define the nth approximation as the solution with $T_l(\tau)$ equal to zero identically for $l\geq 2n$ and

$$T_l(\tau)=\frac{F}{2}(C_le^{-k_l\tau}-e^{-\tau}), \qquad l=0,1,2\ \ldots\ (2n-1) \tag{26}$$

where the C_l's are integration constants to be determined by the boundary condition, eq (23). From eq (24) we have

$$T_l(0)=\int_{-1}^{1}T(0,\mu)P_l(\mu)d\mu, \tag{26a}$$

which, in view of eq (23), gives $T_l(0)=0$. Therefore, $C_l=1$. Putting eq (4) and (24) with finite upper limit into eq (20) and using eq (26), one gets

$$J(\tau,\mu)=\frac{F}{4}\sum_{l=0}^{2n-1}\omega_lP_l(\mu)e^{-K_l\tau}. \tag{27}$$

The formal solution, corresponding to eq. (21), in this case is

$$T(\tau,+\mu)=e^{-\tau}\int_0^{\tau}J(\tau,\mu)e^{\tau}d\tau \tag{28a}$$

$$T(\tau,-\mu)=e^{-\tau}\int_0^{\tau}J(\tau,-\mu)e^{\tau}d\tau. \tag{28b}$$

Putting the expression eq (27) of $J(\tau,\mu)$ into eq (28), and then from eq (17) (with $\mu_0=1$) we obtain the final solution

$$f(\tau,+\mu)=\frac{F}{4}\sum_{l=0}^{2n-1}(2l+1)P_l(\mu)(e^{-k_l\tau}-e^{-\tau})+\pi F\delta(\mu-1)e^{-\tau} \tag{29a}$$

$$f(\tau,-\mu)=\frac{F}{4}\sum_{l=0}^{2n-1}(-1)^l(2l+1)P_l(\mu)(e^{-k_l\tau}-e^{-\tau}). \tag{29b}$$

When $n \to \infty$, the two expressions of eq (29) reduce to the identical form

$$f(\tau,\mu)=\frac{F}{4}\sum_{l=0}^{\infty}(2l+1)P_l(\mu)-k_i r, \qquad -1\leq\mu\leq 1, \tag{30}$$

which is the exact solution, eq (12).

3.2. Exact Boltzmann Equation

In this section we will get the solution of the exact integral equation eq. (18), with boundary condition, eq. (19), by the SH approximation. We develop the functions $T(\tau,\mu)$ and $p(\mu,\mu')$ in eq. (18) in series of Legendre polynomials as in section 3.1, and reduce the whole equation into an infinite series of Legendre polynomials with functions of τ as coefficients by means of the recursion formula and the orthogonality relation of Legendre polynomials. Since the resulting infinite series is an identity in μ, we can equate coefficients of the $P_l(\mu)$'s and thus obtain the infinite system

$$l\frac{dT_{l-1}}{d\tau}+S_lT_l+(l+1)\frac{dT_{l+1}}{d\tau}=\frac{F}{2}e^{-\tau/\mu_0}\omega_l P_l(\mu_0) \qquad l=0,1,2\ldots. \tag{31}$$

where

$$S_l=2l+1-\omega_l, \qquad S_0=0. \tag{32}$$

The associated homogeneous system of eq. (31) is, in matrix form,

$$\begin{pmatrix} 0 & D & 0 & 0 & \cdots \\ D & S_1 & 2D & 0 & \cdots \\ 0 & 2D & S_2 & 3D & \cdots \\ 0 & 0 & 3D & S_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{pmatrix}\begin{pmatrix} T_0 \\ T_1 \\ T_2 \\ T_3 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \quad \text{or} \quad [\Delta T]=0, \tag{33}$$

where $D=d/d\tau$, $[\Delta]=[\Delta(D)]$ and $[T]$ is a column matrix.

We define the nth approximation as the solution obtained by retaining the left upper corner submatrix of $2n$ rows and $2n$ columns in eq. (33), and putting all the other elements equal to zero. This finite system of $2n$ linear differential equations with constant coefficients can be easily solved. All the T_l's are solutions of the differential equation

$$\Delta_n(D)T_l=0 \qquad l=0,1,2,\ldots(2n-1), \tag{34}$$

where

$$\Delta_n(D)=\begin{vmatrix} 0 & D & 0 & 0 & \cdots & \cdot \\ D & S_1 & 2D & 0 & \cdots & \cdot \\ 0 & 2D & S_2 & 3D & \cdots & \cdot \\ 0 & 0 & 3D & S_3 & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & S_{2n-1} \end{vmatrix}. \tag{35}$$

It can be shown that the roots of $\Delta_n(D)=0$ are in pairs with same absolute value but opposite in sign. One pair of the roots is obviously $D=0$, and let us denote the other $n-1$ pairs of nonvanishing roots by k_α with $k_{-\alpha}=-k_\alpha$ and $\alpha=\pm 1,\pm 2, \ldots \pm(n-1)$. Then the general solution of eq. (34) is

$$T_l=\sum_{\alpha=-(n-1)}^{n-1} A_{l\alpha}e^{-k_\alpha\tau}+B_l\tau+C_l, \qquad l=0,1,2 \ldots (2n-1), \tag{36}$$

where the constants are related by

$$A_{l\alpha}=\rho_l(k_\alpha)A_{0\alpha}, \qquad A_{0\alpha} \text{ arbitrary} \tag{37a}$$

$$\rho_{l+1}(k_\alpha)-\frac{2l+1-\omega_l}{k_\alpha(l+1)}\,\rho_l(k_\alpha)+\frac{l}{l+1}\,\rho_{l-1}(k_\alpha)=0, \quad \rho_0=1 \tag{37b}$$

$$B_0 \text{ arbitrary}, \quad B_1=B_2=\ldots=B_{2n-1}=0 \tag{37c}$$

$$C_0 \text{ arbitrary}, \quad C_1=-B_0/(3-\omega_1), \quad C_2=C_3=\ldots=C_{2n-1}=0. \tag{37d}$$

For the nth approximation, the particular integral is obtained from the first $2n$ equations of the nonhomogeneous system, eq. (31). We get

$$T_l=g_l e^{-\tau/\mu_0} \qquad l=0,1,2 \ldots (2n-1), \tag{38}$$

where

$$g_l=\frac{F}{2}\left[\frac{P_{2n}(\mu_0)\,(2n)!}{\mu_0^{2n}\Delta_n(1/\mu_0)}\,\rho_l(1/\mu_0)-P_l(\mu_0)\right].$$

The complete solution is the sum of eq. (36) and (38). Thus we have

$$T_l=\sum_{\alpha=-(n-1)}^{n-1} A_{l\alpha}e^{-k_\alpha\tau}+B_l\tau+C_l + \frac{F}{2}\left[\frac{P_{2n}(\mu_0)\,(2n)!}{\mu_0^{2n}\Delta_n(1/\mu_0)}\,\rho_l(1/\mu_0)-P_l(\mu_0)\right]e^{-\tau/\mu_0}. \tag{39}$$

The constants in the solution obey the conditions of eq. (37a) to (37d),

which leaves only $2n$ independent arbitrary constants to be determined by the boundary conditions.

We shall approximate the boundary conditions of eq (19) by the $2n$ equations [4]

$$\int_0^1 P_l(\mu)T(0,\mu)d\mu=0, \quad \int_{-1}^0 P_l(\mu)T(t,\mu)d\mu=0, \qquad l=1,3\ldots(2n-1)$$

or

$$\sum_{k=0}\epsilon_{lk}T_k(0)=0, \quad \sum_{k=0}^{2n-1}\beta_{lk}T_k(t)=0 \qquad l=1,3\ldots(2n-1), \tag{40}$$

where

$$\epsilon_{lk}=\frac{2k+1}{2}\int_0^1 P_k(\mu)P_l(\mu)d\mu, \quad \beta_{lk}=\frac{2k+1}{2}\int_{-1}^0 P_k(\mu)P_l(\mu)d\mu.$$

The system, eq (40), provides the $2n$ equations to determine the $2n$ independent arbitrary constants, $A_{0\alpha}[\alpha=\pm1,\pm2\ldots\pm(n-1)]$, B_0 and C_0. The function $J(\tau,\mu)$ defined by eq (20) is, in this approximation,

$$J(\tau,\mu)=\frac{1}{2}\sum_{l=0}^{2n-1}\omega_l P_l(\mu)T_l(\tau)+\frac{F}{4}e^{-\tau/\mu_0}\sum_{l=0}^{2n-1}\omega_l P_l(\mu)P_l(\mu_0). \tag{41}$$

Putting the expression, eq (39), for T_l into eq (41) and making use of eq (37a), (37c), and (37d), we have

$$J(\tau,\mu)=\frac{1}{2}\Bigg[\sum_{\alpha=-(n-1)}^{n-1}A_{0\alpha}e^{-k_\alpha\tau}\sum_{l=0}^{2n=1}\omega_l P_l(\mu)\rho_l(K_\alpha)+B_0\tau+C_0-B_0\omega_1\mu/(3-\omega_1)$$

$$+\frac{FP_{2n}(\mu_0)(2n)!}{2\mu_0^{2n}\Delta_n(1/\mu_0)}e^{-\tau/\mu_0}\sum_{l=0}^{2n-1}\omega_l\rho_l(1/\mu_0)P_l(\mu)\Bigg]. \tag{42}$$

We then get the final solution by substituting the expression, eq (42), into eq (21) and in turn into eq (17)

$$f(\tau,+\mu)=\frac{1}{2}\Bigg[\sum_{\alpha=-(n-1)}^{n-1}\frac{A_{0\alpha}(e^{-k_\alpha\tau}-e^{-\tau/\mu})}{1-\mu k_\alpha}\sum_{l=0}^{2n-1}\omega_l\rho_l(k_\alpha)P_l(\mu)$$

$$+B_0\tau+\left(C_0-\frac{3B_0\mu}{3-\omega_1}\right)(1-e^{\tau/\mu})+\frac{FP_{2n}(\mu_0)(2n)!}{2\mu_0^{2n}\Delta n(1/\mu_0)}\ \frac{e^{-\tau/\mu_0}-e^{-\tau/\mu}}{1-\mu/\mu_0}$$

$$\times\sum_{l=0}^{2n-1}\omega_l\rho_l(1/\mu_0)P_l(\mu)\Bigg]+\pi F\delta(\mu-\mu_0)e^{-\tau/\mu_0} \tag{43a}$$

[4] Here we choose the equations with odd l, because we have found that the odd-l choice, among other reasonable choices, gives the best convergence in the problem of isotropic scattering, for which the exact solution is known.

$$f(\tau,-\mu)=\frac{1}{2}\Bigg[\sum_{\alpha=-(n-1)}^{n-1} A_{0\alpha}\frac{(e^{-k_\alpha\tau}-e^{-k_\alpha t-\frac{t-\tau}{\mu}})}{1+\mu k_\alpha}\sum_{l=0}^{2n-1}(-1)^l\omega_l\rho_l(k_\alpha)P_l(\mu)$$

$$+B_0(\tau-te^{-(t-\tau)/\mu})+\left(C_0+\frac{3B_0\mu}{3-\omega_1}\right)(1-e^{-(t-\tau)/\mu})$$

$$+\frac{FP_{2n}(\mu_0)(2n)!}{2\mu_0^{2n}\Delta_n(1/\mu_0)}\frac{e^{-\tau/\mu_0}-e^{-t/\mu_0-(t-\tau)/\mu}}{1+\mu/\mu_0}\sum_{l=0}^{2n-1}(-1)^l\omega_l\rho_l(1/\mu_0)P_l(\mu)\cdot \tag{43b}$$

3.3. Comparison with Gaussian Quadrature Method [5]

The GQ approximation of the exact integral equation (18) has been worked out by Chandrasekhar [18]. Using his expression of I_j in the quadrature formula

$$\int_{-1}^{1} I(\tau,\mu')p(\mu,\mu')d\mu'=\sum_{j=-n}^{\mu} a_j I_j p(\mu,\mu_j),$$

where the a_j's are the gaussian weights, and the μ_j's are the roots of $P_{2n}(\mu)=0$, one gets an expression for $J(\tau,\mu)$ which can be shown to be identical with eq (42). To prove the identity of these two expressions, we have to redefine the arbitrary constants in the following way:

$$A_{0\alpha}=\frac{F}{2}L_\alpha,\quad B_0=\frac{F}{2}\left(1-\frac{\omega_1}{3}\right)L_0,\quad C_0=\frac{F}{2}L_n,$$

and to show that the $k\alpha$'s are the same and

$$\frac{P_{2n}(\mu_0)(2n)!}{\mu_0^{2n}\Delta_n(1/\mu_0)}=H(\mu_0)H(-\mu_0)\cdot \tag{44}$$

In the SH method, our system of equations (34) can be written in the matrix notation

$$\beta T=0$$

with

$$\beta_{lm}=\left(1-\frac{\omega_l}{2l+1}\right)\delta_{lm}+\frac{D}{2l+1}[(m+1)\delta_{l,m+1}+$$

$$m\delta_{l,m-1}],\quad {}^{l}_{m}=0,1,2\ \ldots\ (2n-1).$$

In the GQ method, Chandrasekhar's system [6] of $2n$ equations can

---

[5] Chandrasekhar's way of defining successive approximations of the SH method for the isotropic case (reference [2]) can be shown to lead essentially to the same explicit form of the *nth* approximation as our procedure. The equivalence of the SH and the GQ methods persists also, if one uses odd Legendre polynomials for the latter.

[6] The difference in sign between Chandrasekhar's system (reference [18], eq (95) with $m=0$) and the system which we will put down in the following is due to the fact that he measures angles from the negative instead of the positive x-axis.

be written as $\alpha I=0$ with

$$\alpha_{ij}=(\mu_j D+1)\delta_{ij}-\frac{a_j}{2}\sum_{l=0}^{2n-}\omega_l P_l(\mu_i)P_l(\mu_j), \quad {}^{i}_{j}=\pm 1,\pm 2 \ldots \pm n.$$

Now if we define a matrix S with elements

$$S_{il}=\tfrac{1}{2}(2l+1)P_l(\mu_i), \qquad l=0,1,2 \ldots (2n-1),$$

$$i=\pm 1,\pm 2 \ldots \pm n,$$

the inverse matrix S^{-1} can be shown to have elements

$$S^{-1}_{lk}=a_k P_l(\mu_k), \qquad l=0,1,2 \ldots (2n-1),$$

$$k=\pm 1,\pm 2 \ldots \pm n.$$

Then it is easy to show that

$$S\beta S^{-1}=\alpha. \tag{45}$$

The characteristic roots $k\alpha$ of the SH and the GQ method are determined by $|\beta|=0$ and $|\alpha|=0$ respectively. Since β and α are connected by the transformation equation, eq (45), it is well-known that the characteristic roots must be identical.

To prove the identity eq (44), we need the explicit expression for $H(\mu_0)H(-\mu_0)$ as given by Chandrasekhar, i. e.,

$$H(\mu_0)H(-\mu_0)=(-1)^n\prod_{i=1}^{n-1}\frac{1}{1-\mu_0^2 k_i^2}\prod_{j=1}^{n}\frac{\mu_0^2-\mu_j^2}{\mu_j^2}.$$

Now since the μ_i's are the $2n$ roots of $P_{2n}(\mu)=0$, it is clear that

$$\prod_{j=1}^{n}\frac{\mu_0^2-\mu_j^2}{\mu_j^2}=\frac{P_{2n}(\mu_0)}{(-1)^n P_{2n}(0)}=\frac{2\cdot 4\cdot 6 \ldots 2n}{1\cdot 3\cdot 5 \ldots (2n-1)}P_{2n}(\mu_0).$$

Since the k_i's are the $2n-2$ roots of $\Delta_n(D)/(-D^2)=0$,

$$\prod_{i=1}^{n-1}(1-\mu_0^2 k_i^2)=(\mu_0^2)^{n-1}\prod_{i=1}^{n-1}\left(\frac{1}{\mu_0^2}-k_i^2\right)=\frac{\mu_0^{2n-2}\Delta_n(1/\mu_0)/(-1/\mu_0^2)}{(-1)^{n-1}[3\cdot 5\cdot 7 \ldots (2n-1)]^2},$$

where the denominator is the coefficient of the highest power of D in $-\Delta_n(D)/D^2$. Combining these results we get eq (44) immediately.

Since we have proved that the GQ approximation of Chandrasekhar gives an expression for J identical in form with our eq (42), it follows that the final solution $f(\tau,+\mu)$ and $f(\tau,-\mu)$ must be also identical in form with our eq (43a) and (43b). However, the numerical values of the arbitrary constants are slightly different because of the different ways of approximating the boundary condition in the two methods. Consequently, for any finite approximation, the two methods will give slightly different numerical values for the final solution, too. Table 5.1 gives the forward scattering in the first three approximations for both methods, and the corresponding approximations for the approxi-

mate integral equation. The parameters used in constructing table 5.1 coincide with the first case of Goudsmit-Saunderson's table I [19] (i. e., log $\xi=4$, $\mu=0.0025$.)

TABLE 5.1.[a] *Values of* $\frac{4}{F}[f(t,+\mu)-\pi F\delta(\mu-1)e^{-t}]$ *from eq (43a)*

μ	$n=1$			$n=2$			$n=3$		
	GQ	SH	Ap [b]	GQ	SH	Ap [b]	GQ	SH	Ap [b]
0	2.292	2.430	1.0	−3.60	−4.01	−1.26	3.35	3.89	1.29
0.1	------	------	------	−4.39	−4.65	−1.76	4.13	4.47	1.86
.2	------	------	------	−4.85	−4.95	−2.04	4.05	4.20	1.82
.3	------	------	------	−4.86	−4.83	−2.01	2.99	3.01	1.05
.4	------	------	------	−4.31	−4.17	−1.59	1.08	1.07	−0.34
.5	------	------	------	−3.07	−2.85	−0.67	−1.24	−1.17	−1.97
.6	------	------	------	−1.03	−0.79	.81	−3.09	−2.84	−3.09
.7	------	------	------	1.93	2.13	2.96	−3.11	−2.62	−2.55
.8	------	------	------	5.94	6.02	5.85	0.67	1.33	1.30
.9	------	------	------	11.10	10.98	9.58	10.87	11.49	10.69
1.0	6.256	6.075	3.9	17.54	17.11	14.23	30.90	31.04	28.47

[a] Parameters: $\omega_1=2.9941$, $\omega_2=4.9741$, $\omega_3=6.9343$, $\omega_4=8.702$, $\omega_5=10.778$, $t=19.46$.
[b] The values in these three columns are from the solution of the approximate integral equation [eq (29a) instead of eq (43a)].

4. Reduction of the Anisotropic Problem to a Quasi-Isotropic One

4.1. Forward Scattering

Theoretically the expression eq (43a) gives the forward scattering to any degree of accuracy one wants. But if the single scattering function is extremely forward, one has to carry the approximation to a large n in order to get some sensible result. This means a tremendous amount of numerical work, so it is not very desirable in practice.

Now since the Goudsmit-Saunderson solution eq (30) gives a pretty good approximation for forward scattering, we can consider this as the first approximation of a perturbation treatment. First we rewrite the integral equation eq (5) as follows:

$$\frac{\partial f}{\partial\tau}+f=\frac{1}{2}\int_{-1}^{1}f(\tau,\mu^1)p(\mu,\mu')d\mu'+\epsilon(1-\mu)\frac{\partial f}{\partial\tau}. \tag{46}$$

The parameter ϵ, which is inserted here to indicate that the term is small, will be set equal to unity eventually. Let

$$f=f_1+\epsilon f_2.$$

Putting this into eq (46), we get one equation with terms free from ϵ

$$\frac{\partial f_1}{\partial\tau}+f_1=\frac{1}{2}\int_{-1}^{1}f_1(\tau,\mu')p(\mu,\mu')d\mu' \tag{47}$$

and another one involving ϵ (with ϵ put equal to unity)

$$\mu\frac{\partial f_2}{\partial\tau}+f_2=\frac{1}{2}\int_{-1}^{1}f_2(\tau,\mu^1)p(\mu,\mu')d\mu'+(1-\mu)\frac{\partial f_1}{\partial\tau}. \tag{48}$$

If f_1 and f_2 satisfy the boundary conditions

$$f_1(0,\mu)=\pi F\delta(\mu-1), \quad -1\leq\mu\leq 1 \tag{49a}$$

$$f_2(0,\mu)=0, \quad \mu>0 \tag{49b}$$

$$f_2(t,\mu)=-f_1(t,\mu), \quad \mu<0 \tag{49c}$$

f will satisfy the original boundary condition eq (6), with $\mu_0=1$. Up to this point everything is still exact. Equations (47), (48) and (49) are just another way of defining the original problem given by eq (5) and (6).

The solution of eq (47) with boundary condition eq (49a) is just the Goudsmit-Saunderson eq (30). With f_1 given, one then wishes to find the correction term f_2 satisfying eq (48), (49b), and (49c). Here we will make the approximation by taking the finite series solution eq (29) for f_1 instead of eq (30) in eq (48), and then solve the problem to the same SH approximation with the corresponding approximation in the boundary condition. This approximate solution for f_2 is simply the difference of the two series solutions eq (43a) and (29a), with $\mu_0=1$ and $\tau=t$. Thus we have for forward scattering

$$f(t,\mu)=\text{Goudsmit-Saunderson solution, eq (30)},+[(43a)-(29a)] \tag{50}$$

Table 5.2 gives both the Goudsmit-Saunderson value (or f_1) and the correction (or f_2) for the special case which is treated in table 5.1. In calculating f_2 one does not need to carry the approximation to a large n as in the case of calculating the original f, because the extremely anisotropic part has been taken care of by the first approximation f_1. Here the correction term is calculated with $n=3$, which is, of course, just the difference of the last two columns of table 5.1. One notices from table 5.1 that in the region with positive values (the negative values in the table are meaningless), the percentage difference between the solutions for the exact and the approximate integral equations decreases steadily with the increase of approximation, and for the third approximation it is only about 8 percent for $\mu=1$. Therefore, the Goudsmit-Saunderson value there probably would not be more than 8 percent off from the exact value. Our correction in table 5.2 is about 2 percent at $\mu=1$ or $\theta=0°$.

TABLE 5.2.[a] *Forward scattering*

θ	Goudsmit-Saunderson solution[b]	Correction term[c]
°		
0	120.9	2.57
6	72.6	2.45
12	26.3	2.08
18	8.3	1.55
24	2.7	0.97
30	1.1	.44
45	0.1	−.08

[a] Parameters have same values as in table 5.1.
[b] Values from reference 19.
[c] Difference of last two columns in table 5.1.

Another simple approximation for f_2 is to make use of the fact that the single scattering function is extremely forward, which means the function $p(\mu,\mu')$ has a sharp peak at $\mu'=\mu$ approximately.[7] So we develop the function f_2 under the integral of eq (48) in a Taylor series around $\mu'=\mu$, that is

$$f_2(\tau,\mu')=f_2(\tau,\mu)+\frac{\partial f_2(\tau,\mu)}{\partial\mu}(\mu'-\mu)\ldots. \tag{51}$$

Then we get

$$\frac{1}{2}\int_{-1}^{1} f_2(\tau,\mu')p(\mu,\mu')d\mu'=f_2(\tau,\mu)-\mu k_1\frac{\partial f_2}{\partial\mu}\cdots, \tag{52}$$

where $k_1=1-(\omega_1/3)$. Using (52), eq (48) becomes

$$\mu\frac{\partial f_2}{\partial\tau}+\mu k_1\frac{\partial f_2}{\partial\mu}=(1-\mu)\frac{\partial f_1}{\partial\tau}. \tag{53}$$

For a first approximation, we omit the second term of eq. (53), since $k_1<<1$. Then the solution which satisfies the boundary conditions, eq (49b) and (49c), is

$$f_2(\tau,\mu)=\begin{cases}\dfrac{1-\mu}{\mu}f_1(\tau,\mu), & \mu>0\\[2ex] \dfrac{1-\mu}{\mu}f_1(\tau,\mu)-\dfrac{1}{\mu}f_1(t,\mu), & \mu<0\end{cases}$$

For the second approximation, we let

$$f_2=g_1+g_2 \tag{54}$$

with $g_1>>g_2$. We find the solution that satisfies eq (53) and (54), and the boundary conditions eq (49b) and (49c) to be

$$g_1=\frac{1-\mu}{\mu}f_1(\tau,\mu),\quad -1\le\mu\le 1,$$

$$g_2=-k_1\int_0^\tau\frac{\partial g_1}{\partial\mu}d\tau+\begin{cases}0, & \mu<0,\\ k_1\displaystyle\int_0^\tau\frac{\partial g_1}{\partial\mu}d\tau-\frac{1}{\mu}f_1(t,\mu), & \mu<0.\end{cases}$$

Then we have the final solution

$$f(\tau,\mu)=f_1+f_2=\begin{cases}\dfrac{f_1(\tau,\mu)}{\mu}-k_1\displaystyle\int_0^t\left(\frac{1-\mu}{\mu}\frac{\partial f_1}{\partial\mu}-\frac{f_1}{\mu^2}\right)d\tau, & \mu>0, \quad (55a)\\[2ex] \dfrac{f_1(\tau,\mu)-f_1(t,\mu)}{\mu}+k_1\displaystyle\int_\tau^t\left(\frac{1-\mu}{\mu}\frac{\partial f_1}{\partial\mu}-\frac{f_1}{\mu^2}\right)d\tau, & \mu<0, \quad (55b)\end{cases}$$

[7] The peak is usually not at $\mu'=\mu$, but a little distance off. For example, in the case we used for constructing table 5.1, $\mu'=1.000082\mu$ for values of μ^2 not too near unity.

where f_1 is the solution, eq (30). The forward scattering is given by eq (55a) with $\tau=t$, i. e.,

$$f(t,\mu)=\frac{f_1(t,\mu)}{\mu}-k_1\int_0^t\left(\frac{1-\mu}{\mu}\frac{\partial f_1}{\partial\mu}-\frac{f_1}{\mu^2}\right)d\tau, \quad \mu>0. \tag{56}$$

The result given by eq (55) diverges for $\mu=0$. Since as $\mu\to 0$ the terms neglected in eq (53) might be larger than the ones retained, the whole approximation breaks down. But away from $\mu=0$ the solution, eq (55) might give a reasonable approximation.

A third way of getting an approximate expression for f_2 is by an iteration method. Using f_1 as the first approximation for f, we can calculate the function J from eq (7). Then we obtain $f(t,+\mu)$ by eq (9a). The result is

$$f(t,+\mu)=f_1(t,+\mu)-\frac{F}{4}\sum_{l=0}^{\infty}(2l+1)k_lP_l(\mu)\frac{1-\mu}{1-\mu k_l}(e^{-k_lt}-e^{-t/\mu}). \tag{57}$$

4.2. Back Scattering

The expression eq (43b) gives the back scattering, but here again it is impractical for cases with a very anisotropic single scattering function. The perturbation treatment that we are going to give for this case is based upon the extremely small cross-section of the back scattering.

First we make a Taylor development of the scattering function p (cos α) defined by eq (1) around any backward direction, say cos $\alpha=$cos α_0 with $\alpha_0>\pi/2$, i. e.,

$$p\ (\cos\alpha)=p\ (\cos\alpha_0)+p'\ (\cos\alpha_0)(\cos\alpha-\cos\alpha_0)\ \ldots. \tag{58}$$

Rearrange the terms to make a series of Legendre polynomials

$$p\ (\cos\ \alpha)=\epsilon_n\sum_{l=0}^{n}\omega_l^*P_l\ (\cos\alpha), \tag{59}$$

where $\omega_0^*=1$. Though the function p (cos α) is extremely forward, it is much more isotropic in back scattering. Therefore, we need only a few terms of the development, eq (58) or (59), i. e., n is small. The factor ϵ_n depends upon n and α_0, and it is small for small n. From eq (59) we get a development for $p(\mu,\mu')$ similar to eq (4)

$$p(\mu,\mu')=\epsilon_n\sum_{l=0}^{n}\omega_l^*P_l(\mu)P_l(\mu'). \tag{60}$$

Putting eq (60) into eq (18), we have (setting $\mu_0=1$)

$$\mu\frac{\partial T}{\partial\tau}+T=\frac{F}{4}p(\mu,1)e^{-\tau}+\frac{\epsilon_n}{2}\sum_{l=0}^{n}\omega_l^*P_l(\mu)\int_{-1}^{1}T(\tau,\mu')P_l(\mu')d\mu'. \tag{61}$$

Now multiply eq (61) by exp $(-s\tau)d\tau$ and integrate τ from 0 to t.

Also let $\varphi(s,\mu)=\int_0^t d\tau e^{-s\tau}T(\tau,\mu)$. Then eq (61) becomes

$$\mu[T(t,\mu)e^{-st}-T(0,\mu)]+s\mu\varphi(s,\mu)+\varphi(s,\mu)$$

$$=\frac{F}{4}p(\mu,1)\frac{1-e^{-(1+s)t}}{1+s}+\frac{\epsilon_n}{2}\sum_{l=0}^{n}\omega_l^* P_l(\mu)\int_{-1}^{1}\varphi(s,\mu)'P_l(\mu')d\mu'. \quad (62)$$

In first approximation we will omit the sum in eq (62). Then we put $s=-1/\mu$, and introduce the proper boundary condition eq (19), for $T(\tau,\mu)$. Thus we get

$$T(0,\mu)=0;T^{(1)}(t,\mu)=-\frac{F}{4}\frac{1-e^{(1-\mu)t/\mu}}{1-\mu}p(\mu,1)e^{-t/\mu},\quad \mu>0 \quad (63a)$$

$$T(t,\mu)=0,T^{(1)}(0,\mu)=\frac{F}{4}\frac{1-e^{(1-\mu)t/\mu}}{1-\mu}p(\mu,1),\quad \mu<0. \quad (63b)$$

$T^{(1)}(0,\mu)$ is, of course, the first approximation for back scattering.

In order to get the second approximation for $T(\tau,\mu)$, we have to get the first approximation of $\varphi(s,\mu)$. One just substitutes the two expressions, eq (63a) and (63b), into eq (62), still omitting the sum, and gets

$$\varphi^{(1)}(s,\mu)=\frac{F}{4}\frac{p(\mu,1)}{1+s\mu}\left[\frac{1-e^{-(1+s)t}}{1+s}+\begin{cases}\dfrac{1-e^{(1-\mu)t/\mu}}{(1-\mu)/\mu}e^{-t/\mu-st} & \mu>0\\ \dfrac{1-e^{(1-\mu)t/\mu}}{(1-\mu)/\mu} & \mu<0\end{cases}\right] \quad (64)$$

Putting eq (64) into the integral in eq (62), and carrying out the integration in μ', we get the second approximation by again setting $s=-1/\mu$ and using the boundary condition for $T(\tau,\mu)$

$$T^{(2)}(0,\mu)=T^{(1)}(0,\mu)-\frac{\epsilon_n}{2\mu}\sum_{l=0}^{n}\omega_l^* P_l(\mu)\left[\int_{-1}^{1}\varphi^{(1)}(s,\mu')P_l(\mu')d\mu'\right]_{s=-1/\mu},\quad \mu<0 \quad (65)$$

One can go to higher approximations in a similar manner.

For a semi-infinite medium, $t\to\infty$, the first approximation of back scattering becomes

$$T^{(1)}(0,\mu)=\frac{F}{4}\frac{p(\mu,1)}{1-\mu},\quad \mu<0. \quad (66)$$

Now if we make the development eq (58) around the direction where we wish to calculate the back scattering, that is, put $\cos\alpha_0=\mu$,

$$p(\mu,1)=p(\mu) \quad (67)$$

exactly. Thus we get finally

$$T^{(1)}(0,\mu)=\frac{F}{4}\frac{p(\mu)}{1-\mu},\quad \mu<0. \quad (68)$$

With $t \to \infty$, the expression eq (64), reduces to a very simple form from which one can calculate $T^{(2)}(0,\mu)$ by eq (65).

To give some idea about the order of magnitude of the first approximation and the convergence of the method, we have calculated a few values for the semi-infinite case with the Rutherford scattering function

$$p(\cos\alpha) = \frac{4\beta(1+\beta)}{(1+2\beta-\cos\alpha)^2}, \tag{69}$$

which is normalized according to the definition, eq (1). β is the screening constant, usually much smaller than unity. Using eq (69), we get the first approximation immediately from eq (68)

$$T^{(1)}(0,\mu) = \frac{F\beta(1+\beta)}{(1+2\beta-\mu)^2(1-\mu)} \simeq \frac{F\beta}{(1-\mu)^3}, \quad \mu < 0. \tag{70}$$

If the second approximation is written in the form

$$T^{(2)}(0,\mu) = T^{(1)}(0,\mu)(1+k_n\beta), \tag{71}$$

the constant k_n, which determines the convergence, changes with the number of terms taken in the development, eq (58), besides the first constant term. Table 5.3 gives the values of k_n for $n \leq 4$ at $\mu = -1$ and $\cos\alpha_0 = -1$, too. The last two columns, giving the difference between and the ratio of consecutive k_n's, show that the k_n probably would not diverge as n increases.

TABLE 5.3. *k_n of eq (71)*

n	k_n	$k_{n+1}-k_n$	k_{n+1}/k_n
0	0.69	1.89	3.72
1	2.58	2.31	1.89
2	4.89	2.26	1.46
3	7.15	2.06	1.29
4	9.21	----------	----------

From a physical point of view, one would expect that such a development is good only for thin foils. However, the results of table 5.3 seem to show that this development may be useful for thick foils, too. Further numerical work would be necessary, however, to make this conclusion safe. Physically, this conclusion would mean that in back-scattering a large number of small angle deflections is less probable than a small number of large deflections.

The calculation for finite t is a little longer, but there is no essential difficulty. Only some of the integrals involved must be understood in the sense of Cauchy principal value. We have calculated the first and second approximation only for $n=0$ and $\mu=-1$. We get

$$T^{(1)}(0,-1) = F\beta(1-e^{-2t})/8$$

$$T^{(2)}(0,-1) = T^{(1)}(0,-1)\left\{1+\frac{\beta}{1-e^{-2t}}\left[ln2+E_1(2t)-2E_1(t)e^{-t}-(\gamma+lnt)e^{-2t}\right]\right\}, \tag{72}$$

where γ is the Euler constant and $E_1(x)=\int_x^\infty e^{-u}\frac{du}{u}$. For the case we treated in constructing table 5.1, t is of the order 20. Then eq (72) gives practically the same result as the case of infinite t.

Equation (55b) gives another approximation for back-scattering, i. e.,

$$f(0,\mu)=-\frac{f_1(t,\mu)}{\mu}+k_1\int_0^t\left(\frac{1-\mu}{\mu}\frac{\partial f_1}{\partial\mu}-\frac{f_1}{\mu^2}\right)d\tau,\quad \mu<0. \tag{73}$$

The function $f_1(t,\mu)$ in the form of the infinite series eq (30) is not very practical for numerical calculation, since for extremely forward single scattering function one has to take a tremendous number of terms. For the case of small t, several authors [9, 11, 12] have made developments of this function which are more suitable for numerical calculation than the series. However, they all made the small-angle approximation, so their developments are not suitable for $\mu<0$ as required by eq (73). We shall give another development here that approximates better for smaller μ.

The function $f_1f_1(\tau,\ \mu)(\tau,\mu)$ is a solution of eq (47) and (49a). We first develop $f_1(\tau,\mu)$ in a Taylor series of τ, i. e.,

$$f_1(\tau,\mu)=\pi F\delta(\mu-1)+\tau\left(\frac{\partial f_1}{\partial\tau}\right)_{\tau=0}+\frac{\tau^2}{2}\left(\frac{\partial^2 f_1}{\partial\tau^2}\right)_{\tau=0}\ldots, \tag{74}$$

in which we have made use of the boundary condition, eq (49a). Putting eq (74) into eq (47) and equating coefficients of different powers of τ, we get

$$\left(\frac{\partial f_1}{\partial\tau}\right)_{\tau=0}=\frac{F}{4}p(\mu)-\pi F\delta(\mu-1)$$

$$\left(\frac{\partial^2 f_1}{\partial\tau^2}\right)_{\tau=0}=\frac{F}{4}\left\{\frac{1}{2}\int_{-1}^{1}p(\mu,\mu')p(\mu')d\mu'-2p(\mu)\right\}+\pi F\delta(\mu-1), \tag{75}$$

where $p(\mu)$, which is written for $P(\mu,1)$, is just the single scattering function defined by eq (1). Combining eq (74) and (75), we get

$$f_1(t,\mu)=\left(1-t+\frac{t^2}{2}\right)\pi F\delta(\mu-1)$$
$$+\frac{F}{4}\left\{tp(\mu)+\frac{t^2}{2}\left[\frac{1}{2}\int_{-1}^{1}p(\mu,\mu')p(\mu')d\mu'-2p(\mu)\right]\right\}. \tag{76}$$

The first term inside the curly brackets of eq (76) is the well-known single scattering tail for a thin foil. To evaluate the integral, we notice that the integrand is the product of two extremely peaked functions with one peak at $\mu'\doteqdot\mu$ and the other at $\mu'=1$. So we break the integral at $\mu'=(1+\mu)/2$, and develop the slowly varying function in both intervals into Taylor series around the peak of the other function. Using the single scattering function, eq (69), and keeping terms of first power of β, we get, combining with other terms

of eq (76),

$$f_1(t,\mu)=\left(1-t+\frac{t^2}{2}\right)\pi F\delta(\mu-1)+\frac{Ftp(\mu)}{4}$$
$$\times\left\{1+\frac{\beta t}{1-\mu}\left[2(3+\mu)ln1/2\beta-(3+\mu)ln\frac{2}{1-\mu}\right.\right.$$
$$\left.\left.-\frac{27\mu^3-15\mu^2-27\mu+47}{4(1-\mu)^2}\right]\right\}. \tag{77}$$

In calculating eq (77) we have taken the Taylor series up to the term with third derivative. The above result is incorrect if $\beta/(1-\mu)$ is not small. The deviation from single scattering, given by the term with square brackets, reduces to

$$2\beta t(ln1/\beta-ln2-1/2) \tag{78}$$

for $\mu=-1$ in our approximate expression eq (77). For this special value of μ, the integral in eq (76) can be evaluated exactly. We found the exact value for the deviation up to the first power of β to be $2\beta t(\ln 1/\beta-1)$ instead of eq (78).

We can get an expression for back scattering, similar to eq (57) for forward scattering, by iteration. The only difference in calculation is that here we use eq (9b) instead of eq (9a). Thus we get

$$f(0,-\mu)=\frac{F}{4}\sum_{l=0}^{\infty}(-1)^l\omega_l P_l(\mu)\frac{1-e^{-(1+\mu k_l)t/\mu}}{1+\mu k_l},\quad \mu>0. \tag{79}$$

4.3. Final Remark

Making use of the development, eq (76), one can show, in the case of forward scattering, that for small t the solution, eq (56), reduces to

$$f(t,\mu)\cong[1-t(1-k_1)]\pi F\delta(\mu-1)+\frac{Ftp(\mu)}{4\mu}, \tag{80}$$

and the solution, eq (57), reduces to

$$f(t,\mu)\cong(1-t)\pi F\delta(\mu-1)+\frac{Ftp(\mu)}{4\mu}. \tag{81}$$

The two expressions, eq (80) and (81), are approximately the same, since $k_1<<1$. In the case of back-scattering the three kinds of approximations lead to the same expression

$$f(0,\mu)=-\frac{Ftp(\mu)}{4\mu},\quad \mu<0$$

for small t.

5. References

[1.] L. Gratton, Soc. Astron. Ital. **10,** 309 (1937).
[2.] S. Chandrasekhar, Astrophys. J. **99,** 180 (1944).
[3.] N. Bohr, Kgl. Danske Videnskat. Selokat. Mat-Fys. Medd. XVIII, 8 (1948).
[4.] W. Bothe, Z. Physik **54,** 161 (1929).
[5.] W. Bothe, Ann. Physik **6,** 44 (1949).
[6.] H. A. Bethe, M. E. Rose, and L. P. Smith, Proc. Am. Phil. Soc. **78,** 573 (1938).
[7.] E. J. Williams, Proc. Roy. Soc. (London) **169,** 531 (1939).
[8.] S. Goudsmit and J. L. Saunderson, Phys. Rev. **57,** 24 (1940).
[9.] G. Moliere, Z. Naturforsch. **3a,** 78 (1948).
[10.] G. Wentzel, Ann. Physik **69,** 335 (1922).
[11.] H. S. Snyder and W. T. Scott, Phys. Rev. **76,** 220 (1949).
[12.] S. T. Butler, Proc. Phys. Soc. (London) **A63,** 599 (1950).
[13.] T. Teichmann, Multiple Scattering of High Energy Charged Particles in Thin Foils. (Palmer Physical Laboratory), Princeton University, Princeton, N. J.
[14.] H. W. Lewis, Phys. Rev. **78,** 526 (1950).
[15.] J. W. Weymouth, Multiple scattering in a semi-infinite medium, thesis (Ph. D.), (University of Calif., 1951).
[16.] C. N. Yang, Phys. Rev. **84,** 599 (1951).
[17.] M. E. Rose, Phys. Rev. **58,** 90 (1940).
[18.] S. Chandrasekhar, Radiative Transfer, section 48 (Clarendon Press, Oxford, 1950).
[19.] S. Goudsmit and J. L. Saunderson, Phys. Rev. **58,** 36 (1940).

Note added in proof. Among recent papers we mention the useful article, Multiple elastic scattering of charged particles, by E. J. Saletan (Palmer Physical Laboratory, Princeton University, Princeton, N. J., NYO 3990, August 1952), which aims to present the theory in a form useful to experimentalists; H. A. Bethe, Moliere's theory of multiple scattering, Phys. Rev. **89,** 1256 (1953); and L. V. Spencer, Phys. Rev. **90,** 146 (1953).

257899—54——5

6. Multiple Scattering of Electrons and Positrons in a Gas

By Gerhart Groetzinger [1]

Introduction

In measuring the momentum of a particle by means of the average curvature of its track in a magnetic cloud chamber, difficulties arise because of the obscuring effect of multiple scattering. In an investigation of this problem it was found, however, that a careful analysis of the track allows one to obtain not only an estimate of the magnetic curvature, but also the extent of the multiple scattering [1].[2] If the relation between multiple scattering and the momentum is known, measurement of the multiple scattering thus provides an independent estimate of the momentum. In general, the value of the momentum obtained from the mean curvature is better at higher energies, and that obtained from multiple scattering better at lower energies. Equally, the magnetic cloud chamber is a useful tool for the determination of an experimental scattering law. If, for example, the scattering of electrons is to be studied, it is sufficient to introduce a beta-ray emitter into the chamber and thus have a source of electrons in a certain momentum range. This is, however, not the best way to obtain such a law because the determination of the momentum of an individual particle is subject to errors introduced by multiple scattering. These errors will, however, cancel to a certain extent because in a set of determinations the momentum will be in some cases underestimated, in some cases overestimated.

A determination of the experimental scattering law has been carried out for the case of electrons and positrons in a momentum range between approximately 2,000 and 10,000 Gauss-cm in the manner discussed subsequently [1].

In order to obtain an estimate of the momentum and of the multiple scattering, a track as projected onto a plane perpendicular to the magnetic field has to be divided into several sections in the manner indicated in figure 6.1. $A_0, A_1, \ldots, A_n$ and A_{n+1} are dividing points separated by chords of equal length x, which is very close to the length of a section of the track. The n angles between the $(n+1)$ successive chords are designated as $\omega_1, \omega_2, \ldots \omega_n$, while the angles between the tangents at successive dividing points of the track are designated at $\varphi_0, \varphi_1, \ldots, \varphi_n$. The angles φ_i are the ones usually referred to in the theories of multiple scattering. The angles ω_i are introduced here, since they can be obtained from a track with greater accuracy.

It follows from the theory of multiple scattering that the angles ω_i and the angles φ_i obey a normal distribution about the true magnetic

---

[1] Lewis Laboratory, National Advisory Committee for Aeronautics, Cleveland, Ohio.
[2] Figures in brackets indicate the literature references on p. 69.

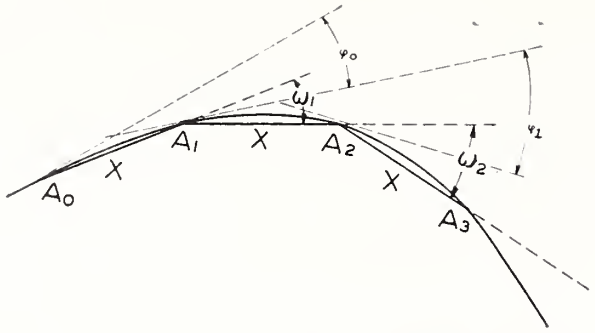


FIGURE 6.1. *Geometrical analysis of a cloud-chamber track used in determining multiple scattering.*

mean deflection $\mu = x/\rho$ with a variance of ψ^2 and σ^2, respectively, where ρ is connected with the momentum p of the particle and the magnetic field H in the chamber by the relation

$$eH\rho = pc. \tag{1}$$

It can be shown that

$$\psi^2 = 2\sigma^2/3. \tag{2}$$

In case the true mean deflection (momentum) is unknown, the best estimate for it, based on a measurement of the angles ω_i of one track (which is done in this investigation) is the sample mean

$$\bar{\omega} = (1/n)\sum_{i=1}^{n} \omega_i. \tag{3}$$

The best estimates for the variance ψ^2 of the ω_i's are the sample variances

$$(\Omega^2)_{Av} = (1/n)\sum_{i=1}^{n} (\omega_i - \mu)^2 \tag{4}$$

for the case in which the true mean is known or

$$(\omega^2)_{Av} = [1/(n-1)]\sum_{i=1}^{n} (\omega_i - \bar{\omega})^2 \tag{5}$$

for the case in which the true mean is unknown. Correspondingly the best estimate for σ^2 is

$$s^2 = 3(\omega^2)_{Av}/2. \tag{6}$$

For a large number of sections (i. e., $n \to \infty$), $\bar{\omega}$ approaches μ, and $(\Omega^2)_{Av}$ and $(\omega^2)_{Av}$ approach ψ^2.

For a normal (multiple scattering) distribution of the quantities ω_i the variable $(n-1)(\omega^2)_{Av}/\psi^2$ is distributed "χ^2" with $(n-1)$ degrees of freedom for particles that are of the same nature, but may have different momenta, provided that the tracks of the particles are divided into the same number of sections of length x.

As in our case the true mean (correct momentum) is unknown, use has to be made of eq (5) for the determination of the extent of multiple scattering. The use of the cloud chamber makes it possible to eliminate the plural and single scattering tail by excluding deflections ω_i in excess of 2.8 times the root mean square angle $[(\omega^2)_{Av}]^{\frac{1}{2}}$. The difference between the variances of this truncated distribution and those of the normal distribution is negligible.

The experimental procedure employed was varied slightly during the course of the investigation, but the basic techniques are given in the following description of the study of electrons from a P^{32} source [3] in the lower energy range (2,000 to 7,000 gauss-cm). In this case, the minimum length of the tracks was 12 cm, the average length 14 cm. They were obtained in a horizontal cloud chamber with a diameter of 24 cm, and an illuminated region 2.5 cm deep, filled with argon at a pressure of 1 atm at 25° C. The source was mounted about 4 cm from the rim of the chamber in the center of the illuminated region. The magnetic field was produced by a set of Helmholtz coils and ranged from 330 to 355 gauss for different exposures.

Table 6.1 gives the minimum path length in 1 atm of argon for which true multiple scattering is said to occur according to Molière's theory [2]. For our experiments we adopted a section length $x=2$ cm, so that the condition of multiple scattering is satisfied even for the electrons of the highest energies. It was this consideration, rather than the difficulty of measuring that puts a lower limit on the length of the sections.

The angles ω_i were measured as follows: The image of the track on the 35-mm photographic negative was projected on a sheet of drawing paper at an enlargement giving the original size of the track, and 2-cm sections were then laid off along the image of the track by dividers. The sheet containing the set of division points was mounted on a drafting table, the points were connected by straight lines representing the chords, and the angles between successive chords were measured by means of a drafting machine.

TABLE 6.1. *Minimum path length in argon (1 atm, 25° C)*

v/c	x_{min} (cm)	v/c	x_{min} (cm)
0.5	0.46	0.90	1.33
.6	.63	.95	1.47
.7	.83	.97	1.54
.8	1.04	1.00	1.61

The experimental error due to optical distortions, emulsion distortion, and inaccuracy of angular measurement was estimated as follows: Circular arcs of radii 5, 10, 15, and 25 cm were drawn on a sheet of drawing paper and photographed in various orientations in a horizontal plane in the position of the sensitive portion of the cloud chamber, using the same optical arrangement employed originally to photograph the tracks. Each circle was then measured independently by three observers, in the same manner as the cloud-chamber tracks. The resulting error, the apparent root mean square angular scattering deflection, was averaged for all observers for each radius, and found to be almost independent of the radius. The average angle was 1.05°.

In a separate investigation concerning the accuracy that can be obtained in measuring angular deflections, the same circular arcs were also measured directly on the films by means of a microscope

[3] The P^{32} source and the Ru^{106} source mentioned later were obtained from the Isotope Branch of the AEC, Oak Ridge, Tenn.

with a mechanical stage, adapted for accurate angular measurements, the magnification used being 51. It was found that this method, while more laborious, will make the experimental error less than 0.3°.

The error due to the turbulence in the chamber is negligible. A few high energy cosmic-ray meson tracks that occurred incidentally in the course of the experiment were investigated. They appeared to be slightly curved. By measuring the deviation from a straight line an upper limit of the spurious curvature due to turbulence can be determined. The deviation was measured by means of a micrometer arrangement at points along the track 1 cm apart, and a circle was fitted to the resulting plot of these deviations. The smallest radius of curvature so obtained was larger than 250 cm, which corresponds to a negligible error not exceeding 0.3° per 2-cm section.

The experimental results are shown in figure 6.2, in which the individual root mean square multiple scattering deflections $s=\left(\frac{3}{2}\right)^{1/2}[(\omega^2)_{Av}]^{1/2}$ are plotted against $H\rho$, each of the 132 points representing one track. Ninety percent of these tracks have 6 to 8 sections, the rest 9 to 13. One hundred points corresponding to apparent momenta above $H\rho=2{,}000$ gauss-cm were fitted to a smooth curve by the method of least squares. The lower limit for the momenta was chosen so as to avoid unduly large errors resulting from our only approximately correct manner of projecting and measuring the tracks. σ^2 can be approximated very well by a second-degree polynomial in $1/(H\rho)^2$. Therefore the least-square curve was obtained by fitting the experimental mean square angles s^2 to such a polynomial.

As mentioned before, the random experimental error due to optics, photography, and personal error in measurement will introduce a spurious root mean square deflection $s_E=1.05°$. On the assumption that the multiple scattering deflections and the "error deflection" s_E are statistically independent, one can set the corrected root mean square deflection

$$s_c=(s_L^2-s_E^2)^{1/2}. \tag{7}$$

This correction was applied to the least-square curve. The result is

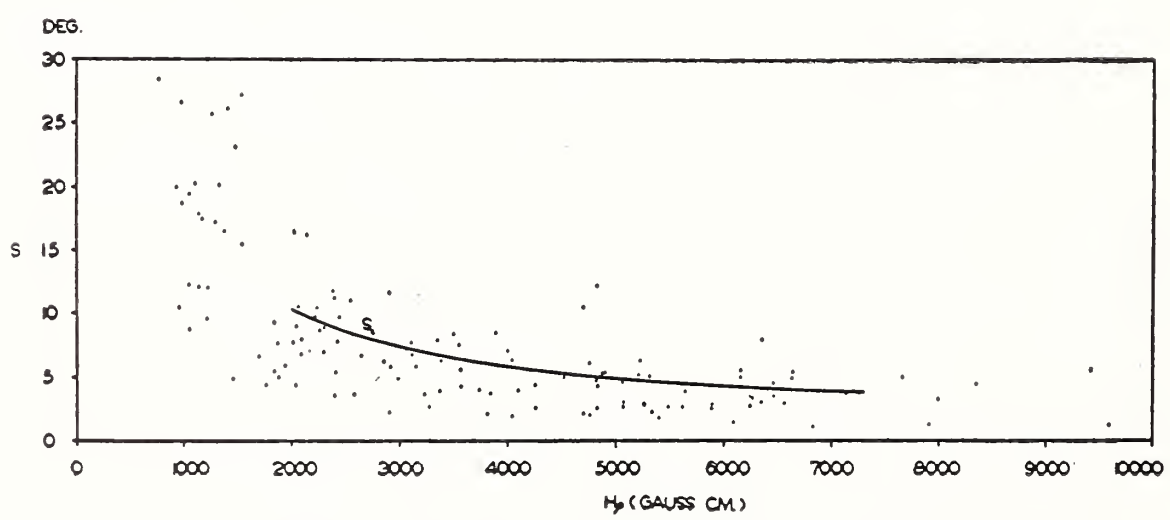


FIGURE 6.2. *Multiple scattering of electrons in 1 atm of argon at 25°C.*

Individual rms scattering deflections s in degrees for 2 cm path length plotted against measured $H\rho$, and corrected curve s_c.

indicated by the solid curve of figure 6.2, which shows s_c as a function of $H\rho$. The curve was broken off at $H\rho$=7,300 gauss-cm, which corresponds to the upper limit of the P^{32} beta-spectrum. Actually there are a few points with an apparently higher $H\rho$ (due to statistical fluctuations) that were included in the fitting of the curve.

Finally, it must be mentioned that sections of track, over which the multiple scattering deflection was more than 2.8 times larger than s for the given momentum, as well as sections with a noticeable single deflection (>0.1 rad.), were excluded from consideration in order to vouchsafe the Gaussian form of the angular scattering distribution. In fact, this exclusion came about almost automatically, since such large deflections were quite rare, and, moreover, tended to make the track ill-defined and difficult, if not impossible, to measure at the position where they occurred.

The dispersion of the experimental points about the "least squares" curve s_L is indicated in figure 6.3. To each experimental point, with a certain value of apparent $H\rho$ and a root mean square scattering angle, s, there corresponds for the same $H\rho$ a point s_L on the "least squares" curve. For every number $k \geq 0$, there is a certain fraction of points such that $s \geq k s_L$. This fraction is plotted for the 100 points above $H\rho$=2,000 gauss-cm, as a function of k (curve A). Theoretically this fraction is given by the integral of the χ^2 distribution of $(n-1)$ degrees of freedom from $(n-1)k^2$ to ∞, where n is the number of angles ω_i per track. For the tracks considered here this number was between five and six. Therefore the integrated χ^2-distributions of four and five degrees of freedom are also plotted in figure 6.3 as functions of k (curves B and C). It is seen that for most values of k, the experimental curve falls well between the two theoretical

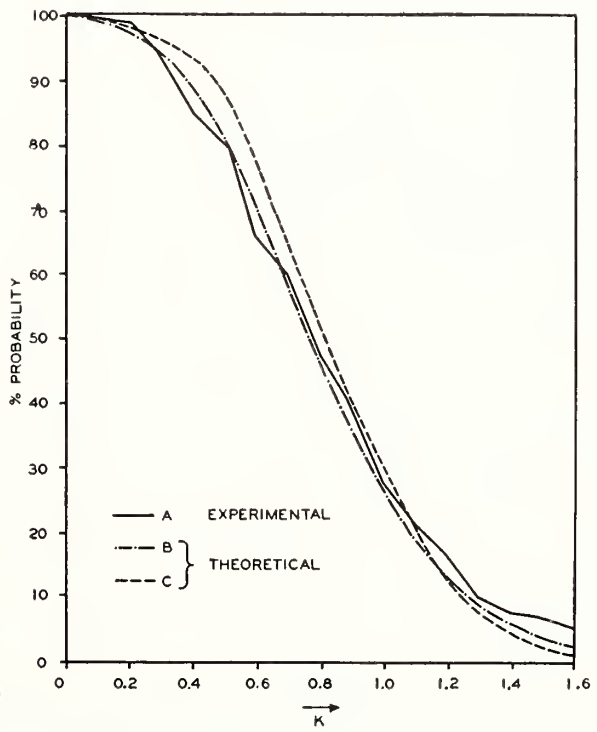


FIGURE 6.3. *Statistical fluctuations of multiple scattering.*

Curve A (experimental), fraction of tracks from figure 6.2 with $s \geq k s_{L.SQ}$ plotted against k; curves B and C (theoretical), integrals of x^2 distributions of n–1 degrees of freedom, respectively, between the limits of $(n-1)\ k^2$ and ∞, plotted against k (with n=6, and n=5, respectively).

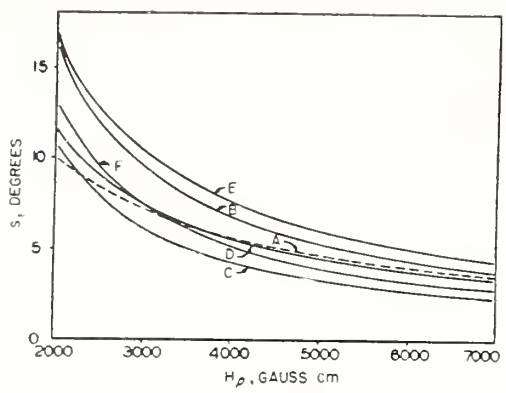


FIGURE 6.4. *Comparison of experimental root-mean-square angle in 1 atm of argon with that following from various theories of multiple scattering.*

Curve A, experimental; curve B, Bothe; curve C, Williams ($\phi_{max}=\left(\frac{1}{2}Q\right)^{\frac{1}{2}}$, $\phi_{min}=(mc/p)\cdot(Z^{\frac{1}{3}}/137)$); curve D, Williams-Bethe ($\theta_{max}=0.1$ radian, $\theta_{min}=(mc/p)\cdot(Z^{\frac{1}{3}}/181)$; curve E, Goudsmit and Saunderson (Thomas-Fermi potential); curve F, Molière.

curves, which confirms that

$$[(n-1)s^2/s_L{}^2] \tag{8}$$

is distributed χ^2, and indirectly that the angular scattering is distributed normally. Only for values of $k\gtrsim 1.3$ is the experimental frequency somewhat too high, which may indicate the presence of a small single-scattering tail.

Figure 6.4 shows the experimental scattering law together with the results of various theories of multiple scattering. In all theories the variance σ^2 (mean-square angular deflection) can be represented as a product of two factors Q and G with

$$Q=\frac{4\pi N x e^4 Z^2}{p^2 v^2} \tag{9}$$

being the same for all theories. Bothe [3] replaces G by the constant value of 4.25,[4] which he derived experimentally (curve B). Curve C shows the results of Williams [4] theory with $G=\log(\varphi_{max}/\varphi_{min})=\log\frac{\sqrt{Q/2}}{\alpha}$ with α being $mcZ^{1/3}/137\ p$. Curve D represents Bethe's [5] modification of William's theory in which he puts for θ_{min}, 0.1 radian and for θ_{min}, $0.757\ \alpha$ (the angles θ being spatial scattering angles). Curve E represents the results of the theory of Goudsmit and Saunderson [6] with G being log $(150/137\alpha)$ and finally curve F displays the results of the theory of Molière [2] with

$$e^G/(G)^{1/2}=1.31Q^{1/2}\theta_{min} \text{ with } \theta_{min}=1.142\alpha\left[1.13+\left(\frac{3.76Z^2}{137^2\beta^2}\right)\right]^{\frac{1}{2}}.$$

The theory of Snyder and Scott [7] gives in this energy range similar results as Molière's. The probable error in the experimental curve A is estimated to be 10 percent.

In order to obtain the experimental scattering law for electrons of higher momentum a pressure of 2 atm of argon and a magnetic field of approximately 680 gauss were used. The 54 electrons studied under these conditions resulted from the decay of 30 sec Rh^{106} (maximum beta-ray energy 3.55 Mev), which in turn is a decay product of Ru^{106}. In these and subsequent experiments, the correction for the experimental errors was done in a somewhat different manner than before. The individual scattering variances were first corrected before they

[4] According to more recent experiments this value has to be replaced by a value of 2.75 (private communication).

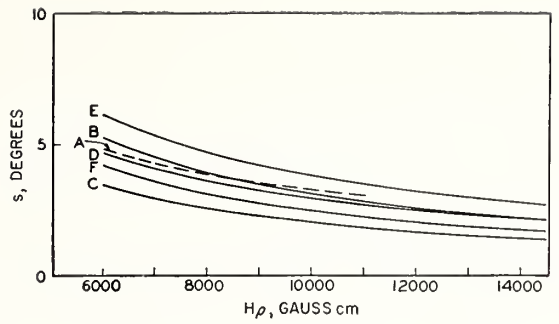


FIGURE 6.5. *Comparison of experimental root mean-square angle in 2 atm of argon with that following from theories of multiple scattering.*

Curve A, experimental; curve B, Bothe; curve c, Williams ($\phi_{max}=\left(\frac{1}{2}Q\right)^{\frac{1}{2}}$, $\phi_{min}=(mc/p)\cdot(Z^{\frac{1}{3}}/137)$); curve D, Williams-Bethe ($\theta_{max}=0.1$ radian, $\theta_{min}=(mc/p)\cdot(Z^{\frac{1}{3}}/181)$; curve E, Goudsmit and Saunderson (Thomas-Fermi potential); curve F, Molière.

were fitted to the curve s_L versus $H\rho$ instead of the resultant curves being corrected as was done previously. Figure 6.5 shows the experimental scattering law (curve A), and the results of the theories discussed before. Again the theories seem to overestimate the energy dependence of multiple scattering.

Scattering of Positrons

A comparison of the multiple scattering of electrons and positrons is of interest in view of two recent investigations in which new differences between the particles are reported. First: Du Mond [8] and Hedgran and Lind [9] have reported results indicating that the mass of the positron is smaller than that of the electron by about one part in ten thousand. Second: Seliger [10] observed that the ratio of the back-scattering coefficients β^- and β^+ for electrons and positrons resulting from beta-decay with a maximum energy of about 0.5 Mev is approximately 1.3 in several media of atomic number between 4 and 82. It follows from theoretical considerations that large angle single scattering for relativistic particles is more pronounced for electrons than for positrons [11]. One might expect that this effect would account for the observed difference in the back-scattering but will hardly affect the multiple scattering since tracks displaying large angle deflections are eliminated from considerations. However, W. Miller [12] remarks that back-scattering also results mainly from the "cumulative effects of small angle single scatterings" but obtains nevertheless a theoretical ratio β^- to β^+ of 1.16. Consequently, it is possible that a similar difference might occur also for the case of multiple scattering.

Positrons of momenta between 2,000 and 6,000 gauss-cm were investigated in 1 atm of argon with an applied magnetic field of approximately 340 gauss, varying somewhat from exposure to exposure. For the investigation of the scattering of positrons of momenta between 5,000 and 9,000 gauss-cm, a magnetic field of approximately 680 gauss, varying somewhat for different exposures, and a pressure of 2 atm were used. A source was prepared by bombarding iron with deuterons in the University of Chicago 33-inch cyclotron. The main positron emitter obtained in this way is the 21 min isomer of Mn^{52} with a maximum energy of 2.66 Mev [13]. The bombarded piece of iron was put into a cylindrical tube extending four centimeters into the chamber and closed inside the chamber by a thin aluminum foil. The source was then used for one to two hours after a bombardment.

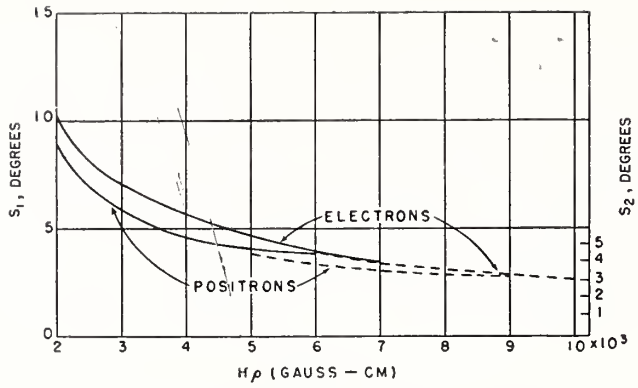


FIGURE 6.6. *Root mean-square angle of multiple scattering for electrons and positrons as a function of $H\rho$.*

The solid curves, in connection with scales s_1, refer to electrons and positrons in 1 atm of argon, and the dashed curves in connection with scale s_2, refer to electrons and positrons in 2 atm of argon. The scales are matched on the basis of theoretical pressure dependence of multiple scattering.

The upper solid curve in figure 6.6 represents in connection with scale s_1 the experimental mean square angle of scattering for electrons of momenta between 2,000 and 7,000 gauss-cm in 1 atm of argon and the upper dashed curve in connection with scale s_2 the experimental mean square angle of scattering for electrons of momenta between 6,000 and 10,200 gauss-cm in 2 atm of argon. In connection with scale s_1 the dashed curve represents the root mean square angle of scattering to be expected on the basis of theoretical considerations in 1 atm of argon. Strictly speaking, the two scales s_1 and s_2 should be matched by taking into account the pressure dependence of both factors Q and G whose product is equal to the mean square scattering angle σ^2. However, since the pressure dependence of G varies from theory to theory and is at most slight, the matching was done by taking into account only the pressure dependence of Q, which is the same for all theories. The two lower curves represent the results obtained from an analysis of tracks of 115 positrons with momenta between 2,000 and 6,000 gauss-cm obtained in 1 atm of argon and 52 tracks of positrons of momenta between 5,000 and 9,000 gauss-cm obtained in 2 atm of argon. The matching of the high and low energy curves in the overlapping momentum range provides an estimate of the errors involved in the method of studying multiple scattering. Arguments based on the matching of these curves necessarily neglect possible systematic errors. Such errors are, however, of little importance in the comparison of electron and positron scattering and the curves obtained for these two cases are sufficiently different from each other to indicate that the rms angle of multiple scattering of electrons is greater by an amount of the order of 10 percent than that of positrons of the same momentum.

The nature of back-scattering is sufficiently different from multiple scattering that it is difficult to make a quantitative comparison of the present results with those of Seliger. It should be noted, however, that the effect is in the same direction and of the same order of magnitude.

This work was done in conjunction with M. J. Berger, W. Humphrey Jr., L. Leder and F. L. Ribe.

References

[1] G. Groetzinger, M. J. Berger, and F. L. Ribe, Phys. Rev. **77,** 584 (1950); G. Groetzinger, W. Humphrey, and F. L. Ribe Phys. Rev. **85,** 78 (1952).
[2] G. Molière, Z. Naturforsch, **3a,** 78 (1948).
[3] W. Bothe, Handbuch der Physik, vol. 22, II, p. 1 (Verlag Julius Springer, Berlin, 1933).
[4] E. J. Williams, Proc. Roy. Soc. **169,** 531 (1939), Phys. Rev. **58,** 292 (1940).
[5] H. A. Bethe, Phys. Rev. **70,** 821 (1946).
[6] S. Goudsmit and J. L. Saunderson, Phys. Rev. **57,** 24 (1940); Phys. Rev. **58,** 36 (1940).
[7] H. S. Synder and W. T. Scott, Phys. Rev. **76,** 220 (1949).
[8] J. W. M. Du Mond, Phys. Rev. **81,** 468 (1951).
[9] A. Hedgran and D. A. Lind, Phys. Rev. **82,** 126 (1951).
[10] H. H. Seliger, Phys. Rev. **78,** 491 (1950).
[11] See, e. g., N. F. Mott and H. S. W. Massey, The theory of atomic collisions, p. 81 (Oxford University Press, New York, N. Y., 1945).
[12] W. Miller, Phys. Rev. **82,** 452 (1951).
[13] K. Way, L. Fano, R. Scott, and K. Thew, Nuclear data, NBS Circular 499 (1950).

Discussion

Dr. W. T. Scott, Smith College, Northampton, Mass.: I would like to ask whether you consider the root mean square angle a good measure of the multiple scattering in view of the fact that it is strongly affected by statistical errors resulting from the relatively small number of large deflections. Would it not be better to use mean absolute angles?

Dr. Groetzinger: Sections that contained large single deflections as well as those in which the multiple scattering deflection exceeded the root mean square angle by more than a factor of 2.8 were excluded. This made it possible to use the root square mean angle without encountering the difficulties mentioned by you. Our primary purpose, of course, was to cut off the single scattering tail of the distribution since we were interested only in the part connected with multiple scattering, which is Gaussian.

Dr. Scott: Does not the fact that you use a truncated rather than a complete Gaussian distribution affect the root mean square angle for multiple scattering?

Dr. Groetzinger: Very little. The Gaussian is cut off sufficiently far out that the root mean square angle differs only slightly from that of the whole distribution.

7. Scattering of Electrons (Experimental)

By W. Bothe [1]

Nuclear Single Scattering of Negatrons

This is meant to be a general, though brief, survey of the present experimental situation in the fields of electron and positron scattering by nuclei and electrons, single as well as multiple scattering, with a light emphasis on recent work that is not yet published or otherwise less known in this country.

From a theoretical point of view, single scattering deserves the higher interest. It is a fact that, though work in this field started some 30 years ago, the interest in single scattering of electrons was, later on, overshadowed by problems of scattering of nucleons. This has experimental as well as theoretical reasons. It is rather difficult experimentally to realize conditions under which pure single scattering is guaranteed, the mass of the electron being so small that plural or multiple scattering and energy losses are appreciable even in very thin layers of matter, except for very high energies. On the other hand the theoretical formulae for scattering of electrons are rather complicated and awkward to handle.

Coulomb scattering of electrons by nuclei, which shall be considered first, is described by the well-known Mott formula, which is a series of generally very poor convergence. First approximations in αZ of the Mott formula have been derived but turned out to be useless except for the lightest nuclei. As late as 1948 McKinley and Feshbach [1] [2] undertook the tedious work of computing fourth order expressions for the scattering cross sections of higher nuclei for electrons. These, combined with straightforward calculations which Bartlett and Watson [2] had made a few years before for the Hg nucleus, offer a sound basis for comparison with experiment. It must be kept in mind that all these theoretical predictions are based on the Dirac theory of the electron, radiation effects being neglected, which ought to be allowed according to prevailing theoretical views of a few years ago. Therefore, experiments on single scattering of electrons may be considered as a touchstone for the adequacy of the Dirac theory, and their fundamental importance cannot easily be overestimated.

There is no point in discussing all the earlier experimental work, which has been done chiefly by the cloud chamber method, because the results were very contradictory and confusing. Repeatedly, deviations both sides of the Rutherford or Mott formulae have been claimed; even discrepancies in order of magnitude have been reported. There is no doubt that in nearly all of this work the experimental difficulties and errors, especially those inherent in the cloud chamber

[1] Institut für Physik im Max-Planek-Institut für med. Forsehung and I. Physikalisehes Institut der Universitat, Heidelberg.

[2] Figures in braekets indicate the literature references on p. 81.

method, have been greatly underestimated. To give only one example: Several cloud chamber experiments, including some performed in our laboratory, seemed to indicate that nuclear large angle scattering is connected with energy losses far greater than can be accounted for by ordinary bremsstrahlung. This point was checked 2 years ago by analyzing scattered electrons by means of a magnetic spectrograph [3]. Figure 7.1 shows the experimental arrangement: A specially constructed monochromater, M, selecting a fairly monochromatic collimated beam out of the β-radiation of a RaE source, P, the scattering foil, F, and the spectrograph, Sp, with a G. M. counter for analyzing the electrons scattered at about 110°. Figure 7.2 shows the results for different thicknesses of the Al-foil, and for different electron energies. It turns out that, with foils not very thin, the electron line is broadened on the low energy side to a surprisingly large extent, but no energy loss is observed if sufficiently thin foils are used. Obviously, the influence of the general slowing down of electrons in passing through the gas of a cloud chamber has frequently been underestimated.

Now a very brief account will be given of the few experimental investigations of the last years that seem to have yielded reliable information on single nuclear scattering. The careful work that Van

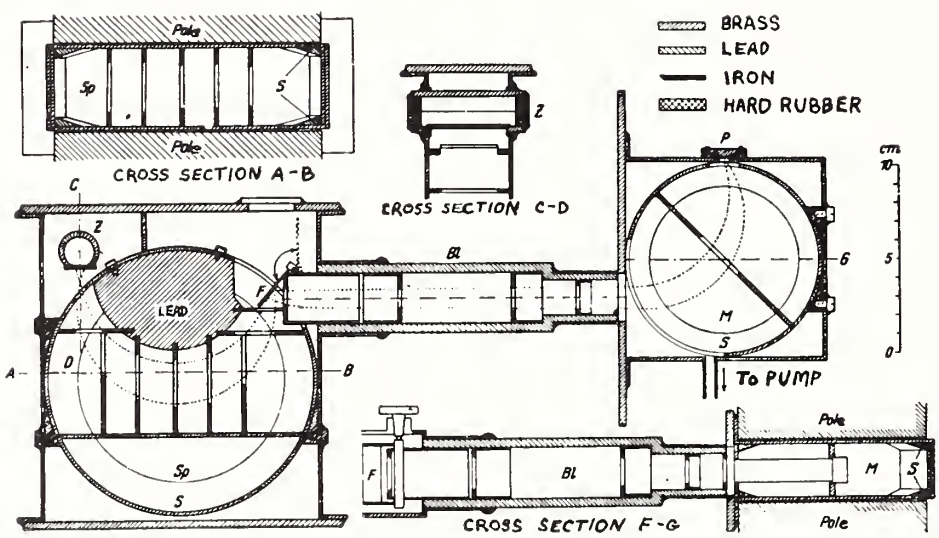


FIGURE 7.1. *Arrangement for analyzing singly scattered electrons.*

P, RaE-source; M, monochromator; F, scattering foil; Sp, analyzer; Z, G. M. counter.

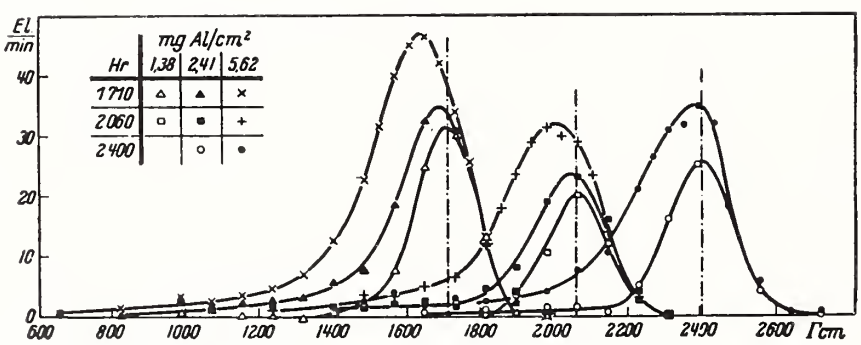


FIGURE 7.2. *Momentum distribution of electrons scattered from thin Al-foils.*

Intensities are reduced to thickness of the thinnest foil.

de Graaff, Buechner and Feshbach [4] did in 1946–47, which is well known in this country, covered the range of energies from 1.3 to 2.3 Mev and the angular range from 20° to 50° (in part 60°). It was the first time one of the modern high voltage plants had been utilized for this kind of work. Monoenergetic electrons were produced by a belt generator and allowed to enter a cylindrical scattering chamber which, at its center, contained the scattering foil. Scattered electrons were measured with a small ionization chamber whose window was covered with filters of sufficient thickness to absorb all stray radiation. These filters have proved, in our own experiments, to be essential. Without them the results were severely distorted by stray radiation. The results confirmed the Mott theory within some 10 percent.

The experiments that Schulze-Pillot [5] performed at Heidelberg covered about the same range of energies (1 to 2.4 MeV) but went down to scattering angles as small as 12°. The arrangement (fig. 7.3) had cylindrical symmetry. It consisted of a magnetic lens monochromator, R_1, furnishing a hollow cone-shaped beam, and a scattering chamber R_2 with fixed angular range. Primary and scattered electrons were counted by a point counter covered with suitable filters. Again agreement with the Mott theory within 10 percent was stated.

So far the situation seems quite clear. But the picture changes as one passes to larger scattering angles. The first indication that with a scattering angle of about 110° the Mott formula is no longer adequate was obtained, as a byproduct, with the arrangement shown in figure 7.1 [6]. The dependance of single scattering on the atomic number of the scatterer was investigated for energies of a few 100 kev. The accuracy was poor due to lack of intensity, but it could be clearly shown that, at this angle, the Z-dependence is less pronounced than predicted by theory. Much better work is now being done by Kinzinger [7] in our laboratory, using an arrangement similar to that used by Van de Graaff and his coworkers (fig. 7.4). The magnetic analysis of the scattered electrons has proved to be unnecessary, so they are directly counted by G. M. counters arranged around the cylindrical scattering chamber. Figures 7.5–7 show some of the results so far obtained. Here the observed ratio of cross section of the atom Z to that of Al is plotted versus the scattering angle (Z=28, 47 and 79). The theoretical curves are also drawn. The electron energy was 245 kev. In each case there is perfect agreement with theory for angles up to about 60°. This is just as far as the experiments of Van de Graaff and coworkers went, so their results for energies of more

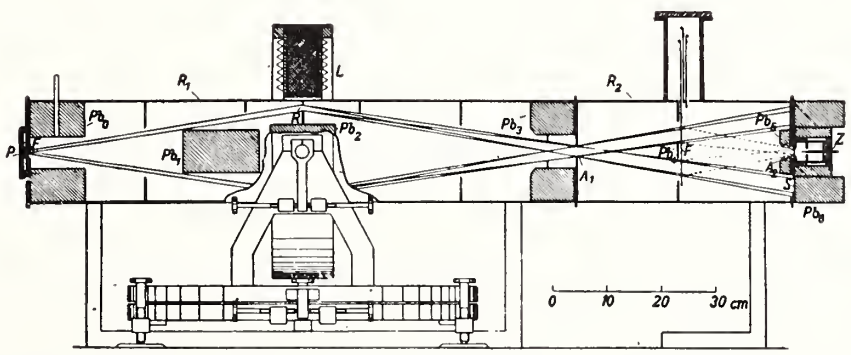


FIGURE 7.3. *Arrangement for single scattering at small angles.*

P, Rn-source; L, monochromator lens; F, scattering foil· Z, point counter.

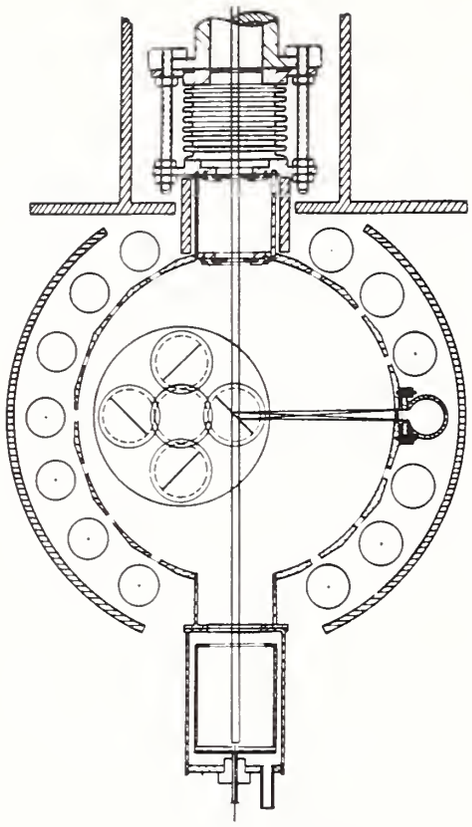


FIGURE 7.4. *Arrangement for single scattering at large angles.*

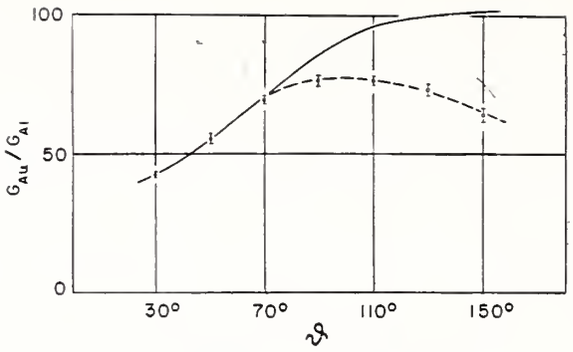


FIGURE 7.5. *Ratio of differential scattering cross sections of Au to that of Al versus scattering angle.*

The electron energy is 245 kev. Full curve is theoretical.

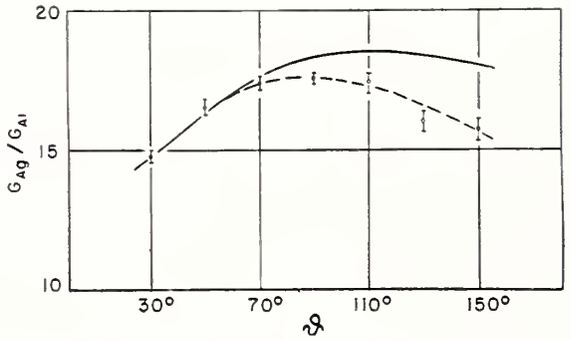
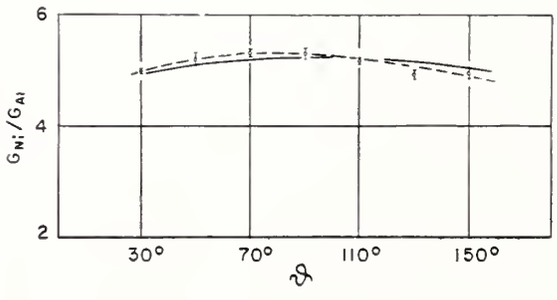


FIGURE 7.6. *Ratios of differential scattering cross sections of Au, Ag, and Ni to those of Al versus scattering angle.*

The electron energy is 245 kev. Full curve is theoretical.

FIGURE 7.7. *Ratio of differential scattering cross sections of Au to that of Al versus scattering angle.*

The electron energy is 245 kev. Full curve is theoretical.



than 1 Mev still hold good in the 100 kev region. But for larger scattering angles, the experimental ratio σ_Z/σ_{Al} more and more lags behind its theoretical value, the more so the higher Z. This is most clearly shown in figure 7.8, which gives the ratio $\sigma/\sigma_{Rutherf.}$ for two fixed angles, 90° and 150° as a function of Z; for Al the theoretical value of this ratio has been assumed. In fact, figure 7.9 shows that for Al the angular distribution of scattered electrons is exactly that predicted by theory, so there is little doubt left that for Al the Mott theory holds good even with energies as low, and scattering angles as large as used here (the absolute cross-section measurements of Kinzinger are not yet finished). In any case it turns out that for the higher atoms the angular distribution is steeper than predicted by

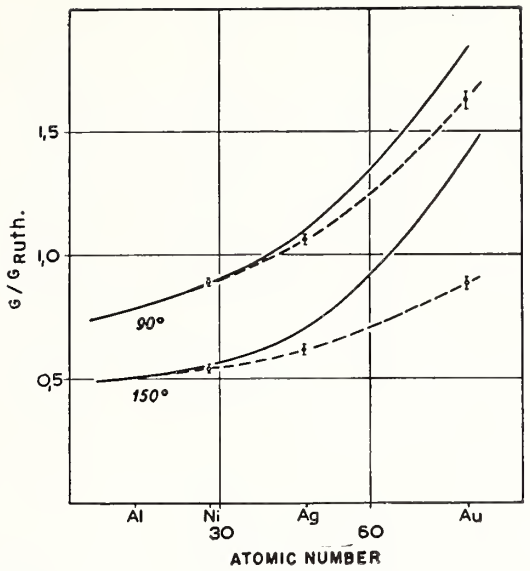


FIGURE 7.8. *Ratios of differential scattering cross sections to Rutherford cross sections for scattering angles 90° and 150° versus atomic number.*

The electron energy is 245 kev. The full curves are theoretical.

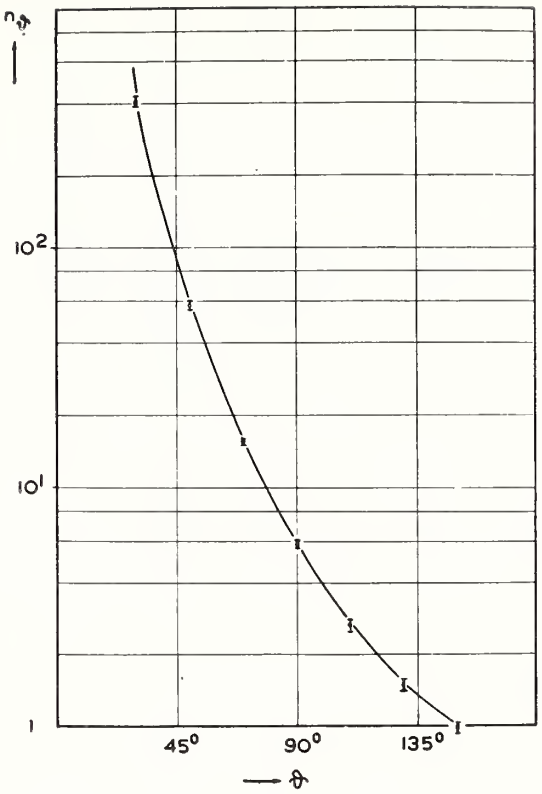


FIGURE 7.9. *Angular distribution of singly scattered 245 kev-electrons from Al.*

The curve is theoretical, adapted at 90°.

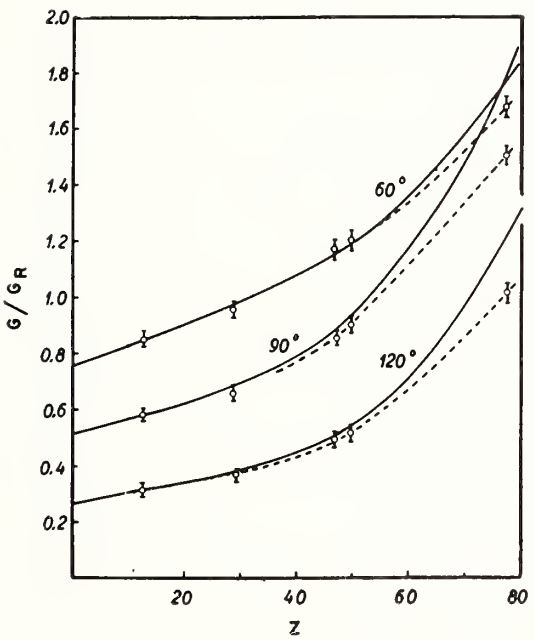


FIGURE 7.10. *Ratios of differential cross sections to Rutherford cross sections for scattering angles 60°, 90°, and 120° versus atomic number.*

The electron energy is about 2 Mev. The full curves are theoretical (H. Reich).

theory. Practically the same picture has been obtained with electrons of about 400 kev.

Similar results (fig. 7.10) have been obtained, quite independently, by Paul and Reich [8] in Goettingen with electrons of about 2 Mev produced by a small betatron.[3]

The total outcome of this group of experiments can be summarized as follows: For energies from 0.2 to 2.5 Mev and for Al the Mott formula is confirmed with an accuracy of about 10 percent or better over the whole range of scattering angles. With heavier elements, the same holds for angles up to about 50°; but for larger angles the observed cross section is smaller than given by the Mott formula, the

[3] The authors were kind enough to let me see their results before publication.

deviation increasing with scattering angle and atomic number. The energy dependance of this effect seems not to be very pronounced. These deviations deserve some interest. Without going into any detail it can be said that it is doubtful whether they can be explained by the well known complications, such as screening effects, crystal interferences, the magnetic moment of the nucleus, and ordinary bremsstrahlung. With regard to screening and interference effects from the electronic shell the theoretical situation seems not quite clear. Bartlett and Welton [9] as well as Mohr [10] have made calculations indicating considerable effects of this kind even for large scattering angles. But whereas Bartlett and Welton seem to have some doubt whether these calculations are reliable in the large angle region, the results are taken seriously by Mohr. In any event the calculations made by Mohr, when applied to our case, would point to the wrong direction. Another point that may be worth mentioning is that exchange of the beam electron with one of the atomic electrons has been neglected so far in screening theories. This does not seem to be serious since this exchange effect turns out to be negligible as long as the beam energy is larger than the binding energies of all the target electrons, which is the case in all the experiments mentioned before. So the possibility must be kept in mind that the failure of the Mott formula might have some deeper cause. Of course, one will be very reluctant to conclude that something is wrong with the Dirac theory, but a solution might possibly be found in the following direction. The Schwinger radiation correction [11] has so far only been calculated to the first approximation in α, and so far would be small in our case, but McKinley and Feshbach [1] have shown that, in developing the Mott formula into a series in α, even the 4th order approximation is not quite sufficient for the highest nuclei. In the same way the Schwinger correction, when extended to higher orders, might become considerably larger. In any event this point ought to be cleared up before farther reaching conclusions are drawn from the experiments described here.

Quite a different point comes up as one moves on to electron energies of the order of 10 Mev or more. When the wavelength of the beam becomes comparable with nuclear dimensions, nuclear interference effects are to be expected, similar to the Ramsauer effect with slow electrons when their wavelength is comparable with the dimensions of the atom. This is the interpretation given by Lyman, Hansen and Scott [12] to their very interesting recent experiments on single scattering of 16 Mev-electrons ($\lambda/2\pi = 1.2 \times 10^{-12}$cm) on the basis of calculations made by Rose, Elton, Parzen, and Acheson [13]. In principle, it must be possible in this way to obtain direct information on the charge distribution within the nucleus. A detailed report on this question will come from a more competent side in the course of this conference. Let me only mention two points. Calculations on this effect are necessarily based on the assumption that for scattering by an atom with *point* nucleus the Mott formula is strictly valid. But the single scattering experiments in the 1 Mev region described before have thrown some doubt on this assumption. Apart herefrom it seems that one has always to take into account the possibility that in experiments in this energy region the "smallest length" might begin to play a part, so that the basis of the Mott theory will become somewhat shaky.

Multiple Scattering of Negatrons

Multiple scattering in itself is not of very fundamental interest, but has gained considerable practical importance as a means for determining particle masses from cloud chamber or photographic tracks. In the last 30 years many formulae, of increasing intricacy, have been derived for multiple scattering (Bothe, Williams, Goudsmit and Saunderson, Molière, Snyder and Scott, Blanchard and Fano [14] and others). Numerous experiments have been performed in order to check these formulae, either using thin foils, or by evaluating cloud chamber tracks. The latter procedure has the advantage that large single deflections (not included in multiple scattering theory) are easily eliminated. As an example of this kind of experiment the careful investigation of Groetzinger, Berger, and Ribe [15] may be mentioned. Best fit was stated with the Molière and the Snyder-Scott formulae. Multiple scattering by thin foils has been studied by Kulchitsky and Latyshev [16] with 2.2 Mev electrons. Both the Williams, and the Goudsmit-Saunderson formulae fit pretty well, except for the highest elements, where the theoretical values of the width of angular distribution are high by about 13 percent. Similar experiments with electrons of a few 100 kev done in our laboratory by O. Knecht [17] yielded similar results (fig. 7.11). Here foils of approximately equal number of atoms per cm^2 have been compared, whereas Kulchitsky and Latyshev have compared foils of equal scattering power. Again the width of angular distribution as calculated after the Williams formula is too large. This can, at least partly, be explained by the fact that the Williams formula applies to the *projected* angles of deflection, whereas the experiments of Kulchitsky and Latyshev, and of Knecht deal with the *spacial* distribution of deflections. The Molière theory of spacial distribution fits the results of Knecht perfectly (fig. 7.11).

The derivation of reliable multiple scattering formulae is rather complicated, and becomes still more complicated when nonsymmetrical experimental arrangements are considered. Asymmetric problems arise in connection with single scattering experiments, where corrections for the contribution of multiple scattering are nearly unavoidable. Therefore, Knecht has also made some measurements on multiple scattering in oblique foils. An example of his results is given in figure 7.12, where the "isophots" are drawn for a 730 kev

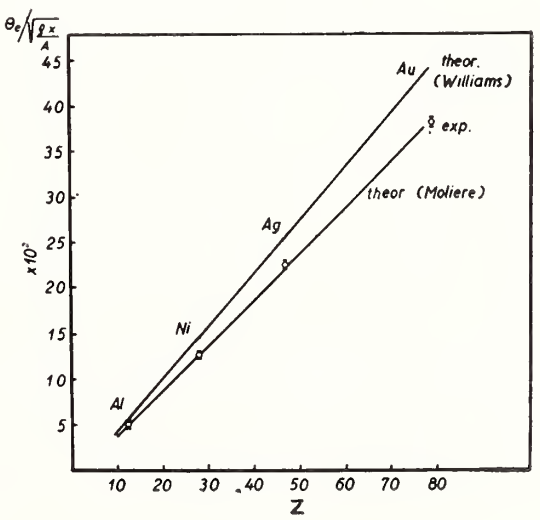


FIGURE 7.11. *e-width of multiple scattering of 830 kev-electrons for 1 g-atom/cm² versus atomic number.*

Measurements were made with $g/A \approx 1.7 \cdot 10^{-4}$ g-cm^{-2}.

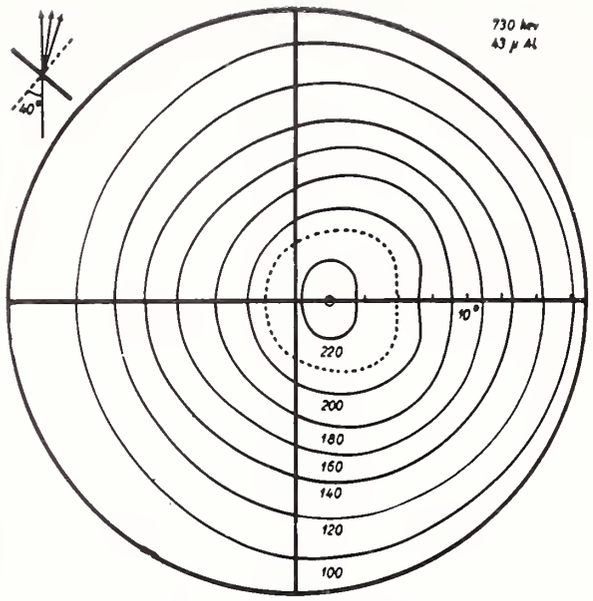


FIGURE 7.12. *Angular distribution of 730 kev-electrons after passing through a 43 μ Al foil inclined by 40°.*

The curves represent cones of constant intensity.

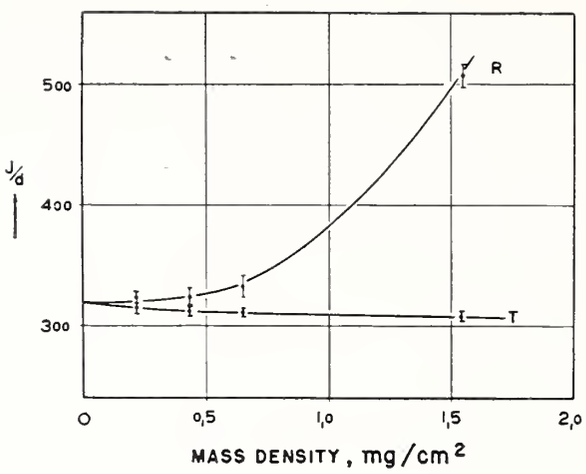


FIGURE 7.13. *Intensity per unit mass of the foil, of 300 kev-electrons scattered at 90° from a Au foil inclined by 45°, versus mass of the foil.*

R reflection position, T transmission position of the foil.

electron beam passing through a 43 μ Al foil inclined at 40°. The intensity maximum is shifted in the direction away from the foil. This shift, ψ, can be calculated under simplifying assumptions; the calculated value is $\psi = \lambda^2 tg\alpha - \frac{7}{4}\lambda^4 tg^3\alpha \ldots$ (λ=most probable multiple deflection, α=inclination of the foil). This is in satisfactory agreement with the experimental results. Calculations of this kind may serve as a starting point for generalizing the correction formula given by Chase and Cox [18] for the multiple scattering contributions in single scattering experiments. Calculations along this line, though far from being satisfactory in every respect, have yielded two interesting results. The first one is that the Chase-Fox formula should hold good for scattering angles as large as 90°, provided the "symmetrical transmission position" of the foil is chosen. On the other hand, if the foil is in the "symmetrical reflection position," the calculated contribution of multiple scattering becomes much larger. This is a qualitative explanation of the "reflection-transmission-effect" first observed by Chase and Cox. Reich [19] in Goettingen and Kinzinger [7] in our laboratory have studied this effect in some more detail. Figure 7.13 gives an example of Kinzinger's results. It seems that in the transmission position the scattered intensity per unit foil thickness even *decreases* with increasing thickness, the effect of multiple scattering being negative. This strange fact has so far not found an explanation.

Another form of nonsymmetrical multiple scattering, the "back-scattering" from thick layers, is of practical interest, too. Figure 7.14 shows the energy distribution curves of back-scattered 370 kev electrons taken with the arrangement of figure 7.1 [20]. It can be seen that in the region of large energy losses the nature of the back-scattering substance is of little consequence, whereas electrons with nearly the full primary energy are much more abundant from heavy

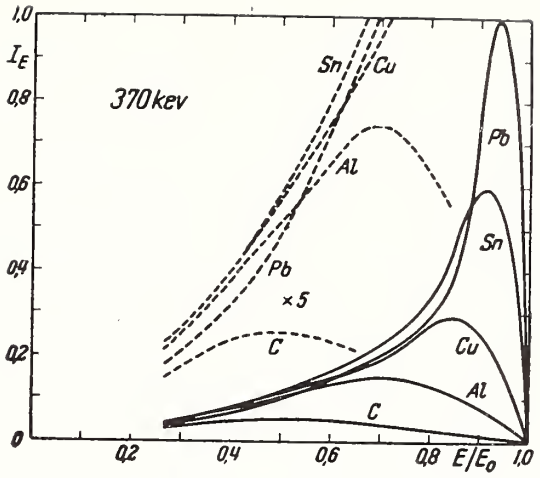


FIGURE 7.14. *Energy distribution of back-scattered 370 kev electrons.*

The angle of incidence is about 28°, the angle of emergence is about 50° in the plane of incidence.

than from light elements. This clearly demonstrates the importance, in all scattering work, of lining chambers with low atomic material, and of proper filtering. The theory of back-scattering is still in a rudimentary state, though a few approaches have been made [21].

Nuclear Scattering of Positrons

In the classical (Rutherford) approximation, negative and positive electrons would be expected to behave exactly the same way. But effects of relativity and spin ought to reduce considerably the scattering cross sections of heavy nuclei for positrons as compared with negatrons [22]. Qualitative experimental evidence of this effect has been given by Lipkin and White [23] who, by a direct comparison of negatron and positron scattering, have observed a ratio of cross sections of more than 3 for 1-Mev particles in platinum at 58°.

In multiple scattering such large differences cannot be expected, because here one has mostly to deal with great numbers of small deflections, for which the difference in scattering cross sections is not so pronounced according to theory. It is, therefore, a little surprising at first sight that Seliger [24] has observed marked differences, up to 30 to 40 percent, in the back-scattering coefficients of positrons and negatrons. Rough calculations made by Miller [25] on the basis of a very crude theory of back-scattering lead at least to the observed order of magnitude. I think we will hear more about this interesting subject from Dr. Seliger.

Scattering by Electrons

The most complete theory of negatron-negatron collisions, though necessarily in the first approximation only, has been given by Møller [26]. Experiments on this subject are still very scarce. Groetzinger, Leder, Ribe and Berger [27], in a cloud chamber investigation, have stated agreement with the Møller formula, though other formulae are not excluded. Here, again, counter experiments appear more promising. Page [28] has used two counters in coincidence, each of them receiving one of the two fast electrons present after collision. His results fit the Møller theory within 10 percent, which proves that the spin terms are essential. On the other hand a surprising discrepancy has been reported by Deutschmann [29] in connection with cloud chamber experiments on cosmic ray showers, and on slow sec-

ondaries (δ-rays) produced by cosmic ray particles. The differential energy distribution of slow secondaries ought to follow a E^{-2}-law, since the Rutherford approximation is quite sufficient here. This has been confirmed by Deutschmann for single cosmic ray particles, presumably mesons, but for shower particles, which are mostly electrons, the exponent was found to be 2.8 instead of 2 (figure 7.15). No explanation can so far be given.

Collisions between a fast positron and a negatron at rest offer some points of special interest. The first point is that after the collision the two particles can be identified by means of a magnetic field, therefore head-on collisions can be distinguished from grazing collisions, which in the case of two equal particles is impossible. It is obvious that head-on collisions, though much less frequent, are most interesting, because they are most sensitive to changes in theoretical assumptions. The second point is that positron-negatron collisions offer a check on the Dirac hole theory of the positron. As first shown by Bhabha [30] on the basis of the Møller theory, Diracs conception requires, apart from direct scattering, the existence of an exchange process consisting in annihilation of the original positron and negatron and creation of a new pair. This exchange process ought to have a nearly isotropic angular distribution, so changing considerably the overall angular distribution. Though this additional effect will be overcompensated by interference between the two processes, still it should be possible, by examining the angular distribution of scattered positrons, to decide whether the exchange process does exist or not. A preliminary cloud chamber study of positron-negatron collisions has been made by Ho Zah-wei [31] in our laboratory. Continuation of this work by the Heidelberg group [32] has yielded ampler statistic material. The cloud chamber was filled with methane, and positrons of 100 to 400 kev were sorted out from the Cu^{64} radiation. Altogether 2900 m of well-defined positron tracks were evaluated. Figure 7.16 presents the

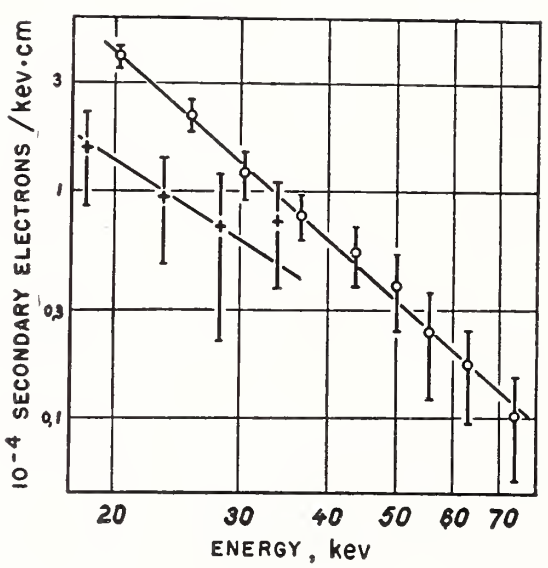


FIGURE 7.15. *Differential energy distribution of slow secondaries from cosmic rays.*

Upper curve, shower particles; lower curve, single particles (M. Deutschmann).

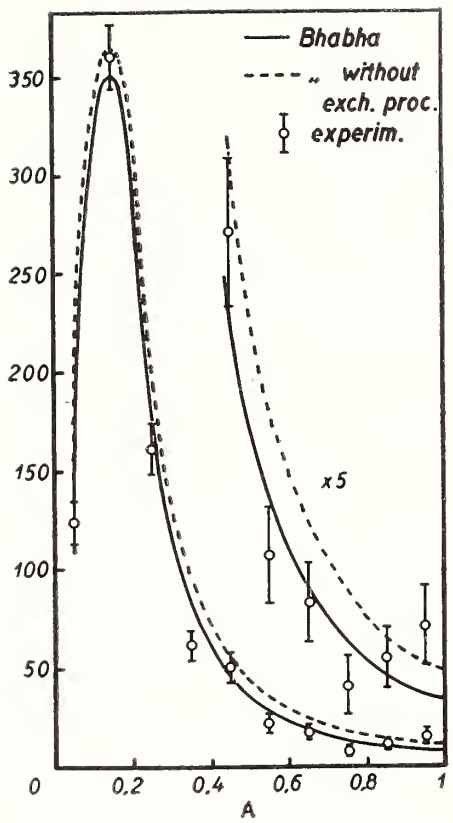


FIGURE 7.16. *Distribution of relative energy transfer A in positron-negatron collisions, for positrons of 100 to 400 kev.*

The full curve has been calculated from the complete Bhabha formula. In the dashed curve the exchange process and interference effects have been neglected.

result. Here the number of observed collisions is plotted for each interval $\Delta A = 0.1$, where A is the relative energy transfer, which, of course, is closely connected to the scattering angle. The full curve has been computed from the complete Bhabha formula. The agreement, even in absolute values, is as close as can be expected. In computing the dashed curve the exchange process and interference effects have been disregarded. The deviations from experimental values are clearly greater and more systematic. Though we have learned to be very careful in interpreting absolute cloud chamber results, one can say that these results lend support to the hole theory of the positron.

Nevertheless, more accurate measurements on positron scattering are highly desirable with regard to the fundamental importance of these problems. For the rest, it is always possible that quite unexpected features of electron interaction may be revealed by experiments of this kind. In the last-mentioned series of cloud chamber photographs one very clear case of a collision between one positron and *two* negatrons has been observed. The energy-momentum balance seems to be fulfilled in toto, but the process cannot be split up into two independent collisions between two particles without violating conservation of momentum in each case, so a genuine "triple collision" must have occurred.

References

[1] W. A. McKinley, H. Feshbach, Phys. Rev. **74,** 1759 (1948).
[2] J. B. Bartlett, R. E. Watson, Proc. Am. Acad. Arts Sci. **74,** 53 (1940).
[3] W. Bothe, Z. Naturforsch. **4a,** 88 (1949).
[4] R. J. Van de Graaff, W. W. Buechner, H. Feshbach, Phys. Rev. **69,** 452 (1946); **72,** 678 (1947).
[5] G. Schulze-Pillot, W. Bothe, Z. Naturforsch. **5a,** 440 (1950).
[6] W. Bothe, Z. Naturforsch. **5a,** 8 (1950).

[7] E. Kinzinger, Z. Naturforsch. **7a,** 390 (1952); **8a,** 312 (1953).
[8] W. Paul, H. Reich, Z. Physik **131,** 326 (1952).
[9] J. H. Bartlett, T. A. Welton, Phys. Rev. **59,** 281 (1941).
[10] C. B. O. Mohr, Proc. Roy. Soc. London [A] **182,** 189 (1943).
[11] J. Schwinger, Phys. Rev. **76,** 790 (1949).
[12] E. M. Lyman, A. O. Hansen, M. B. Scott, Phys. Rev. **79,** 228 (1950) [A]; **81,** 309 (1951) [A].
[13] M. E. Rose, Phys. Rev. **73,** 279 (1948); L. R. B. Elton, Phys. Rev. **79,** 412 (1950); G. Parzen, Phys. Rev. **80,** 355 (1950); L. K. Acheson, Phys. Rev. **82,** 488 (1951).
[14] W. Bothe, Z. Physik **4,** 161 (1921); E. J. Williams, Proc. Roy. Soc. London [A] **169,** 531 (1939); S. Goudsmit, J. L. Saunderson, Phys. Rev. **57,** 24 (1940); **58,** 36 (1940); G. Moliere, Z. Naturforsch. **2a,** 133 (1947); **3a,** 78 (1948); H. S. Snyder, W. T. Scott, Phys. Rev. **76,** 220 (1949); C. H. Blanchard, U. Fano, Phys. Rev. [A] **82,** 767 (1951).
[15] G. Groetzinger, M. J. Berger, F. L. Ribe, Phys. Rev. **77,** 584 (1950).
[16] L. A. Kulchitsky, G. D. Latyshev, Phys. Rev. **61,** 254 (1942).
[17] O. Knecht, unpublished.
[18] C. T. Chase, R. T. Cox, Phys. Rev. **58,** 246 (1940).
[19] H. Reich, Z. Physik **130,** 144 (1951).
[20] W. Bothe, Z. Naturforsch. **4a,** 542 (1949).
[21] W. Bothe, Ann. Physik **6,** 44 (1949).
[22] N. F. Mott, H. S. W. Massey, The theory of atomic collisions, p. 81 (Oxford 1949).
[23] H. J. Lipkin, M. G. White, Phys. Rev. **79,** 892 (1950).
[24] H. H. Seliger, Phys. Rev. **78,** 491 (1950).
[25] W. Miller, Phys. Rev. **82,** 452 (1951).
[26] C. Møller, Ann. Physik **14,** 531 (1932).
[27] G. Groetzinger, L. B. Leder, F. L. Ribe, M. J. Berger, Phys. Rev. **79,** 454 (1950).
[28] L. A. Page, Phys. Rev. **81,** 1062 (1951).
[29] M. Deutschmann, Z. Naturforsch. **2a,** 61 (1947).
[30] H. J. Bhabha, Proc. Roy. Soc. London [A] **154,** 195 (1936).
[31] W. Bothe, Ho Zah-wei, Nachr. Akad. Wiss. Göttingen 1946, p. 59.
[32] O Ritter, C. Lieseberg, H. Maier-Leibnitz, A. Papkow, K. Schmeiser, W. Bothe, Z. Naturforsch. **6a,** 243 (1951).

8. Back-Scattering of Positrons and Electrons

By H. H. Seliger [1]

Introduction

There has been relatively little recent experimental work on the back-scattering of fast electrons in matter and none on the back-scattering of positrons. Brand [1][2] has studied the back-scattering cathode rays up to 32 kev in energy, and Bothe [2] has extended the measurements to 680 kev, using radioactive sources in a magnetic analyzer.

Recently the author reported an excess of electron back-scattering over positron back-scattering [3]. This excess of electron back-scattering over positron back-scattering is to be expected indirectly on the basis of the single scattering cross sections as calculated by Bartlett and Watson [4] for electrons and Massey [5] for positrons. A further study of the back-scattering process has provided several very interesting results from which one is able to deduce a qualitative understanding of the general back-scattering process.

Experimental Procedure

The geometry of the previous experiment (see [3] is indicated schematically in figure 8.1, a. If one defines an angle θ as the angle between the plane containing the front surface of the backing and the plane containing the axis of the detector and the source, then in figure 8.1, a, $\theta=90°$.

The present experiments can be divided into two parts. In the first part the back-scattering coefficients for both positrons and electrons as functions of the atomic number Z of the backing material were measured for a 2π solid angle, in contrast with the solid angle of roughly 1 steradian in figure 8.1, a. These 2π back-scattering coefficients were determined in 4π absolute beta counters [6]. The geometry of the measurement is indicated schematically in figure 8.1, b. In the second part the angular distributions of the back-scattered electrons and positrons, both in number and in energy, were measured as functions of Z. The arrangement of figure 8.1, a, was modified so that the solid angle subtended was only 0.01 steradian. The source approximated a point source, and the backing on which the source was mounted was constructed so that it could rotate, with its axis of rotation lying in the plane of the front surface of the backing and perpendicular to the line containing the axis of the detector. Thus by rotation of the backing relative to the counter one could sample the back-scattered radiation at various angles θ. Estimates of the

[1] National Bureau of Standards, Washington, D. C.
[2] Figures in brackets indicate the literature reference on p. 90.

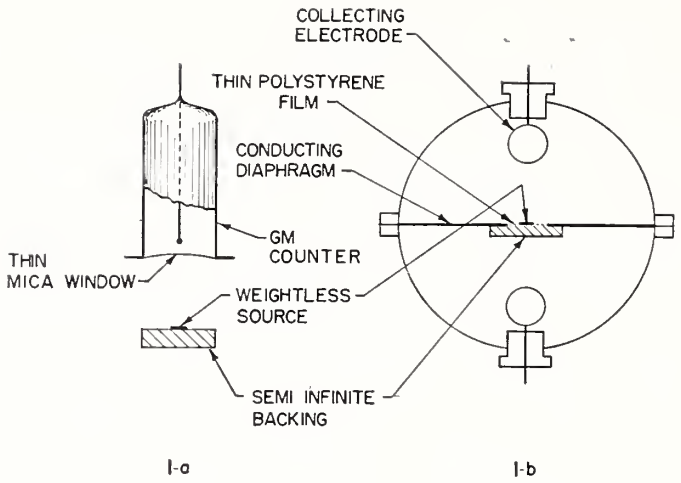


FIGURE 8.1. *a, Schematic drawing of geometrical arrangement of the previously reported experiments; b, schematic drawing of 4π counter arrangement.*

energy distribution of the back-scattered electrons were made by means of absorption curves in aluminum. These absorption curves, extrapolated to zero total absorber, also gave the correction for the absorption due to the thickness of air and counter window.

P^{32}(E_{max}1.71 Mev) and Na^{22}(E_{max} 0.58 Mev) were used as sources of negative and positive electrons, respectively. The back-scattering was measured for lucite, aluminum, copper, silver, platinum, and lead.

2π Back-scattering Coefficients

The 2π back-scattering coefficients were measured in the following manner: The absolute disintegration rate, N_0, of the source was determined in the 4π counter by the method outlined in reference [6]. Then the source, mounted on a thin polystyrene film, was placed on the polished face of the backing, and the counting rate in the top half of the counter was measured.

This is given by

$$N_t' = \frac{N_0}{2}[1+\beta(1-\tau)^2], \tag{1}$$

where N_t' is the counting rate in the top half of the 4π counter, β is the back-scattering coefficient, and τ is the fractional absorption in the mounting film, ($\tau \ll 1$). The factor $(1-\tau)$ is taken squared because a particle must pass through the mounting film twice if it is back-scattered. The procedure for determining τ is also given in reference [6]. β, defined by (1), can be written as

$$\beta = \frac{1}{(1-\tau)^2}\left[\frac{N_t' - \frac{N_0}{2}}{N_0}\right], \tag{2}$$

$\tau \sim 0$ for P^{32} electrons, so that β reduces to

$$\beta = \frac{2N_t'}{N_0} - 1. \tag{3}$$

In the case of Na^{22} positrons τ is not zero and a further correction is necessary because of the associated γ and annihilation radiation. It is possible that a γ or annihilation quantum accompanying a positron that is stopped in the backing will produce a count in the top half of the 4π counter, thereby making β^+ appear larger than the true value. The presence of cascade or annihilation radiation does not affect the determination of N_0 since the entire process occurs well within the resolving time of the proportional counter and amplifier for the original beta particle ionization pulse.

A new set of 4π counter equations was derived, following the same line of reasoning as in the original paper on 4π counting, but this time taking into account the γ and annihilation radiation from the Na^{22} positrons. The final result for β gave

$$\beta=\frac{1}{(1-\tau)^2}\left[\frac{N'_t-\frac{N_0+N'_b}{2}}{\frac{N_0-N'_b}{2}}\right], \tag{4}$$

where N'_b, the counting rate in the bottom half of the 4π counter with the semi-infinite backing in place, is due entirely to the effect of the photons. The total correction, including that for film absorption and for γ and annihilation radiation, amounted to roughly 5 percent.

Another possible correction involved in the case of the Na^{22} positrons is introduced by the finite probability that a positron will be annihilated in motion. However, even in the case of lead, this probability is of the order of 1 or 2 percent [7]. For lower Z it would be still smaller. No corrections were made for this effect.

Angular-Distribution Measurements

An arrangement was used here whereby the point source was evaporated on a thin film and supported below the detector. The backing could be brought up directly in contact with the thin film, and then backing and film could be rotated so as to sample the radiation at various angles θ. A complete absorption curve in aluminum was measured for the source mounted on the thin film with no backing behind it. The angular distribution of this unscattered or direct radiation was effectively isotropic from $\theta=+10°$ to $\theta=-10°$, the limiting angular range of the experimental arrangement. A slight asymmetry due to difficulty in exact centering of the source was present, but the average of the values for plus and minus θ removed this effect. Next, similar absorption curves were measured for $\theta=\pm 10°$, $\pm 20°$, $\pm 30°$, $\pm 60°$, and 90° with the various backings in position. At each angle, θ, the difference between the latter set of readings and the former or unscattered readings gave the net back-scattered radiation reaching the detector through that particular absorber. These differences plotted as functions of the thickness of aluminum absorber represent the absorption of the net back-scattered radiation. Corrections were made in the case of Na^{22} for the γ-ray background.

Results

Except for quantitative values, the positrons exhibited the same characteristics as electrons so that the discussion will be limited to electrons.

The final results for the 2π back-scattering coefficients as measured in the 4π counter are shown by the solid lines in figure 8.2. The dashed-line curves are the results of the previous measurement with the geometry of figure 8.1, a. The differences in back-scattering between the two types of geometries and especially the lower values of β at high Z and higher values of β at low Z measured over the 2π solid angle point up the anisotropy of the angular distribution of the back-scattered radiation. The magnitude of this anisotropy is shown more clearly in figure 8.3, where the back-scattering coefficient β is plotted as a function of the angle θ. Here $\beta(\theta)$ is defined by

$$\beta(\theta)=\frac{N(\theta)}{N_u}, \tag{5}$$

where $N(\theta)$ is the net extrapolated back-scattered radiation at the angle θ, and N_u is the extrapolated value of the unscattered radiation, both at zero total absorber.

From the $\beta(\theta)$ curves of figure 8.3 one can infer the results of figure 2, since

$$\overline{\beta(\theta)}=\int_0^{\pi/2}\beta(\theta)\cos\,\theta d\theta.$$

Therefore $\overline{\beta(\theta)}<\beta(90°)$ for high Z and $\overline{\beta(\theta)}>\beta(90°)$ for low Z. In addition to the values of $N(\theta)$, the absorption curves in aluminum of the net back-scattered radiation give very important results concerning the relative energy degradation of the back-scattered particles as functions of θ and of Z. Figures 8.4 and 8.5 show curves

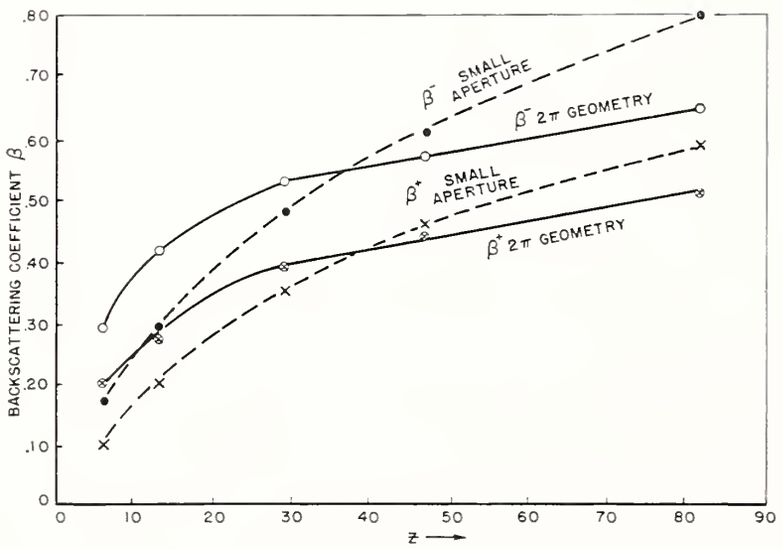


FIGURE 8.2. *Backscattering coefficients for positrons and electrons as functions of Z, measured with 2π geometry (solid curves) compared with previously measured backscattering coefficients.*

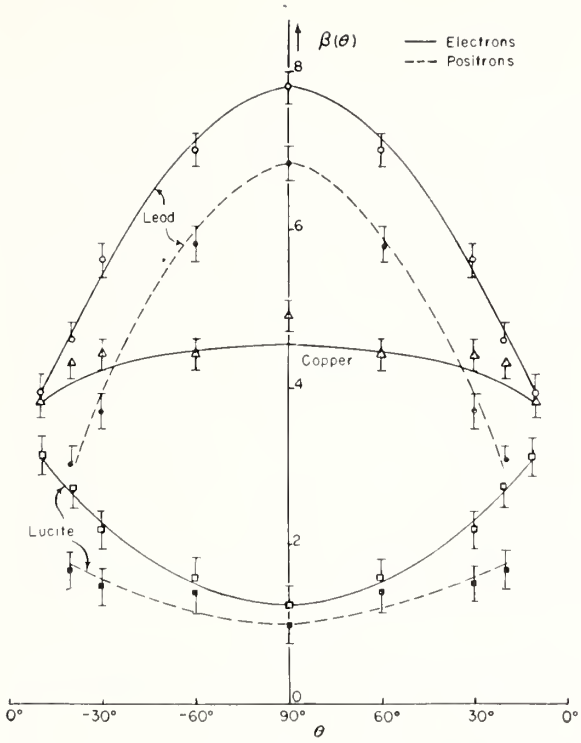


FIGURE 8.3. *Angular distribution of the backscattering coefficients for positrons and electrons for high, medium, and low Z backing materials.*

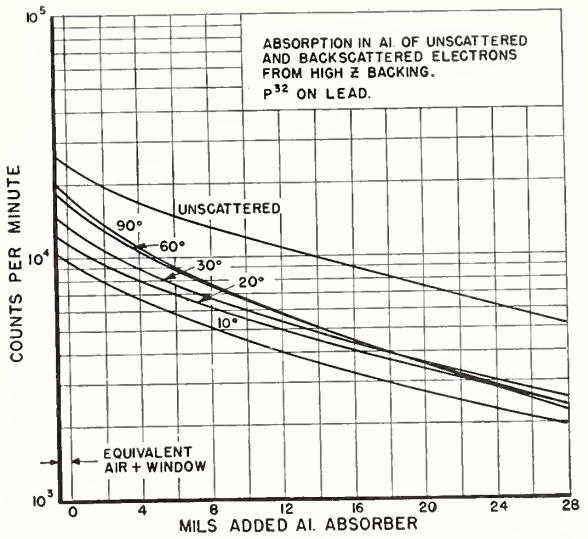


FIGURE 8.4. *Absorption in aluminum of the net backscattered P^{32} electrons from a lead backing.*

The unscattered absorption curve is shown for comparison.

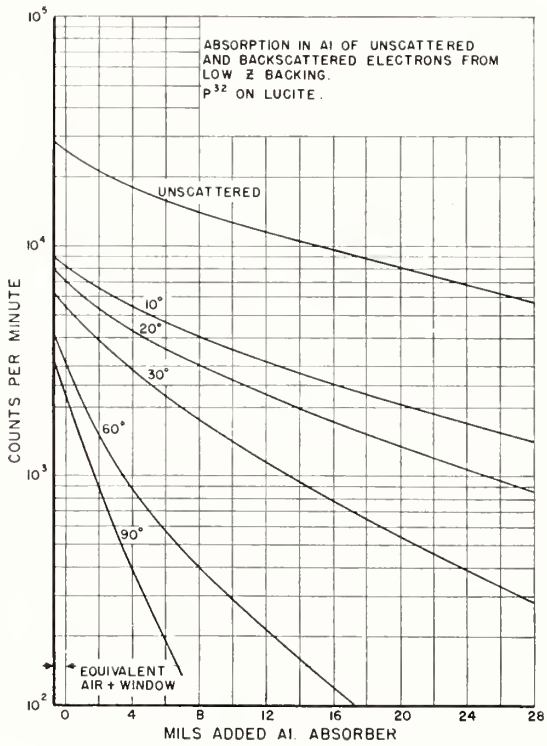


FIGURE 8.5. *Absorption in aluminum of the net backscattered P^{32} electrons from a Lucite backing.*

The energy loss can be estimated by comparison with the unscattered absorption curve.

representing the absorption of P^{32} electrons back-scattered from lead and from Lucite, respectively.

It is seen that electrons emerging at $\theta=90°$ from low-Z materials are tremendously degraded in energy, while those emerging at $\theta=10°$ are only slightly degraded. In the case of high-Z materials the electrons emerging at $\theta=90°$ are only slightly different in energy from those emerging at 10°. The relative energy degradation as a function of θ and Z is brought out more clearly by figure 8.6, where the fractional transmission from lead, copper, and Lucite backing is plotted for $\theta=10°$ and $\theta=90°$. Here one can see that the particles emerging at small angles θ have a similar energy spectrum to the original unscattered electrons, while the spectrum of those emerging at large angles θ is quite different and has a large Z dependence.

Discussion

An important feature and a fortunate simplification in the interpretation of the present experimental results is the fact that the backscattering is approximately energy independent in the energy range considered. This was shown in abundant detail by Brand[1] and Bothe[2] for monoenergetic electrons and by Burtt[8], Zumwalt[9], and the author[3] for heteroenergetic radioactive sources with widely differing end points. C. H. Blanchard and U. Fano have shown[10] that the shape of the distribution function that describes the angular distribution of an electron inside a scatterer with respect to its original direction at any point along its path depends on the energy loss E_0-E and the original energy E_0 principally as the ratio $(E_0-E)/E_0$. Although a 100-kev electron beam will penetrate deeper into a scatterer than a 50-kev electron beam, both beams will lose their original "sense" of direction, or become isotropic at the same fractional loss in energy. Therefore, the same fraction of each should be reflected in any given Z element. However, the ratio of elastic to inelastic scattering is proportional to Z, so that an electron in a high-Z material will become isotropic with much less energy loss than it would in a low-Z material, thereby increasing its probability of escape from the high Z material. The angular distribution of diffusing electrons emerging from any plane surface follows a cosine ϕ law, where the angle ϕ is defined by $\phi=(90°-\theta)$.

A qualitative estimate of the diffusion effect can be obtained by normalization of the experimental curves to a cosine distribution at $\theta=90°$. In figure 8.6 this is represented by the dashed concave downward curves. The difference between the experimental or solid-line curves in figure 8.6 and the cosine distribution curves is given by the dashed concave upward curves. This distribution is due to the incompletely diffused or sidescattered electrons. Electrons initially incident and subsequently emerging at small angles with the surface lose only a small fraction of their original energy in being deflected, and it is to those electrons that the adjective "side scattered" is applied. Since the ratio of scattering to absorption cross sections is still proportional to Z, the sidescattered electrons will lose less energy in the high-Z material than in the low-Z material. However, since little energy is lost even in low-Z materials, the

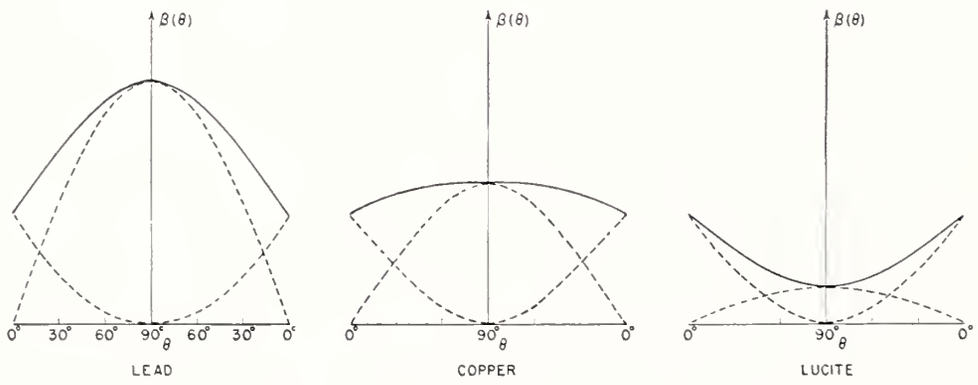


FIGURE 8. 6. *Relative contributions to β (θ) of the diffusion effect (concave downward dashed curves) and the side-scattering effect (concave upwards dashed curves), assuming a cosine distribution for the diffusion effect.*

The solid lines are taken from figure 8.3 for electrons.

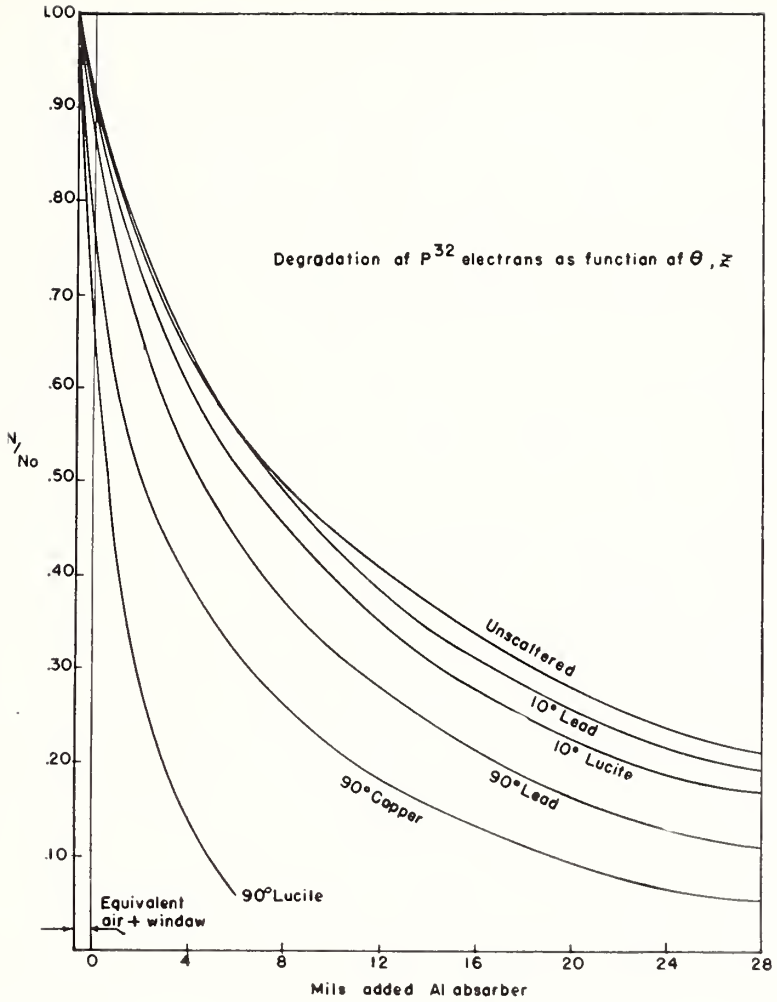


FIGURE 8.7. *Relative transmission of P^{32} backscattered electrons from low, medium and high Z for $\theta=10°$ and $\theta=90°$.*

degradation for the sidescattering effect will be very much less than the degradation for the complete diffusion effect. In figure 8.7 the 10°-transmission curves show that electrons sidescattered from Lucite are only slightly less energetic than those sidescattered from lead, while at 90° electrons diffusing from Lucite are almost totally degraded, being very much lower in energy than electrons diffusing from lead.

Figure 8.6 indicates that the diffusion effect increases materially with Z, while the side-scattering remains relatively constant. It can be argued that at small angles of incidence and emergence, while the mean deflection necessary for emergence will be reached after many more collisions in a low-Z material than in a high-Z material, the electron has roughly the same probability of emerging from a low-Z material as from a high-Z material, although it will be slightly less energetic upon emerging from the low-Z material.

W. Miller[11] has made calculations of the ratio β^-/β^+, using Bartlett and Watson's and Massey's cross-section calculations in a classical multiple-scattering theory originally given by Bothe[12] for electrons. He estimates $\beta^-/\beta^+=1.16$ for mercury. The experimental results show $\beta^-/\beta^+=1.3$ for $Z=80$, which is in fair agreement with Miller's results, especially since the initial conditions in Bothe's derivation are not exactly those of the present experiments. In any case, an excess of electron back-scattering over position back-scattering is to be expected.

The isotropic initial conditions of the experiments are those most commonly encountered in actual practice in radioactivity measurements. It is therefore quite important to recognize that the back-scattering is anisotropic, so that the proper back-scattering correction can be made, depending on the geometry used. The curves of figure 8.2 should be especially helpful to those working with C^{14}, S^{35} and other low-energy beta emitters, where a 2π windowless counter is often used.

The author thanks Misses L. Cavallo and S. V. Culpepper for their valuable assistance in making these measurements, and Drs. U. Fano and C. H. Blanchard for many stimulating and enlightening discussions.

References

[1] J. O. Brand, Ann. Physik **5,** 609 (1936).
[2] W. Bothe, Z. Naturforsch. **4a,** 542 (1949).
[3] H. H. Seliger, Phys. Rev. **78,** 491 (1950).
[4] J. H. Bartlett and R. E. Watson, Proc. Am. Acad. Arts Sci. **74,** 53 (1940).
[5] H. S. W. Massey, Proc. Roy. Soc. London **A181,** 14 (1942).
[6] H. H. Seliger and L. Cavallo, NBS J. Research **47,** 41 (1951).
[7] W. Heitler, Quantum Theory of Radiation, p. 231, 2d ed. (Oxford Univ. Press, Oxford, England, 1944).
[8] B. P. Burtt, Nucleonics [2] **5,** 28 (1949).
[9] L. R. Zumwalt, Mon C 397 (1950).
[10] C. H. Blanchard and U. Fano, Phys. Rev. **82,** 767 (1951).
[11] W. Miller, Phys. Rev. **82,** 452 (1951).
[12] W. Bothe, Ann. Physik **6,** 44 (1949).

9. Range of Electrons in the Energy Interval 0.5 to 1.3 Mev

By Jerome Fleeman [1]

I would like to report on the results of some range-of-electron experiments that have been performed at the National Bureau of Standards. The work is part of a program in the diffusion and penetration of electrons in materials. In this report, we shall focus particular attention on the value of the extrapolated range of electrons in the energy interval 0.5 to 1.3 Mev with regard to the Z dependence of the range. It is the usual custom in physics to express the range of electrons as a function of the energy of the incoming electrons when measured in aluminum. Our problem has been to extend the measurements of range to include materials of varying Z.

The range of electrons was measured in the usual manner. Monoenergetic electrons, obtained from an electron accelerator, were collimated to a beam 1.0 in. in diameter, and incident upon metallic absorbers. The absorbers were placed perpendicular to the incident beam, and the transmitted beam was measured by a parallel plate ionization chamber. The plates of the ionization chamber were made of polystyrene of thickness 2½ mg/cm^2. To measure the range, absorbers of various thickness were used, and the ionization current produced in the chamber was determined. A plot was made of the ionization current as a function of the thickness of absorber. These graphs are each characterized by an initial rise in the ionization current to a maximum value after which there is a linear decrease of current for increasing absorber thickness to a region where the ionization current decreases slowly to zero current. By extrapolation of the linear portion of the ionization curve to zero current one obtains the thickness corresponding to the extrapolated range.

We have measured the ionization current after absorption in beryllium, aluminum, copper, cadmium, and gold. The ranges have been determined for three energies 0.5, 0.9, and 1.3 Mev., and the results of these measurements are shown in figures 9.1 through 5 respectively.

The feature that I would like to point out in these curves is that the extrapolated range R_p is nearly equal to the greatest range for light materials as shown in figure 9.1. However, as Z increases, it will be noted that the curves trail out until for gold the extrapolated range is located well below the greatest range. The term "greatest range" means roughly the thickness corresponding to zero current. However, for currents near zero we are almost at the limit of measurement of the apparatus as determined by background currents and fluctuations in our meters. As a result the "greatest range" isn't too well deter-

[1] National Bureau of Standards. Now at Brookhaven National Laboratory, Upton, N. Y.

mined experimentally. It is more reasonable to compare the extrapolated range with the greatest range, say, as determined by integration of the Bethe-Bloch formula, R_{bb}. Figure 9.6 shows the ratio of the measured range R_p divided by R_{bb} plotted as a function of the atomic number Z. This ratio represents the reduction in the theoretical range as a result of multiple scattering. It is to be noted that the ratio is only slightly less than unity for beryllium and equal to about 0.3 for gold. The dotted curve is drawn so as to give a reasonable fit to all the points.

Dr. Blanchard has recently completed some calculations on the shape of the ionization-depth curve from which we can get values for the reduction in the range. It turns out that the curve of figure 9.6 is drawn a little too high for beryllium and a little too low for gold. For beryllium the curve should have gone through 0.94, and actually it is drawn for 0.96, and for gold it should have gone through 0.32, and it actually goes through 0.30. However, this fit is as good as can be expected in view of the difficulties of the calculation as well as the measurement.

Finally, figure 9.7 is a plot of the Z dependence of the extrapolated range in which the points are plotted for the three energies used in the experiment. We find that we can match our results for the range by use of the formula

$$R_p = \frac{667}{Z^3} E_0 \delta(\delta + 1), \qquad \text{mg/cm}^2,$$

where

$$\delta = \frac{E_0 + mc^2}{E_0 + mc^2}, \qquad E_0 \text{ in Mev.}$$

Actually, this formula fits our data only moderately well at low energies and at low Z, but it exhibits a good fit at the higher energies and for the intermediate Z's. We have been able to determine range formulas that fit our data with a smaller error. However, these turn out to be more complicated and the one used has the virtue of simplicity.

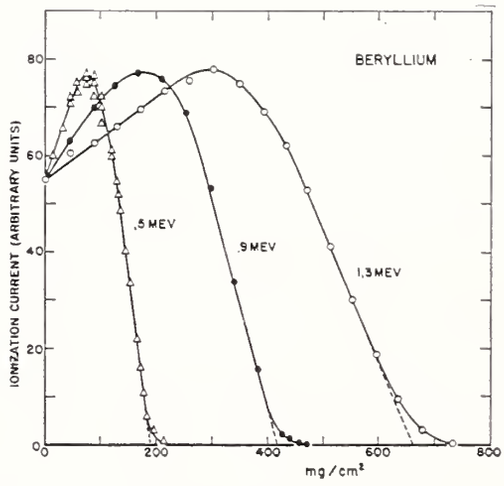


FIGURE 9.1.

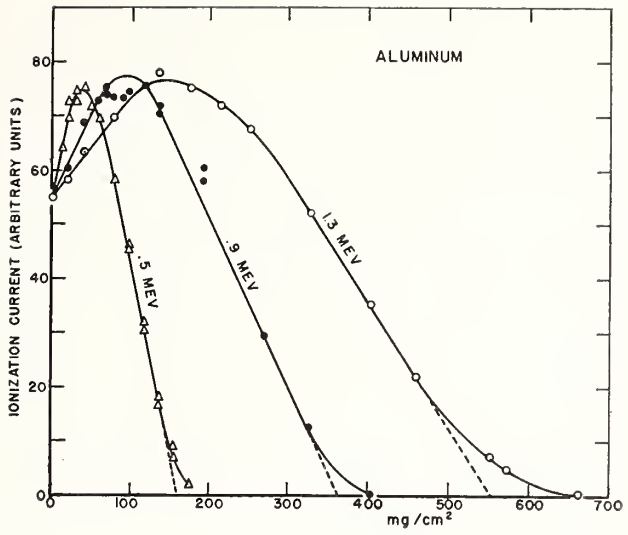


FIGURE 9.2.

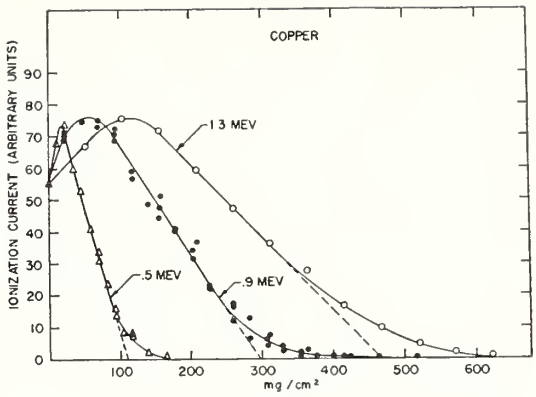


FIGURE 9.3.

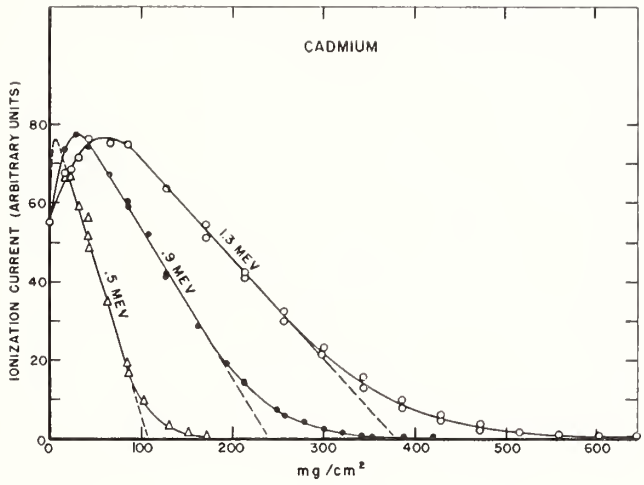


FIGURE 9.4.

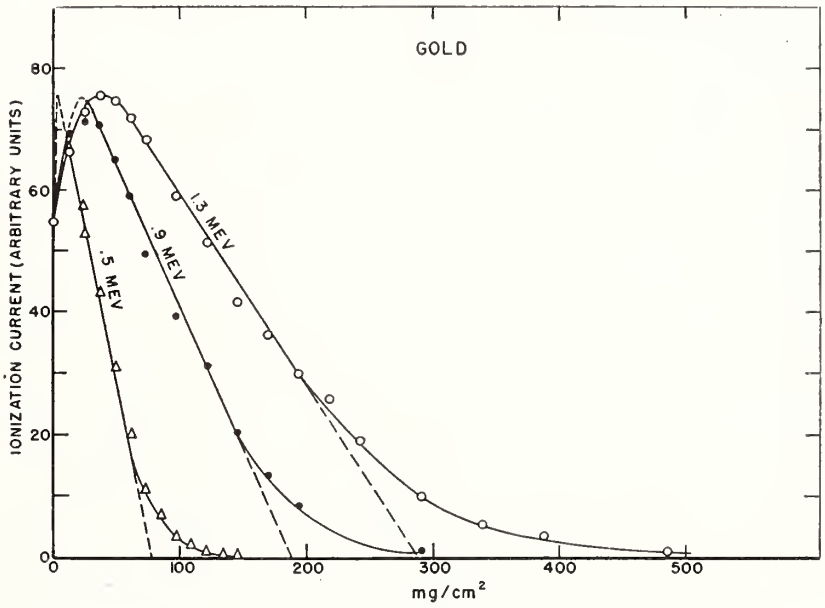


FIGURE 9.5.

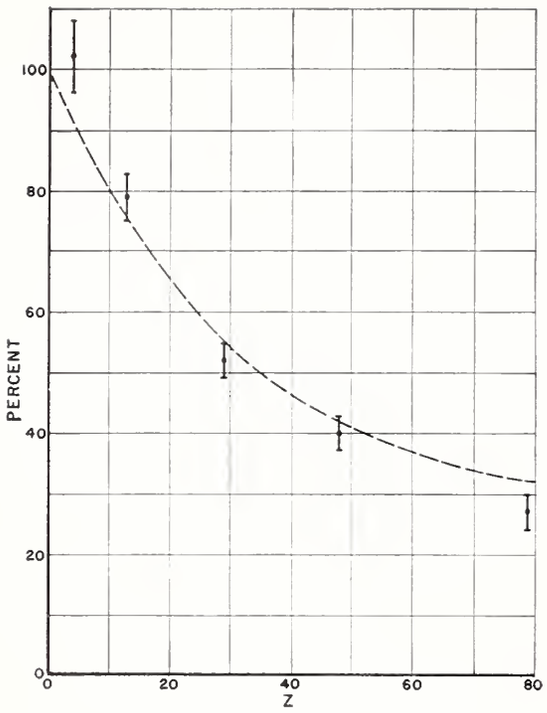


FIGURE 9.6.

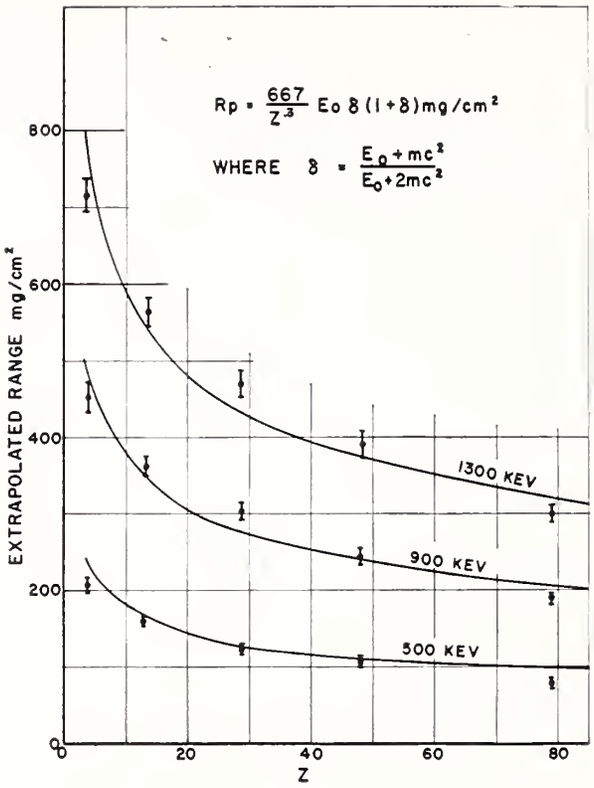


FIGURE 9.7.

10. Measurements of Electron Scattering With the Electron Microscope

By C. E. Hall [1]

The intensity variations in an electron microscope image are influenced by two principal factors: The electron optical properties of the imaging system, and the scattering processes occurring in the object. In seeking an adequate description of the significant scattering phenomena, we are interested mainly in beam potentials from about 30 to 100 kv, in specimen thickness of about 1000 A or less, and in the character of the transmitted beam at angles of less than about 10^{-2} radian. Quantitative experimental evidence of this kind has been almost entirely lacking, although qualitative evidence is present in all electron micrographs. Theoretical calculations for scattering at these small angles have been presented on several occasions. The first serious attempt to account for image intensities through the application of scattering theory was made by Marton and Schiff in 1941.[2] Later works of von Borries,[3] Boersch,[4] and of Hillier and Ramberg [5] are also particularly to be noted. Whether the theoretical scattering cross sections can be used with confidence under practical circumstances cannot be decided, however, without experimental evidence. The work to be described is concerned with the measurement of effective scattering cross sections made with the electron microscope itself, under practical operating conditions.

The primary consideration in electron microscope image formation is that, owing to spherical aberration, only electrons contained within a relative aperture of about 10^{-2} radian or a little less will reach the image plane sufficiently close to their proper image point to form a sharply defined image. The intensity at large angles reaches the image plane as a diffuse background intensity or is lost against the walls and diaphragms of the instrument. The incident intensity is generally of considerably smaller relative aperture than the effective aperture of the objective lens. Within the effective aperture of the objective lens we recognize two components of intensity: A transmitted intensity, I_T, of very small aperture approaching that of the incident beam, and a scattered intensity, I_K. We assume that in traversing a layer dw, where w is the mass per unit area, that a fraction, k, of the intensity lost to I_T is scattered within the effective aperture of the lens and so contributes to I_K while the remaining fraction $(1-k)$ is permanently lost to the imaging beam. Similarly, it is assumed that the probability that an electron will be permanently lost from I_K in a layer dw is the same as for those in I_T, ignoring the possible angular dependence of such an event within the very small

[1] Massachusetts Institute of Technology, Cambridge, Mass. This investigation was supported in part by research grant G-3396 from the National Institute of Health, Public Health Service.

[2] L. Marton and L. I. Schiff, J. Applied Phys., **12,** 759 (1941).

[3] B. von Borries, Z. Naturforsch. **4a,** 51 (1949).

[4] H. Boersch, Z. Naturforsch. **2a,** 615 (1947).

[5] J. Hillier and E. M. Ramberg, Z. angew. Physik **2,** 19 (1950).

effective aperture of electron microscope objectives. With these assumptions the following equations[6] have been derived for the intensities I_T and I_K:

$$I_T = I_0 e^{-S_T w}. \tag{1}$$

and

$$I_K = I_0 e^{-(1-k) S_T w} - I_0 e^{-S_T w}, \tag{2}$$

where S_T is the specific cross section in square centimeters per gram, and w is the mass thickness of the film traversed in g cm^{-2}. I_T produces an image of high resolution superimposed on one produced by I_K of lower resolution like those seen in dark-field images. The total intensity, I_c is $I_T - I_K$, or

$$I_c = I_0 e^{-S_c w}, \tag{3}$$

where

$$S_c = (1-k) S_T.$$

Equations (1), (2), and (3) are essentially empirical and required experimental confirmation. They have been verified as to form by measuring the relative values of I_T, I_c, and I_0 through SiO films ranging in thickness up to 80×10^{-6} gm-cm^{-2}. Specific cross sections, S_c, for elements may be measured from a photographic determination of I_c / I_0 for films of known mass thickness, w. In the initial experiments w was determined from the increase in weight of glass cover slips placed beside electron microscope grids on which films of various materials were deposited by vacuum evaporation. Fluctuations in weight of the glass cover slips, probably due to their slightly hygroscopic nature, however, rendered them unreliable for the determination of small weight increments. Aluminum foil was later found to be much more reliable for the purpose.

A new method of measuring the scattering power of thin metallic films is shown in figure 10.1. This figure shows a 2×10 in. photographic plate recorded at 20,000 x. The first three frames are micrographs of polystyrene spheres on a collodion film, shadowed with Ge at an angle of 3 to 1. The mass thickness of the metallic deposit was obtained from the weight increment of an aluminum foil placed beside the specimens in the vacuum evaporator. Each frame is recorded near the edge of an opaque nickel grid wire so that the relative background intensity I_B can be measured and subtracted from the intensities in the image field. The final 3-in. portion of the plate is a series of stepped exposures made by displacing the plate at 10-sec intervals with the specimen removed. A plot of the photo-

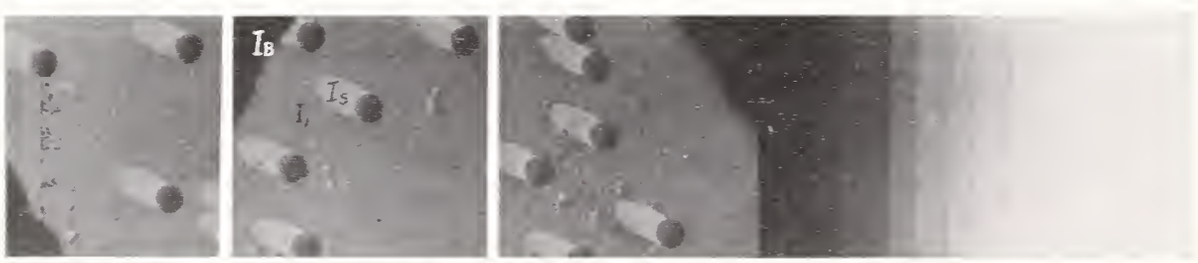


FIGURE 10.1. *Typical photographic plate of* Ge-*shadowed polystyrene spheres from which the scattering power of the metal film can be measured.*

[6] C. E. Hall, J. Applied Phys. **22**, 655 (1951).

graphic density of the steps versus their exposure time provides a graph from which relative intensities in any frame can be found from the corresponding photographic densities. If I_s is the intensity in the shadow of a sphere (after subtraction of background) and I_1 the intensity through the Ge film, then according to eq (3), a plot of $\log_e I_s/I_1$ versus w should be a straight line whose slope is S_c for Ge. A plot of the results for three different film thicknesses is shown in figure 10.2. The line should, of course, go through the origin. The three points within circles were made with the lens previously described. A point obtained from the Philips microscope, which has a lens of different structure, is also shown in the graph for comparison at one film thickness. Cross sections probably do not vary greatly for different lenses since the effective relative aperture is not sensitive to differences in the spherical aberration constant. The method shown in figure 10.1 has also been used for Cr, since the values originally reported for Ge and Cr were most suspect owing to the instability of the glass cover slips.

Another experiment for the determination of the electron scattering of polystyrene is shown in figure 10.3. Images of polystyrene spheres about 2600 A in diameter like those shown in figure 10.1, but without shadowing, were recorded along with a stepped exposure. The intensities at distances from the edges of the spheres were measured with a microphotometer slit width of 30 A to obtain the intensities I_1 and I_2 (minus background) as indicated in figure 10.3. Plotted circles with dots represent measurements from one edge and circles with crosses from the opposite edge. The plot should be a straight line but the devation is large for x greater than about 1500 A. The reason for the discrepancy is not known, but it could be accounted for by a gradual increase in background into the particle. There is thus some uncertainty as to where a line should be drawn for the determination of S_c for polystyrene. The maximum slope yields a value of $S_c = 5.5 \times 10^4 \text{cm}^2\text{g}^{-1}$, as indicated. (The density of polystyrene [7] is

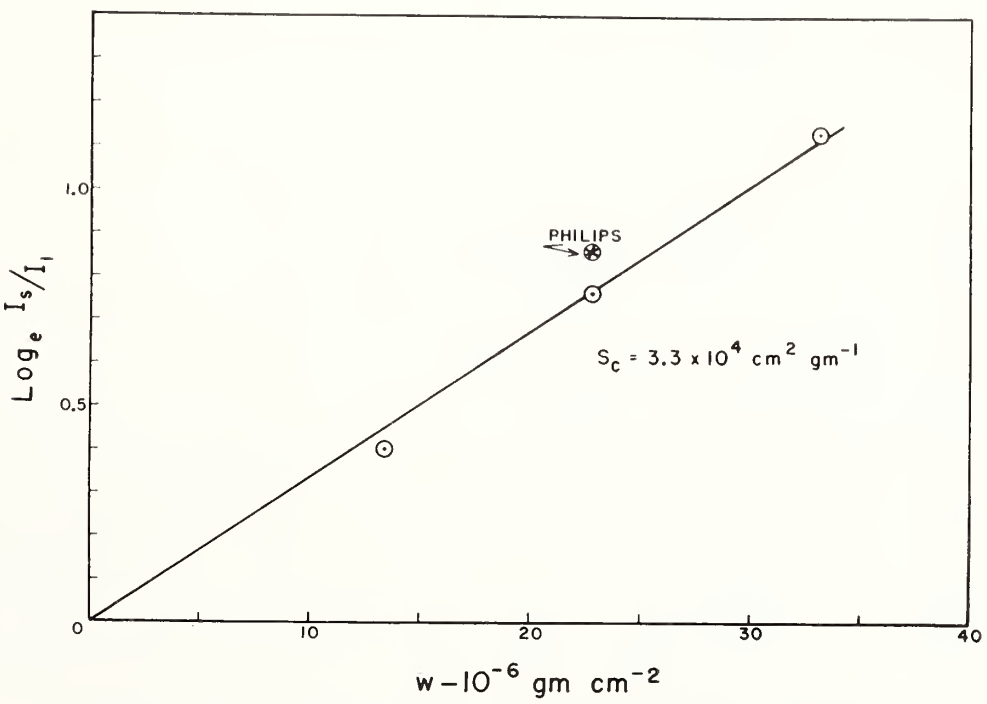


FIGURE 10.2. *Test of eq (3) and measurement of S_c for* Ge.
Scattering power of germanium films 65 kv.

[7] D. G. Sharp and J. W. Beard, J. Biol. Chem. **185,** 247 (1950).

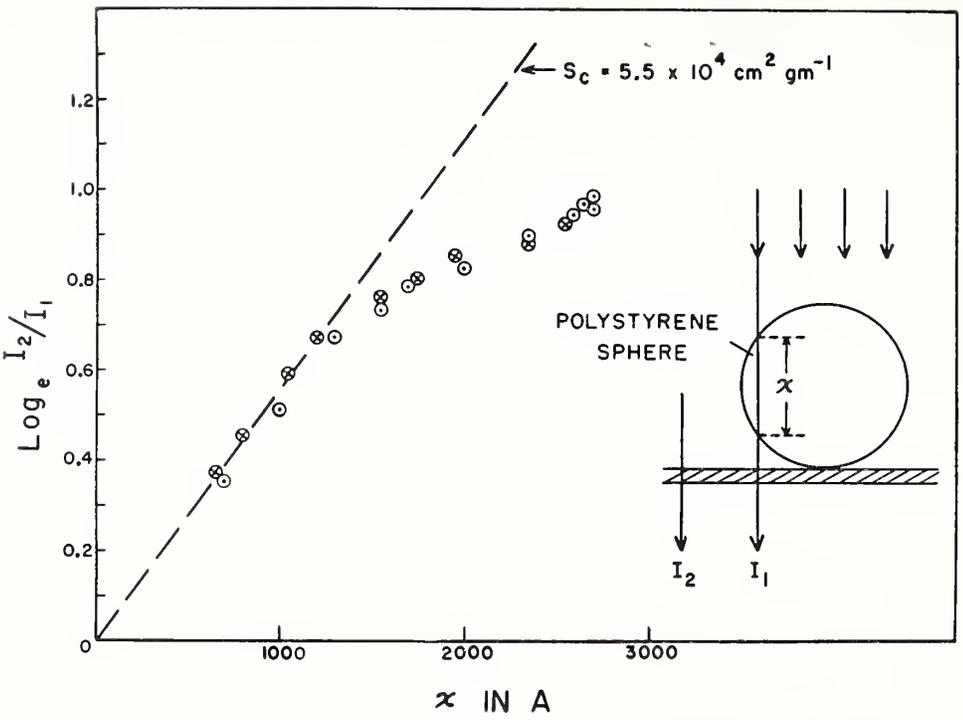


FIGURE 10.3. *Results from a microphotometer tracing of the image of a polystyrene sphere.*

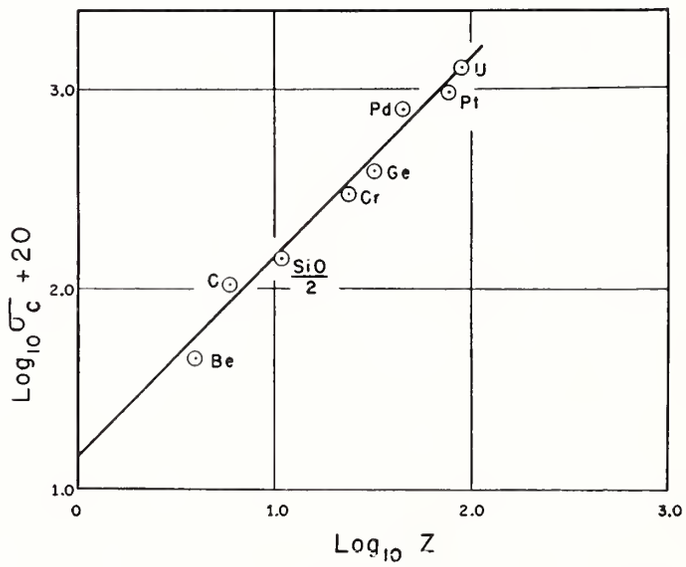


FIGURE 10.4. *Dependence of σ_c on atomic number*

The line has been drawn for a slope of unity.

1.05 and the composition is $(CH)_x$).

In table 10.1 the current data for the scattering cross sections of various materials are listed. The values for carbon were taken from the polystyrene tracings. The magnitude of S_c is relatively insensitive to atomic number, while the cross section per atom, σ_c, increases steadily with Z, as would be expected. Since previous theories indicate that σ_c should increase approximately as Z^n, where n is a number close to unity, the data in table 10.1 have been plotted on a log-log scale in figure 10.4. To the order of accuracy of the measurements, the data are representable by the equation

$$\sigma_c = 0.14 \times 10^{-18} Z \text{ cm}^2, \tag{4}$$

although the data are not sufficiently accurate to fix the value of n reliably within 10 or 20 percent.

TABLE 10.1. *Electron scattering constants 65 kv*

Substance	Z	S_c 10^4 cm²/g	σ_c 10^{-18} cm²
Be	4	3.0	0.45
C	6	5.5	1.1
SiO	14+8	3.7	1.4
Cr	24	3.5	3.0
Ge	32	3.3	3.9
Pd	46	4.5	8.0
Pt	78	3.0	9.8
U	92	3.3	13.0

Since all the preceding measurements were made at 65 kv, it is desirable to test the dependence of effective cross sections on beam potential. At present, only one series of measurements have been made. Records like that shown in figure 10.1 were made of a single Ge film of 23×10^{-6}g cm^{-2} at 40, 60, 80, and 100 kv with a Philips microscope operated with a large objective diaphragm. The results are given in table 10.2 and show about a 40-percent decrease in cross section in going from 40 to 100 kv. It is not possible to deduce the effect of voltage on image quality from the dependence of σ_c on voltage alone since, as was pointed out previously (see footnote 6), the relative magnitudes of I_T, I_K, and background will also vary with beam potential.

TABLE 10.2. *Germanium film* 23×10^{-6} *g cm*$^{-2}$

Beam potential	Relative S_c $\log_e Is/I_1$
kv	
100	0.6
80	.7
60	.9
40	1.0

In a recent publication, Hillier and Ramberg have calculated the angular dependence of cross section for representative elements according to formulas given by von Borries and others. At an aperture angle of 10^{-2} radian the cross sections in cm² per gram are relatively insensitive to atomic number, in qualitative agreement with the results in table 10.1, but the magnitudes of their cross sections at this aperture are between 8×10^4 and 1.5×10^5cm²g^{-1}, which are roughly 2 to 3 times greater than the experimental values in table 10.1. Hillier and Ramberg made their calculations for 50 kv, but the difference in beam potential cannot account for the discrepancy. The reasons for the differences between theory and experiment are not apparent at present.

In concluding, the author wishes to emphasize that the experimental evidence is far from adequate as it now stands. There is need for measurements of the factor, k, for various film thicknesses and for various materials, to test whether there is justification for assuming the factor to be a constant within limits. Further data are also

desirable concerning the dependence of cross sections on aperture. state of chemical combination and on beam potential. If these and other problems are worked out satisfactorily, there should be the possibility, in certain instances, of measuring amounts of materials in electron microscope specimens through microdensitometry.

Discussion

Dr. D. Gabor, Imperial College, London, England: Have you any evidence regarding the coherency of the transmitted part?

Dr. Hall: Yes, I think so. If there is a heavy particle on a film, you can still see clear-cut Fresnel fringes around it. However, some of the films are so thick that it is barely possible to distinguish an opaque edge through them because of the high background intensity. One can see Fresnel fringes through films that electron microscopists would consider quite "thick." I believe it is only the intensity I_T which can produce such fringes, and the cross section for decay of this component is roughly three times as great as those given for the intensity within the effective aperture.

Question: Did you use an aperture of 10^{-2} radian?

Dr. Hall: Yes, that is approximately what I would consider the effective aperture of the objective. The actual physical aperture of the lens is of course much larger.

Question: Don't you get an overlap (of velocities) when you go from a very thick to a very thin area? When you take a microphotometer tracing at an edge, you actually get a fairly high intensity just inside the edge of a very thin portion.

Dr. Hall: I lumped all the scattered intensity together without regard to whether it has altered significantly in velocity or not. The evidence is that the probable energy losses are about 20 electron volts and that they are mostly within the effective aperture of the instrument. If greater losses occur, they are probably outside the effective aperture and would simply contribute to background.

Dr. F. A. Hamm, General Aniline and Film Corporation: I presume you averaged your pnotometer tracing for the background over quite a large distance. It seems to me you would get a significant difference in the value of the transmission just outside the thick part or just inside the thin part.

Dr. Hall: The microphotometer traces are taken at the edge of a nickel screen, which is completely opaque. The intensity inside the shadow of the edge is all background. The decrease of background intensity from the bright side into the shadow is gradual, and the slope is practically the same on both sides of the edge. I have taken the useful image intensity as the sudden drop at the opaque edge of the nickel screen.

Dr. Hamm: I am surprised that you did not get a little dip in the curve.

Dr. Hall: The image of the edge is over-focused because the microscope was focused on the level of the film, which is some distance below the edge. Since the aperture of illumination is finite, this results in some unsharpness in the image of the edge.

11. Experimental Study of Limits Imposed by Plural Scattering in Electron-Diffraction Studies

By S. G. Ellis [1]

Introduction

It has been known for a long time that, in the study of thin films by electron diffraction, the contrast of the rings with respect to the background increases as the accelerating voltage is increased, and is reduced as the film thickness is made larger. The aim of the present study was to obtain quantitative information on this phenomenon.

Experimental Procedure

A preliminary requirement is to ensure that the electron diffraction camera does not influence the results either by increasing the ring width, or by increasing the background. The latter effect can occur if the electron scattering from the walls of the instrument is not suppressed by suitably disposed diaphragms. The camera employed was optically similar to that described by Hillier and Baker.[2] In addition to the diaphragm system they describe, another diaphragm was placed below the specimen to prevent electrons scattered in the specimen from reaching the walls of the camera.

The contrast was measured in the following manner: Using the standard method for establishing the characteristic curve of the photographic emulsion, it is possible to determine the current density at the peak of the most intense line in the pattern; the background current density, I'_B, is also found by extrapolating the background to a point below the peak of the line. The difference of these current densities, I'_R, is the peak ring current density. I'_R/I'_B is then taken as the peak contrast C' of the ring. Using Kodak medium lantern slide plates, it was found that the rings could be observed satisfactorily provided $C \geqq 0.1$.

Films of aluminum and thallium chloride were evaporated upon thin films of collodion. The thicknesses of the films were found by multiple-beam interferometry. Electron diffraction patterns were then obtained from these films at 50, 100, and 150 kv, and were analysed as described above.

The collodion films did not contribute significantly to the electron scattering in the range of total thickness employed here, the region of major interest being that in which the contrast falls to low values.

Results

In general, an increase in electron speed produced an increase in contrast, and this was often very marked (table 11.1). On the other

[1] RCA Laboratories Division, Princeton, N. J.
[2] J. Hillier and R. F. Baker, J. Appl. Phys. **17**, 12 (1946).

TABLE 11.1.

V_{kv}	Aluminum		Thallium chloride	
	t=1000 A	2000 A	1000 A	2000 A
50	C'=2.3	0.55	1.3	0.2
100	5.8	1.9	2.5	0.65
150	10.1	3.8	3.2	1.2

hand, the contrast fell rapidly with increasing film thickness. The film thickness, $T_{\max}$, for which the contrast C' becomes 0.1, is the greatest film thickness that will yield useful electron diffraction patterns with the photographic emulsion employed in this work. The variation of $T_{\max}$ with accelerating voltage V is shown in table 11.2.

TABLE 11.2. $T_{\max}$ in A

V_{kv}	Aluminum	Thallium chloride
50	2900	2500
100	4600	3700
150	5600	4400

It will be seen that for voltages up to 150 kv, $T_{\max}$ increases less rapidly than the voltage. This is in qualitative agreement with the results of Möllenstedt.[3]

While these results are suggestive, they do not warrant generalization since the specimens encountered in electron diffraction studies seldom have the form of thin films. With this in mind, I'_R and I'_B were measured, independently for different film thicknesses, t. Elementary considerations suggest that

$$I'_R = I'_0 k'_1 e^{-Kt} \cdot t, \tag{1}$$

and the experimental results supported this view. Here, I'_0 is the current density in the incident beam, k'_1 measures the probability of diffraction, and K gives the attenuation of the primary beam due to all causes. The background current presents a more difficult problem. From some work which E. G. Ramberg did on the subject, it seemed possible that it might be approximated by the expression

$$I'_B = I'_0 k'_2 e^{-Kt}(t + \alpha t^n). \tag{2}$$

This was found to be the case. The parameter k'_2 measures the probability of incoherent scattering through twice the Bragg angle, and α and n are parameters dependent on the nature of the film. The term αt^n gives the contribution of the plural scattering. From the experimental results it was found possible to obtain k'_1/k'_2, K, α, and n. The results for aluminium are given in table 11.3, for which $n=5$ to the nearest integer.

[3] G. Möllenstedt, Nachr. Wiss. Göttingen **1**, 83 (1946).

Table 11.3.

V_{kv}	50	150
K cm^{-1}	2.22×10^{5}	1.37×10^{5}
α cm^{-4}	4.46×10^{19}	6.45×10^{18}
k_1'/k_2'	2.9	8.4
K^{-1} A	450	730

Conclusions

The experimental results show that the contrast decreases with increasing film thickness because I_R' falls off very rapidly. I_B' also falls with increasing t, but this occurs at higher values of film thickness where I_R' is already very small. In a polycrystalline wedge of aluminium, for example, the major contribution to the rings, at 150 kv, comes from the part with thickness between 250 and 2500 A. The major contributions to the background comes from the part with thickness between 1000 and 5000 A.

Equations (1) and (2) can be integrated for polycrystalline wedges and spheres. If the maximum thickness of these bodies is taken to be very large, it is found that detectable rings (C′ 0.1) will be obtained only if k_1'/k_2' is large enough.

For speciments which have very thick regions, the calculations suggest that the contrast will increase less rapidly than the voltage for voltages up to 150 kv. It does increase in all cases, however, where the specimen has no very thick regions and its thickness is of the order K^{-1} the contrast increases very rapidly with voltage. In reflection studies the implication is that the contrast will increase very slowly with voltage if the specimen is smooth, and a little more rapidly if the specimen is rough. This again is in general agreement with our experience.

The results are of interest from two points of view. In the first place, by studying the variation of ring intensity and background intensity with film thickness, it is possible to compare the coherent and incoherent scattering in the same film and to form some idea of the contribution of plural scattering. Secondly, we can estimate the advantages that may be expected from using higher accelerating voltages on electron diffraction cameras. Broadly speaking, up to 150 kv we may expect a rather general increase in contrast in the patterns but not a very spectacular increase in the number of specimens that can be examined. In fact, careful specimen preparation may often give more improvement in contrast than an increase in accelerating voltage.

Discussion

Dr. C. E. Hall, Massachusetts Institute of Technology, Cambridge, Mass.: Do you have any ideas concerning the velocity distribution in the background?

Dr. Ellis: No, it was not measured in these experiments. However, there was no lens between the specimen and the photographic plate so that the current distribution was not changed by the chromatic aberration of the lenses.

Dr. A. C. van Dorsten, Phillips Research Laboratory, Eindhoven, Holland: How would the crystal size influence your measurements?

Dr. Ellis: This is a difficult question to analyze because the ring width depends not only on the crystal size but also on the refraction due to inner potential which in turn depends on the crystal size and orientation. Thallium chloride was chosen as a specimen because it gives very narrow rings compared with aluminium, presumably because there are smaller refractive deviations. The variation of ring width with voltage was not the same with these two specimens. The contrast increased more rapidly with voltage in the case of aluminium, in which the refractive effects were probably larger. Aluminium is considered to be a more typical specimen, and the conclusions reached are based largely on the results from the aluminium films.

12. Electron Diffraction and Average Vibrational Motion of Gas Molecules

By J. Karle [1]

In the past few years quantitative measurements have replaced visual estimates of the scattered electron intensities due to interatomic interference (molecular scattering) [1, 2, 3, 4, 5][2]. As a consequence, improvements have been made in the accuracy of measuring equilibrium interatomic distances, and, in addition, it has been possible to evaluate the average vibrational motion of these distances.

The experimental features of the quantitative method involve the use of a rotating sector and a microphotometer. The rotating sector removes the steeply falling background from the scattering pattern and therefore accentuates the maxima and minima of the molecular scattering. It is possible to treat the molecular scattering data to obtain an intensity curve representing essentially the scattering from bare nuclei. It was shown by Debye [6] that the Fourier transform of the scattering from bare nuclei gives the probability distribution for the interatomic distances according to the formula

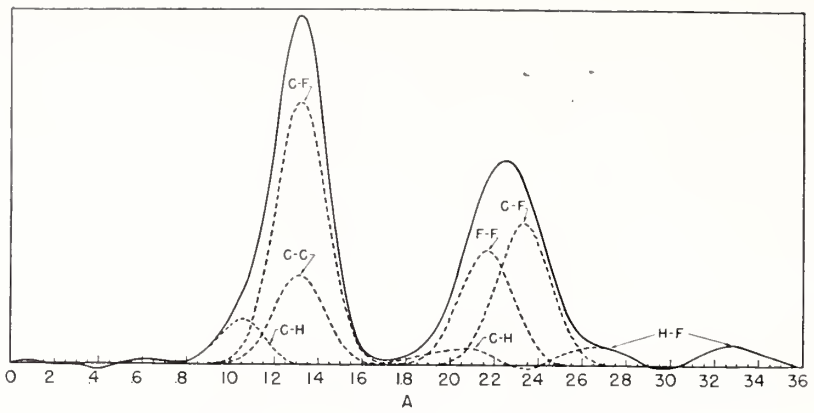
$$\sum_{i,j} c k_{ij} \frac{P_{ij}(r)}{r} = \frac{2}{\pi} \int_0^\infty s I_{\mathrm{mol}}(s) \sin sr \, ds, \tag{1}$$

where P_{ij} is the probability distribution for the distance, r, between the ith and jth atoms, c is a scale factor equal to unity if the areas under the probability curves are unity, k_{ij} is a constant equal to the product of the atomic numbers of the ith and jth atoms, I_{mol} is the experimental molecular scattering and s is equal to $4\pi \sin(\theta/2)/\lambda$, where θ is the scattering angle and λ is the electron wavelength. Generally, the probability distributions for interatomic distances have a Gaussian shape, as predicted from theoretical considerations. [7]. In these cases the maxima of the Gaussian distributions give the equilibrium interatomic distances occurring in the molecule, and the shapes of these distributions are related to the average deviations of these distances from equilibrium owing to the vibrational motion of the molecule.

The application of eq (1) to the study of several molecules is illustrated in figures 12.1 to 12.5, inclusive. The radial distribution curve for $CH_2CF_2^5$ is shown in figure 12.1. It should be noted that this curve, a Fourier sine transform computed from experimental data, is non-negative as is to be expected from its interpretation as related to the probability distribution of the distances occurring in the molecule. The curve is decomposed into the component distances on the basis that the components have a gaussian shape, and that their areas are related in terms of the atomic numbers of the atoms involved.

[1] U. S. Naval Research Laboratory, Washington 25, D. C.

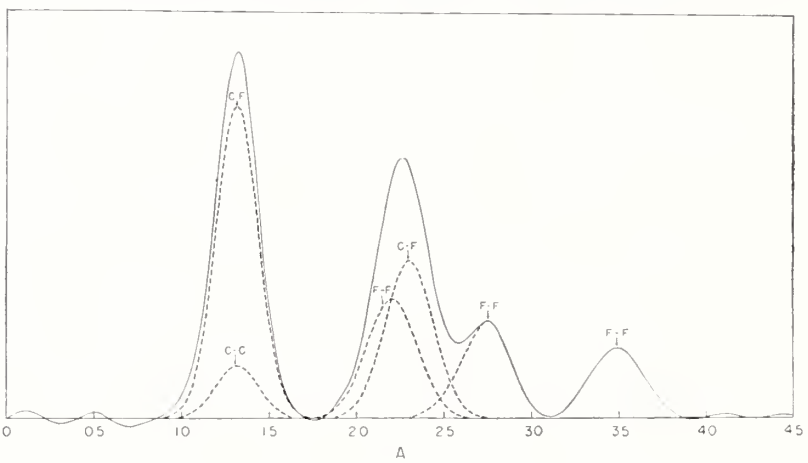
[2] Figures in brackets indicate the literature references on p. 110.



DISTANCE	C-H	C-C	C-F	C-H	F-F	C-F	H-F	H-F
nZ_iZ_j	4	12	36	4	27	36	6	6
ACTUAL AREA	4 2	12.1	360	4 1	272	360	54	7

FIGURE 12.1. *Radial-distribution curve for* CH_2CF_2 *showing component distances.*

All radial-distribution curves illustrated here are related to true probability distributions by means of a folding theorem.



DISTANCE	C-F	C-C	F-F	C-F	F-F	F-F
nZ_iZ_j	36	6	27	36	27	27
ACTUAL AREA	359	60	272	361	268	270

FIGURE 12.2 *Radial-distribution curve for* CF_2CF_2 *showing component distances.*

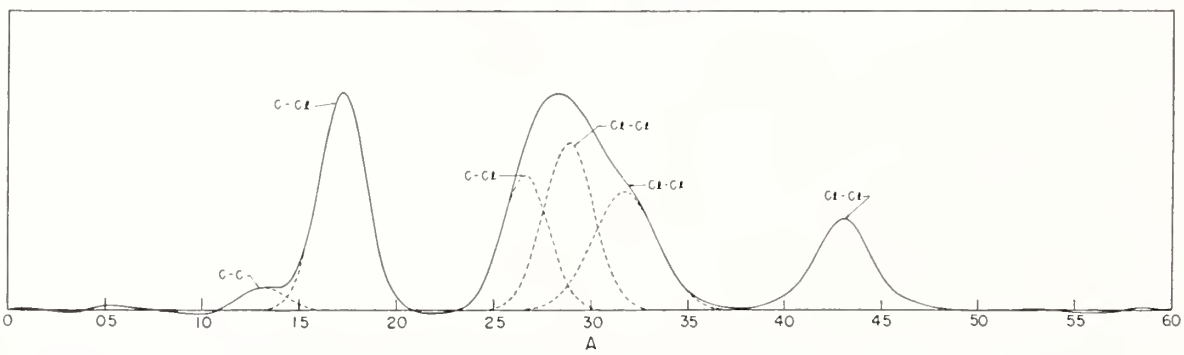


FIGURE 12.3. *Radial-distribution curve for* C_2Cl_4 *showing component distances.*

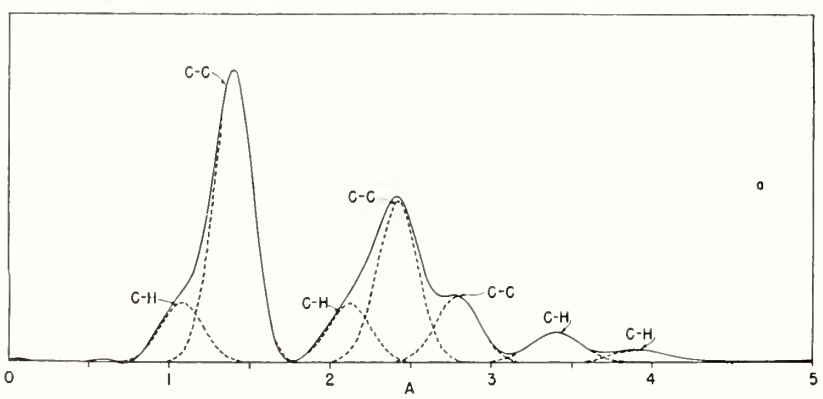


FIGURE 12.4. *Radial-distribution curve for benzene showing component distances.*

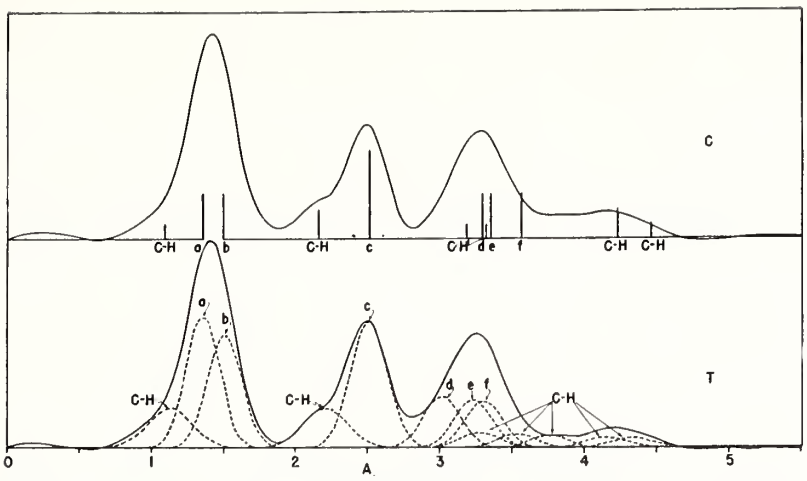


FIGURE 12.5. *Radial-distribution curve for cyclooctatetraene showing component distances.*

The upper curve shows the unsatisfactory distribution of distances at the third peak for the crown model. The lower curve shows the correct tub form.

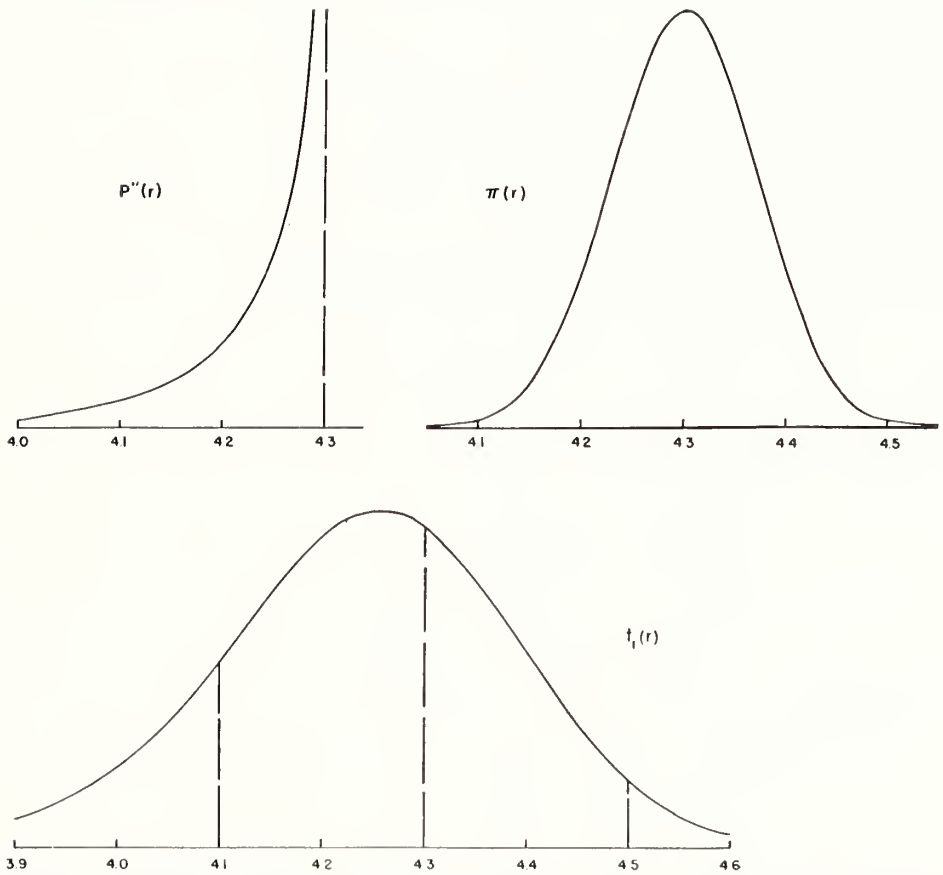


FIGURE 12.6. *Theoretical distribution function, $t_1(r)$, for trans-distance affected by torsional oscillation and general vibration of the molecular frame.*

The separate distribution functions for torsional oscillation, P(r), and for the contribution from the vibration of the molecular frame, H(r), are shown. The dotted line at 4.3 A indicates the shift of the maximum of the $t_1(r)$ curve from the equilibrium position. These distributions are related to true probability distributions by means of a folding theorem.

These two criteria are quite restrictive. Distances involving hydrogen appear clearly resolved in figure 12.1, and the structure and internal motion of the molecule are readily obtained. The same analytic features are seen in the radial distribution curves for CF_2CF_2 [5], figure 12.2, CCl_2CCl_2 [8], figure 12.3, benzene [9], figure 12.4, and cyclooctatetraene [9], figure 12.5.

In special cases, such as for molecules with internal torsional oscillation [10], the probability distributions are not Gaussian and their analysis is somewhat more complicated As indicated in figure 12.6, the probability distribution for the *trans*-position is asymmetric,

being a combination of the over-all vibration of the molecular frame and the internal torsional oscillation. It is also seen from figure 12.6, that the equilibrium distance is displaced from the maximum of the probability distribution. Figure 12.7 shows the approximate shape of the potential hindering rotation in a symmetric di-substituted ethane such as 1,2-dichloroethane. The radial distribution curve for 1,2-dichloroethane [11] is shown in figure 12.8. The *trans*-Cl–Cl peak at about 4.30 A shows the characteristics exhibited in figure 12.6. A detailed study of the radial-distribution curve and the theoretical intensity curves showed that the equilibrium value for the *trans*-Cl–Cl distance was about 0.02 A larger than the maximum of the *trans*-Cl–Cl peak. Experimental uncertainties prevented an accurate evaluation of the torsional motion in this molecule. However, a similar analysis applied to other molecules may give more information. A Cl–Cl peak is seen to appear at about 3.35 A in figure 12.8. This corresponds to the *gauche*-form wherein a Cl atom is rotated about 120° from the trans-position. The *gauche*-form is found to occur to the extent of 27-percent, which is in good agreement with spectroscopic measurements based on the temperature variation of infrared bands [12] and dipole moment measurements [13].

It is of interest to compare the values for the average vibrational motion of the interatomic distances obtained from electron diffraction with those which may be computed from a force model and a measurement of the fundamental frequencies. For performing the spectroscopic computations, the following relations are required [4, 7],

$$\overline{l_{ij}^2}^{\frac{1}{2}} = \sum_n C_n^2 / 2B_n, \tag{2}$$

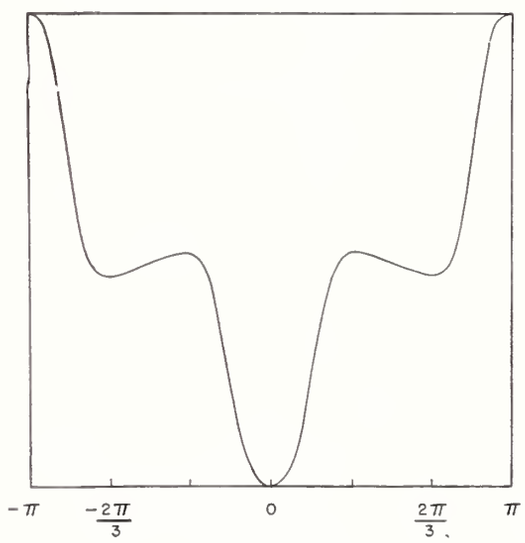


FIGURE 12.7. *Idealized potential function restricting internal rotation in 1,2-dichloroethane.*

Only the regions around the minima affect the electron diffraction patterns.

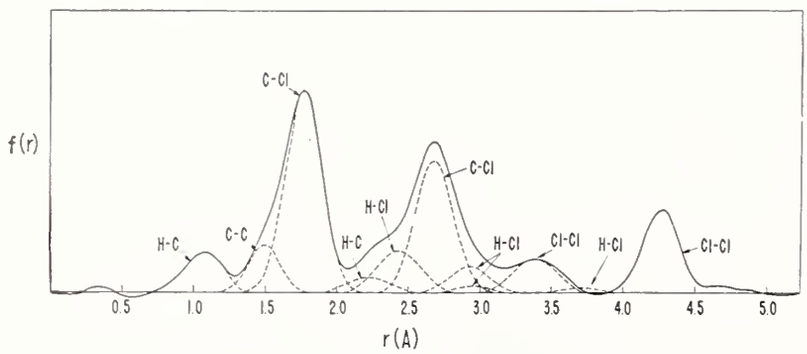


FIGURE 12.8 *Radial distribution curve for 1,2-dichloroethane showing component distances.*

$$B_n = \frac{4\pi^2 \nu_n}{h} \tan \frac{h\nu_n}{2kT}, \tag{3}$$

$$C_n = [\xi_{ij}(\alpha_{in} - \alpha_{jn}) + \eta_{ij}(\beta_{in} - \beta_{jn}) + \zeta_{ij}(\gamma_{in} + \gamma_{jn})]/r_{ij}, \tag{4}$$

where $\overline{l_{ij}^2}^{\frac{1}{2}}$ is the root mean square amplitude of vibration projected on the line connecting the atoms at equilibrium, ν_n is the frequency of the nth normal vibration, h is Planck's constant, k is the Boltzmann constant, T is the absolute temperature, ξ_{ij}, η_{ij}, and ζ_{ij} are the components of the equilibrium distance, r_{ij}, projected on the x, y, z axes of some cartesian coordinate system, and α_{in}, β_{in}, γ_{in} are the coefficients of the nth normal coordinate in the transformation of the x, y, z components of the displacement of the ith atom to normal coordinates.

Some of the very few results obtained so far are shown in table 12.1 The spectroscopic computation for CCl_4 was based on a central force model which has since been shown to be incorrect [14]. The lack of agreement with the electron diffraction results was to have been expected and the spectroscopic computations should be repeated with a general force model. In the case of CO_2, there is no question about the validity of the spectroscopic computations and the good agreement with the electron diffraction results is gratifying.

TABLE 12.1. *Comparison of electron-diffraction and spectroscopic determinations of the average vibrational amplitudes, in* CCl_4 *and* CO_2

	r_0	$\overline{l_{ij}^2}^{\frac{1}{2}}$	
		Electron-diffraction	Spectroscopic
CCl_4			
C-Cl	1.770±0.010	0.041±0.005	0.055
Cl-Cl	2.877±0.010	.054±0.005	.071
CO_2			
C-O	1.162±0.010	0.034±0.003	0.034
O-O	2.310±0.020	.040±0.007	.041

At present we are comparing $\overline{l^2}^{\frac{1}{2}}$ values from spectroscopic and electron-diffraction methods for CX_2–CX_2-type molecules. A preliminary result is shown in table 12.2. For two somewhat different interpretations of the fundamental frequencies the values are obtained for $\overline{l_{\mathrm{Cl\text{-}Cl}}^2}^{\frac{1}{2}}$ in the CCl_2-group. They are seen to be considerably different. In this case the electron-diffraction result falls between the spectroscopic ones and the question of the correct fundamental frequencies is not resolved. However, in the complete spectroscopic analysis, the amplitudes of all the distances in the molecule will be computed and compared with the electron-diffraction results. In general, such comparisons should present new information concerning assignments of fundamental frequencies and choices for force models.

TABLE 12.2. *Spectroscopic and electron-diffraction investigations of the average amplitude for the* Cl-Cl *distance in a* CCl_2*-group of tetrachloroethylene*

	ν_{11}	ν_{12}	Spectroscopic $\overline{l^2_{\text{Cl-Cl}}}^{\frac{1}{2}}$
			A
Bernstein	913	318	0.076
Torkington	782	387	.063
Electron-diffraction = 0.070 A $\overline{l^2_{\text{Cl-Cl}}}^{\frac{1}{2}}$			

References

[1] C. Finbak, Avhandl. Norske Videnskaps-Akad. Oslo, I Mat. Naturv. Kl., No. 13 (1937).
[2] P. P. Debye, Physik, Z. **40,** 66, 404 (1939).
[3] O. Hassel and H. Viervoll, Acta Chem. Scand. **1,** 149 (1947).
[4] J. Karle and I. Karle, J. Chem. Phys. **18,** 957 (1950).
[5] I. Karle and J. Karle, J. Chem. Phys. **18,** 963 (1950).
[6] P. Debye, J. Chem. Phys. **9,** 55 (1941).
[7] R. W. James, Physik Z. **33,** 737 (1932).
[8] I. Karle and J. Karle, J. Chem. Phys. **20,** 63 (1952).
[9] I. Karle, J. Chem. Phys. **20,** 65 (1952).
[10] J. Karle and H. Hauptman, J. Chem. Phys. **18,** 875 (1950).
[11] J. Ainsworth and J. Karle, J. Chem. Phys. **20,** 425 (1952).
[12] H. J. Bernstein, J. Chem. Phys. **17,** 258 (1949).
[13] Mizushima, Morino, Watanabi, Simanouti and Yamaguchi, J. Chem. Phys. **17,** 591 (1949).
[14] J. Rosenthal, Phys. Rev. **46,** 730 (1934).

13. Fundamental Problems of Theoretical Electron Optics

By W. Glaser [1]

The image on either the fluorescent screen or on the photoplate of the electron microscope is given by the distribution of density of electron current existing there. It is the fundamental problem of theoretical electron optics to determine this distribution of current density from given data pertaining to the object and the image-forming field. Figure 13.1 shows this situation schematically. Electron rays as near to monochromatic as possible strike upon the object and are scattered. In the plane $z=z_0$ immediately behind the object, there exists a distribution of current-density with the z-component $J_z(z_0,x_0,y_0)$.

The determination of this function is the task of the theory of scattering, which forms one of the subjects of this symposium. The specific problem of electron optics, with which we shall deal, is the calculation of the distribution $J_z(z,x,y)$ in the different reference planes from the distribution $J_z(x_0,y_0,z_0)$ in the object plane. It is true that a direct relation between these two distributions exists according to geometrical optics, but does not exist from the point of view of wave mechanics, which is valid here. Such a relation, however, exists according to Schrödinger's equation for the corresponding wave functions ψ from which the current density is deduced through the well-known relation

$$I_z=\frac{h}{2im}\left(\psi^*\frac{\partial\psi}{\partial z}-\psi\frac{\partial\psi^*}{\partial z}\right). \tag{1}$$

The wave mechanical treatment of scattering yields the wave function in the object plane. From that we must calculate the solution of Schrödinger's equation in the reference plane.

Demonstration of an Object-True Image Formation Based on Schrödinger's Equation

Generalization of Kirchhoff's Diffraction Formula in Space Containing a Field

In light optics, provided the index of refraction is constant, and in wave mechanics, in the case of the field-free space, our problem is solved with satisfying accuracy by Huygens-Kirchhoff's principle. Together with my collaborator, P. Schiske, in Vienna, we have established a formula that has the same significance for the paraxial range

[1] Technische Hochschule, Wien, Austria.

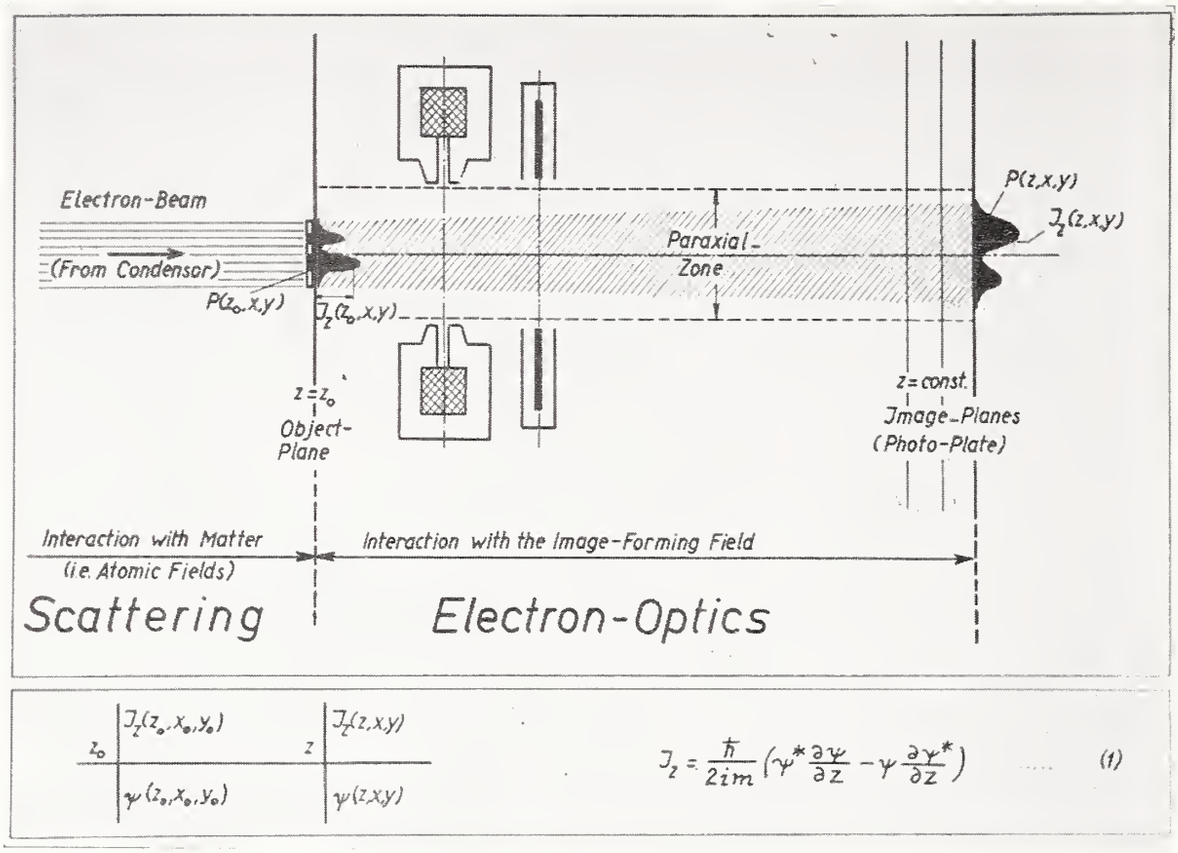


FIGURE 13.1. *Electron-optical image-formation wave mechanically considered.*

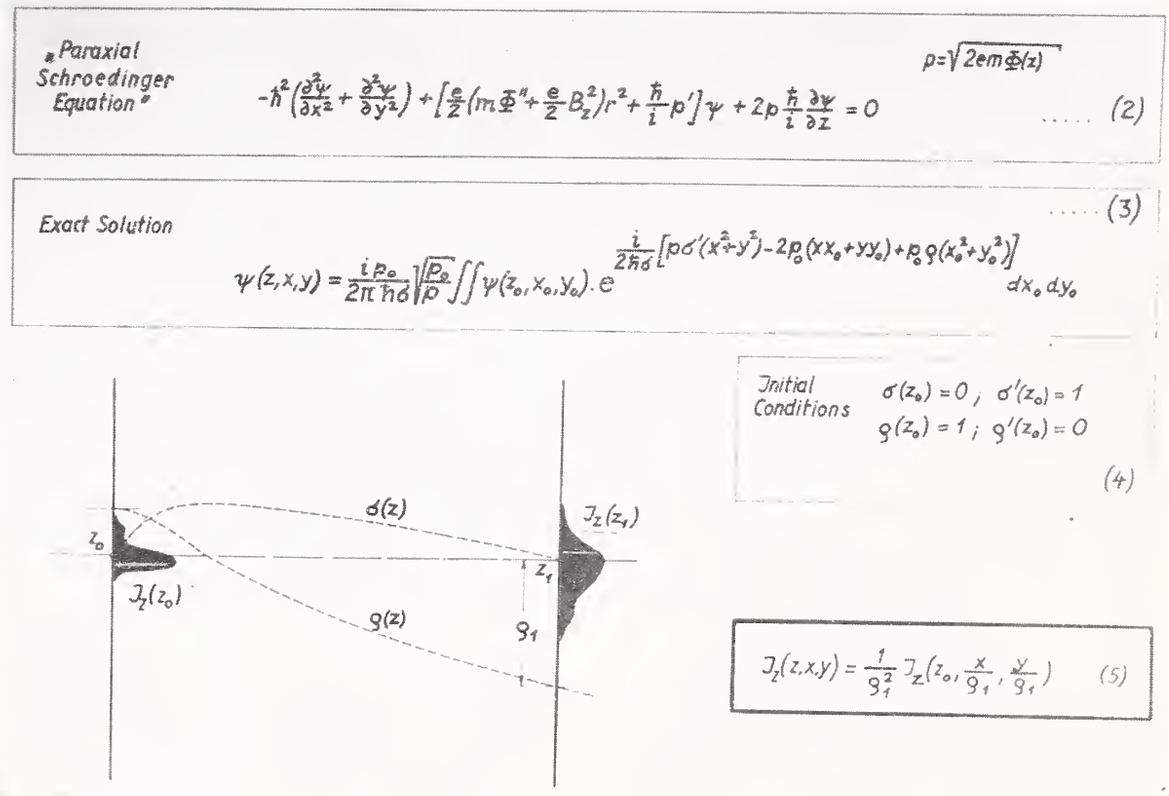


FIGURE 13.2. *Proof of object-true image formation resulting from Schrodinger's equations.*

of a field of rotational symmetry. This paraxial Schrödinger equation reads as follows

$$-\hbar^2\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right)+\left[\frac{e}{2}\left(m\Phi''+\frac{e}{2}B_z^2\right)r^2+\frac{\hbar}{i}p'\right]\psi+2p\frac{\hbar}{i}\frac{\partial\psi}{\partial z}=0, \quad p=\sqrt{2em\Phi}. \tag{2}$$

This equation admits an exact integration and leads to the solution

$$\psi(z,x,y)=\frac{{p_0}^{3/2}}{2\pi i\hbar\delta\sqrt{p}}\iint\psi(z_0,x_0,y_0)\exp\left(\frac{i}{2\hbar\sigma}[p\sigma'(x^2+y^2)-\right.$$

$$\left.2p_0(xx_0+yy_0)+p_0\rho(x_0^2+y_0^2)]\right)dx_0dy_0. \tag{3}$$

Here $\psi(z_0,x_0,y_0)$ is the given wave function in the object plane, while ρ and σ represent the solutions of an ordinary differential equation of second order. This equation coincides with the known differential equation for paraxial trajectories. ρ and σ fufill the initial conditions

$$\sigma(z_0)=0;\ \sigma'(z_0)=1$$

$$\rho(z_0)=1;\ \rho'(z_0)=0 \qquad \text{(figure 13.2)}. \tag{4}$$

According to formula (3) and to formula (1) we are able to determine the current density in the different reference planes as long as the function σ not vanish.

Let z_1 be next zero of σ, the first being the object position $z=z_0$. Moving the reference plane into the plane $z=z_1$ and proceeding to the limit $z\to z_1$, we obtain by means of (3) the relation

$$I_z(z,x,y)=\frac{1}{\rho_1^2}I_z\left(z_0,\frac{x}{\rho_1},\frac{y}{\rho_1}\right) \tag{5}$$

for the current density. Choosing in the image plane $z=z_1$ a ρ_1-times greater length unit, than in the object plane and measuring the intensity there in a ρ_1^2-times greater unit, we learn that the distribution of current density in the image plane is completely identical with the distribution in the object plane. The image position z_1 and the magnification ρ_1 coincide with the corresponding attributes of the geometrical electron optics. Herewith is demonstrated the object-true image formation based on the wave-mechanics.

Fresnel's Diffraction Phenomena

In each reference plane other than the image plane $z=z_1$, there is in general a Fresnel-diffraction phenomenon. Since for the magnetic bell-shaped field—with which L. Marton and his coworkers have made such interesting applications [1, 2]—the trajectories ρ and σ are known exactly, we are able to evaluate our formula for various objects. As an example we chose, because of its simplicity, the image formation of a narrow fringe (zinc oxide crystal or split diatom) by parallel illumination (fig. 13.3). In this case the Fresnel diffraction fringes were first observed by H. Boersch, J. Hillier, and E. Ruska [3, 4, 5]. In accordance with our general considerations, the distribution of intensity in the

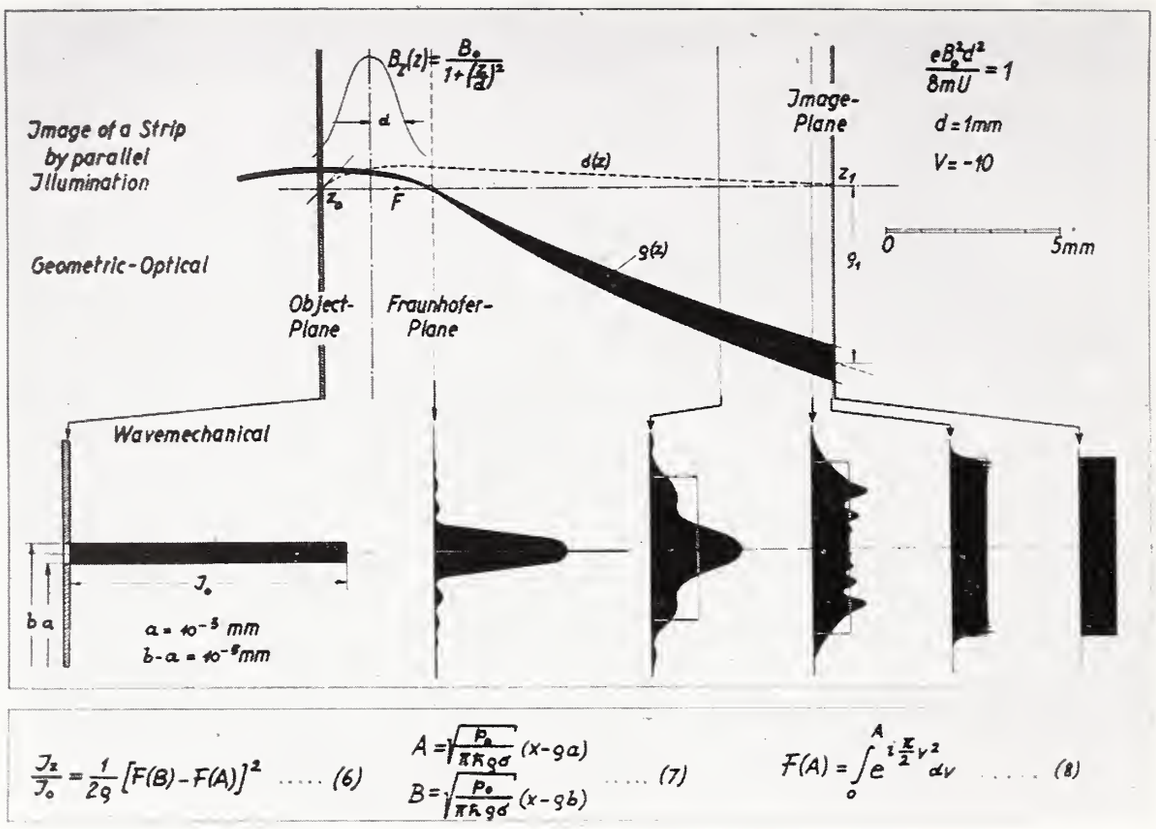


FIGURE 13.3. *Fresnel's diffraction phenomena.*

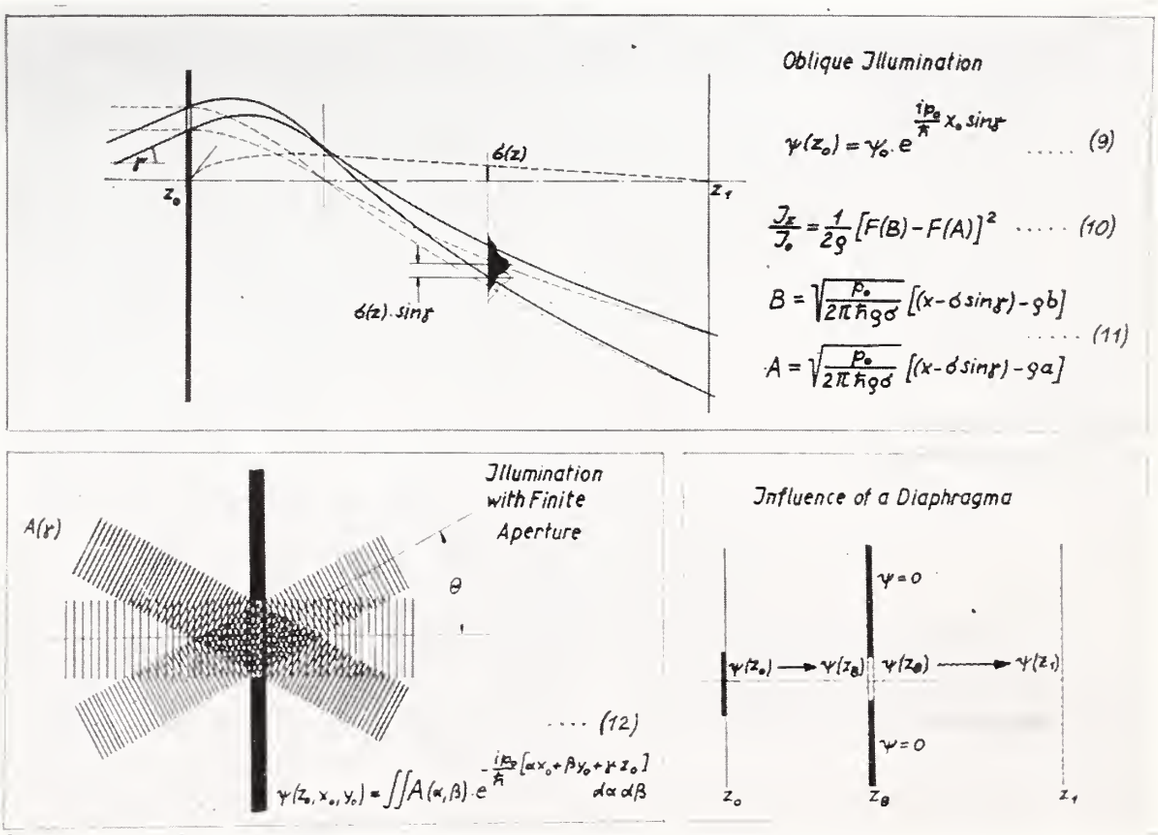


FIGURE 13.4. *Fresnel's diffraction phenomena.*

image plane is an image of the corresponding intensity distribution in the object plane. The intensity distribution in a few different reference planes have been evaluated numerically. It is to be seen distinctly, that with the approach to the image plane, the diffraction fringes come closer to each other until at the boundary of the beam, as determined by geometric optics, they form a sharp edge in the image plane. In the plane, where ρ vanishes, we have a Fraunhofer diffraction phenomenon. We call this plane (generally different from the focal plane) the "Fraunhofer plane." Only in the case where the object lies in a field-free space does this plane coincide with the focal plane, as in the light optics. We imagine now that this slit will be illuminated by a parallel beam with the inclination γ. Instead of $\psi(z_0,x_0,y_0)=$const. we must now write

$$\psi(z_0, x_0, y_0) = \psi_0 e^{\frac{2p_0}{\hbar} x_0 \sin \gamma}. \tag{9}$$

From this calculation it follows that in the different reference planes $z=$const there exist the same distributions of intensity as in the first case, but they are displaced by the distance $\delta(z,\gamma)=\sigma \sin \gamma$ If we illuminate the object by a few more parallel beams, the inclinations of which vary in the aperture Θ, we have to superpose the corresponding intensities in the different planes of reference (fig. 13.4).

Also the influence of an aperture, for example, the aperture of contrast, can be comprehended in the way that one first ascertains the wave function in the aperture plane according to formula (3). The wave function in the image plane will be obtained by a second application of our formula, whereby we have to integrate only over the aperture.

Since every wave function can be represented by superposition of plane waves

$$\psi(z_0, x_0, y_0) = \iint A(\alpha, \beta) e^{-\frac{ip^0}{\hbar}(\alpha x_0 + \beta y_0 + \gamma z_0)} \, d\alpha d\beta, \tag{12}$$

we are able to extend our considerations to the general case. Thus by means of our formula we can translate the corresponding considerations of wave theory in light optics.

Our formula can immediately be extended to an image forming system with axial astigmatism, since the paraxial Schrödinger equation admits exact integration also in such a field.

One can further establish a consequent wave-mechanical theory of aberrations, either generalizing the Kirchoff principle or proceeding from the paraxial Schrödinger equation and using one of the well-known perturbation methods of wave mechanics. The calculations pertaining to this are still in progress.

Question of Resolving Power

The formula

$$\delta = \frac{0.6\lambda}{n \sin \gamma} \tag{1}$$

for the limit of resolution according to Helmholtz [5a] and Lord Rayleigh [6], is for a self-luminous point, radiating equally in all directions (fig. 13.5). However, from the point of view of the wave

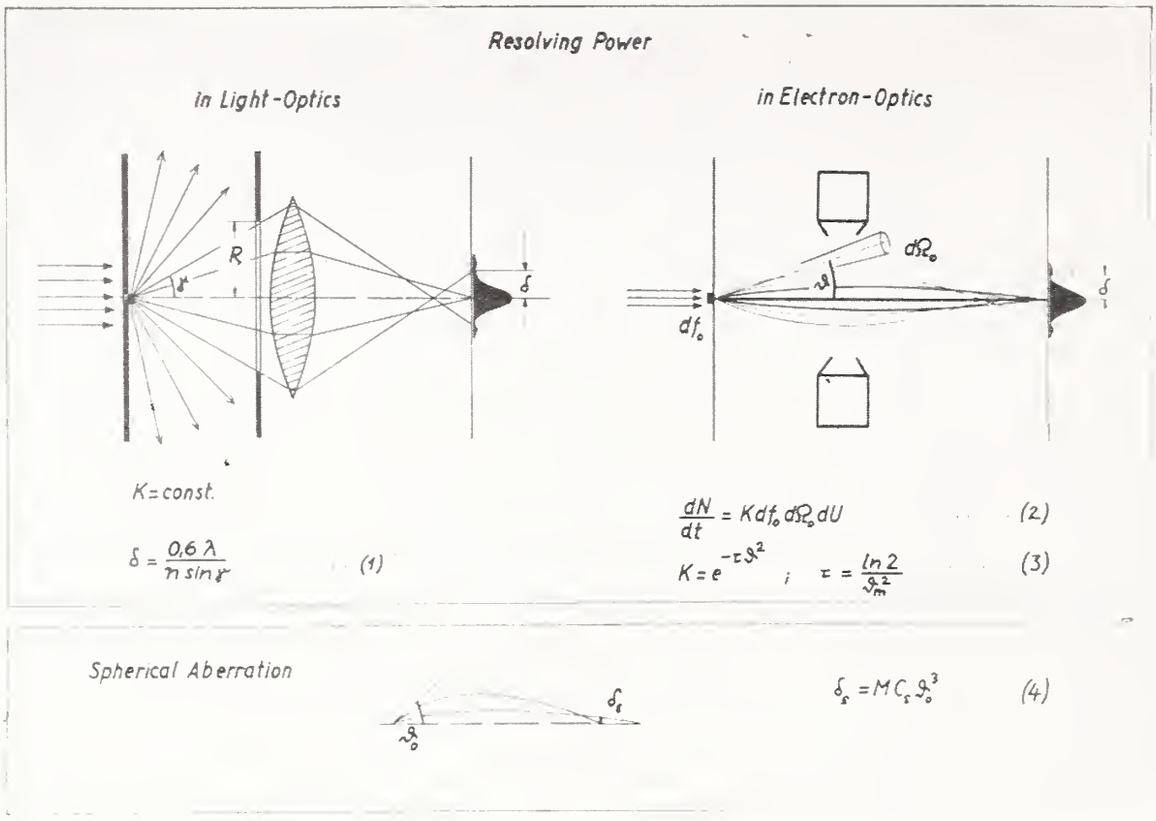


FIGURE 13.5. *Resolving power.*

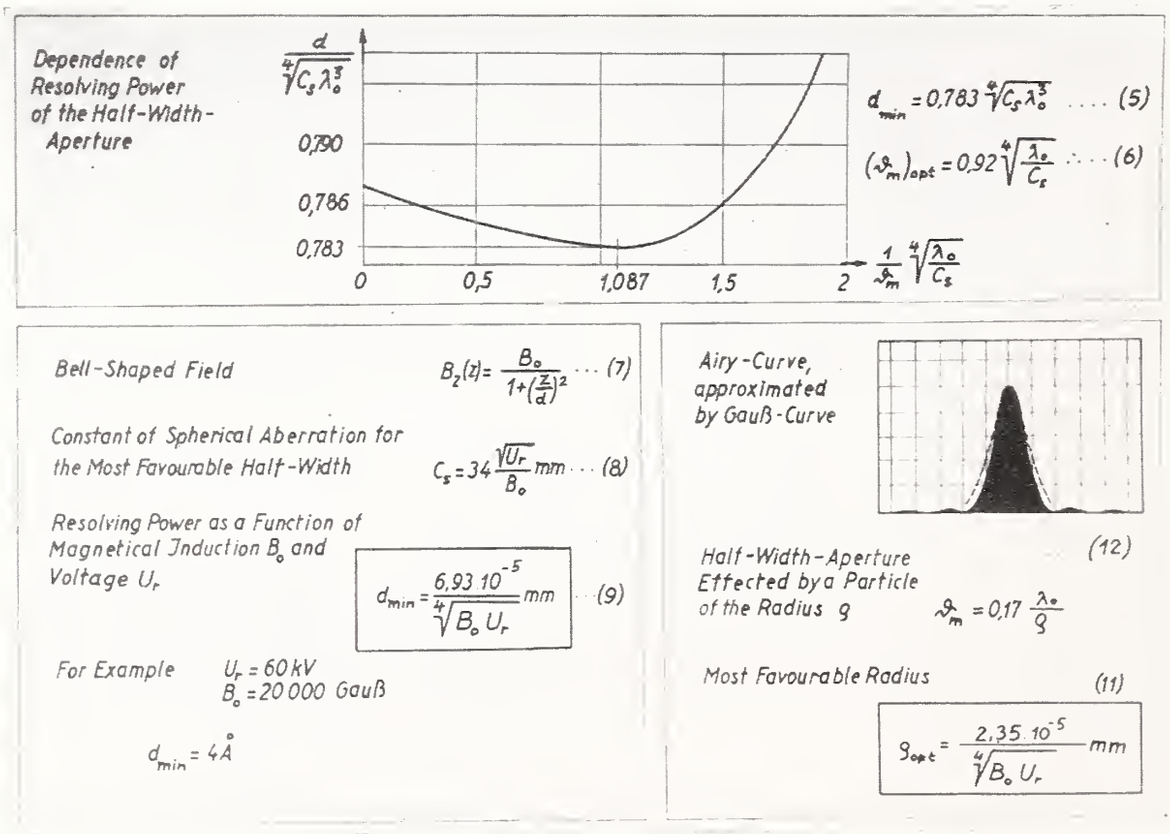


FIGURE 13.6. *Resolving power.*

theory, it is not permissible to speak about a radiating "point." It follows, especially from the work by M. v. Laue [7] on the degrees of freedom of radiation bundles, that there are only radiating surfaces which send out neither exactly monochromatic nor exactly parallel bundles. Nevertheless, in light optics it is possible on account of the great wavelength to realize a point source by illuminating a small aperture in an opaque screen.

Since in electron optics the wavelength is a great deal shorter than the atomic radius, it is impossible in this way to realize an object point radiating equally in all directions. There is a further difference in that the existence of an aperture in light optics is essential, whereas in electron optics one usually works without an aperture. Both facts are intimately connected. Objects commonly used in electron optics to measure the resolution power are large relative to the de Broglie wavelength, and scatter the greater part in the direction of illumination. In spite of these essential differences, this resolving formula has been repeatedly and uncritically transferred from light optics into electron optics.

We should proceed more correctly in the following way [8]. Let us suppose the surface element df_0 of the object emits in unit time into the solid angle $d\Omega$ the number

$$\frac{dN}{dt}=Kdf_0d\Omega_0dU \tag{2}$$

of electrons with velocities in the range dU. The radiation characteristics K is dependent upon the scattering angle and has to be taken from the theory of scattering. Without dealing with details, we can use the fact that K decreases rapidly with the scattering angle. In order to interpret it schematically we have used for our calculations the formula

$$K=K_0e^{-\tau\theta_0^2}. \tag{3}$$

The exponent τ characterizes the rapidity of decrease with the scattering angle. It is related to the half-aperture by the formula

$$\tau=\frac{ln\ 2}{\theta_m^2} \tag{3}$$

We further used the fact that the image space of high magnification electron lenses is practically field free. Consequently, it is possible to apply P. Debye's optical formula for the wave function of a bundle of rays with sperical aberration

$$\delta_s=MC_s\theta_0^3 \tag{4}$$

to electron optics. From these numerical calculations, which were carried out by my coworkers Krammer and Titze, we obtained the resolving power as a function of the halfaperture, the plot of resolving power against halfaperature is represented in figure 13.6. We see that the resolving power is a minimum where the halfaperture is

$$(\theta_m)_{\text{opt}}=0.92\sqrt[4]{\frac{\lambda_0}{C_s}}. \tag{6}$$

We get C_s for the bell-shaped field, choosing the most favorable distance of pole pieces from the formula

$$C_s = 34 \frac{\sqrt{U_r}}{B_0} \text{ mm}. \tag{8}$$

Substituting this value C_s and the de Broglie wavelength into

$$d_{\min} = 0.783 \sqrt[4]{C_s \lambda_0^3}, \tag{5}$$

we get the optimum resolving power as a function of the voltage U_r and the magnetic field strength B_0 in the form

$$d_{\min} = \frac{6.93.10^{-5}}{\sqrt{B_0 U_r}} \text{ mm}. \tag{9}$$

For the value $U_r = 60$ kv and a magnetic field strength of 20,000 gauss, we obtain from that formula $d_{\min} = 4$A.

We must remember, however, that the optimum conditions were assumed in three places: (1) optimum halfaperture; (2) the bell-shaped magnetic field, favorable in itself with respect to spherical aberration; and (3) optimum distance of pole pieces. Naturally, a perfect rotationally symmetrical field was under consideration. But 9 years ago we showed quantitatively the deciding effect of an axial astigmatism on the resolving power [8a], which 4 years ago J. Hillier and E. Ramberg [9], in an elegant way, were able to eliminate in great part. On the other hand a little better value could result by another choice of reference plane. A comparison of our estimation of the resolving power with the commonly used linear or quadratic superposition of spherical aberration, and diffraction-discs (giving a little more unfavorable values) was not carried out because this procedure seems to be without a physical meaning as in the optical case where leading opticians, among them Capski [10] and Straubel [11] pointed this out many years ago.

Relation Between Particle Size and Resolving Power

We can use the above results to estimate approximately the particle radius most desirable for a determination of the resolving power. Let us suppose that we have a disc of the radius ρ, which we take to be perfectly opaque for simplicity. The radiation characteristic of such an object is given by the diffraction pattern of a circular disc and corresponds to the well-known Airy distribution. In proximity to the main maximum, which is solely important, we replace the Airy distribution by the Gauss curve (3) and we obtain the relation between particle radius and halfaperture

$$\theta_m = 0.17 \frac{\lambda_0}{\rho}. \tag{12}$$

From the above calculated optimum halfaperture, we now obtain the optimum radius for the determination of the resolving power

given by

$$\rho_{opt}=\frac{2.35\cdot 10^{-5}}{\sqrt[4]{B_0 U_r}}\ mm. \tag{11}$$

Such an optimum particle size checks physically in that very small particles yield large apertures and therefore a great deal of spherical aberration, whereas larger particles yield smaller apertures and therefore larger diffraction errors. Naturally, we should have in mind the conditions necessary for the validity of the result. The more closely the condition of opaqueness is approximated, the more accurate the estimation will be. For small particles, consisting only of a few heavy atoms, the above formula for the most favorable particle size can be a certain guide only as long as the exact law of scattering of such particles is not known.

Meaning of Focal Length in Electron Optics

In light optics the relation between object position and image position and the magnification is given by the formula

$$M=\frac{y_1}{y_0}=\frac{f_0}{z_0}=\frac{z_1}{f_1}. \tag{1}$$

We will call it the "Newtonian Equation for Image Formation." In electron optics the corresponding relation is far more complicated and given by formula

$$M=\frac{u(z_1)}{u(z_0)}=\frac{w(z_1)}{w(z_0)}, \tag{2}$$

whereby u and w represent any two independent electron trajectories. If object and image are in the field-free space, as is commonly the case for electrostatic lenses, then it is possible to replace the electron trajectories by their asymptotes

$$M=\frac{u_\infty+z_1u'_\infty}{u_{-\infty}+z_0u'_{-\infty}}=\frac{w_\infty+z_1w'_\infty}{w_{-\infty}+z_0w'_{-\infty}}=\frac{f_0}{z_0}=\frac{z_1}{f_1} \tag{3}$$

This leads to a Newtonian equation for image formation, the focal distance of which is defined by figure 13.7. For distinctness we will call it "virtual" focal distance. Another definition of focal length is represented in the same figure. We speak in this case of the "real" focal-length, and we shall see that this is the most important definition in electron microscopy. If object and image are in the image forming field, the Newtonian equation is not generally valid and both focal distances just defined are without meaning. It is, therefore, not permissible in all cases to operate with focal distances and lens equations as it happens repeatedly in electron optics (Nevertheless, the existence of an optical image formation, which is our sole concern in electron optics is not affected by this fact).

One can try as well as possible to represent the general electron optical equation for image formation (2), at least in the neighborhood of a certain point, by a Newtonian equation (1) [12]. This is indeed possible and one is led to the concept we have called (together with O. Bergmann) the "Osculating Newtonian Image Formation". This

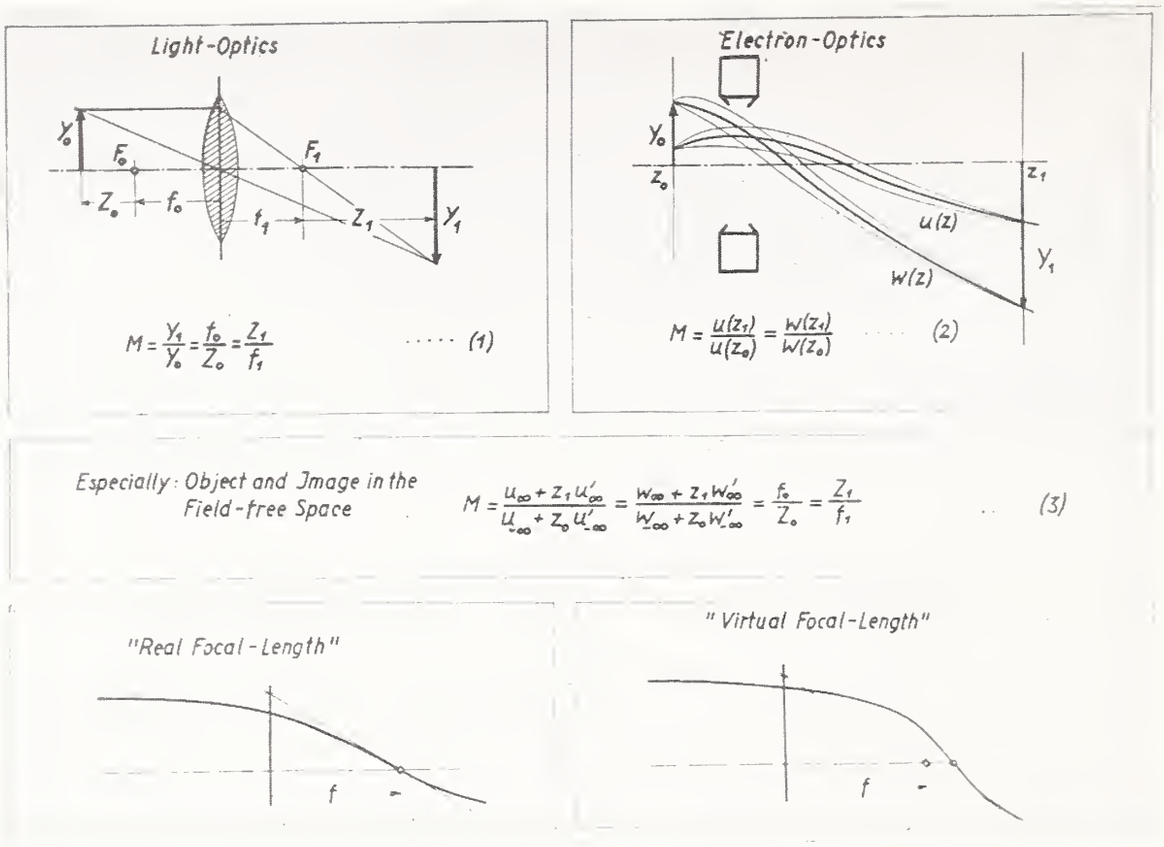
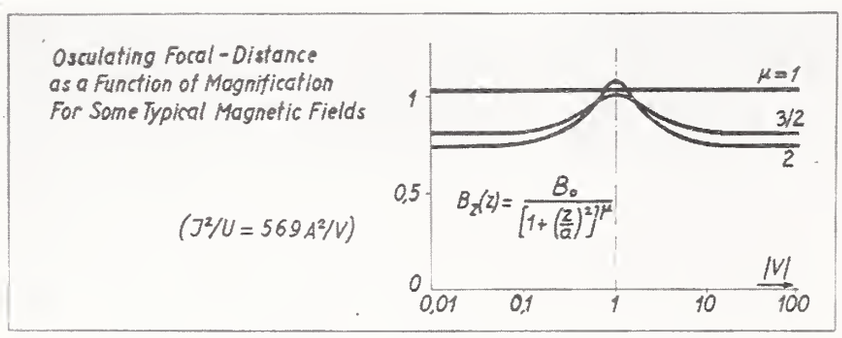


FIGURE 13.7. *The meaning of focal length.*



	FOCAL-DISTANCE REAL	VIRTUAL	OSCULATING
Generally.	○	○	● Is Dependend of Magnification
High Magnification in All Fields	●	○	● Identical with Real FD
Extrem Short Fields	● Identical	● Identical	●
Object and Image in Field-Free Space	○	● Identical	●
Newtonian Fields	●	○	● Identical with Real FD

○ Cannot Be Used

● Can Be Used

FIGURE 13.8. *The meaning of focal length.*

is characterized by the fact that it represents the general electron optical image formation by Newtonian equation (1). But, in contrast to light optics, the osculating focal distance depends on magnification, in the neighborhood of which it approximates the electron optical image formation. Only for certain special image-forming fields—we called them "Newtonian fields"—the osculating focal distance does not depend on the chosen magnification. In this case the general electron optical image formation is represented, whatever the object position may be, by the same lens equation. Some time ago, together with E. Lammel, we determined all Newtonian fields [13]. The simplest field of this type is the simple bell-shaped magnetic field. The focal distance of this field, and generally of every Newtonian field, is identical with the above-defined real focal distance.

In figure 13.8 the dependence of the osculating focal distance upon the magnification for a certain value of I^2/U^κ is represented for the field of the ring circuit and for two bell-shaped fields with $\mu=1$ and $\mu=2$ [14]. While for the simple bell-shaped field, which belongs to the Newtonian class, the focal distance is constant, the osculating focal distance of the field with $\mu=2$ and of the ring circuit is dependent upon the magnification and varies particularly in the neighborhood of the magnification 1. From this representation one sees that the field of the ring circuit is not a Newtonian field, as was repeatedly asserted. It is also obvious from the figure that for high magnifications, the graph of osculating focal distance becomes a straight line, parallel to the z-axis, i. e., it becomes constant. This can be shown generally. In this way we are able to give an answer to the question, what is the physical meaning of the above-quoted real and virtual focal distance? The answer is that for a small magnification, if object and image are in the field, these focal distances do not have a meaning as far as the equation of image formation is concerned. Only if the field is Newtonian, the common lens equation is valid. In this case the real focal distance must be taken. However, for the highest magnifications, such as are of interest in electron microscopy, and where the object plane is near to the real focal point, the real focal distance is identical with the osculating focal distance for this case, and, therefore, is to be used in the representation of electron optical image formation. At the same time the question of which of the two definitions of the focal distance is the correct one is settled. There are limiting cases, mutually exclusive where one or the other shall be used.

Electron Optical Image Formation as a Problem of Eigenvalue

The determination of paraxial electron trajectories in the general equation for image formation by means of numerical integration is laborious and wearisome. One can, however, avoid these calculations by formulating the electron image formation problem as an eigenvalue problem in which the lens power or the current intensity in the objective represents the eigenvalue [14a]. This point of view corresponds precisely to the procedure in practice, because here is also given the object position $z=z_0$ and the position $z=z_1$ of the screen and one regulates the current intensity in the objective so that a sharp image formation results upon the screen. Translated into mathematics this

means one postulates that the solution $\sigma(z)$ of the differential equation of the paraxial trajectories

$$\sigma'' + k^2 F(z)\sigma = 0;\ F(z) = \frac{B\sigma_z^2}{B_0^2};\ k^2 = \frac{eB_0^2 a^2}{8mu} \tag{1}$$

has zeros at $z = z_0$ and $z = z_1$

$$\sigma(z_0) = 0;\ \sigma(z_1) = 0. \tag{3}$$

This is the case only for a certain eigenvalue k^2, which is calculated.

For determination of k^2 there can be used any one of the numerous methods for approximate calculation of eigenvalues; for example the Rayleigh-Ritz method, etc. It has been shown, that even the simple Rayleigh method (4) (fig. 13.9) yields a very accurate equation for image formation in the form

$$k^2 = f(z_0, z_1) \tag{6}$$

if one puts in (4) as an approximate electron trajectory (i. e., as an approximate eigenfunction $\sigma(z)$) the known trajectory of the bell-shaped field (fig. 13.9) shows the graph of the real focal distance of the ring circuit determined in this way. In order to test the accuracy of the method, three values were also used, which had been determined by means of a numerical path integration, which was carried out as accurately as possible.

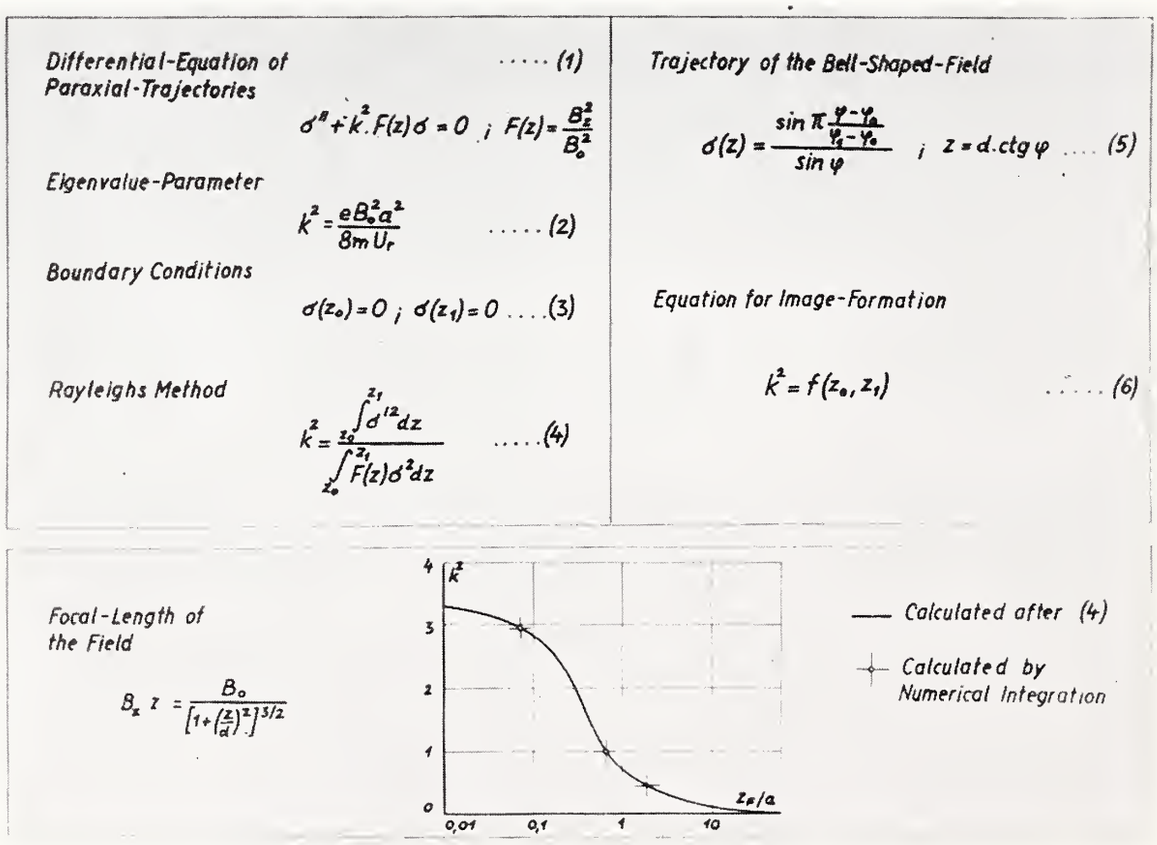


FIGURE 13.11. *Geometric-optical distribution of electron density.*

Image of a point, not on axis, for a typical reference plane. Number mean I_z/I_0 (percent).

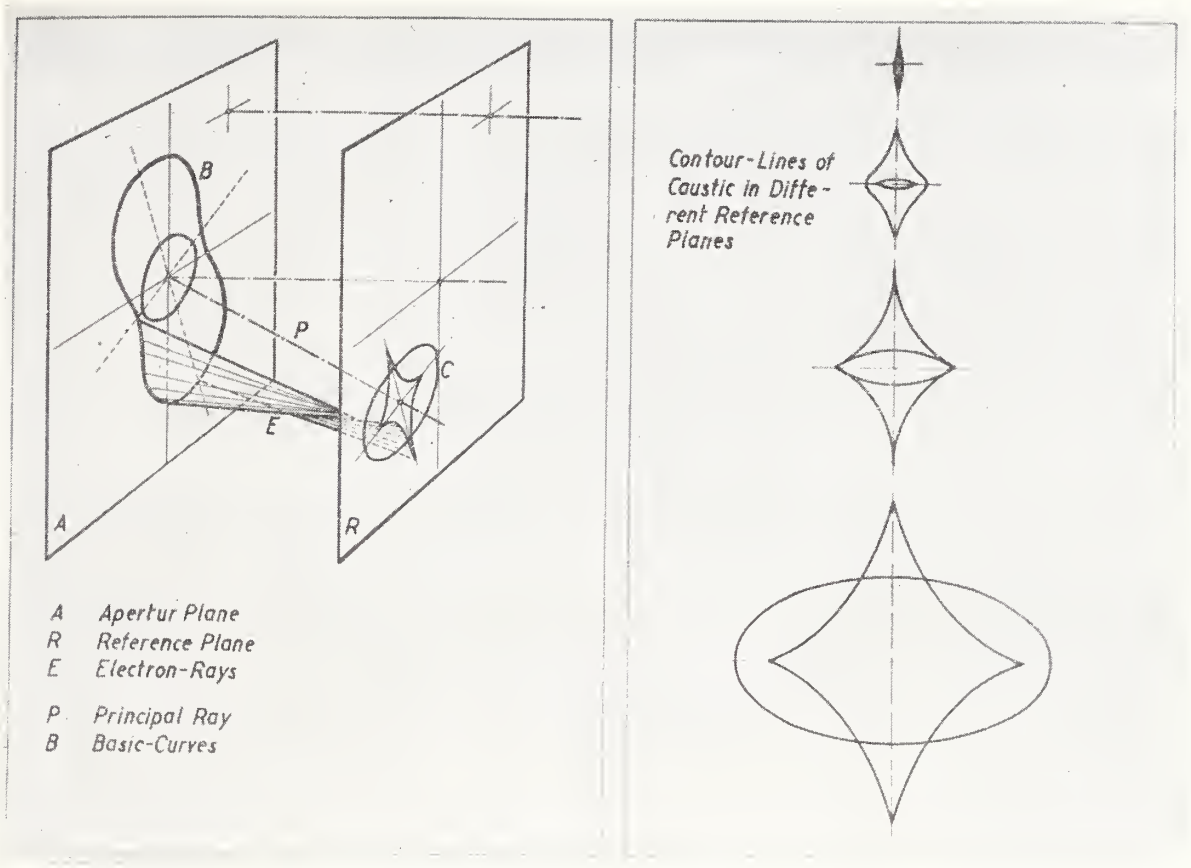


FIGURE 13.9. *Electron-optical image formation as a problem of eigenvalue.*

Geometrical Optical Image of a Luminous Point

In finishing my exposition, I would like to say a few words about the calculation of the electron intensities in image space from the point of view of geometrical optics. The pattern of concentration formation of rays leaving a certain object point is a sharp point only to the first order. In general, however, it is a more complicated pattern, which is called the "caustic". It is the geometrical locus of the points of intersection of neighboring trajectories in the beam and is a kind of surface with two jackets. Figure 13.11 shows the formation of the caustic. To every cross section of the caustic, two "basic curves" generally correspond in the aperture plane. The general form of the caustic has been determined with my workers H. Grümm and G. Hofmann. The main result is that the basic shape of the caustic is the same as in light optics. However, on account of the anisotropic abberations of magnetic lenses, its position in space is different and the caustic surface as a whole is rotated through a certain angle. For the bell-shaped field of which the aberration coefficients are known as a result of calculations, together with E. Lammel, the contour lines were calculated by H. Grümm and the caustic represented in actual size as a model of plaster. The lines of equal electron density in different reference planes, the analog of the isophots in light optics were also calculated.* These graphs determine directly the real geometric optical image on the photoplate by the image formation of a point, by means of an electron lens. It is our intention to determine the corresponding image by means of wave mechanics where a certain

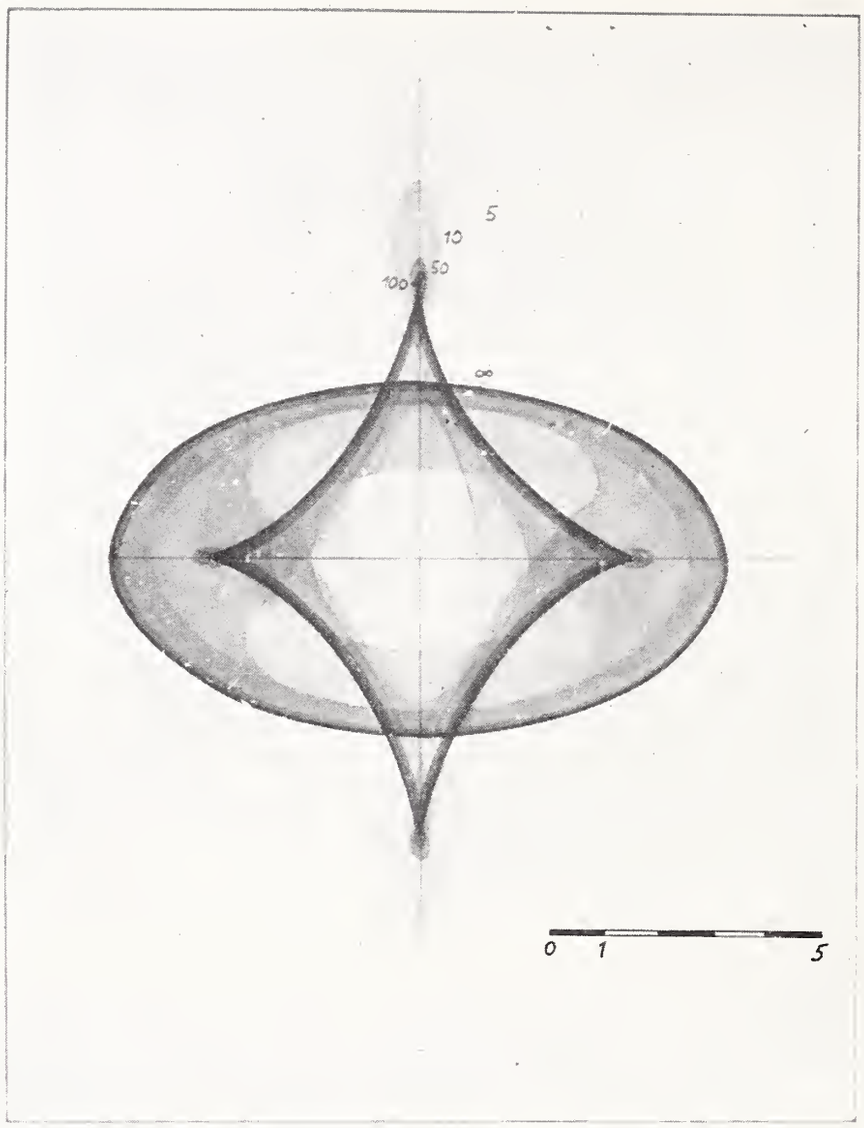


FIGURE 13.10. *Caustic of electron lenses.*

atom or a small particle with its actual radiation characteristic shall be chosen as physical object.

References

[1] L. Marton and R. G. E. Hutter, Phys. Rev. **161,** 65 (1944).
[2] L. Marton and K. Bol, J. Appl. Phys. [6] **18,** 522–529 (1947).
[3] H. Börsch, Naturwissenschaften **28,** 710 (1940).
[4] J. Hillier, Phys. Rev. **58,** 842 (1940).
[5] E. Ruska, Kolloid.-Z. **105,** 43–52 (1943).
[5a] H. Helmholtz, Pogg. Ann. 1874.
[6] Lord Rayleigh, Phil. Mag. **42,** 167 (1896).
[7] M. v. Laue, Ann. Phys. **44,** 1197 (1944).
[8] W. Glaser, Z. Phys. **121,** 647 (1943); Acta Phys. Austriaca **III,** 38 (1949).
[8a] W. Glaser, Z. f. Phys. **120,** 1 (1942).
[9] J. Hillier and R. Ramberg, J. Appl. Phys. **18,** 48–71 (1947).
[10] S. Czapski, Z. Instrumentenk. **8,** 203 (1888).
[11] R. Straubel, Abhandl kgl. bayr. Akad. Wiss. II. Cl, XVIII (München, 1893).
[12] W. Glaser and O. Bergmann, Z. agnew. Math. Phys. **I,** 363 (1950).
[13] W. Glaser and E. Lammel, Ann. Phys. **5,** 367 (1941).
[14] W. Glaser and F. Lenz, Ann. Phys. **6,** 19 (1951).
[14a] P. Funk and W. Glaser, Z. f. Phys. **102,** 603 (1936).
[14b] W. Glaser, Ann. d. Phys., 213 (1950).

[15] W. Glaser, Comm. Congres Intern. Paris 24 (1951).
[16] W. Glaser and H. Grümm, Optik **7,** 96–120 (1950).
[17] G. Hofmann, Diplomarbeit, Prag-München (1950).
[18] W. Glaser and H. Lammel, Arch. Electrotech. **37,** 347–356 (1943).
[19] H. Grümm, Optik. **9,** 281 (1952).

*Note added October 21, 1953: For a more detailed treatment, see W. Glaser, Grundlagen der Elektronenoptik (Springer, Wien, 1952).

Discussion

Dr. M. E. Haine, Associated Electrical Industries, England: I wonder if Dr. Glaser would tell us what the actual value of spherical aberration constant was?

Dr. W. Glaser: Two tenths to one millimeter, corresponding to the values of voltage and magnetic field strength, according to our general formula in our lecture.

Dr. D. Gabor, Imperial College of Science and Technology, City and Guild College, London, England: The reference plane is the Gaussian plane?

Dr. Glaser: Yes, the reference plane is the Gaussian image plane. This choice is the most natural one, and the general formula can be worked out in the easiest way in this case. But calculations concerning other reference planes are still in progress. It must yet be considered that our scattering characteristic is a hypothetical one, and, therefore, it does not seem to be so important to take into account other reference planes also.

Dr. Gabor: Well, we will see how that compares with the figures of Dr. Marechal, who determined the optimum position of reference plane in light optics.

Dr. Glaser: It is not possible to transfer considerations of light optics to electron optics directly. In light optics, the resolution is commonly calculated for a spherical wave, which radiates in all directions. Spherical aberration is not considered. This case is not realized in electron optics. Here you have a natural decay of intensity with scattering angle, and therefore you can work without an aperture. In light optics, the radiation characteristics commonly are not taken into account.

Dr. Gabor: Well, except in very high apertures. There are some calculations that will help you.

Dr. Glaser: Yes, of course, I know the interesting papers of H. Hopkins and others. But, although they are fully adequate to light optics, they can't help me in electron optics, because they do not take into account the spherical aberration, which is of deciding influence in electron optics.

Dr. V. E. Coslett, University of Cambridge, England: I did not understand the concept that you put forward of the optimum size of the test particles with respect to resolution. From the figures you gave there, it appeared that the optimum size was distinctly smaller than the resolution you obtained.

Dr. Glaser: Our calculations have shown that there exists, on account of spherical aberration, an optimum aperture. The aperture is dependent upon the size of the scattering particle. Therefore, you get an optimum size of two particles in order to distinguish them in the best way. In light optics, where you have corrected objectives, the resolution is the better the greater the aperture, and therefore

the particle size is smaller. Condition is, of course, a sufficient contrast.

Dr. Coslett: From a practical point of view, your resolution here is a separation of two objects. You take a photograph and say the resolved distance is so and so. On your figures the actual particle diameter needs to be two-thirds of that size.

Dr. Glaser: Yes, the diameter is two thirds of the distance of the particles. This condition cannot be perfect for small particles that contain only a few atoms. In this case, we must take into account the real scattering characteristics of such a particle.

14. Über die Bildentstehung im Elektronenmikroskop [1]

von H. Boersch [2]

Auflösungsvermögen und Kontraste sind, abgesehen von Präparationsfragen, die Hauptprobleme der mikroskopischen Abbildung. Beide Problemkreise sind eng miteinander verknüpft. So wird z. B. durch Änderung der Objektivapertur gleichzeitig Auflösungsvermögen und Kontrast geändert. Diese Untersuchung ist nun hauptsächlich dem Kontrastproblem gewidmet, allerdings endet sie, fast zwangsläufig im Auflösungsproblem.

Die Bildung der Kontraste im Elektronenmikroskop kann unter gewissen Umständen sehr einfach auf den Intensitätsverlust durch Streuabsorption [1–16] [3] zurückgeführt werden. Diese Art der Kontrastbildung ist für die Objekte gegeben, deren *Ausdehnung gross* gegen das Abbe'sche Auflösungsvermögen ist. Im anderen Extremfall, wenn das Objekt klein gegen das Abbe'sche Auflösungsvermögen ist, führt die Methode der Streuabsorption jedoch zu unrichtigen Resultaten. In diesem Fall muß zunächst die Lichterregung in der hinteren Brennebene des Objektivs bestimmt werden. Durch Interferenz ergibt sich hieraus die Lichterregung im Bilde.

Vom Standpunkt der Bilderzeugung durch Interferenz muß natürlich auch der vorerwähnte Fall des Kontrastes durch Streuabsorption erklärt werden können. Durch hier nicht wiedergegebene Rechnungen hat sich der Verfasser versichert, daß im Fall *ausgedehnter* Amplituden- und Phasengitter Übereinstimmung zwischen beiden Methoden besteht.

Der Fall *kleiner* Objekte ist für die Elektronenmikroskopie insofern von besonderem Interesse, als zwar die Auflösung von Atomabständen mit schnellen Elektronen noch nicht gelungen ist, aber vom Standpunkt der Weiterentwicklung des instrumentellen Auflösungsvermögens durchaus möglich erscheint.

Zwar sind über die Kontraste von Atomen schon einige Untersuchungen veröffentlicht [7, 8, 17–19]. Solange jedoch keine vollständige Lösung dieses schwierigen theoretischen Problems vorliegt scheint es von Interesse, an einigen Extremfällen die zu erwartenden Kontraste zu studieren und gewisse Grenzen aufzustellen, an denen der theoretische und experimentelle Fortschritt in der Behandlung des Problems ermessen werden kann.

An dem Beispiel der Abbildung von einfachen Gittern und Kreisschirmen werden die Beziehungen zwischen der Beugungs- und der Bild-Amplitude dargestellt und auf die Abbildung von Kristallgittern und Einzelatomen übertragen. Hierbei wird vorausgesetzt, daß die Ausdehnung der Atome, gemessen an ihrer Wirkung auf den Elektronenstrahl, klein ist gegen die Atomabtände, und daß nur der Beugungsfehler das Auflösungsvermögen beschränkt. Der Einfluß der sphärischen Aberration kann durch Verringerung der Objektiv-

[1] Vortrag gehalten auf dem Symposium on Electron Physics des National Bureau of Standards in Washington am 7.11.51.

[2] Physikalisch-Technische Bundesanstalt, Braunschweig.

[3] Figures in brackets indicate the literature references on p. 144.

apertur unterdrückt werden. Hierbei tritt selbstverständlich eine entsprechende, aber nicht sehr ins Gewicht fallende, Verminderung des Auflösungsvermögens ein. Außer dem Wunsch nach klaren und übersichtlichen Verhältnissen besteht ein weiterer Grund für dieses Vorgehen darin, daß die Beugungserscheinungen von Atomen und Kristallgittern relativ gut bekannt sind, so daß die Nutzanwendung dieses Verfahrens auf der Hand liegt.

Die folgenden Abschnitte sind den kontraststeigernden Verfahren gewidmet, nämlich der Übertragung des Phasenkontrast-Verfahrens nach Zernicke, des Amplituden-Kontrast-Verfahrens und des Filterverfahrens auf die Elektronenmikroskopie. Besonders eingegangen wird auf das Elektronenfilter-Verfahren und die experimentelle Durchführung dieser Methode im Netz- und Linsenfilter. In speziellen Fällen hat diese Methode bereits zu beachtlichen Kontraststeigerungen geführt.

Besondere Bedeutung besitzt der Kontrast für die Abbildungsdosis. Aus statistischen Betrachtungen folgt nämlich daß das Objekt mit einer bestimmten Elektronendosis bestrahlt werden muß, um eine Abbildung durchzuführen. Diese Abbildungsdosis ist kontrastabhängig.

Andererseits wird das Objekt durch die Elektronenbestrahlung zerstört. Dieser Einwand wurde schon gleich zu Beginn der Entwicklung des Elektronenmikroskops erhoben. Obwohl diese pessimistische Einstellung damals nicht durch entsprechende Erfahrungen gestützt werden konnte, bedeutete sie einen sehr unangenehmen psychologischen Faktor in der Entwicklung des Elektronenmikroskops zum Übermikroskop. Es war daher sehr beruhigend, als sich zeigte, daß dieser Einwand in dem damals erhobenen Ausmaß nicht zutraf, sondern daß man Bilder erhielt, die nicht durch die Beobachtung selbst verändert erschienen. In der Folgezeit hat sich jedoch ergeben, daß der Optimismus, der sich daraufhin als Gegenspieler der erwähnten pessimistischen Einstellung bemerkbar machte, auch nicht ganz zu recht bestand. Das Objekt wird tatsächlich durch die Beobachtung im Elektronenmikroskop verändert, wie die in den letzten Jahren sich häufenden Beobachtungen zeigen. Diese Objektänderung bedeutet in Fall der Abbildung von Atomen einen Platzwechsel des Atoms.

Dieser Platzwechsel des Atoms ist ebenfalls von der Dosis abhängig. Es wurde daher versucht, einen Eindruck von der Größenordnung der maximalen Platzwechseldosis zu gewinnen.

Damit nun die Abbildung eines Atoms zustande kommt, darf es natürlich seinen Platz während des Abbildungsvorganges nicht wechseln. Die Platzwechseldosis muß also größer als die Abbildungsdosis sein. Falls diese fundamentale Bedingung verletzt ist, tritt eine neue Art der Auflösungsbegrenzung, nämlich durch Platzwechsel, ein. Hier äußert sich auch die volle Bedeutung des Kontrastes für das Auflösungsvermögen, denn der Kontrast bestimmt in erster Linie die notwendige Abbildungsdosis.

Ob eine Abbildung der leichten Atome der organischen Chemie, die ja von besonderem Interesse wäre, mit schnellen Elektronen möglich ist, kann auf Grund der jetzigen Unterlagen nicht entschieden werden, zumindest ist sie, wie sich auf Grund der Diskussion ergibt, unter bestimmten Bedingungen fraglich. Es bedarf daher noch weiterer intensiver Arbeit der Theoretiker und Experimentatoren, um diese wichtige Frage definitiv zu entscheiden.

1. Kontraste von Schirmen

a. Amplitudenschirm

Zunächst wurde der Kontrast eines absorbierenden Schirms bestimmt, dessen Radius r klein gegen das Abbe'sche Auflösungsvermögen d_A ist. Dieser Schirm befindet sich im Mittelpunkt eines freien Objektfeldes vom Radius R, das groß gegen das Abbe'sche Auflösungsvermögen ist. Es gilt also:

$$\lambda \ll r \ll d_A = \lambda/\vartheta_0 \ll R \tag{1}$$

(λ = Wellenlänge, ϑ_0 = Objektivapertur). Durch den Schirm wird die Amplitude der einfallenden Strahlung ($u_0 = 1$) um den Faktor $p < 1$ geschwächt. Aus der Kirchhoff'schen Formel ergibt sich unter diesen Umständen die Amplitude der Erregung im Beugungsbild (hintere Brennebene des Objektivs) zu:

$$\phi = \frac{2\pi R^2}{\lambda} \frac{I_1(2\pi R\vartheta/\lambda)}{2\pi R\vartheta/\lambda} + (p-1) \frac{2\pi r^2}{\lambda} \frac{I_1(2\pi r\vartheta/\lambda)}{2\pi r\vartheta/\lambda} \tag{2}$$

($I_1(x)$ = Bessel-Funktion 1. Art, 1. Ordnung, ϑ = Streuwinkel). Die Erregung setzt sich also aus zwei Gliedern zusammen:

$$\phi = \phi_R + \phi_r \tag{2a}$$

von denen das erste nur von R, das zweite nur von r abhängig ist. Bei der weiteren Bilderzeugung nimmt gemäß der Voraussetzung Gl.(1) das gesamte erste Glied der Gl.(2a) an der Bilderzeugung teil während vom zweiten Glied wegen der Voraussetzung Gl.(1) ($\vartheta < \vartheta_0 \ll \lambda/r$) nur der konstante Teil des Zentralmaximums $\phi_r(0)$ zur Bilderzeugung beiträgt. Die für die weitere Bilderzeugung maßgebende Erregung in der hinteren Brennebene des Objektivs kann also durch den Ausdruck:

$$\phi = \phi_R + (p-1) r^2 \pi/\lambda = \phi_R + \phi_r(0) \tag{3}$$

dargestellt werden.

In der Gauss'schen Bildebene ergibt sich die Amplitude der Erregung aus der Anwendung der Kirchhoff'schen Beugungsformel auf Gl.(3) zu:

$$u_{\text{Abb}} = 1 + \phi_r(0) \frac{2\pi\lambda}{d_A^2} \cdot \frac{I_1(2\pi\rho/d_A)}{2\pi\rho/d_A} \tag{4}$$

für den Fall der Abbildung 1:1 (ρ = Abstand vom Mittelpunkt des Bildfeldes). Für $\rho = 0$ wird der maximale Bildkontrast:

$$\frac{\Delta I(0)}{I} = \frac{2\Delta u(0)_{\text{Abb}}}{u_{\text{Abb}}} = 2\pi\phi_r(0) \frac{\lambda}{d_A^2} \tag{5}$$

($I = |u_{\text{Abb}}|^2$ = Bildintensität).

b. Phasenschirm

Ein nichtabsorbierender Schirm in der Objektebene, der nur die Phase um den Winkel $\varphi \ll 1$ schiebt, ergibt unter sonst gleichen Voraussetzungen folgende Beziehungen:

$$\Phi_r(0) = (e^{i\varphi} - 1) r^2 \pi / \lambda \approx i \sin \varphi \cdot r^2 \pi / \lambda \tag{6}$$

$$u_{\mathrm{Abb}} = 1 + \frac{2\pi\lambda}{d_A^2} \frac{I_1(2\pi\rho/d_A)}{2\pi\rho/d_A} \tag{7}$$

$$\frac{\Delta I(0)}{I} = 2\pi |\phi_r(0)| \frac{\lambda}{d_A^2} \cdot \sin \varphi/2 \tag{8}$$

Der Phasenschirm führt also, gleichen Betrag der Streuamplitude bzw. gleiche Strevintensität vorausgesetzt, zu einem um den Faktor $\sin \varphi/2$ geringeren Kontrast als der Amplitudenschirm. Falls, wie in dem Phasenkontrastverfahren von Zernicke [20], in der Beugungsfigur die Phase des Primärstrahls (ϕ_R) gegenüber der Phase der Beugungsfigur (Φ_r) des Phasenschirms um 90° verschoben wird, werden die Bildkontraste des Phasenschirms um den Faktor $1/\sin\varphi/2$ verstärkt und durch die gleichen Beziehungen wie die des Amplitudenschirms beschrieben. Z. B. lautet der Maximalkontrast in diesem Fall

$$\frac{\Delta I(0)}{I} = 2\pi |\phi_r(0)| \lambda / d_A^2 \tag{5a}$$

c. Anwendung auf die Kontraste von Einzelatomen.

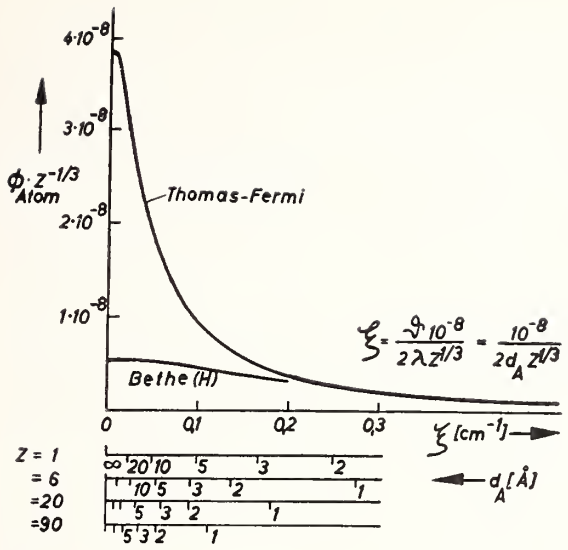
Für die Abbildung von Einzelatomen ist die Kenntnis ihrer Streuverteilung (Φ_{Atom}) bezüglich Amplitude und Phase notwendig. Im allgemeinen werden Amplitude und Phase eine Funktion des Streuwinkels sein. Zur Vereinfachung der Rechnung möge daher zunächst angenommen werden, daß die Phase im Beugungsbild so korrigiert werden kann, als ob die Beugungsfigur von einem nur absorbierenden Teilchen stammt. Der auf diese Weise berechnete Kontrast ist der maximale Kontrast, den das Atom erzielen kann. Als Beispiel ist in Figur 14.1 die Streuamplitude des Thomas-Fermi-Atoms dargestellt. Demnach ist bei sehr kleinen Streuwinkeln ϑ bzw. sehr großen Werten des Abbe'schen Auflösungsvermögens d_A die Amplitude der Streuverteilung konstant, so daß die Voraussetzungen der vorher behandelten Abbildung kleiner Schirme für diesen Fall zutreffen.

Die Streuamplitude des Thomas-Fermi-Atoms lautet für $\vartheta = 0$ bzw. $\zeta = 0$:

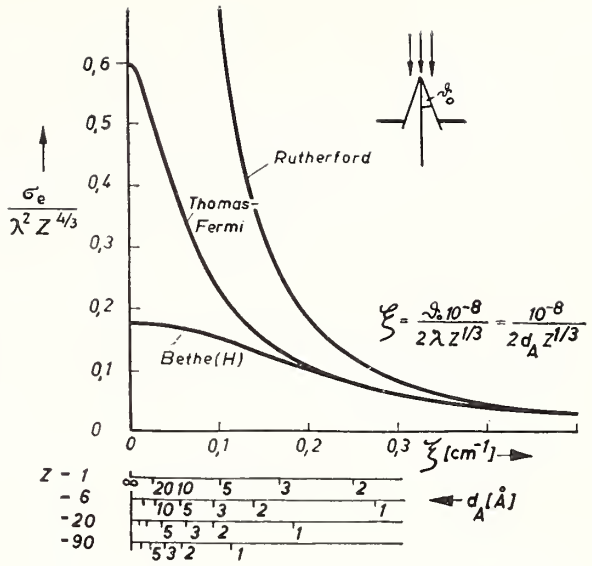
$$\Phi_{\mathrm{Atom}}(0) = \frac{Z^{1/3} e^2 m}{2h} \cdot F_e(0) = 3{,}8 \cdot 10^{-8} Z^{1/3} \tag{9}$$

Dieser Ausdruck in Gl. (5) eingesetzt, ergibt den maximalen Kontrast des Thomas-Fermi-Atoms:

$$\frac{\Delta I(0)}{I} = 24 \cdot 10^{-8} Z^{1/3} \lambda / d_A^2 \tag{10}$$



FIGUR 14.1 *Streuamplitude des Atoms.*



FIGUR 14.2 *Elastischer Streuquerschnitt des Atoms.*

Voraussetzungsgemäß kann diese Beziehung jedoch nur gelten, solange Φ einigermaßen konstant ist, also schätzungsweise in dem Bereich $\phi = \phi(0)$ bis $\phi = \phi(0)/2$. Dies ist nach Figur 14.1 für $\zeta \leq 0{,}05$ bzw. $Z^{1/3} \geq 10^{-7}/d_A$ der Fall. Für $d_A = 3$Å gilt also Gl.(10) nur für $Z > 37$. Um für kleinere Ordnungszahlen zu einer Kontrastabschätzung zu kommen, möge angenommen werden, daß die in die Objektivöffnung gestreuten Elektronen diese mit einer konstanten mittleren Amplitude $\overline{\Phi_{\text{Atom}}(0)}$ ausleuchten. Diese Amplitude ergibt sich aus der Differenz des totalen elastischen Streuquerschnitts $\sigma_e(0)$ und des elastischen Streuquerschnitts $\sigma_e(\vartheta_0)$ für die Objektivöffnung ϑ_0 bzw. für den entsprechenden Wert d_A zu:

$$\overline{\Phi_{\text{Atom}}(0)} = \left(\frac{\sigma_e(0) - \sigma_e(\vartheta_0)}{\vartheta_0^2 \pi} \right)^{1/2} \tag{11}$$

Unter diesen Voraussetzungen ergibt sich der maximale Bildkontrast des Beugungsfehlerscheibchens:

$$\frac{\Delta I(0)}{I} = \frac{2\sqrt{\pi}}{d_0} \left(\sigma_e(0) - \sigma_e(\vartheta_0) \right)^{1/2} \tag{12}$$

$\sigma_e(0)$ und $\sigma_e(\vartheta_0)$ können Figur 14.2 und Tabelle 14.1 entnommen werden.

TABELLE 14.1. *Elastischer* (σ_e) *und unelastischer* (σ_u) *Streuquerschnitt und der Kontrastgewinn s durch Filterung bei dünnen, ausgedehnten und amorphen Objekten.* (*Fig. 14.2, Gl.23, 24*) ($\lambda = 0{,}05$A, $d_A = 3$A, $\vartheta_0 = \lambda/d_A$)

Z	1	6	20	90
$\sigma_e(0)/\lambda^2$	0,18	6,5	32	241
$\sigma_e(\vartheta_0)/\lambda^2$	0,12	2,7	19	180
$\sigma_u(0)/\lambda^2$	3,2	14	47	210
$s = \frac{\sigma_u(0) + \sigma_e(\vartheta_0)}{\sigma_e(\vartheta_0)}$	29	6,3	3,5	2,2

2. Kontraste von Gittern

a. Amplitudengitter

Durch ein quadratisches Kreuzgitter mit der Gitterkonstanten d werde in der Objektebene die Amplitude der Lichterregung:

$$u=1+\Delta u \cos 2\pi x/d+\Delta u \cos 2\pi y/d \tag{13}$$

hervorgerufen ($\Delta u \ll 1$). Diese Lichtverteilung bedingt nach der Fourier-Analyse ein Interferenzdiagramm, von 4 Interferenzstrahlen, deren Hauptmaxima im Abstand $\vartheta_g=\lambda/d$ vom Primärstrahl liegen. Die maximale Amplitude dieser Interferenzstrahlen ist

$$\phi(\vartheta_g)=\frac{\Delta u_0}{2}\cdot\frac{L_x L_y}{\lambda} \tag{14}$$

(L_x, L_y=Kantenlänge des Gitters in Richtung der Achsen x, y). Wenn alle Interferenzmaxima an der Abbildung teilnehmen ($\vartheta_g<\vartheta_0$), wird in der Gauss'schen Bildebene die gleiche Lichtverteilung wie in der Objektebene (Gl.13) entstehen. Der maximale Bildkontrast ist dann gegeben durch:

$$\frac{\Delta I_g}{I}=\frac{8\Delta u}{u}=16\phi(\vartheta_g)\frac{\lambda}{L_x L_y} \tag{15}$$

b. Phasengitter

Durch das Phasengitter werde nur die Phase φ der Lichterregung in der Objektebene periodisch beeinflußt. Für $\varphi_0 \ll 1$ gelts:

$$u=1+i\varphi_0 \cos 2\pi x/d+i\varphi_0 \cos 2\pi y/d \tag{16}$$

Entsprechend Gl. (14) wird die Maximalamplitude der Interferenzstrahlen:

$$\phi(\vartheta_g)=i\frac{\varphi_0}{2}\cdot\frac{L_x L_y}{\lambda} \tag{17}$$

Die Interferenz dieser Strahlen zusammen mit dem Primärstrahl führt im Bilde zu der in Gl. (16) beschriebenen Erregung. Es treten also keine Bildkontraste auf. Erst eine Phasenschiebung der Interferenzstrahlen um 90° gegenüber dem Primärstrahl nach dem Verfahren von Zernicke ergibt gemäß Gl. (15) die gleichen Maximalen Bildkontraste wie die des Amplitudengitters:

$$\frac{\Delta I_g}{I}=16/\phi(\vartheta_g)/\frac{\lambda}{L_x L_y}=8\varphi_0 \tag{18}$$

c. Anwendung auf die Kontraste von Kristallgittern

Die Intensitäts- bzw. die Amplitudenverteilung der Beugungserscheinung eines dreidimensionalen Kristallgitters wird durch die Laue'-sche Gleichung beschrieben. Es werde vorausgesetzt, daß die z-Achse

eines einfachen kubischen Raumgitters in Richtung des Primärstrahls liegt, und daß die Zahl der Gitterebenen in dieser Richtung sehr gering ist. Infolgedessen wird das nullte Hauptmaximum dieser Periode so breit, daß die Interferenzen 1. Ordnung der beiden anderen Perioden noch in einem angenähert konstanten Bereich dieses 0. Hauptmaximums liegen. Es ergibt sich die maximale Beugungsamplitude in den Interferenzmaxima 1. Ordnung:

$$\phi(\vartheta_g) = \phi_{\mathrm{Atom}} N_x N_y N_z \tag{19}$$

wobei N_i die Anzahl der Atome in Richtung der Achsen i bedeutet. Falls nur die Interferenzmaxima 0. und 1. Ordnung das Bild aufbauen, ergibt sich natürlich die durch Gl. (13) gekennzeichnete einfache Lichtverteilung. Der maximale Bildkontrast [3,8] des Kristallgitters ergibt sich dann durch Einsetzen von Gl. (19) in Gl. (15) zu:

$$\frac{\Delta I_g}{I} = 16\,\phi_{\mathrm{Atom}} \frac{\lambda}{d^2} \cdot N_z \tag{20}$$

In Tabelle 14.2 sind einige Zahlenwerte des Kontrastes für den Fall $N_z = 1$ angegeben. Voraussetzung ist hierbei allerdings, daß das Verfahren von Zernicke angewendet wurde, d. h. daß die Phase der Interferenzstrahlen in optimaler Weise gegen den Primärstrahl verschoben ist.

Tabelle 14.2. *Maximaler Kontrast der Beugungsfehlerscheibchen von Atomen infolge elastischer und unelastischer Streuung (Index u), (Gl. 10, 12, 25),* $\lambda = 0{,}05$ Å, $d = d_A = 3$ Å).

Z	1	6	20	90
$\frac{\Delta I(0)}{I}$ (Einzelatom)	0, 015	0, 02	0, 22	0, 45 0, 6
$\frac{\Delta I_g}{I}$ (Kristallgitter)	0, 031	0, 06	0, 37	0, 9
$\left(\frac{\Delta I(0)}{I}\right)_u$ (Einzelatom)	0, 0028	0, 012	0, 043	0, 19

3. Phasenschiebung

In einer früheren Untersuchung [7, 8] wurde die Möglichkeit der Phasenschiebung nach Zernicke [20] im ideal abbildenden Elektronenmikroskop durch Folien mit innerem Potential bzw. durch zusätzliche elektrostatische Felder diskutiert. Auch wurde darauf hingewiesen [8, 19] daß bei nichtidealer Abbildung die sphärische Aberration und die extrafokale Abbildung ebenfalls Phasenschiebungen herbeiführen.

Die sphärische Aberration mit der Aberrationskonstanten $C_ö$ hat bekanntlich im Zusammenhang mit der Beugung eine Auflösungsbegrenzung zur Folge Bei optimaler Objektivapertur wird die Auflösungsgrenze:

$$d_ö = c_1 C_ö^{1/4} \lambda^{3/4} \tag{21}$$

($c \sim 1$). Diese Beziehung kann auch so gedeutet werden, als ob der Strahl maximaler Ablenkung durch den Einfluß der sphärischen Aberration um den Winkel $\alpha \sim 90°$ gegen den Primärstrahl in der Phase verschoben wäre.

Auch infolge Defokussierung um die Strecke Δf tritt bekanntlich eine Auflösungsbeschränkung ein. Für $d_{\Delta f} > d_{\ddot{o}}$ wird bei optimaler Bemessung der Objektivapertur die Auflösungsgrenze:

$$d_{\Delta f} = c_2 \Delta f^{1/2} \lambda^{1/2} \tag{22}$$

($c_2 \sim 1$). Sie kann ebenfalls als Folge eienr Phasenschiebung von ca. 90° des maximal gestreuten Strahls gegen den Primärstrahl gedeutet werden. Öffnungsfehler und extrafokale Abbildung führen also im Grenzfall und für die extrem abgebeugten Strahlen zu den gewünschten Phasenschiebungen.

4. Amplitudenschiebung

Falls die Intensität des Primärstrahls in der Frauenhofer'schen Beugungsfigur durch ein Absorptionsplättchen um den Faktor $r < 1$ geschwächt wird ergibt sich nach Gl. (5, 15) innerhalb der Voraussetzungen eine Verstärkung des Kontrastes um den Faktor $1/\sqrt{r}$.

5. Unelastische Streuung

In den vorstehenden Ausführungen wurde nur die Wirkung der kohärenten (elastischen) Streuvorgänge auf die Bildkontraste berücksichtigt. Unberücksichtigt blieb die Wirkung der inkohärenten oder unelastischen Streuvorgänge. Nun ist nach den bisher vorliegenden Untersuchungen [5–9] der mittlere Streuwinkel der unelastischen Streuung wesentlich kleiner als der elastischen, sodaß unter den üblichen Bedingungen und bei dünnen Objekten (Einzelstreuung) fast sämtliche unelastisch gestreuten Elektronen in die Objektivöffnung und an den zugehörigen Bildort gelangen und daher nicht zu Kontrasten Anlaß geben. Eine Kontrasterhöhung durch Ausschaltung der unelastisch gestreuten Elektronen kann durch das Elektronenfilter [21] oder durch Verringerung der Objektivapertur [10] erzielt werden. Da die Verringerung der Objektivapertur auf die notwendigen Werte mit einer entsprechenden Verringerung des Auflösungsvermögens gekoppelt ist, soll im folgenden nur die Wirkung des Elektronenfilters behandelt werden. Durch das Filter können sämtliche unelastisch gestreuten Elektronen unabhängig vom Streuwinkel ausgeschaltet werden. Es interessiert daher hier nur der unelastische Gesamtstreuquerschnitt $\sigma_u(0)$. Die in Tabelle 1 angegebenen Werte von $\sigma_u(0)$ wurden aus einer schon früher [8] mitgeteilten Beziehung:

$$\sigma_u(0) = 24 \frac{F \lambda^2}{\Delta U} \tag{23}$$

ermittelt, die das experimentell abgeleitete Widdington'sche Gesetz über den mittleren Energieverlust (gemittelt auch über die elastisch gestreuten Elektronen) mit dem experimentellen Resultat [22, 23] über den mittleren Energieverlust ΔU der unelastisch gestreuten Elektronen miteinander verknüpft. Diese Streuquerschnitte stimmen größenordnungsmäßig mit den theoretisch ermittelten Resultaten von Marton und Schiff [6] und von v. Borries [9] überein. Da noch sehr wenig Elemente vermessen sind, wurden nicht die individuellen

Werte, die sich in der Größenordnung von $\Delta U = 5$—25 eV bewegen, benutzt, sondern ein Mittelwert von 10 eV. Die $\sigma_u(0)$—und $\sigma_u(\vartheta_0)$—Werte von Wasserstoff der Tabelle 14.1 sind der früheren Untersuchung des Verfassers entnommen, die σ_e—Werte der Figur 14.2.

Unter diesen Umständen ist der Kontrastgewinn durch Filterung bei der elektronenmikroskopischen Abbildung dünner Präparate, deren Kontraste durch Streuabsorption zustande kommen, gegeben durch:

$$s = \frac{\sigma_u(0) + \sigma_e(\vartheta_0)}{\sigma_e(\vartheta_0)} \tag{24}$$

In Tabelle 14.1 sind einige Werte für den Kontrastgewinn durch Filterung angegeben für eine Objektivöffnung, die einem Abbe'schen Auflösungsvermögen von 3 Å entspricht. Es sind also besonders bei leichten Substanzen durch die Filterung erhebliche Kontrastgewinne gegenüber den Kontrasten durch Streuabsorption zu erwarten. Bei den Kontrasten durch Streuabsorption ist angenommen, daß es sich um amorphe Objekte handelt. Objekte also, in denen die Atome unabhängig voneinander streuen (Intensitätsüberlagerung) im Gegensatz zur Streuung an kristallinen Objekten [3, 8, 12–16]. Sowie aber die Kontraste nicht durch Streuabsorption sondern durch Interferenz der kohärenten Streuung zustande kommen, ist eine derartige Steigerung der Kontraste durch das Filterverfahren nicht zu erwarten, denn der Maximalkontrast durch inkohärente Streung:

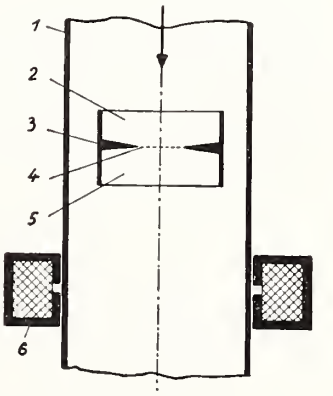
$$\frac{\Delta I(0)}{I} = \pi \frac{\sigma_u(0)}{d_A^2} \tag{25}$$

ist klein gegen den durch elastische Streuung (vgl. Tabelle 2).

5. Elektronenfilter

a. Netzfilter

In einer früheren Untersuchung [21] war über ein Elektronenfilter berichtet worden, das unelastisch gestreute Elektronen aus Elektronenbeugungsdiagrammen und elektronenmikroskopischen Abbildungen ausscheidet. Dieses Filter besteht im wesentlichen aus einer netzartigen Elektrode, die etwa auf Kathodenpotential aufgeladen ist (Figur 14.3) Elektronen, die das Objekt ohne Energieverlust durchsetzt haben, können gegen das Feld soweit anlaufen, daß sie das Netz erreichen, durch die Maschen hindurchtreten und auf der



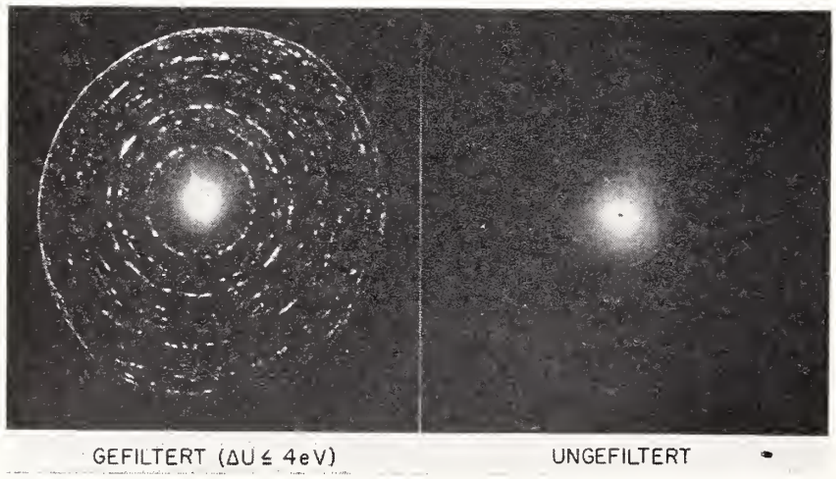
Figur 14.3. *Netzfilter.*
1, Wand des Vakuumgefäßes (Erdpotential); 2, Immersionslinse; 3, 4, Gegenspannungs-Elektrode mit Netz (Kathodenpotential); 5, 6, Immersionslinse und magnetische Linse.

anderen Seite des Netzes durch das Feld wieder auf ihre ursprüngliche Geschwindigkeit beschleunigt werden. Elektronen mit verminderter Geschwindigkeit werden dagegen durch das Feld vor dem Netz reflektiert und so aus dem Strahlengang ausgefiltert.

Neben der Einhaltung gewisser elektronenoptischer Maßnahmen war für die Güte der Abbildung die Qualität des Netzes entscheidend. Es fanden gewebte Drahtnetze mit einen Drahtabstand von ca 0, 1 mm und einer Durchlässigkeit von ca 30% Verwendung. Das Schattenbild dieser relativ grobmaschigen Netze überlagert sich dem Bilde und beeinträchtigt die Bildqualität, besonders aber die Bildkontraste. In der zitierten Untersuchung wurde daher eine "kompensierte Verschiebung" des Bildes vorgenommen, um das Schattenbild des Netzes zu verwischen. Diese Filterungsversuche wurden inzwischen mit einem Netz von 0,02 mm Maschenabstand und einer Durchlässigkeit von ca. 70% weitergeführt.[4] Infolge der kleineren Maschenabstände, der höheren Durchlässigkeit und der gleichmässigeren Netzstruktur genügte schon eine leichte Defokussierung von L_2, um eine Abbildung der Netzstruktur zu unterdrücken, so daß die Methode der kompensierten Verschiebung unter diesen Umständen nicht angewendet zu werden brauchte.

b. Anwendung des Netzfilters

In Figur 14.4 wird die Wirksamkeit des Filters durch die Gegenüberstellung eines gefilterten und eines ungefilterten Elektronenbeugungsdiagramms demonstriert. Da es sich um ein relativ starkes Objekt handelt, ist der Untergrund in dem ungefilterten Diagramm so intensiv, daß die Kontraste der Kristallgitterinterferenzen dadurch außerordentlich reduziert werden. Der Untergrund im ungefilterten Diagramm besteht aus unelastisch gestreuten Elektronen von ca. 15 eV Geschwindigkeitsverlust, die Kristallgitterinterferenzen dagegen aus elastisch gestreuten Elektronen. Im gefilterten Diagramm kommen nur Elektronen zur Wirkung, die einen Geschwindigkeitsverlust von weniger als 4 eV erlitten haben. Dadurch verschwindet der gesamte inkohärente Untergrund und eine außerordentliche Kontraststeigerung ist die Folge. Beide Aufnahmen wurden in gleicher



FIGUR 14.4. *Anwendung des Netzfilters auf das Elektronenbeugungsdiagramm einer relativ dicken Aluminiumoxyd Folie* ($U = 27$ *keV*).

[4] Dieses Netz wurde dem Verf. von Herrn Prof. Dr. Nils Hast, Nobelinstitut för fysik, Stoekholm zur Verfügung gestellt, wofür ihm aueh an dieser Stelle bestens gedankt sei.

Weise entwickelt und kopiert, jedoch wurden die Belichtungszeiten so gewählt, daß der Untergrund in beiden Elektronen-Diagrammen etwa gleiche Schwärzung aufwies. Die Kontrastdifferenzen beider Aufnahmen sind also allein durch die Wirkung des Elektronenfilters und nicht durch Umkopieren verursacht. Der noch verbliebene Untergrund in der gefilterten Aufnahme ist auf die Auflösung von Sekundär-Elektronen aus dem Gitter durch Ionen-Aufprall zurückzuführen [21].

Mit dieser Filtermethode wurde auch das Problem der Kikuchi-Linien und -Bänder in Elektronenbeugungsdiagrammen wieder aufgegriffen. Nach von Laue [24] müssen an der Entstehung derartiger Interferenzerscheinungen inkohärente Streuvorgänge beteiligt sein, die den Phasenzusammenhang der austretenden mit der eintretenden Strahlung aufheben, da sonst nur die bekannten diskreten Interferenzmaxima entstehen könnten. Es blieb aber offen, welcher Art diese inkohärenten Streuvorgänge sind, ob es sich um Anregung von Elektronentermen handle, die im allgemeinen mit Energieverlusten von einigen Volt bis zu einigen 20 V verknüpft sind, oder um die Anregung von thermischen Gitterschwingungen, die nur Energieverluste von einigen Zehntel Volt zur Folge hätten. Nach Emslie [25] sollten nämlich die Kikuchi-Bänder eine unmittelbare Folge der Anregung von Gitterschwingungen sein, während nach von Laue die Kikuchi-Bänder auf die dynamische Wechselwirkung der inkohärenten gestreuten Elektronen mit dem Gitter zurückzuführen wären. Seinerzeit wurde versucht [26], durch Beugung an Eisen- und Diamant-Einkristallen (Debye'sche Temperatur 420°K bzw. 2340°K) [27] und Änderung der Temperatur (+20°C bis −180°C bei Diamant und +20°C bis +800°C bei Flußspat) die Anregungsbedingungen für die thermischen Gitterschwingungen möglichst extrem zu gestalten. Merkliche Unterschiede in den Diagrammen wurden hierdurch nicht hervorgerufen. Mit aller Vorsicht wurde damals daher nur geschlossen daß es auf dem beschriebenen Wege nicht gelingt, die Anregungsbedingungen für die Wärmeschwingungen genügend zu variieren, um eine experimentelle Entscheidung für oder gegen Emslies Erklärungsversuch zu erhalten.

Die jetzt durchgeführten Filterungsversuche fanden an Kikuchi-Diagrammen von sehr dünnen Glimmereinkristallen statt, so daß gleichzeitig eine große Zahl von Laue-Interferenzen auftrat. Es zeigte sich, daß das gesamte Kikuchi-Diagramm mit dem Filter ($\Delta U < 4$ eV) zu beseitigen war und nur das Laue'sche Interferenzfleckensystem bestehen blieb.

Hiermit ist der unmittelbare Nachweis dafür geliefert, daß unelastische (inkohärente) Streuvorgänge am Aufbau des Kikuchi-Diagramms beteiligt sind, und daß das Kikuchi-Diagramm auf Anregung von Elektronentermen mit Energieverlusten von mehreren eV und nicht auf Anregung von thermischen Gitterschwingungen zurückzuführen ist. Die früheren Versuche des Verf. stehen mit diesen Resultaten im Einklang. Damit ist aber eine experimentelle Entscheidung gegen den Erklärungsversuch von Emslie und für die Laue'sche Theorie gefallen.

c. Linsenfilter

In der vorangegangenen Untersuchung [27] (Anmerkung 9) wurde auch darauf hingewiesen, daß Arbeiten im Gange sind, die normale

elektrostatische Einzellinse zu einem Filter weiter-zuentwickeln.[5]

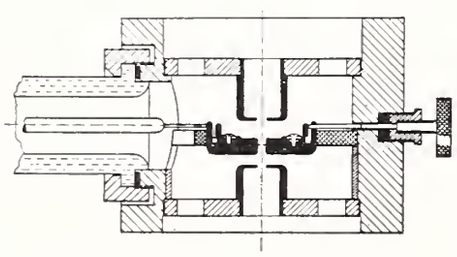
Schon die normale Linse mit negativer Mittelelektrode stellt ein Filter dar. Indessen ist ihre Filterwirkung für diese Zwecke ungenügend, da das Achsenpotential nur auf ca 90% des Kathodenpotentials ansteigt. Es werden also nur Elektronen ausgefiltert, die mehr als 10% Geschwindigkeitsverlust erlitten haben. Das Achsenpotential soll aber etwa Kathodenpotential erreichen und darf es nur um wenige Volt unterschreiten.

Diese Erhöhung des Achsenpotentials kann entweder durch Erhöhung des Elektrodenpotentials um den restlichen Betrag von 10% bei den üblichen Linsen erfolgen oder durch Veränderung dieser Linsen, indem der Durchgriff der Mittelelektrode verringert wird. Da die Einfügung einer Spannungsquelle von mehreren tausend Volt zwischen Kathode und Mittelelektrode leicht Anlaß zu Spannungsschwankungen geben kann, wurde die Verringerung des Durchgriffs bevorzugt, indem der Durchmesser der Öffnung der Mittelelektrode auf 0,1 bis 0,3 mm Durchmesser verringert wurde (vgl. Fig. 14.5). Diese Öffnung befand sich in einer Platinfolie von 0,1 mm Stärke. Um den Felddurchgriff durch diese Öffnung noch weiter herabzusetzen, wurde die Feldstärke vor der Platinfolie durch Vorblenden und Nachblenden größeren Durchmessers (3 mm) verringert. Diese Blenden bilden ihrerseits Potentialgebirge aus, die wie Immersionslinsen wirken. Die Brennweite dieser Linsen kann durch entsprechende Wahl ihres Abstandes von der Platinfolie wunschgemäß eingestellt werden.

Die Brechkraft des resultierenden Linsensystems nimmt wie bei den üblichen elektrostatischen Linsen [29] mit wachsender Spannung zunächst zu, erreicht ein Maximum, geht über Null zu negativen Werten, erreicht dort wieder einen Extremwert und strebt dann wieder über Null einem dritten positiven Extremwert zu usw. Auch bei diesem Linsensystem ist es—wie bei den üblichen elektrostatischen Linsen—vorteilhaft, als Arbeitspunkt ein derartiges Maximum der Brechkraft zu wählen, denn einmal ist dort die Verzeichnung am geringsten (Übergang von kissen- zu tonnenförmiger Verzeichnung), zum anderen besitzt dort aber auch der chromatische Fehler ein Minimum. Diese für die Einzellinse geltenden experimentellen Erfahrungen wurden erstmalig in einer Patentschrift [29] niedergelegt und später von anderer Seite [30] bestätigt.

Durch entsprechende Dimensionierung der Vorblende wurde erreicht, daß sich der Arbeitspunkt dieses Systems im dritten Maximum befindet, etwa 3 V vom Sperrpunkt der Linse entfernt (bei 27 kV Linsenspannung).

Durch diese Maßnahmen sind die Linsenfehler gegenüber denen der üblichen Linsen erheblich gesteigert, z.B. wächst der Öffnungs-



FIGUR 14.5. *Linsenfilter mit zentrierbarer Vorlinse.*

[5] Diese Arbeiten mussten Ende 1948 wegen Demontage des Institutes abgebrochen werden und konnten erst nach Aufbau eines neuen Laboratoriums wieder aufgenommen werden. Inzwischen hat nach Veröffentlichung der zitierten Untersuchug auch die AEG (SDL/Mosbach) ein Linsenfilter entwickelt [28].

fehler derartig, daß die Linse im allgemeinen nicht als hochauflösendes Objektiv verwendet werden kann, sondern ähnlich wie das Netzfilter nur als Projektiv. Für geringere Ansprüche, wie z.B. für die in Figur 14.6 wiedergegebenen Abbildungen, kann das benutzte System selbstverständlich auch als Objektiv verwendet werden. Zweckmäßig ist es jedoch auch bei diesen Aufnahmen die Divergenz der Primärstrahlung, z.B. durch eine vorgeschaltete Linse kurzer Brennweite, auf einen optimalen Wert herabzusetzen. Auch die übrigen Bildfehler werden durch die Verkleinerung der Öffnung der Mittelelektrode erheblich vergrößert so z.B. der besprochene Verzeichnungsfehler, und vor allem der chromatische Linsenfehler. Der chromatische Linsenfehler begrenzt letzten Endes auch das Auflösungsvermögen, wie sich anschaulich sehr einfach aus dem Verhältnis der thermischen Geschwindigkeitsverteilung ($U_{\text{therm}}=$ 0,5÷1 eV bei Wolfram-Kathoden) zu der Minimalgeschwindigkeit im Filter von $\Delta U=3\div10$ eV ergibt. Eine Reduzierung des chromatischen Fehlerscheibchens wäre also durch Reduzierung der thermischen Geschwindigkeitsverteilung möglich, also z. B. durch Senkung der Kathodentemperatur, indem die jetzigen Wolfram-Kathoden durch Kathoden niedrigerer Austrittsarbeit (z.B. BaO) ersetzt würden.

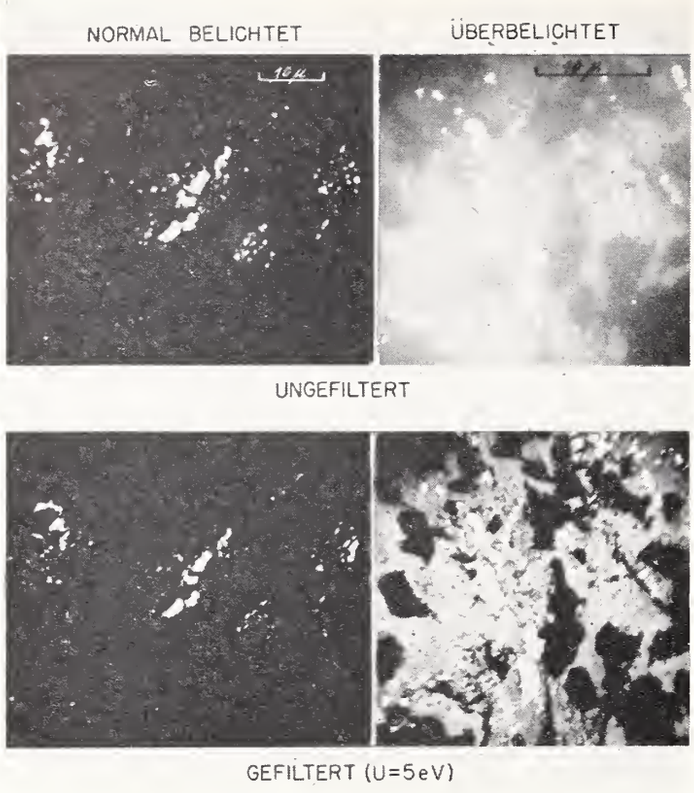
Wie bei allen Linsen war auch hier eine Gesichtsfeldblende notwendig, um Störungen durch Randstrahlen auszuschalten. Diese Gesichtsfeldblende mußte wegen der geringen Ausdehnung des verwendungsfähigen Bildbereichs relativ klein gehalten werden (0,05÷0,1 mm ϕ) und während des Betreibes durch einen Kreuztisch justiert werden können.

Da der Elektronenstrahl durch einen sehr hohen und schmalen Sattel des Potentialgebirges hindurchgefädelt werden muß, ist das Linsenfilter gegen Dezentrierung der Elektroden und gegen geringste Neigungsänderung des Primärstrahls sehr empfindlich. Um derartige Einflüsse zu kompensieren wurde auch die auf Hochspannung befindliche Vorblende auf einen Kreuztisch gesetzt, der während des Betriebes justiert werden konnte.

Unter diesen Voraussetzungen betrug die Zahl der Bildpunkte 150 bis 300 längs eines Bilddurchmessers. Bestimmt wurde diese Zahl aus dem Durchmesser der Bildfehlerscheibchen. Die Brennweite dieser Linse ergab sich im Arbeitspunkt zu etwa $f=2{,}6$ mm. Hierbei befand sich die abgebildete Objektebene noch im abgeschirmten Raum der auf Erdpotential befindlichen Linsenelektrode.

d. Anwendung des Linsenfilters

Die Wirkung des Linsenfilters wird in Figur 14.6 demonstriert. Das Objekt, eine geätzte Goldfolie mit einigen Löchern, ist relativ dick und daher, wie die beiden normal belichteten Aufnahmen auf der linken Seite der Figur 14.6 zeigen, für Elektronen relativ undurchsichtig. Die Wirkung der Filterung macht sich in diesen beiden Aufnahmen nur dadurch bemerkbar, daß die ohnehin schwache Aufhellung in einigen Bereichen außerhalb der Löcher durch das Filter noch weiter reduziert wird. Die eigentliche Wirkung des Filters kommt jedoch erst in den beiden anderen Aufnahmen (rechts) zum Ausdruck, die wesentlich länger mit Elektronen belichtet sind, so daß die Elektronendichte gewisser Bereiche mittlerer Schwärzung in beiden Aufnahmen ungefähr übereinstimmt. Die Wirkung des



FIGUR 14.6. *Elektronenmikroskopische Abbildung einer dicken Goldfolie. Anwendung des Linsenfilters* ($U=27$ *ekV*).

Filters dokumentiert sich in einer außerordentlichen Steigerung der Kontraste und, soweit erkennbar, auch in einer Steigerung des Auflösungsvermögens.

Das Filter ist also nicht nur für dünne Objekte geringer Ordnungszahl sondern auch für dicke Objekte, also wahrscheinlich auch Gewebeschnitte, in dem Bereich der Mehrfach- und Vielfachstreuung zur Konstraststeigerung geeignet. Die Versuche zeigen, daß die Ausschaltung der inkohärenten Streuung in hinreichendem Maße gelingt, so daß die Voraussetzungen für die theoretische Behandlung der Kontraste durch unelastische Streuvorgänge wenigstens in dieser Hinsicht zutreffen.

6. Dosisprobleme

Damit ein Gegenstand abgebildet werden kann, darf er seinen Ort während des Abbildungsvorganges nicht ändern. Auch der Abbildungsvorgang selbst darf nicht Ursache eines solchen Platzwechsels sein. Es liegt hier das gleiche Problem vor, das Heisenberg in seinem berühmten Gedankenexperiment—Abbildung eines freien Elektrons durch einen Lichtquant—zur Begründung der Unschärfe-Relation diskutierte. Zum Unterschied aber das Objekt im Elektronenmikroskop nicht frei, sondern mit einer gewissen Bindungsenergie an seinen Ort gebunden. Die Frage, ob das Objekt abgebildet werden kann, führt daher zu der anderen Frage, ob die Zahl der Elektronen, die im Mittel zur Abbildung der Atome notwendig ist (Abbildungsdosis), kleiner ist als die Zahl der Elektronen, die im Mittel zu einem Platzwechsel des Atoms führt.

a. Abbildungsdosis

Für die Frage nach der Abbildungsdosis (n_B=Anzahl der Elektronen pro Flächeneinheit im Bild) ist der Kontrast von entscheidender Bedeutung; denn aus schwankungstheoretischen Überlegungen [31] folgt, daß:

$$n_B \geq \frac{(k_1 k_2)^2}{\left(\frac{\Delta I}{I}\right)_B^2} \cdot \frac{1}{r_B^2 \pi} \tag{26}$$

sein muß, um den mittleren Kontrast des untersuchten Bildelements $r_B^2 \cdot \pi$ von den statistisch bedingten Kontrastschwankungen des Untergrundes unterscheiden zu können. Die Konstante k_1 ist eine Funktion der Wahrscheinlichkeit W, mit der Bildkontrast von schwankungsmäßig bedingten Untergrundkontrasten unterschieden werden kann. Z. B. ist W=68% für $k_1=1$, W=95% für $k_1=2$, W=99,7% für $k_1=3$. In den hier angeführten Beispielen wurde $k_2=2$ und $k_1=3$ gesetzt, so daß 0,3% der Bildfläche die statistisch für bedingten Kontraste über den Bildkontrasten liegen.

Um den Anschluß an die früheren Überlegungen zu gewinnen, soll als Bildelement die Fläche des Bildfehlerscheibchens des Atoms gewählt werden. Im Falle der Abbildung von Einzelschirmen wird $r_B^2 \pi = \rho_0^2 \pi$ gesekt mit $\rho_0 = 0{,}61 \cdot d_A$, wobei ρ_0 die erste Nullstelle der Besselfunktion bedeutet; im Fall der Gitterabbildung dagegen $r_B^2 \pi = d^2/2$ (halbe Fläche der Elementarzelle). Der mittlere Bildkontrast $(\Delta I/I)_B$ der Gl.(26) ergibt sich durch Mittelung des Bildkontrastes über die untersuchte Bildfläche. Im Fall der Abbildung eines Einzelschirms wird daher:

$$\left(\frac{\Delta I}{I}\right)_B = 0{,}14 \, \frac{\Delta I(0)}{I} \tag{27}$$

und im Fall des Kreuzgitters

$$\left(\frac{\Delta I}{I}\right)_B = 0{,}2 \, \frac{\Delta I_g}{I} \tag{27a}$$

In Tabelle 14.3 ist die so ermittelte Dosis für die Abbildung von Atomen verschiedener Ordnungszahl bestimmt worden.

Um die Dosis, die zur Abbildung notwendig ist, möglichst herabzusetzen, sind also möglichst starke Kontraste notwendig. Mit großem Vorteil wird hier daher das Phasenkontrastverfahren Anwendung finden. Dagegen bringt das Verfahren der Amplitudenschiebung in diesem Fall keinen Vorteil. Es hat zwar eine Reduzierung der Bilddosis um den Faktor 1/r zur Folge, gleichzeitig steigt aber die im

Tabelle 14.3. *Abbildungsdosis für Kontrast-Werte der Tabelle 14.2.* (Gl.(26), $d=d_A=3$ Å)

Z	1	6	20	90
n_B/cm^2 (Einzelatom)	$8,\ 6.10^{21}$	$1{,}4.10^{20}$	$0{,}15.10^{20}$	$0{,}025.10^{20}$
n_B/cm^2 (Kristallgitter)	$2{,}1.10^{21}$	$0{,}78.10^{20}$		

Objekt notwendige Dosis wegen der Absorption des Primärstrahls um den gleichen Faktor, so daß sich beide Wirkungen aufheben.

b. Platzwechseldosis

Objektänderung durch Elektronenbeschuß kann einerseits eine sekundäre Folge der Temperatursteigerung, andererseits eine direkte Folge der Wechselwirkung des Elektronenstrahls mit dem Objekt sein. Abhilfe kann im ersten Fall durch Verminderung der Elektronenintensität oder durch Verbesserung der Wärmeableitungsbedingungen im Objekt geschaffen werden. Im zweiten Fall ist die Objektänderung nur eine Angelegenheit der auf das Objekt fallenden Elektronendosis und daher unabhängig von der Bestrahlungsintensität. Ein klassisches Beispiel dafür ist die direkte Schwärzung von Bromsilberschichten ohne Entwicklung. Andere Beispiele stellen die biologischen Wirkungen dar, wie die Abtötung lebender Organismen oder die Auslösung von Mutationen. Ein weiteres bekanntes Beispiel ist die Dissoziation zweiatomiger Moleküle (H_2, N_2, etc.) durch Anregung von Elektronentermen infolge unelastischer Streuvorgänge. Massey und Mohr [32] haben theoretisch die Dissoziationswahrscheinlichkeit von Wasserstoff in neutrale Wasserstoffatome bis 25 eV berechnet. Für höhere Voltenergien und andere Zerfallsprodukte liegen noch keine Angaben vor.

Da dem Verf. auch keine anderen Angaben über Dissoziations oder Platzwechselwahrscheinlichkeiten bekannt waren, war es notwendig, an einem möglichst einfachen Objektmodell die Dosis zu bestimmen, die das Atom ertragen kann, ohne seinen Platz zu wechseln, also den reziproken Wert des Platzwechselquerschnitts (oder Dissoziations-Querschnitts)

Es mußte hierbei auf die Berücksichtigung der Wirkung unelastischer Streuvorgänge, die zur Anregung von Elektronentermen führen, verzichtet werden, da deren Wirkung in außerordentlichem Maße und in ganz unübersichtlicher Weise von dem speziellen Objektmodell abhängt. Es blieben für diese Betrachtungen also nur die sogenannten elastischen Streuvorgänge übrig.

Aus dem Impuls- und Energiesatz folgt nämlich, daß auch bei der elastischen Streuung des Elektrons um den Winkel ϑ Energie u_M (in Elektronenvolt) auf das Atom der Masse M übertragen wird:

$$u_M = 600 \frac{h^2}{eM} \cdot \frac{\sin^2 \vartheta/2}{\lambda^2} \tag{28}$$

Falls die Bindungsenergie des Atoms u_{Diss} kleiner als die beim Stoß übertragene Energie u_{M} ist, kommt es vom klassischen Standpunkt aus zu einer Abtrennung oder zu einem Platzwechsel des Atoms, d.h., wenn gemäß Gl. (28) $\vartheta > \vartheta_{\mathrm{Diss}}$ ist. Die Platzwechseldosis, d.h. die mittlere Zahl von Elektronen pro Flächeneinheit, die einmal zu einem Stoß $\vartheta > \vartheta_{\mathrm{Diss}}$ führt, ergibt sich aus dem Integral der Rutherford'schen Streuformel von $\vartheta = \vartheta_{\mathrm{Diss}}$ bis $\vartheta = 180°$ zu:

$$n_{\mathrm{Diss}} = 420, \frac{M}{m_{\mathrm{p}}} \cdot \frac{u_{\mathrm{Diss}}}{\lambda^2 Z^2} \tag{29}$$

(m_{p} = Masse des Protons).

c. Diskussion

Die so ermittelte Platzwechsetdosis ist natürlich viel zu groß, da die gesamten unelastischen Streuprozesse nicht berücksichtigt sind. In Übereinstimmung hiermit stehen die experimentellen Erfahrungen von König [33], der eine um mehrere Zehnerpotenzen geringere Platzwechseldosis feststellt. Andererseits ist die Abbildungsdosis zu klein gegenüber den praktischen Werten, da hier noch die Dosis erhöhenden Eigenschaften des photographischen Aufnahmematerials eine Rolle spielen. Sicher sind also die hier benutzten Werte für die Platzwechseldosis zu groß und für die Abbildungsdosis zu klein gegenüber den tatsächlichen Werten. Wenn also die errechneten Werte zu einer Abschätzung des Auflösungsvermögens benutzt werden, so wird sicher ein zu günstiger Wert für das Auflösungsvermögen erzielt. Auch die meisten anderen Vernachlässigungen wirken sich in dieser Weise aus.

Für den Fall des Wasserstoffs z.B. beträgt bei einer Bindungsenergie $u_{\mathrm{Diss}}=4{,}2$ eV und für $\lambda=0{,}05$ Å nach Gl. (29) die Dissoziationsdosis $n_{\mathrm{Diss}}/\mathrm{cm}^2=7{,}4.10^{21}$. Sie ist also von der gleichen Größenordnung wie die in Tabelle 14.3 angegebene Abbildungsdosis. Bei der Bewertung dieser Zahl ist zu berücksichtigen, daß einerseits die Abbildungsdosis wächst, wenn der tatsächliche Wert des Atomabstandes (Wasserstoffradius=0.56 Å) anstelle des Vergleichswerts von 3 Å der Tabelle 14.3 berücksichtigt wird. Außerdem wird es sicher nicht möglich sein, den optimalen Wert des Kontrastes zu erreichen, so daß hierdurch eine weitere Steigerung der Abbildungsdosis eintritt. Ferner ist bei der Dissoziation die Anregung von Elektronentermen durch unelastische Streuvorgänge nicht berücksichtigt. Der gesamte unelastische Streuquerschnitt ist um etwa den Faktor 27 großer als der elastische. Alle diese Faktoren wirken sich einerseits in einer Vergrößerung der Abbildungsdosis andererseits in einer Verkleinerung der Dissoziationsdosis aus, sodaß eine Abbildung von Wasserstoffatomen mit schnellen Elektronen unter den geschilderten Umständen sehr unwahrscheinlich wird.

Für Kohlenstoff in homöpolarer Einfachbindung ($u_{\mathrm{Diss}}=4{,}2$ eV) ergibt sich für $\lambda=0{,}05$ Å nach Gl. 29 eine Dissoziationsdosis $n_{\mathrm{Diss}}=2{,}4.\,10^{21}\ \mathrm{cm}^{-2}$. Die Abbildungsdosis nach Tabelle 14.3 ist um mehr als eine Zehnerpotenz niedriger, so daß die Abbildung von Kohlenstoffatomen zunächst durchaus möglich erscheint. Wenn allerdings die beim Wasserstoff diskutierten Umstände auch im diesem Fall berücksichtigt werden, so erscheint es trotz dieser Tatsache zumindest zweifelhaft, ob eine Abbildung von Kohlenstoffatomen oder von anderen leichten Atomen der organischen Chemie mit schnellen Elektronen möglich ist.

Eine weitere Einschränkung der Wahrscheinlichkeit der Abbildung ergibt sich bei Berücksichtigung tatsächlich vorliegender Abbildungsprobleme. Z.B. ist bei einfachen Kristallgittern die regelmäßige periodische Anordnung der Atome für die Abbildung ganz ohne Interesse. Diese Anordnung kann sehr einfach ohne jedes Abbildungsgerät aus dem Beugungswinkel der Röntgenoder Elektronen interferenzen 1. Ordnung mit beliebiger Genauigkeit bestimmt werden. Interessant wäre für die Abbildung nur die Lage der Fehlstellen oder die Abweichung der Lage bestimmter Atome (z.B. eines Eckatoms) von der mittleren Atomlage, die auf anderem Wege nicht festzustellen ist. Die Fehlstellen sind jedoch mit einer um Zehnerpotenzen geringeren Platzwechsel-Energie an ihren Ort gebunden als

die hier verwendeten Bindungsenergien des Wasserstoffs und des Kohlenstoffs. Da nach Gl. (29) mit der Platzwechselenergie auch die Platzwechseldosis abnimmt, sinkt dementsprechend auch die Wahrscheinlichkeit des Nachweises dieser Fehlstellen. Im Fall der Bestimmung der genauen Lage individueller Atome bedeutet die Forderung nach Erhöhung der Meßgenauigkeit einfach die Erhöhung der Zahl der Meßakte, also der Abbildungsdosis. Es wird also hier der Erfolg der Messung umso ungewisser und unwahrscheinlicher je höhere Ansprüche an die Meßgenauigkeit gestellt werden.

Eine Besserung der Situation tritt dagegen ein, wenn nicht die Lage einzelner Atome sondern ganzer Atomgruppen, z.B. der Atome längs einer Kristallkante, gemessen werden soll. Da es sich hier um ein statistisches Problem handelt, wird die Abbildungsdosis im Verhältnis zur Wurzel aus der Zahl der vermessenen Abstände gesenkt werden können. Auch der Einfluß sterischer Faktoren ist von Vorteil, da er sich in einer Erhöhung der Dissoziationsdosis und damit einer Erhöhung der Abbildungswahrscheinlichkeit auswirkt.

Zusammenfassend muß bemerkt werden, daß die vorstehenden Betrachtungen nur als erster Schritt gewertet werden wollen. Sie sollen nur auf einige Probleme aufmerksam machen, die bei der Abbildung von Atomen mit schnellen Elektronen auftreten können. Eingehendere Untersuchungen werden notwendig sein, um die hier angeschnittenen Fragen einem definitiven Entscheid zuzuführen.

Literatur

[1] E. Ruska, Z. f. Phys. **87,** 580 (1934).
[2] L. Marton, Physica **9,** 959 (1936).
[3] H. Boersch, Ann. Phys. **26,** 631 (1936).
[4] M. v. Ardenne, Elektronenmikroskopie, Berlin 1940 S. 60.
[5] H. Boersch, Jahrb. d. AEG-Forschung **7,** 27 (1940).
[6] L. Marton und L. J. Schiff, J. appl. Phys. **12,** 759 (1941).
[7] H. Boersch, Monatshefte f. Chemie **76,** 112 (1946); **76,** 116 (1946); **78,** 163 (1948).
[8] H. Boersch, Z. f. Naturforschung **2a,** 615 (1947).
[9] B. v. Borries, Z. f. Naturforschung **4a,** 51 (1949).
[10] J. Hillier u. E. G. Ramberg, Z. f. angew. Physik **2,** 273 (1950).
[11] C. E. Hall, J. appl. Physics **22,** 655 (1951).
[12] B. v. Borries u. E. Ruska, Naturwiss. **28,** 366 (1940).
[13] R. D. Heidenreich, Phys. Rev. **62,** 291 (1942); J. appl. Phys. **20,** 993 (1949).
[14] J. Hillier u. R. F. Baker, Phys. Rev. **61,** 722 (1942).
[15] H. Boersch, Z. f. Phys. **118,** 706 (1942); **121,** 746 (1943); Koll. ZS. **106,** 169 (1944).
[16] E. Kinder, Naturwiss. **31,** 149 (1943).
[17] J. Hillier, Phys. Rev. **60,** 743 (1941).
[18] L. J. Schiff, Phys. Rev. **61,** 721 (1942).
[19] O. Scherzer, J. appl. Phys. **20,** 20 (1949).
[20] F. Zernicke, Z. f. techn. Phys. **16,** 454 (1935).
[21] H. Boersch, Optik **5,** 436 (1949); Naturwiss. **35,** 26, 1948.
[22] G. Ruthemann, Ann. Phys. **6,** 113. 1948.
[23] G. Möllenstedt u. F. Heise, Phys. Blätter, **5,** 80, 1949; Optik **5,** 499 (1949); **6,** 251 (1950).
[24] M. v. Laue, Materiewellen und ihre Interferenzen, Leipzig 1944.
[25] A. G. Emslie, Phys. Rev. **45,** 43 (1934).
[26] H. Boersch, Z. f. techn. Physik **18,** 574 (1937).
[27] Nach Landolt-Bornstein, Phys.-Chem. Tabellen II. Erg. Bd. S. 1232, Berlin 1935.
[28] G. Möllenstedt u. O. Rang, Z. f. angew. Phys. **3,** 187 (1951).
[29] H. Boersch u. H. Mahl, DRP v. 15.11.38 Nr. 1733345, pl. 21g, Gruppe 372o
[30] F. Heise u. O. Rang, Optik **5,** 201 (1949); O. Rang, Optik **4,** 251 (1948); F. Heise, Optik **5,** 479 (1949).
[31] H. Boersch, Z. f. Phys. **127,** 391 (1950).
[32] H. W. S. Massey u. C. B. O. Mohr, Proc. Roy. Soc. **A.132,** 258 (1932).
[33] H. König, Z. f. Phys. **130,** 483 (1951).

15. On a Phase-Contrast Theory of Electron-Optical Image Formation

By H. Bremmer [1]

1. Introduction

In electron-microscopy the contrast in the image depends on the scattering undergone by the primary electrons when they pass through the object. This scattering is partly elastic and partly inelastic. In practice the contrast is nearly always produced by the elastically scattered electrons. In fact, for the high voltage applied, the deviation of electrons that suffer inelastical scattering is in general too small for them to be stopped by the aperture of the system.[2] Actually, the aperture is adjusted for a control of the elastically scattered electrons. In order to get an insight into the mechanism of the image formation, it is therefore reasonable to consider the effect of elastic scattering in the first instance.

The usual theories concerning elastic scattering start from the scattering produced by a single atom, the latter being represented by the Thomas-Fermi model. This scattering by a single atom may be computed either by classical or by wave-mechanical methods. The scattering effect of an object consisting of a large number of atoms is thereupon obtained by adding up in some way or other the scattering originating from the single atoms. When dealing with the wave-mechanical method, however, it is not necessary to restrict the Schrödinger equation in question to an isolated atom. This equation may just as well be applied to the entire space occupied by the scattering object; the electrostatic potential $\varphi\,(x, y, z)$ occurring in it should then refer to the complete set of scattering atoms.

The method we have in mind here will clearly show the essential properties of the resulting potential field governing the scattering; these properties may be independent of the atomic structure, atom models and so on. As a matter of fact, the modification imposed by the object on a primary wave e^{ikz} describing electrons arriving in the direction of the symmetry-axis of an electron lens (z-axis), can be described approximately as follows. The phase kz_σ for an undisturbed wave, to be observed at the point $P(x, y, z_\sigma)$ of the object plane $z=z_\sigma$ (this plane is situated immediately behind the object), will be increased in a first approximation by an amount Φ proportional to $\int_{-\infty}^{z_\sigma} \varphi(x, y, \zeta)d\zeta$. The integration, to be extended along a line drawn through P parallel to the incident beam, can actually be limited to the section of this line that is inside the object (continuous line in figure 15.1). The fact that the only object points influencing Φ are situated on such a line, indicates a shadow effect caused by the object.

---

[1] Philips Research Laboratories, N. V. Philips Gloeilampenfabrieken, Eindhoven, Netherlands.
[2] Compare figure 92 in Bodo v. Borries, die Übermikroskopie, Aulendorf 1949, p. 180.

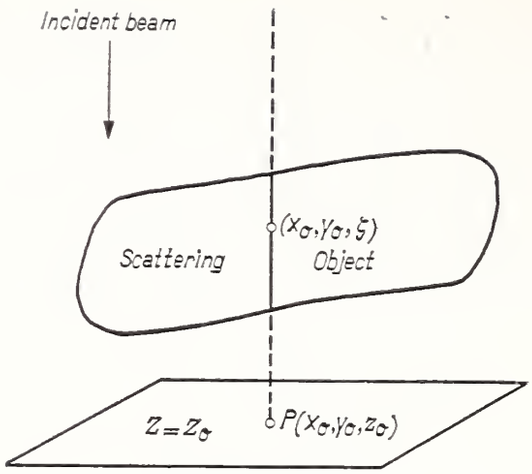


FIGURE 15.1.

The approximation indicated here is the better the faster the electrons are and the less the wave function changes in the object plane over distances comparable to a de Broglie-wavelength. The influence of the scattering object on electron waves proves to be very similar to that of a transparent object on optical waves; the former case leads to phase shifts proportional to $\int \varphi(x, y, \zeta) d\zeta$, the latter to phase shifts proportional to $\int \mu(x, y, \zeta) d\zeta$, when $\mu(x, y, z)$ is the optical refractive index in the object point (x, y, z). The results stated above will be derived in the next sections.

2. Series Development of the Wave Function Describing the Elastically Scattered Electrons

The wave-mechanical picture of a homogeneous beam of electrons, moving in vacuo in the direction of the z-axis, consists of a plane wave $\psi_0 = e^{ikz}$. The wave number $k = 2\pi/\lambda$ depends on the de Broglie-wavelength in vacuo, viz.

$$\lambda = \sqrt{\frac{150}{V}}\, 10^{-8} \text{ cm},$$

V being the accelerating emf in volts.

We shall define the electrostatic potential $\varphi(x, y, z)$ inside the object as the energy of a charge of $+1$ esu in the point under consideration (x, y, z), relative to the energy of the same charge far away from the object. The Schrödinger equation then reads

$$(\Delta + k^2)\psi = -\frac{8\pi^2 me}{h^2}\, \phi\psi. \tag{1}$$

The notation has been chosen so as to show clearly the disturbing effect of the object: Only inside it does the right-hand member differ from zero. An essential feature of eq (1) is that φ can be considered as a function given a priori; this does not hold for inelastic scatterings.

The form of eq (1) is that of the equation for harmonic retarded potentials in Maxwell's theory. We obtain at once the following "solution" if we treat the right-hand member of (1) as if it were a

known function

$$\psi(P)=e^{ikz_p}+\frac{2\pi me}{h^2}\iiint d\tau_Q\varphi(Q)\psi(Q)\,\frac{e^{ikQP}}{QP}. \tag{2}$$

The integral in eq (2) extends over the entire object, a special point $Q(\xi,\eta,\zeta)$ of which has a distance QP to the point of observation P; the volume elements of the object are given by $d\tau_Q=d\xi dnd\zeta$. As a matter of fact, eq (2) constitutes an integral equation [3] in which we have allowed for the role of the primary wave: This would represent the complete solution if there is no scattering object.

The well-known procedure of successive approximations [4] leads to a series of the following form (if convergent) for the complete solution of eq (2):

$$\psi(P)=\sum_{n=0}^{\infty}\psi n(P);$$

$\psi_0(P)=e^{ikz_p}$ represents the primary wave, while any two consecutive terms are connected by the relation

$$\psi_n(P)=\left(\frac{2\pi me}{h^2}\right)\iiint d\tau_Q\varphi(Q)\psi_{n-1}(Q)\frac{e^{ikQP}}{QP}, \qquad (n=1,2,3,\ \ldots).$$

By combining these recurrence relations we arrive at the following explicit form for ψ_n as an $3n$-fold integral

$$\psi_n(P)=\left(\frac{2\pi me}{h^2}\right)^n\int\ldots\int d\tau_1 d\tau_2\ \ldots\ d\tau_n\phi(Q_1)\phi(Q_2)\ \ldots\ \varphi(Q_n)\times$$

$$\frac{\exp ik(\zeta_1+Q_1Q_2+Q_2Q_3+\ \ldots\ +Q_{n-1}Q_n+Q_nP)}{Q_1Q_2Q_2Q_3\ \ldots\ Q_{n-1}Q_nQ_nP}. \tag{3}$$

The integration extends over the $3n$ coordinates ξ_j,η_j,ζ_j of the n points $Q_j(j=1,2,\ \ldots,\ n)$ each of which is situated in a volume-element $d\tau_j=d\xi_j d\eta_j d\zeta_j$ of the object; Q_tQ_s indicates the distance between the two points Q_t and Q_s.

The splitting of the total scattered wave into the terms of the series $\sum_{n=1}^{\infty}\psi_n(P)$ can be interpreted physically as follows. Any of the n volume-elements in (3) points to the scattering caused by such an element. The final expression for ψ_n thus represents the effect of n successive scatterings. Therefore, the contribution described by ψ_n may be termed as nth-order scattering. It is to be emphasized, however, that one or more of the volume-elements concerned with one nth-order scattering may belong occasionally to one and the same atom because so far the atomic structure of the object has not come to the fore at all.

Before discussing a proper approximation of ψ_n, to be worked out in section 4, we shall investigate a limit connected with it.

---

[3] Compare N. F. Mott, Proc. Roy. Soc. London [A] **127**, 659 (1930).

[4] The same method was applied by Born, Z. Physik. **38**, p. 816 (1926), when investigating the behavior at great distance of the wave scattered by a single atom.

3. A lemma concerning the limit of two-dimensional diffraction integrals for infinitely small wavelength

The integral (3) may be considered as referring to n linear integrations in the direction of the z-axis (integration variables $\zeta_1,\zeta_2, \ldots,\zeta_n$) and to n double integrations over the planes $z=\zeta_j (j=1,2, \ldots,n)$ perpendicular to the z-axis (pairs of integration variables ξ_j,η_j). In order to prepare for the approximate evaluation of ψ_n in the next section, we shall now discuss the asymptotic behavior of double integrals such as occurring in eq (3) for k tending to infinity (de Broglie wavelength $2\pi/k$ tending to zero). What we actually need, is an investigation of the limit $I=\lim_{k\to\infty} I(k)$ for

$$I(k)=ke^{-ikd}\int\int_{-\infty}^{\infty} dx_Q dy_Q u(Q)\,\frac{e^{ikQP}}{QP}, \tag{4}$$

in which $D=|z_P-z_Q|$ represents the distance between P and the plane of integration $z=z_Q$ while $u(Q)$ does not depend on k.

The integral is of a type frequently occurring in diffraction theories. By changing over to polar coordinates ρ_Q,φ_Q with origin at the projection P_0 of P on the integration plane $z=z_Q$ (see fig. 15.2) we obtain

$$I(k)=ke^{-ikd}\int_0^{2\pi} d\phi_Q \int_0^{\infty} d\rho_Q \rho_Q u(Q)\,\frac{e^{ikQP}}{QP}. \tag{5}$$

Now we replace ρ_Q by the variable $\zeta=QP-P_0P=(\rho_Q^2+D^2)^{1/2}-D$, which is also connected with the division of the Q-plane into Fresnel zones in the case of a virtual point source at P. After a transformation of the inner integral in eq (5) we get, in succession

$$\begin{aligned} I(k) &= k\int_0^{2\pi} d\phi_Q \int_0^{\infty} d\zeta\, u(Q)e^{ik\zeta} \\ &= -i\int_0^{2\pi} d\phi_Q \int_{\zeta=0}^{\zeta=\infty} u(Q)\,d(e^{ik\zeta}) \\ &= -i\int_0^{2\pi} d\phi_Q \{u(Q)e^{ik\zeta}\}\Big|_{\zeta=0}^{\zeta=\infty} + i\int_0^{2\pi} d\phi_Q \int_0^{\infty} d\zeta\,\frac{\partial u(\phi),\zeta)}{\partial\zeta}\,e^{ik\zeta}. \end{aligned}$$

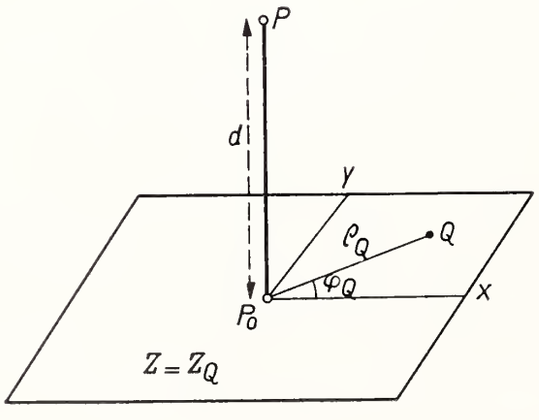


FIGURE 15.2.

In the first term the contribution from $\zeta=\infty$ vanishes if the integration over Q concerns a *finite* part of the Q-plane only. This happens to be in our case due to the finite dimensions of the scattering object. Hence ($\zeta=0$ corresponding to the position P_0 of Q)

$$I(k)=2\pi i u(P_0)+i\int_0^{2\pi} d\varphi_Q \int_0^{\infty} d\zeta \frac{\partial u(\varphi_Q,\zeta)}{\partial \zeta} e^{ik\zeta}, \tag{6}$$

in which the upper limit of integration of the inner integral can also be considered as finite. Therefore, this integral tends to zero for $k\to\infty$ according to Riemann-Lebesque's theorem

$$\operatorname*{Lim}_{k\to\infty} \int_a^b d\zeta X(\zeta) e^{ik\zeta}=0. \tag{7}$$

This relation [5] is of essential importance for the derivation of geometrical optics as the limiting case [5] of wave optics. The function X may become infinite provided its integrability is guaranteed. As a matter of fact the function $\partial u/\partial\zeta$ of (6) behaves like $\zeta^{-1/2}$ as $\zeta\to\infty$ in view of the formula

$$\frac{\partial u}{\partial \zeta}=\left(\cos\varphi\frac{\partial u}{\partial x}+\sin\varphi\frac{\partial u}{\partial y}\right)\frac{(\zeta+d)}{\sqrt{2d\zeta+\zeta^2}}$$

provided $u\ (x,\ y)$ can be expanded into a Taylor series at $x=y=0$. The final result then proved, viz.,

$$\operatorname*{Lim}_{k\to\infty} ke^{-ikD}\iint_{-\infty}^{\infty} dz_Q dy_Q u(Q)\frac{e^{ik.QP}}{QP}=2\pi i u(P_0),$$

may also be given in the alternative form as the asymptotic expression

$$\iint_{-\infty}^{\infty} dx_Q dy_Q u(Q)\frac{e^{ik.QP}}{QP}\sim\frac{2\pi i}{k}e^{ikD}u(P_0)\qquad(k\to\infty). \tag{8}$$

The sign $\sim$ indicates that the ratio of both sides does tend to unity for $k\to\infty$.

4. The asymptotic value of $4n$ for $k\to\infty$

The general formula (8) can be extended to multiple integrals depending on n different Q-planes instead of on a single one. The case of two planes $z=z_1$, and $z=z_2$, for instance, deals with

$$I_2=\iint_{-\infty}^{\infty} dx_{Q_1}dy_{Q_1}\iint_{-\infty}^{\infty} dx_{Q_2}dy_{Q_2}u(Q_1,Q_2)\frac{e^{ik.Q_1Q_2}}{Q_2Q_2}\frac{e^{ik.Q_2P}}{Q_2P}.$$

In order to derive the asymptotic value for $k\to\infty$, we invert the order of integration and substitute for the double integral referring to Q_1 its approximate value according to (8). It follows that

$$I_2\sim\frac{2\pi i}{k}e^{ik|z_1-z_2|}\iint_{-\infty}^{\infty} dx_{Q_2}dy_{Q_2}u(Q_2^*,Q_2')\frac{e^{ik.Q_2P}}{Q_2P},\qquad(k\to\infty),$$

[5] For a proof see, e. g., S. Bochner, Vorlesungen über Fouriersche Integrale, Leipzig 1932, p. 4.

Q_2' (x_{Q_2}, y_{Q_2}, z_1) being the projection of Q (x_{Q_2}, y_{Q_2}, z_2) on the plane $z=z_1$. Another application of (8) to the remaining integration over Q_2 results in

$$I_2 \sim \left(\frac{2\pi i}{k}\right)^2 e^{ik(|z_1-z_2|+|z_2-z_p|)} u(p_1,p_2), \qquad (k\to\infty)$$

when P_1 (x_P, y_P, z_1) and P_2 (x_P, y_P, z_2) are the projections of P on the planes $z=z_1$ and $z=z_2$, respectively. The extension of this formula to an integration over any number of planes $z=z_1, \ldots, z=z_n$ is obvious. In the corresponding expression, viz.,

$$\iint_{-\infty}^{\infty} dx_{Q_1} dy_{Q_1}, \ldots, \iint_{-\infty}^{\infty} dx_{Q_n} dy_{Q_n} u(Q_1, \ldots, Q_n) \frac{e^{ik\cdot Q_1 Q_2}}{Q_1 Q_2} \cdots \frac{e^{ik\cdot Q_{n-1} Q_n}}{Q_{n-1} Q_n} \frac{e^{ik\cdot Q_n P}}{Q_n P}$$

$$\sim \left(\frac{2\pi i}{k}\right)^n \exp\{ik(|z_1-z_2|+\ldots+|z_{n-1}-z_n|+|z_n-z_P|)\} u(p_1, \ldots, p_n), \qquad (k\to\infty) \quad (9)$$

occur the projections $P_1, \ldots, P_n$ of P on all the planes of integration. This formula can be applied straight away to the $2n$ integrations over $\zeta_1, n_1, \ldots, \zeta_n, n_n$ of the multiple integral (3) for ψ_n by substituting $\varphi(P_1) \ldots \varphi(P_n)$ for u $(P_1, \ldots, P_n)$. Taking into account the remaining integrations over $\zeta_1, \ldots, \zeta_n$ we find

$$\psi_n(P) \sim \left(\frac{4\pi^2 ime}{kh^2}\right)^n \int d\zeta_1 \ldots \int d\zeta_n e^{\{ik(\zeta_1+|\zeta_2-\zeta_1|+\ldots+|\zeta_n-\zeta_{n-1}|+|z_P-\zeta_n|)\}} \times \varphi(P_1) \ldots \varphi(P_n), \qquad (k\to\infty). \quad (10)$$

The domain of integration of this multiple integral can be divided into $2n$ subdomains each of which is characterized by fixed signs of the n expressions $|\zeta_i-\zeta_j| = \pm(\zeta_i-\zeta_j)$. Of all the corresponding subintegrals, so far as they are different from zero, only a single one shows an exponent that is independent of all the integration variables. This special subintegral is characterized by the domain of integration

$$z_P > \zeta_n > \zeta_{n-1} \ldots > \zeta_2 > \zeta_1, \quad (11)$$

whence its value becomes

$$\left(\frac{4\pi^2 ime}{kh^2}\right)^n e^{ikz_P} \int_{-\infty}^{z_P} d\zeta_n \varphi(x_p, y_p, \zeta_w) \int_{-\infty}^{\zeta_n} d\zeta_{n-1} \varphi(x_p, y_p, \zeta_{n-1}) \ldots \times \int_{-\infty}^{\zeta_2} d\zeta_1 \varphi(x_P, y_P, \zeta_1). \quad (12)$$

The lower limits of integration may be replaced by a finite number ϵ if the object is situated wholly beyond the plane $z=\epsilon$.

The subintegral (12) is zero for points P in front of the object because $\varphi(x_P, y_P, \zeta_n)$ in the first integral then refers to points outside

the object. For P beyond the object, however, we may compare (12) with other subintegrals, for instance with the following one that has only a single sign, that of $\zeta_3-\zeta_2$, which differs from the corresponding one according to (11) ($\zeta_3<\zeta_2$ instead of $\zeta_3>\zeta_2$)

$$\left(\frac{4\pi^2 ime}{kh^2}\right)^n e^{ikz_P}\int_{-\infty}^{z_P} d\zeta_n \varphi(x_P, y_P, \zeta_n)\int_{-\infty}^{\zeta_n} d\zeta_{n-1}\varphi(x_P, y_P, \zeta_{n-1}) \cdots$$

$$\times\int_{-\infty}^{\zeta_\nu} d\zeta_3 e^{-2ik\zeta_3}\varphi(x_P, y_P, \zeta_3)\int_{\zeta_3}^{\infty} d\zeta_2 e^{2ik\zeta_2}\varphi(x_P, y_P, \zeta_2)\int_{-\infty}^{\zeta_2} d\zeta_1\varphi(x_P, y_P, \zeta_1). \tag{13}$$

A comparison of (12) and (13), while omitting the common factor e^{ikz_P} shows (13) to be infinitely small relative to (12) for $k\to\infty$ owing to the exponential factors still remaining in (13); in fact, these factors cause the limit of (13) to vanish in view of the basic property (7). The same result holds for any other subintegral of (10), each of which shows at least one exponential factor (apart from e^{ikoz_P} Hence it follows that $\psi_n(P)$ is represented asymptotically by the single integral (12). In its turn the latter can be simplified considerably with the aid of the identity

$$\int_{-\infty}^{z} d\zeta_n\varphi(\zeta_n)\int_{-\infty}^{\zeta_n} d\zeta_{n-1}\varphi(\zeta_{n-1})\int_{-\infty}^{\zeta_{n-1}} d\zeta_{n-2}\varphi(\zeta_{n-2}) \cdots \int_{-\infty}^{\zeta_2} d\zeta_1\varphi(\zeta_1)$$
$$= \frac{1}{n!}\left\{\int_{-\infty}^{z} d\zeta\varphi(\zeta)\right\}^n. \tag{14}$$

This identity is easily proved by induction while using the following recurrence relation, which applies to either of the members $K_n(z)$ of

$$\frac{dK_n(z)}{dz} = \varphi(z)K_{n-1}(z). \tag{14}$$

An application of (14) to (12), the only remaining subintegral of ψ_n for $k\to\infty$, leads to the following final asymptotic expression

$$\psi_n(P)\sim\frac{e^{ikz_P}}{n!}\left\{\frac{4\pi^2 ime}{kh^2}\int_{-\infty}^{z_P} d\zeta\varphi(x_P, y_P\zeta)\right\}^n, \qquad (k\to\infty). \tag{15}$$

As to the derivation of (15), we emphasize the essential role played by the reduction of the domain of integration of (10) to that of the subdomain characterized by (11). This reduction for $k\to\infty$ can be interpreted physically by the fact that the source Q_j of the jth scattering should be *beyond* the source Q_{j-1} of the preceding scattering; in other words, *forward* scatterings are only to be taken into account if k tends to infinity.

We remark that (15) might also be derived from the exponential integral (3) with the aid of the saddle-point method. This method (which will be used in section 6) is based on the fact that the main contribution to an exponential integral such as (3) does arise from the vicinity of the saddlepoint S. The situation of this point is generally to be derived by equating to zero the partial differential quotients

of the exponent with respect to the integration variables. In the case of (3), however, this method is less convenient because we have an infinity of saddlepoints. In fact, any set of positions of the points $Q_1, \ldots, Q_n$ on the line drawn through P parallel to the z-axis corresponds to a saddlepoint of the exponent of (3) provided each point Q_j is situated beyond the preceding one Q_{j-1}. The physical meaning of these saddle-point positions is that the scattering is most important when effected in the forward z-direction by n successive volume-elements.

5. The summation over the contributions due to scatterings of different orders

The approximations (13) of the nth order scatterings ($n=1,2,3\ldots$) constitute the terms of an exponential series. A summing of these terms, completed by the zero-order term ψ_0 (primary wave), yields for points P beyond the object

$$\psi(P) \sim e^{i\{kz_P + \phi(P)\}}, \tag{16}$$

with

$$\Phi(P) = \frac{4\pi^2 me}{kh^2} \int_{-\infty}^{z_P} d\zeta \varphi(x_P, y_P, \zeta).$$

The latter expression can be replaced by

$$\Phi(P) = \frac{4\pi^2 me}{kh^2} \int_{-\epsilon}^{0} d\zeta \varphi(x_P, y_P, \zeta), \tag{16a}$$

if the potential field of the object is entirely situated between the planes $z=-\epsilon$ and $z=0$.

The modification imposed by the object thus results in a mere phase shift Φ as far as the approximations for $k \to \infty$ are concerned. The remarkable thing is that the terms of the series

$$e^{i\Phi(P)} = \sum_{n=0}^{\infty} \frac{\{i\phi(P)\}^n}{n!}$$

just represent the nth-order scatterings, that is the contributions resulting from a fixed number n of successive elastic scatterings. A similar interpretation of the individual terms of an exponential series is essential in the corpuscular theory of elastic scattering by Bodo von Borries [6] in which the atoms and not our volume-elements $d\tau_j$ are the scattering units. A related series also occurs in D. L. Dexter and W. W. Beeman's theory of small-angle scattering of X-rays.[7]

The simple form of (16) allows the following remarks:

(a) The coherence of the scatterings of different orders is obvious. In fact, in the case of complete incoherence we should have to add up the squares of the moduli of (15) in order to get the square of the final amplitude of the resulting wave function. This summation leads to the series development of a Bessel function (of order zero and imagi-

[6] Compare S. Bochner, Vorlesungen uber Fouriersche Integrale, Leipzig, 1932, formula (77) on p. 175.
[7] Phys. Rev. **76,** 1782 (1949).

nary argument) instead of that of an exponential function. The modulus then obtained exceeds unity, which signifies that the coherence does cause an attenuation of ψ and not a magnification.

(b) Apart from the factor e^{ikz_P} the individual terms (15) are either positive (for n equal to a quadruple $4m$) or negative (for $n=4m+2$) or positive imaginary ($n=4m+1$) or negative imaginary ($n=4m+3$). The scattered waves thus disintegrate into four groups each of which is composed of contributions having identical phases.

6. The Effect of the Aperture in Electron Microscopy

In the case of electron microscopy we are particularly interested in the distribution of the wave function in the object plane which may be a plane directly beyond the object. For simplicity's sake we take this plane as the coordinate plane $z=0$. The wave function then is represented there by $e^{i\Phi(x,y)}$ with

$$\Phi(x,y)=\frac{4\pi^2 me}{kh^2}\int_{-\epsilon}^{0} d\zeta\,\varphi(x,y,\zeta), \tag{17}$$

as far as the above approximation is concerned.

In the case of an ideal lens without any aberrations and any aperture the corresponding wave function in the point $P(x, y)$ of the image plane equals (apart from a constant factor and a phase factor corresponding to the optical path length between the two conjugated points in object and image plane)

$$e^{i\Phi(x/M,\,y/M)}, \tag{18}$$

M being the magnification. Such an image does not show any contrast, the modulus of the wave function being constant throughout the image plane. The role of the aperture in causing the contrast will now be described in terms of wave theory.

The role of the aperture becomes apparent in the easiest way when it is placed in the back focal plane, that is, the plane conjugated to that containing the source (which has to be supposed at infinity in accordance with our assumption of a parallel beam of primary electrons). The paraxial diffraction theory of lenses leads to values of the wave function at the points (x_1, y_1) of the back focal plane that are proportional to $G(kx_1/f, ky_1/f)$, where $G(x, y)$ is the Fourier transform

$$G(x,y)=\int\int_{-\infty}^{\infty} d\xi d\eta\, e^{i\Phi(\xi,\eta)} e^{-i(x\xi+y\eta)}$$

of the distribution $e^{i\Phi(x,y)}$ valid in the object plane (f=focal distance). This wave function represents the mathematical formulation of the physical ideas in Abbe's primary image.

Now the aperture causes special parts of the frequency spectrum to be eliminated. The contribution $1/(4\pi^2)G(u,v)e^{i(ux+vy)}$ is then suppressed for all values of u and v (the frequencies of $e^{i\Phi}$ in the x- and y-directions) for which the corresponding points $(fu/k, fv/k)$ in the focal plane happens to be situated *outside* the edges of the aperture. Hence the wave function at the point (x, y) of the image plane is

proportional to the following expression instead of (18).

$$\frac{1}{4\pi^2}\iint du\,dv\,e^{\frac{6}{M}(xu+yv)}\int\int_{-\infty}^{\infty} d\xi\,d\eta\,e^{i\Phi(\xi,\eta)-i(\xi u+\eta v)}, \qquad A(fu/k,fv/k)<0, \tag{19}$$

if $A(x,y)<0$ characterizes the domain of the focal plane that is inside the aperture.

An approximate evaluation of (19) is possible with the aid of the saddle-point method. The saddlepoint has to be derived in this case by equating to zero the partial differential quotients of the exponent of (19) with respect to u, v, ξ and η. The corresponding relations read

$$\frac{x}{M}=\xi_s, \qquad \frac{y}{M}=\eta_s, \tag{20a}$$

$$\left(\frac{\partial\Phi}{\partial\xi}\right)_s=u_s, \qquad \left(\frac{\partial\Phi}{\partial\eta}\right)_s=v_s. \tag{20b}$$

According to (20a) the saddlepoint values ξ_s and η_s of ξ and η represent the coordinates of the point P′ in the object plane that is conjugated to $P(x,y)$ in the image plane. Further, the geometrical significance of (20b) is clear from the following construction. At each point (ξ,η) of the object plane $z=0$ we plot a perpendicular in the negative z-direction having a length $\Phi(\xi,\eta)/k$. The locus of the end-points of these perpendiculars constitute a surface $z=-\Phi(x,y)/k$ which may be called the phase surface (σ in fig. 15.3. In this figure the lens has been drawn, for convenience, as an optical lens). From (20b) we infer that the rays parallel to the normal n of σ at P_1' (the point of σ having P' as projection on the object plane) are imaged by the lens in the back-focal point $K(fu_s/k,fv_s/k)$. This is simply verified by observing that the image of any paraxial direction with directional cosines $\cos\alpha$, $\cos\beta$, $\cos\gamma$ is situated at the point $(f\cos\alpha, f\cos\beta)$ in the focal plane, while the cosines of n are given (in the paraxial approximation)

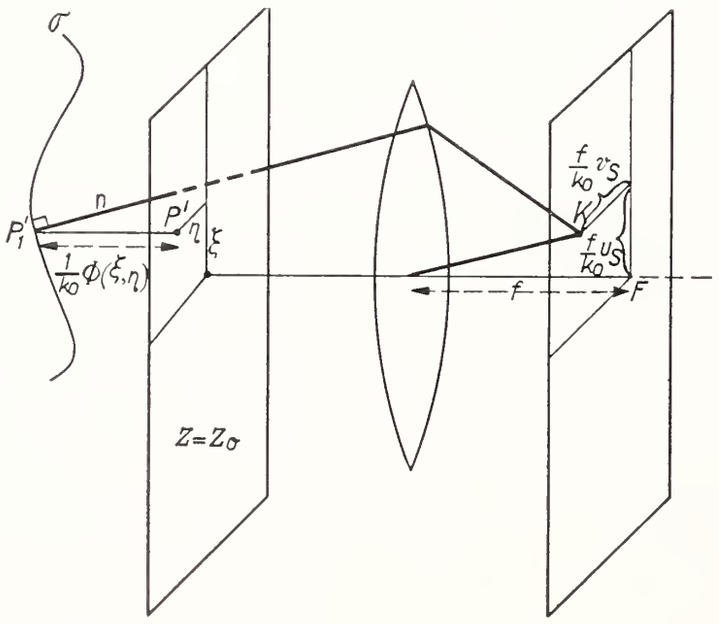


FIGURE 15.3.

by

$$\frac{1}{k}\left(\frac{\partial\Phi}{\partial\xi}\right)_{P'}, \qquad \frac{1}{k}\left(\frac{\partial\Phi}{\partial\eta}\right)_{P'}, \qquad 1.$$

In these considerations ξ_s and η_s appear to be representative for the point P' (conjugated to the point P of observation), u_s and v_s for the focal point of rays parallel to the normal n of the phase surface at the point corresponding to P'. An appreciable numerical value of an exponential integral such as (19) now is only to be expected of the saddlepoint is situated inside the domain of integration. In our case this holds exclusively if $A(fu_s/k,fv_s/k)<0$, that is if K is inside the aperture. If this happens to be we can evaluate the so-called second-order saddlepoint approximation. The latter is here obtained by expanding the exponent into a Taylor series of four variables around the saddlepoint, by stopping this series after the second-order terms and by extending the domain of integration until infinity. It is easily verified that the approximation thus derived is identical with (18). However, if K is outside the aperture, the corresponding approximate value of (19) has to be taken as zero.

The approximation concerned can be summarized as follows: The wave function in the image plane does differ from zero only at special points P, which depend on the normal to the phase surface at the corresponding point P'. The rays parallel to the normal must be focused at a point *inside* the aperture. The angles between these directions and the z axis depend on the phase shift ϕ caused by the object due to elastical scattering. In other words, the phase contrast arising from the elastically scattered electrons reduces those areas of the object plane that contribute to the image formation. This is the wave-mechanical formulation of the statement in corpuscular language, that the scattering mechanism reduces the number of electrons that pass through the aperture. It has thus been shown how the complete imaging mechanism, as far as it depends on elastic scattering, can be described with the aid of wave theories. The wave treatment includes the explanation of the scattering by the object, the formation of Abbe's primary image, and the role of the aperture.

7. Corrections to the phase-contrast theory

The theory outlined above is essentially based on the limit (8) which has been applied to the integrations perpendicular to the z-axis that occur in the multiple integral (3) for ψ_n. We can obtain a better approximation to ψ_n by using more accurate approximations of the two-dimensional integrals

$$\iint dx_Q\, dy_Q\, u(Q)\frac{e^{ik\cdot QP}}{QP}.$$

Such approximations can be obtained from any finite number of terms of the asymptotic series

$$\int_{-\infty}^{\infty}\!\!\int dx_Q\, dx_Q\, u(Q)\frac{e^{ik\cdot QP}}{QP}\sim$$

$$-2\pi\sum_{n=0}^{\infty}\frac{D^{2n+1}}{4^n(n!)^2}\Delta_2^n u(P_0)\left\{-1+\frac{\partial^2}{d(ikD)^2}\right\}^n\left(\frac{e^{ikD}}{^{ik}D}\right), \qquad (k\to\infty), \quad (21)$$

in which $D=|Z_p-Z_Q|$; Δ_2 indicates the two-dimensional Laplace operator, for instance,

$$\Delta_2 u(x, y)=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2},$$

$$\Delta_2^2 u(x, y)=\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)^2 u=\left(\frac{\partial^4}{\partial x^4}+2\frac{\partial^4}{\partial x^2 \partial y^2}+\frac{\partial^4}{\partial y^4}\right)u.$$

The derivation of (21) is completely analogous to that of the series for the corresponding integral

$$-\frac{1}{2\pi}\frac{\partial}{\partial z_P}\iint dx_Q\, dy_Q\, u(Q)\frac{e^{ik\cdot QP}}{QP},$$

which will be published elsewhere.[8] Obviously the expression given in (8) represents the first term of (21). The next few terms are given by

$$\iint_{-\infty}^{\infty} dx_Q\, dy_Q\, u(Q)\frac{e^{ik\cdot QP}}{QP}\sim\frac{2\pi i}{k}e^{ikD}\left\{u(P_0)+\frac{iD}{2k}\left(1+\frac{i}{kD}\right)\Delta_2 u(P_0)\right.$$

$$\left.-\frac{D^2}{8k^2}\left(\frac{3i}{kD}-\frac{3}{k^2D^2}\right)\Delta_2^2 u(P_0)\ldots\right\}\qquad(k\to\infty).$$

With the aid of this series, it is possible in principle to derive an expansion for each ψ_n that shows terms of decreasing orders of magnitude with respect to $k^{-1}=\lambda/(2\pi)$. The first term of such a series is of the order k^{-n} and is represented by (15), the next term is the order of k^{-n-1}, and so on. Contributions proportional to a fixed power k^{-m} occur in each of the terms $\psi_1, \psi_2, \ldots, \psi_m$.

The actual computation of the correction terms is very tedious. We only mention the first two terms of ψ_1 (P beyond the object).

$$\psi_1(P)=i\frac{4\pi^2 me}{h^2 k}e^{ikz_P}\left\{\int_{-\epsilon}^{x_p} d\zeta\,\varphi(x_P, y_P, \zeta)\right.$$

$$\left.+\frac{i}{2k}\int_{-\epsilon}^{z_P} d\zeta(z_P-\zeta)\Delta_2\varphi(x_P, y_P, \zeta)+\ldots\right\}.\qquad(22)$$

This formula illustrates how the first correction term depends on the two-dimensional Laplace operator of the potential field in a plane perpendicular to the symmetry axis. The $k+$th order correction term of ψ_n, which is proportional to k^{-n-k}, is in general composed of contributions of the first k iterated Laplace operators $\Delta_2, \Delta_2^2 \ldots, \Delta_2^k$. These terms may also be connected with the electron density P (x, y, z) according to Poisson's equation

$$\Delta_2\varphi=-\frac{\partial^2\varphi}{\partial z^2}+4\pi e\rho.$$

[8] See H. Bremmer, On the asymptotic evaluation of diffraction integrals with a special view to the theory of optical contrast, Physica (to be published).

It is thus possible to reduce the second term of (22) with the aid of two partial integrations into the following expression if we assume P to be outside the potential field.

$$-\frac{8\pi^3 me^2}{h^2k^2}\, e^{ikz_P}\int_{-\infty}^{z_P} d\zeta(z_P-\zeta)\rho(\zeta).$$

The influence of similar terms may disturb the simple-phase-contrast conditions according to (16).

8. Final Remarks

The theory leading to the phase-contrast formula (16) is directly applicable to all propagation problems concerning the forward scattering caused by a plane wave e^{ikz} that passes through a dispersive medium having a refractive index μ given by

$$\mu^2(x, y, z)=1-\frac{f(x, y, z)}{k^2}$$

($k=2\pi/\lambda$ wave number in the regions where $f=0$). The complete differential equation then reads

$$(\Delta+k^2)\Pi=f(x, y, z)\Pi. \tag{23}$$

Our problem of elastic scattering due to fast electrons is included as the special case for which

$$f(x, y, z)=-\frac{8\pi^2 me}{h^2}\,\varphi(x, y, z).$$

The theory is particularly suited for the investigation of the scattered waves directly beyond the scattering space. The theory of the scattering observed at great distances shows different features because it depends completely on Fourier analysis.[9] Our main result (16) concerning the total wave function beyond the scattering object for great values of k, has the following form for the general problems described by (23)

$$\Pi(P)\sim e^{ikz_P}e-\frac{i}{2k}\int_{-\epsilon}^{0} d\zeta f(x_P, y_P, \zeta). \tag{24}$$

Here again, the individual terms of the exponential series for the second factor do represent the contributions produced by a special number of successive scatterings.

Some other problems to which the above theory may be applied are:

(a) *The scattering of radio waves in the ionosphere.* Here we have

$$f(x, y, z)=\frac{4\pi e^2}{mc^2}N(x, y, z),$$

[9] The case of a two-dimensional object has been treated thoroughly by Booker, Ratcliffe and Shinn, Phil. Trans. Roy. Soc. of London [A] **242** (1950), p. 579–609.

N being the number of electrons per unit volume;

(*b*) *The optical scattering by media having a refractive index* μ *slightly different from unity.* In this case we should substitute

$$f(x, y, z) = -k^2\{\mu^2(x, y, z) - 1\}.$$

The fundamental formula (24) here becomes

$$\Pi(P) \sim e^{ikz_P} e^{\frac{ik}{2}\int_{-\epsilon}^{0} d\zeta\{\mu^2(x_P, y_P, \zeta)-1\}}.$$

For a nondispersive medium (μ independent of k) the terms corresponding to the different orders of scattering are now proportional to increasing powers of k instead of decreasing powers in the case of scattering by fast electrons. Therefore, the dominating order of scattering is very different in either case.

(c) *The elastic scattering of X-rays.* The exact differential equation to be applied here [10] is a bit more complicated than (23) so that the above theory has to be modified accordingly.

The theory may possibly be extended to problems with a complex-valued f, in which the imaginary part of f corresponds to an absorption. This may be important in order to include the effects of inelastic scattering in our electronic problem.

[10] See M. von Laue. Röntgenstrahl-Interferenzen, Berlin 1941, equation (26.17) on p. 246.

16. Theoretical Study of Quality of Images

By A. Marechal [1]

Introduction

The following study of the quality of images has been performed for immediate application to light optics.[2] Although the orders of magnitudes of the resolving power and wavelengths, and the way in which the objects produce absorption or retardation are different in light and electron optics, the results give an account of the effects of aberrations in electron optics.

In light optics the objects to be considered are composed of luminous sources that can be either perfectly incoherent (photography, astronomy, etc.) or partially coherent as is the case in microscopy. We have studied the two extreme cases of perfect incoherence and perfect coherence of phase, for which the basic principles of the formation of the images are very simple.

Formation of Images

Let us consider first the case where the object is incoherent; the various points act as independent sources. Let $O(y,z)$ be the repartition of luminance in the object (fig.16.1), and let

$$D(y',z')=|A(y',z')|^2=\left|\iint_P \exp\left[j\frac{2\pi}{\lambda}(\Delta+\beta y'+\gamma z')\right]d\beta d\gamma\right|^2$$

be the repartition of illumination in the diffraction image of a single point. Δ is the distortion of the wave front in the exit pupil; λ is the wavelength; y' and z' are coordinates in the image; and β and γ are angular coordinates in the exit pupil.

The repartition of light will be

$$I(y',z')=\iint_0 D(y'-y,z'-z)O(y,z)dydz.$$

Various objects can be studied for example: (1) Isolated bright point, (2) isolated dark point, (3) isolated bright line, (4) isolated dark line, and (5) edge of a bright area.

For each of those cases the function $O(y,z)$ is different. If we represent by a surface the function $D(y'z')$ (sometimes called the diffraction solid), the contrast of the image of an isolated bright point will be obviously related to the maximum illumination in the center of the diffraction solid. The case of an isolated dark point will be comple-

---

[1] Institut d'Optique, Paris, France.

[2] The detailed computations are to be found in a paper presented to the Symposium on Optical Image Evaluation (NBS C526, paper No. 2), or in a book, Diffraction, to be published by the Revue d'Optique.

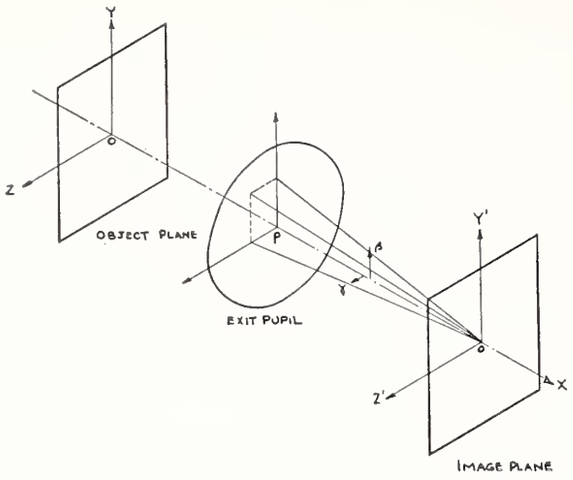


FIGURE 16.1

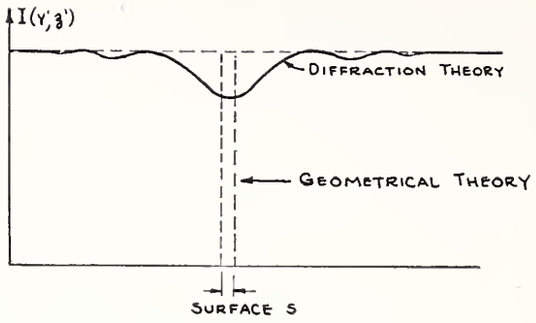


FIGURE 16.2.

mentary to the preceding. Given a uniformly illuminated area, presenting a dark spot of small surface s the image will be obtained by subtracting from a uniform illumination the "diffraction solid" multiplied by a factor proportional to s. In that case, the contrast will also be related to the maximum illumination in the center of the solid (fig. 16.2).

Let us now consider the case of a bright line parallel to the Oz' axis; $D(y'z')$ is zero everywhere, except for $0<y'<\epsilon$. The resulting illumination in the image will be:

$$I(y',z')=\epsilon\int_{-\infty}^{+\infty}D(y',z'-z)dz=\epsilon\int_{-\infty}^{+\infty}D(y',z')dz'=\epsilon S(y'),$$

where $S(y')$ is the surface of section of the diffraction solid by a plane of ordinate y'. It could be shown very easily that in cases 4 and 5 the contrast will also be related to that quantity.

If now we consider the case where the illumination is no longer incoherent but coherent, the same treatment leads to similar conclusions, provided we represent the image of an isolated point by a repartition of complex amplitudes and not of intensities.

The Losses of Contrast Due to Small Aberrations

The problem is now to express the effects of small geometrical aberrations, producing various distortions of the wavefront. The case of points is the simplest, because it is easy to express the maximum illumination in the image.[3] The cases of lines, edges of areas, etc., would need a supplementary integration on the diffraction solid: but, thanks to Dr. M. Duffieux's work,[4] it is easy to transform those integrals into integrals performed on the exit pupil. It is then possible to express easily the effects of small aberrations; if those aberrations are known by the distortion $\Delta(\beta,\gamma)$ of the wavefront, the repartition of amplitude on the exit pupil will be, for a single object point

$$\exp j\frac{2\pi\Delta(\beta,\gamma)}{\lambda}=1+j\frac{2\pi\Delta}{\lambda}-\frac{2\pi^2}{\lambda^2}\Delta^2.$$

[3] A. Marechal, Rev. Opt. **26**, 257 (1947).
[4] P. M. Duffieux L'integrale de Fourier et ses applications a l'Optique (Besancon, 1945), and Rev. Opt. (1947 and 1948)

The contrast for various cases can then be expressed by the following expressions, provided the aberrations are small:

Points:

(a) Bright point, or dark point incoherent background,

$$1-\frac{4\pi^2}{\lambda^2}\left[\iint \Delta^2 dw-\left(\iint \Delta dw\right)^2\right];$$

(b) Dark point in coherent background,

$$1-\frac{2\pi^2}{\lambda^2}\iint \Delta^2 dw.$$

Lines:

(a) Bright or dark line in incoherent illumination,

$$\int (\beta_2-\beta_1)^2\left[1-\frac{4\pi^2}{\lambda^2}\left(\int_{\beta_1}^{\beta_2}\frac{\Delta^2 d\beta}{\beta_2-\beta_1}-\left(\int_{\beta_1}^{\beta_2}\frac{\Delta d\beta}{\beta_2-\beta_1}\right)_2\right]\right]d\gamma;$$

(b) Coherent illumination,
Bright line,

$$(\beta_2-\beta_1)^2\left[1-\frac{4\pi^2}{\lambda^2}\left(\int\frac{\Delta^2 d\beta}{\beta_2-\beta_1}-\left(\int\frac{\Delta d\beta}{\beta_2-\beta_1}\right)^2\right)\right];$$

Dark line,

$$(\beta_2-\beta_1)\left[1-\frac{2\pi^2}{\lambda^2}\int\frac{\Delta^2 d\beta}{\beta_2-\beta_1}\right].$$

We notice that, within the approximations used, the losses of contrast vary as the squares of the aberrations. The losses of contrast will be inversely proportional to the squares of the tolerances which we shall now determine.

The Case of Third Order Aberrations

The application of the preceding formula to the case of third order aberrations leads to various expressions for loss of contrast. If we apply them to the Rayleigh limit for various aberrations (distortion of the wave front equals to $\lambda/4$), we find losses of contrast varying from 0.1 to 0.3. In fact those expressions allow us to define a more precise tolerance than the Rayleigh limit, by writing a priori that the loss of contrast is *exactly 0.2*. The various tolerances obtained are tabulated on table 16.1.

In the case of spherical aberration we have quoted the maximum distortion of the wavefront (refered to a sphere centered in the paraxial image) for the two cases where we either use the paraxial image, or the best image obtained by a small axial displacement from the paraxial focus. The second tolerance is obviously the value to be used practically. The tolerance obtained is of the order of $1\cdot\lambda$ for the different objects or illuminations. Consequently if α' is the half angular aperture of the bundle of rays, the tolerable longitudinal spherical aberration will be $4\lambda/(\alpha')^2$.

TABLE 16.1. *Values of third-order aberrations leading to a loss of contrast of 0.2*

Aberrations	Incoherent illumination			Coherent illumination				
	Point	Line		Dark point	Dark line		Bright line	
Defect of focusing	0.25λ	0.29λ		0.17λ	0.22λ		0.24λ	
Astigmatism $\left(\phi=\frac{\pi}{4}\right)$	.17λ	.18λ		.25λ	∞		∞	
3rd order spherical:								
Paraxial focus	.25λ	.28λ		.22λ	.30λ		.27λ	
Best focus	.95λ	1.04λ		.9λ	1.0λ		.92λ	
3rd order coma:		$\phi=0$	$\phi=\pi/2$		$\phi=0$	$\phi=\pi/2$	$\phi=0$	$\phi=\pi/2$
Paraxial image	.20λ	0.29λ	1.3λ	.28λ	0.26λ	∞	0.19λ	∞
Best image	.60λ	.85λ	1.3λ	.85λ	.66λ	∞	.47λ	∞

In the case of coma we have mentioned by ϕ the azimuth of the comatic flare with respect to the object line, in the case of spherical aberration we have determined the tolerance for the image obtained in the paraxial focus, or for the best image which is laterally shifted from that focus. Let us consider for example a coherent bright line perpendicular to the direction of the comatic flare $\left(\phi=\frac{\pi}{2}\right)$. The distortion of the wave front can be 0.47λ on both sides of the paraxial sphere, which means that the maximum geometrical comatic flare (which will not be perciptible by the coexistence of diffraction effects) could be as a maximum $(3 \times 0.47\lambda)/\alpha'$.

In the case of astigmatism, ϕ is the angle between the linear object and one of the focal lines when one of the focal lines is parallel to the object ($\phi=0$ or $\pi/2$). The effect of astigmatism is zero when we focus on that line and any displacement from that position has to be considered as a defect of focusing. In the case of $\phi=\pi/4$ the best focusing is located midway between the lines and the tolerances are expressed for that case. We notice that the tolerances on astigmatism are much more severe than the tolerances on coma or spherical aberration: The slightest astigmatisms produce a noticeable loss of contrast (except when the tolerances are infinite, which means that the aberration has no effect in the specified conditions). We will see in the case of larger aberrations that the astigmatism has also the worse effect.

The mechanical computation of diffraction patterns

When the aberrations are either very small or very large it is possible to compute the distribution of energy in diffraction patterns by using mathematical expansion.[5,6] We have already mentioned that the losses of contrast can be expressed as functions of various pupillar integrals that can also be easily computed in the case of small aberrations. When the aberrations are neither very small nor very large (in the transition between diffraction and geometrical optics) these mathematical procedures may fail. It is then useful to perform the integration expressed in the formula for the case of a point source by means of a special mechanical device (fig. 16.3).

[5] B. R. A. Nijboer, (1942).
[6] Van Kampen, Physica **XIV,** 580 (1949).

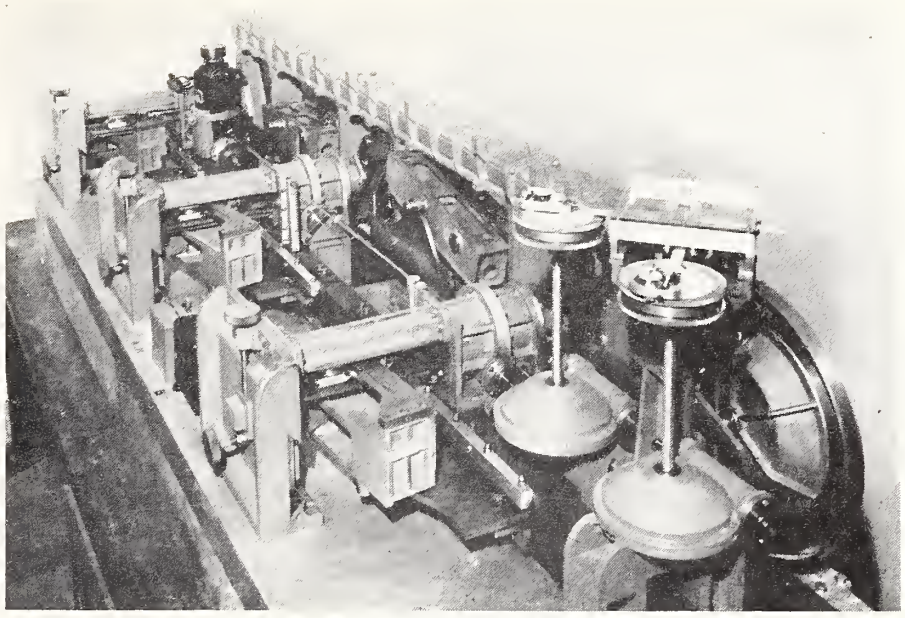


FIGURE 16.3.

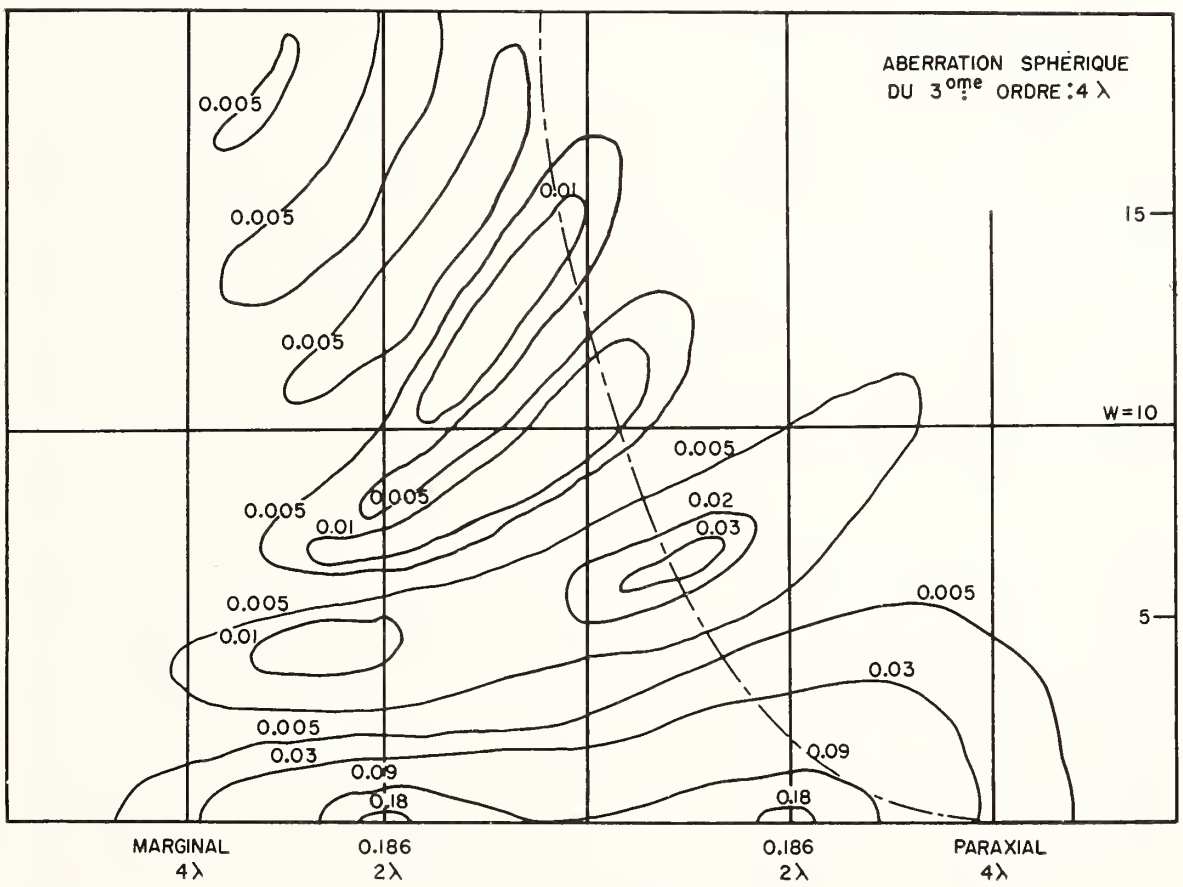


FIGURE 16.4.

The principle of such a machine has already been described elsewhere.[6] The real and imaginary parts of the complete integral are developed by two integrating wheels whereas the distortions of the wavefront (of any order) are given by cams, amplifier levers, adding tape, etc. The machine has been used for solving the following problems:

(a) The distribution of energy in the presence of third order aberrations. The goal of that study was mainly the knowledge of the

[6] A. Marechal, Rev. opt. **27**, 73 (1948); and J. Opt. Soc. Amer. **37**, 982 (1947).

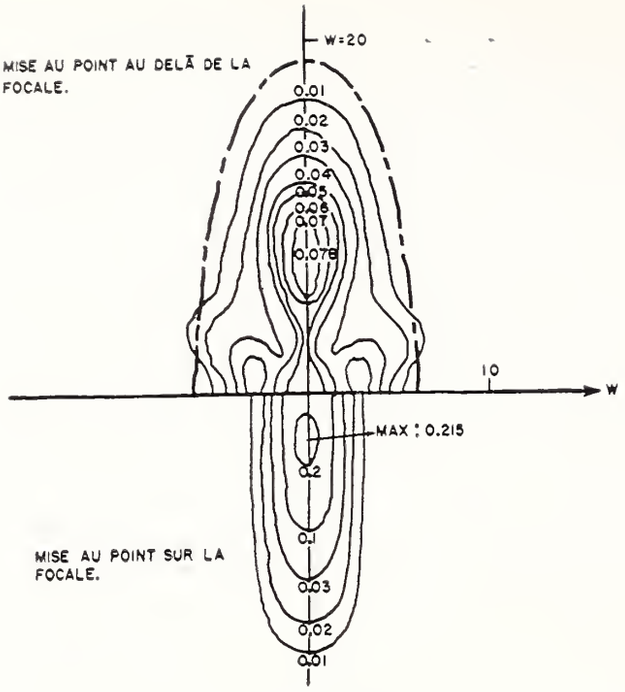


FIGURE 16.5.

transition between diffraction patterns and geometrical caustics. The transition is very rapid in the case of astigmatism when the focusing is done on a focal line. The diffraction pattern is already similar to the geometrical pattern even when the aberration is as small as $\lambda/4$. The transition is more gradual for coma, and still more gradual for spherical aberration.

(b) Determination of the proper aperture of correction of geo metrical aberration when the aberrations are of the order of 2 or times the Rayleigh limit.

(c) Special studies of practical cases of microscope objectives, Schmidt cameras, etc.: Fig. 16.4 represents the example of the study of spherical aberration of 4λ. Curves of equal illumination are drawn in a meridional plane and are able to give the illumination in various planes of focussing. Fig. 16.5 represents the example of astigmatism of $\pm 0.5\lambda$ for two different foci. The lower part of the figure represents the repartition of light in the plane of a geometrical focal line, the upper part in a plane located midway between a focal line and the "circle of least confusion". We notice that even with that weak astigmatism of $\pm 0.5\lambda$ the loss of illumination is very severe: The astigmatism has a very bad effect on the contrast of point sources.

It is possible with that machine to study the effect of any aberration, isolated or superimposed and to obtain an evaluation of the quality of the image.

Discussion

DR. L. MARTON, National Bureau of Standards, Washington, D. C.: How far do you think these can be directly translated into electron optics?

DR. MARECHAL: What we have studied is the two extreme cases of perfect coherence illumination and perfectly incoherent illumination. Dr. Glaser said that what is not taken into account is the way that one point diffracts. If we consider the case of perfectly incoherent

illumination, I would like to ask Dr. Glaser whether we can't consider that the illumination is practically uniform in the aperture open to the electrons, and, in the reverse case, in the case of coherent illumination, those computations take into account the weight of diffraction of the structure of the object by taking a division of amplitudes and taking account of the diffraction by the division of the amplitudes in the object plane.

Dr. D. Gabor, Imperial College, London, England: Do you mean the diffraction by the distribution in the object plane?

Dr. Marechal: Yes, we suppose that the amplitude in the object planes are represented, and we take account of the diffraction of that aberration even to the entrance pupil. So, for example, if we have a certain structure we would give the entrance pupil two points, two vibrations. That is what we have taken into account, incoherent illumination.

Now, in the case of incoherent illumination we have only supposed that the division of amplitude on the pupil is uniform for any of the points.

Dr. W. Glaser, Technische Hochschule, Wien, Austria: We have often discussed with Dr. Marechal his lecture at Paris in September, and it is possible to use, I think, very useful calculations in electron optics the influence of aberrations, because we get general formulas that can be translated in this form, but there is one difference. The wavelengths are much greater.

Since the wavelengths of light are great and it is possible to get some accurate forms, but some wavelengths are so much smaller than the light wavelength that it is not comparable with mechanical devices, and therefore I think it would be necessary to generalize on this. It would be a great deal more work in order to apply it to electron optics.

Dr. Marechal: I would like to say only that, of course, the wavelength is very much less than for ordinary optics, but in any case the structure of the images depends only on the number of the wavelengths of aberrations that you have got.

If you get, for example, ten wavelengths aberrations, you are practically in the geometrical case so it probably would not be any use to try to compute those diffraction patterns, and, moreover, in the question of diffraction of a given single-point source in the pupil, this might be complicated if that diffraction pattern should be composed of many oscillations, but that would also be a case probably less interesting than would be the case of a large disk. The object would then be great with respect to the resultant pattern. When the object is of the order of the resulting pattern or smaller, we probably get only one oscillation of the maximum.

17. Lentilles Electroniques

Par P. Grivet [1]

I. Introduction

Les propriétés essentielles des lentilles électroniques sont solidement établies et clairement comprises; leur connaissance a permis de délimiter le domaine où chaque type de lentille doit être utilisé. Nous les supposerons connues du lecteur qui, au besoin, pourra se reporter à une revue récente de la question [1].[2] Aujourd'hui il faut aller plus loin et nombre d'applications délicates requièrent une connaissance détaillée et approfondie des propriétés optiques de chaque modèle, pour l'adapter au mieux à sa fonction. En général, les qualités optiques d'une lentille sont définies par les éléments cardinaux de Gauss: distances focales, position des foyers, qui suffisent pour déterminer l'image d'un objet quelconque en grandeur et en position. Les coefficients d'aberrations ne sont que plus rarement utiles: lorsqu'il devient nécessaire d'estimer la finesse ou la fidèlité de l'image. Le problème de l'heure est donc de déterminer les éléments cardinaux (cf. appendice I) connaissant la structure géométrique de la lentille (dimensions et positions des électrodes) et les données électromagnétiques: tension d'accélération des électrons, tensions appliquées aux électrodes ou courants magnétisants (aimantation d'excitation s'il s'agit d'aimants permanents).

Des méthodes sûres existent depuis longtemps pour résoudre ces problèmes, mais elles exigent un travail considérable de calcul numérique, qui excède en général les possibilités des laboratoires ordinaires, même lorsqu'ils sont bien équipés, et pourvus de machines à calculer: un bureau de calcul spécialisé serait nécessaire et il devrait consacrer à chaque problème quelques semaines voire quelques mois. Une solution de ce genre est de peu de secours surtout dans le stade initial du choix d'une lentille. C'est pour subvenir à ce défaut que la théorie se développe aujourd'hui dans une nouvelle direction. Le but est de représenter par des formules simples les propriétés de toute une famille de lentilles en partant des données premières. Ce sont quelques solutions apportées par ces recherches récentes dans ce sens, que je vais passer ici en revue; et pour mieux mettre en évidence la progression des idées, je commencerai par les questions les plus rebelles, dont la solution relève encore largement des anciennes méthodes.

L'élaboration de ce travail, m'a amené à quitter la division traditionelle entre électrostatique et magnétique. C'est pourquoi j'ai changé le titre initial de l'article: Electrostatic lenses, qui est devenu electron lenses.

---

[1] Laboratoire de Radioélectricité, Université de Paris, France.
[2] Figures in brackets indicate the literature references on p. 196.

II. Machines Simples

Description du Champ

Le calcul d'une lentille comporte deux étapes:

Partant de la structure de la lentille, on calcule le champ agissant sur les électrons,

Connaissant le champ, on calcule les trajectoires puis les éléments cardinaux.

Nous avons rappelé qu'il n'est pas nécessaire de donner une expression détaillée des trajectoires et qu'on se contente de les caractériser optiquement par les éléments cardinaux. De même il n'est pas utile de connaître la répartition du champ dans tout le volume de la lentille, mais seulement au voisinage de l'axe. On montre (cf. référence [1], p. 56, 64 et 70) que les éléments cardinaux résultent de la résolution de l'une ou l'autre des équations différentielles suivantes:

Lentilles électrostatiques:

Équation des rayons

$$\frac{d^2r}{dz^2}+\frac{\varphi'}{2\varphi}\frac{dr}{dz}+\frac{\varphi''}{4\varphi}r=0 \tag{1}$$

Équation des rayons réduits

$$r=R\varphi^{-1/4} \qquad \frac{d^2R}{dz^2}+\frac{3}{16}\left(\frac{\varphi'}{\varphi}\right)^2R=0 \tag{2}$$

Lentilles magnétiques:

$$\frac{d^2r}{dt^2}+\left(\frac{e}{8m\varphi_0}\right)B^2(z)r=0 \tag{3}$$

Il suffit donc de connaître avec précision, les fonctions $\varphi\ (z)$ et $\varphi'\ (z)$ ($\varphi(z)$, étant le potentiel sur l'axe) dans le cas électrostatique, $B(z)$ champ sur l'axe dans le cas magnétique.

Mesure du Champ sur l'Axe

Dans les cas où la détermination est difficile et échappe encore au calcul—nous en donnerons deux exemples—il est indispensable de mesurer le champ et il serait alors très utile de réduire les mesures à leur plus grande simplicité, c'est-à-dire à la mesure directe de $\varphi(z)$, $\varphi'(z)$ ou $H(z)$ et à cette mesure seulement.

Il est difficile de mesurer directement $\varphi(z)$ et $\varphi'(z)$ dans le cas électrostatique. On peut il est vrai y parvenir à 1: L'aide d'une cuve électrolytique, mais on est obligé de recourir à une cuve profonde, où l'on plonge une reproduction de la moitié de la lentille, coupée par un plan méridien; et alors les électrodes sont encombrantes et difficiles à réaliser; l'utilization d'une cuve plus simple en forme de coin est possible [2] mais la mesure est alors indirecte et délicate. Au contraire, dans le cas du champ magnétique $B(z)$ et $B'(z)$ sont faciles à mesurer avec l'appareil de Le Poole [3] sous la forme pratique que lui a donnée FERT [4] et qui livre directement l'enregistrement des courbes $B(z)$ et $B'(z)$.

C'est pourquoi, au laboratoire de l'auteur, Septier essaie actuellement d'employer l'analogie champ magnétique—champ électrique en sens inverse des habitudes acquises (figure 1): au lieu d'étudier le champ magnétique d'une lentille magnétique, à la cuve électrolytique [5, 6] il essaie de mesurer magnétiquement—par une méthode dérivée de celle de Le Poole—le champ électrostatique dans une lentille électrostatique. Pour cela les électrodes sont représentées 10 à 20 fois agrandies par des feuilles de métal très perméable (fer Armco, ou mieux hypernick) et sont exciteés magnétiquement par un courant (figure 17.1) qui correspond à une tension électrostatique V (en unités Giorgi rationalisées) par les formules:

$$E = mB \ (m, \text{ coefficient arbitraire}),$$

$$-V = m\mu_0 nI \ ((nI) \text{ ampères-tours magnétisants.})$$

Ici on peut se placer dans une région où le fer est très perméable et l'aimantation faible; l'échauffement du circuit magnétique par hystérésis est faible dans ce domaine si l'on pratique une mince coupure méridienne dans le circuit magnétique ou si l'on emploie un matériau magnétique isolant. On peut simplifier les mesures en employant un courant d'excitation alternatif; $\varphi(z)$ et $\varphi'(z)$ sont alors mesurés par simple effet d'induction dans une bobine d'exploration, longue pour $\varphi(z)$, courte pour $\varphi'(z)$.

Les Cas Difficiles

Deux exemples typiques nous montreront la nature des difficultés auxquelles se heurte la théorie:

1. Lentille Magnétique à Blindage Saturé

Dans ce cas la coure $B(z)$ a la forme simple d'une cloche et Glaser a montré dès 1941, qu'on pouvait la représenter par l'équation $B = B_0/[1+(z^2/a^2)]$ (cf. [1] p. 73) et qu'il était facile d'en tirer les éléments cardinaux et les aberrations. Mais la solution s'arrête là car, bien que cette courbe soit simple, il est difficile de la relier aux

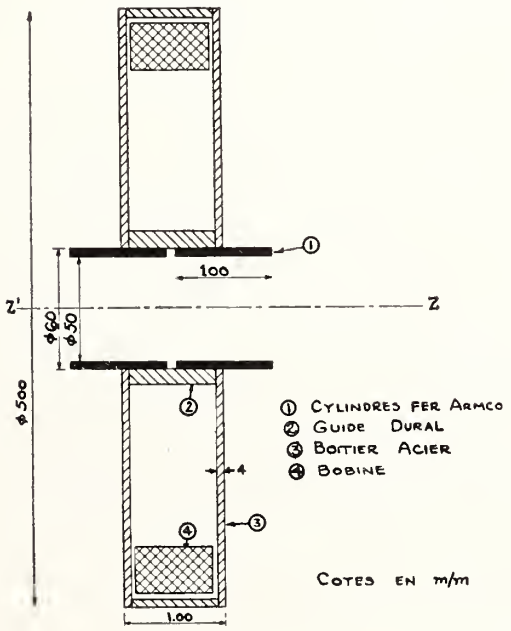


Figure 17.1

dimensions du circuit magnétique aux caractéristiques du fer et au courant magnétisant. On se heurte à une difficulté classique, mais encore insurmontée: Comment décrire simplement la saturation du fer?

2. Lentille Electrostatique à Immersion

C'est là un organe fondamental de l'optique électronique: le canon à électron. Les difficultés très réelles dues à la charge d'espace ne semblent pas près d'être résolues par la théorie et conventionnellement on les néglige en envisageant la lentille sans charge d'espace. C'est un cas qui se trouve rigoureusement réalisé dans le microscope électronique à émission et dans le transformateur d'image, mais qui n'est qu'une approximation—mal définie—dans l'oscillographe. Malgré cette simplification draconienne on ne connaît cependant pas encore de solution complète, car on se heurte à deux difficultés.

1. La courbe $\varphi(z)$ est compliquée comme on en peut juger d'aprés le relevé de Duchesne [7] (figures 17.2 et 17.3), et par conséquent, l'équation des rayons est difficile à traiter directement. Mais simultanément la courbe φ'/φ est compliquée aussi et l'équation réduite

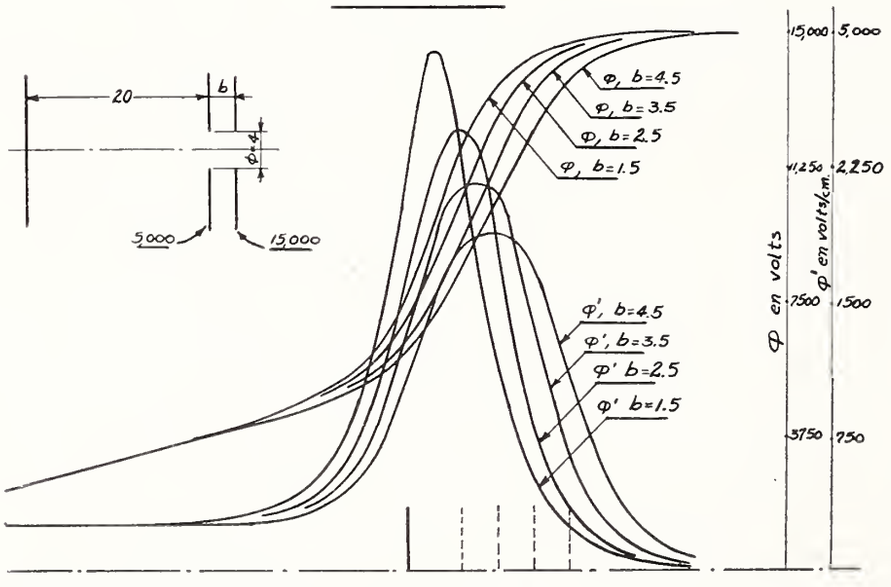


FIGURE 17.2. *Potentiel et gradient de potentiel pour differentes valeurs de l'écartement b.*

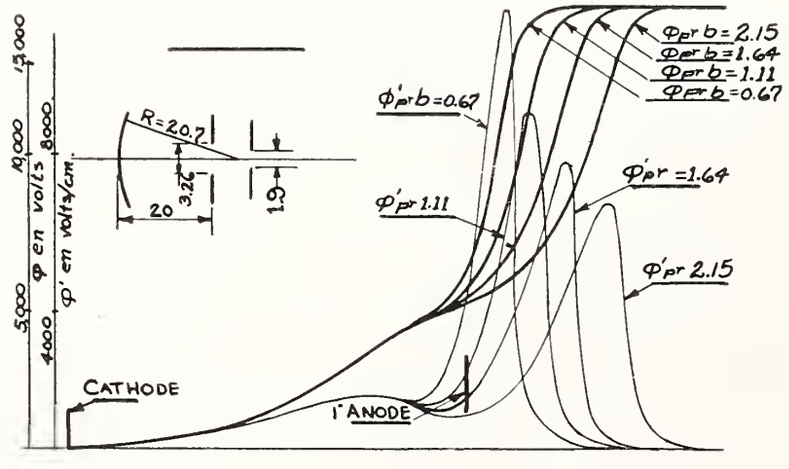


FIGURE 17.3. *Potentiel et gradient de potentiel pour differentes valeurs d'écartement b.*

n'apporte pas d'avantages parce qu'au voisinage de la cathode $\varphi(z)$ est petit et le quotient φ'/φ est important bien que φ paraisse négligeable.

2. Dans la région de la cathode la marche des rayons est très sensible à une petite variation de la force focalisante, c'est-à-dire à de petites variations de tension appliquée à la première électrode, et à de légers changements de forme de cette électrode ou de la cathode; ces petites perturbations ont de grosses répercussions, parce que la vitesse des particules est faible. Ainsi Duchesne [7] a montré que la courbure de la cathode employée couramment dans les télescopes électroniques, pour réduire la courbure de champ, avait une influence notable sur la mise au point de l'image et son grossissement; c'est ce qui ressort bien de la comparaison des figures (2) relative à une cathode plane et (3) pour laquelle la cathode était sphérique.

C'est pourquoi aucun calcul complet de l'objectif à immersion n'a été encore publié. On ne connaît que des solutions partielles, certaines fort utiles comme celle proposée par Jacob [8]; cet auteur a observé à la cuve qu'avec les proportions que l'on donne habituellement aux canons d'oscillographe, on avait de manière très approchée $\varphi(z)=sh(Kz)$ entre la cathode et le cross-over et il a remarqué que l'action de cette seule région déterminait des propriétés importantes des oscillographes: finesse et luminosité du spot.

Les difficultés que nous venons d'analyser sont purement théoriques; bien que le fonctionnement des lentilles à immersion soit difficile à analyser, ces organes fournissent d'excellentes images, comme on pourra en juger, par les images des figures 17.4 et 17.5 obtenues par Septier & Gauzit avec leur microscope à émission; les images d'une source *ionique* obtenues avec le même appareil, ne sont pas encore aussi bonnes, mais comme le montre la communication de Couchet, Septier & Gauzit au présent Congrés il s'agit surtout là de difficultés auxiliaires d'observation.

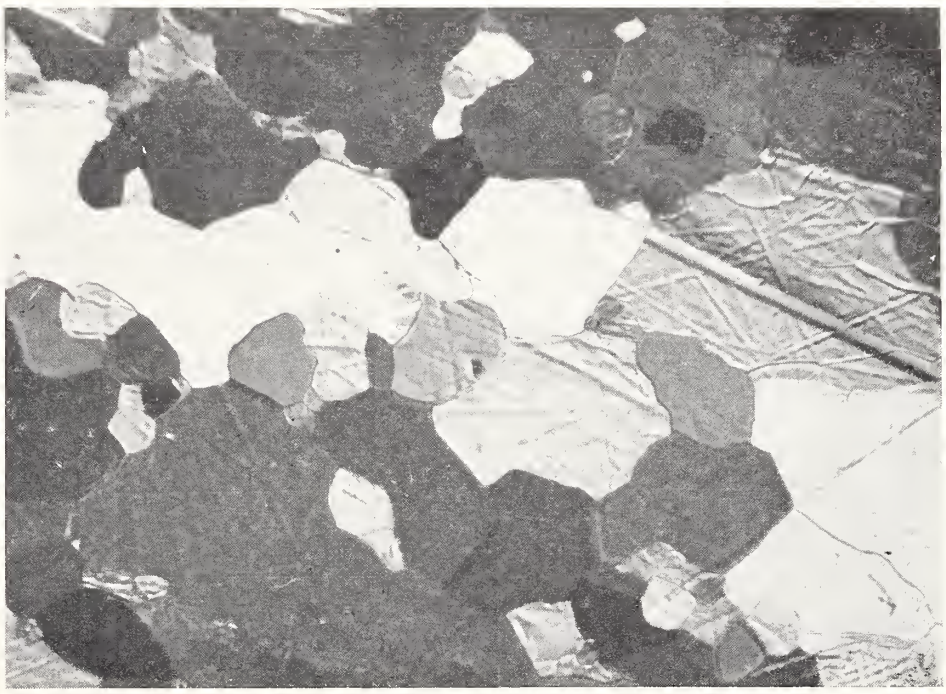


FIGURE 17.4. *Molybdéne thorié.*

Une machine à intégrer l'équation des trajectoires reste donc utile dans de nombreux cas. Il est certain que les grandes machines à calculer classiques satisfont à cette exigence, mais elles sont peu accessibles et il serait intéressant de disposer d'une machine simple et rapide. L'auteur en collaboration avec Y. Rocard [9] a esquisse un projet que Septier étudie actuellement. C'est une machine analogique basée sur la similitude entre l'équation qui donne la tension le long d'une ligne électrique alimentée à une extrémité par une source sinusoïdale de pulsation ω (figure 17.6) et les équations

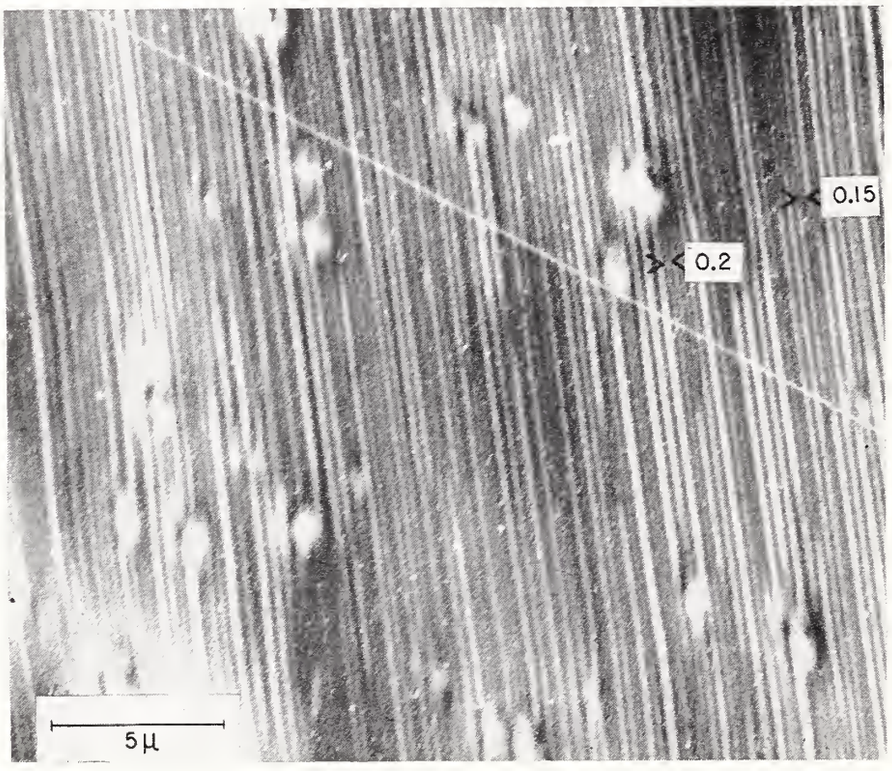


FIGURE 17.5. *Flux du baryum le long des lignes de glissement sur la surface de fer.*

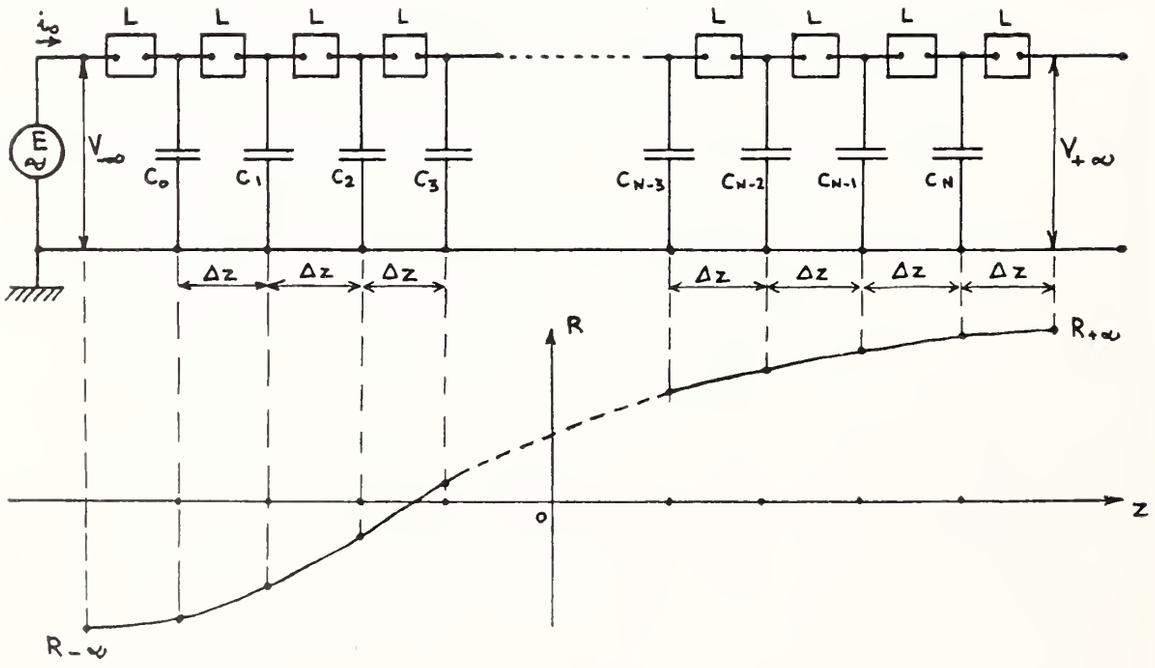


FIGURE 17.6.

(1), (2), ou (3). L'équation du potentiel V le long d'une ligne formée par exemple d'une self longitudinale répartie avec la densité $L(x)$ par unité de longueur et d'une capacité transversale répartie avec la densité $C(z)$ par unité de longueur, s'écrit:

$$\frac{d^2V}{dz^2}-\frac{L'(z)}{L(z)}\frac{dV}{dz}+LC\omega^2V=0. \tag{4}$$

Essayons par exemple de copier électriquement l'équation (2); nous voyons qu'il suffit de constituer la ligne avec une self par unité de longueur $L(z)$ et une capacité par unité de longueur $C(z)$ tel que:

$$L(z)=L, \qquad \omega^2LC(z)=\frac{3}{16}\left(\frac{\varphi'(z)}{\varphi(z)}\right)^2=T^2(z). \tag{5}$$

Naturellement on renonce en pratique à une variation continue de $L(z)$ et $C(z)$ et on emploie une ligne artificielle, discontinue contenant de nombreuses cellules L_K, C_K. Une solution simple consiste à choisir $L_K=L$ constant, et les éléments C_K variables, la fonction $T(z)$ est alors représentée par une suite de capacités C_K facile à mesurer donc à ajuster aux valeurs voulues et on obtient $R(z)$ en mesurant la répartition de $V(z)$ le long de la ligne artificielle; l'espace objet correspond à la sortie de la ligne et un rayon principal parallèle à l'axe y est représenté par une extrémité ouverte; l'espace image correspond au côté source de la ligne et dans cette région la pente du rayon émergent est proportionnelle à $\omega L_K i_0$ (i_0 courant d'alimentation).

La même machine peut servir au calcul des coefficients d'aberations par la méthode des perturbations. La fonction de perturbation correspondant à l'aberration choisie soit $\epsilon(z)$ est représentée par une tension $u(z)$ injectée entre le pied de chaque capacité et la terre comme on peut le déduire facilement de la référence [9].

III. Lentille Indèpendante à Trois Diaphragmes

Champ sur l'Axe

La lentille indépendante à trois électrodes, a été calculée complètement par Regenstreif [10]; cet auteur a réussi à déterminer $\varphi(z)$ à partir des dimensions de la lentille et de la tension appliquée entre l'électrode centrale et les électrodes extérieures par une méthode semi-empirique. La lentille similaire formée par 3 morceaux de cylindre de révolution de même diamètre, coaxiaux peut être calculée rigoureusement et la fonction $\varphi(z)$ y apparaît comme une combinaison linéaire de fonctions relatives au cas simple de deux cylindres. Regenstreif a transposé cette idée, au cas des diaphragmes plans et a essayé si l'on ne pourrait pas simplement représenter la fonction de la lentille totale, comme une combinaison linéaire des trois fonctions $\psi(z)$ qui chacune représenterai rigoureusement un diaphragme unique, en l'absence de ses deux voisins: $\Psi(z)$ est bien connu et s'écrit

$$\Psi(z)=a+bz+cz\ \text{arctg}\ \frac{z}{R}, \tag{6}$$

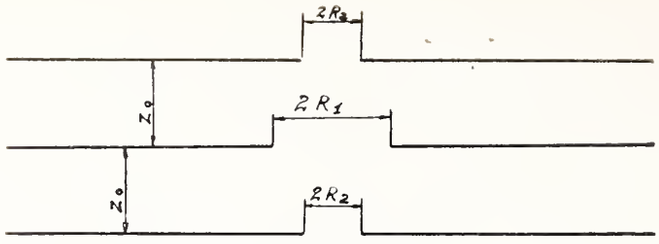


FIGURE 17.7.

R étant le rayon du diaphragme considéré. On essaie donc

$$\varphi(z)=A\psi(z-z_0)+B\psi(z)+C\psi(z+z_0). \tag{7}$$

La fonction φ ainsi batie satisfait rigoureusement à l'équation de Laplace, mais ne s'adapte qu'approximativement aux conditions aux limites. Regenstreif détermine les constantes A, B, C, a, b, c pour que le champ soit bien représenté au centre de la lentille région d'action déterminante car c'est là que les électrons sont les plus lents donc les plus sensibles aux forces déviatrices. Regenstreif obtient ainsi

$$\left.\begin{aligned}
\varphi(z)&=a+b\left[(z+z_0)\ \text{arctg}\ \frac{z+z_0}{R_2}+(z-z_0)\ \text{arctg}\ \frac{z-z_0}{R_z}\right.\\
&\qquad\left.-2z\ \text{arctg}\ \frac{z}{R_1}\right]\\
a&=V_1-2b\left(R_1+z_0\ \text{arctg}\ \frac{z_0}{R_2}\right)\\
b&=-\frac{V_2-V_1}{2R_1+2z_0\ \text{arctg}\ \frac{z_0}{R_1}}
\end{aligned}\right\}\tag{8}$$

R_1 étant le rayon du diaphragme intérieur porté au potentiel V_1, R_2 celui des deux diaphragmes extérieurs portés au potentiel V_2, z_0 distance mutuelle des diaphragmes (figure 17.7).

Il est ainsi parvenu à des formules maniables pour la fonction $\varphi(z)$; leur comparaison aux mesures faites à la cuve électrolytique dans ces cas typiques montra qu'elles représentaient fidèlement le potentiel. Il est à remarquer que la précision requise est beaucoup moins grande que pour l'objectif à immersion: Une erreur de quelques pour cent est de peu d'importance car les électrons ne tombent pas à une vitesse inférieure à 20% de leur vitesse maxima, tandis que près d'une cathode ils auraient une vitesse équivalente qui peut descendre à moins de 1/100 de celle qu'ils auront au passage de l'anode.

Éléments Cardinaux

La formule du potentiel (8) est trop compliquée pour qu'on puisse l'introduire dans l'équation (1) ou (2) avec quelques chances de succès. Regenstreif s'inspirant d'un travail de Rudenberg [11] a assimilé la courbe ϕ (z) à l'ensemble de trois arcs de parabole tangents. On peut alors intégrer l'équation (1) par des fonctions sinus hyperbolique ou cosinus hyperbolique, morceaux de courbes que l'on raccorde ensemble tangentiellement à la frontière des régions paraboliques.

Les calculs sont faciles mais longs; ils ont l'avantage de livrer des formules maniables car on constate que les données physiques n'interviennent dans les résultats finaux que par l'intermédiaire d'un paramètre x, unique, donc très commode

$$x=\frac{\varphi(0)}{\varphi(z_0)}=\frac{V_1+\dfrac{V_2-V_1}{1+\dfrac{z_0}{R_1}\operatorname{arctg}\dfrac{z_0}{R_1}}}{V_1+(V_2-V_1)\left[1-\dfrac{(R_2/2R_1)}{1+\dfrac{z_0}{R_1}\operatorname{arctg}\dfrac{z_0}{R_1}}\right]}, \tag{9}$$

x est simplement le rapport du potentiel au centre ϕ (0) à celui qu'on observe sur l'axe à l'aplomb des électrodes extérieures ϕ (z_0).

Parmi les nombreux rèsultats de Regenstreif nous choisirons les exemples suivants:

1. La distance focale des lentilles de convergence faible ou moyenne (figure 17.8)

$$\frac{f}{z_0}=\frac{8}{3}\frac{x}{(1-x)^2} \tag{10}$$

2. La distance focale des lentilles fortes f (figure 17.9) et la position de leur foyer Z_F

$$\frac{f}{z_0}=\frac{0{,}72}{\sin\,(0{,}707\,\log_e x+0{,}355)}, \tag{11}$$

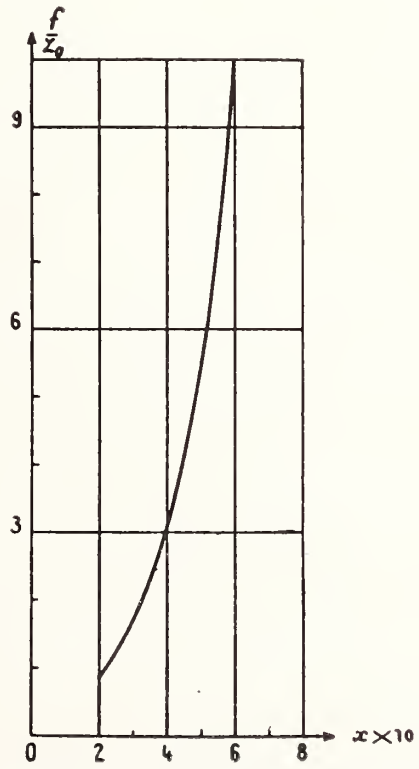


FIGURE 17.8. *Distance focale d'une lentille à une électrode centrale mince (lentilles faibles).*

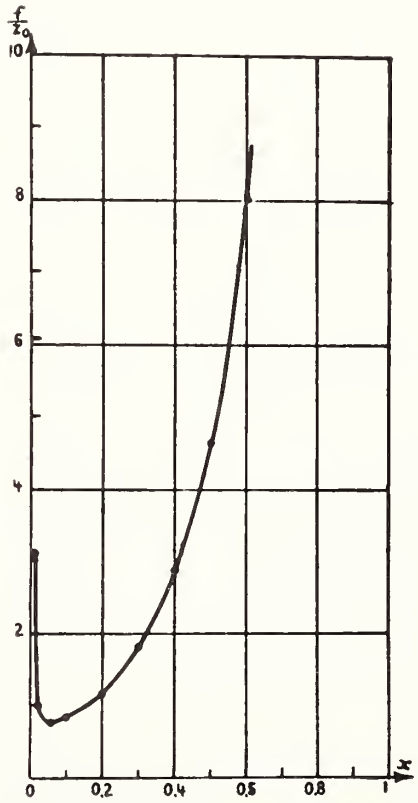


FIGURE 17.9. *Distance focale d'une lentille à une électrode centrale mince (lentille fortes).*

$$\frac{Z_F}{z_0}=1+0{,}764\,\frac{\sin\,(0{,}707\,\log_e x-0{,}887)}{\sin\,(0{,}707\,\log_e x+0{,}355)}. \tag{12}$$

3. Le potentiel critique V_c de l'électrode centrale pour lequel la lentille devient miroir

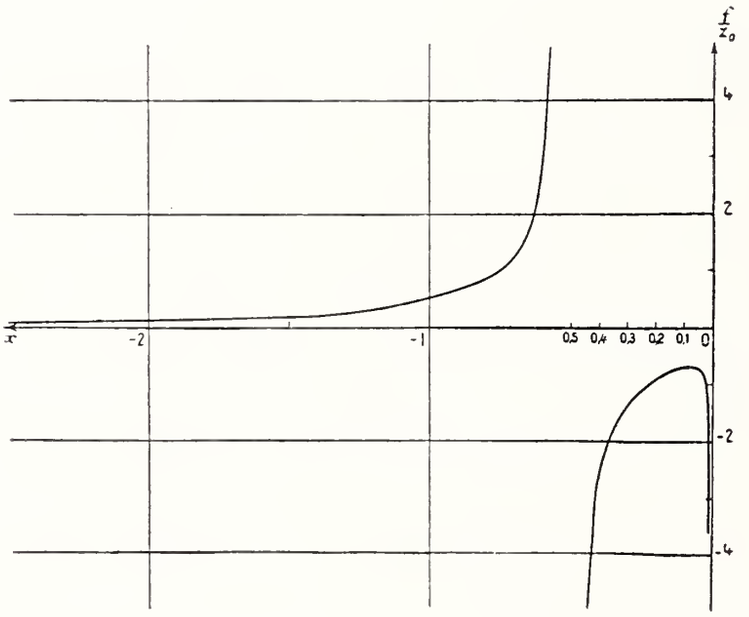
$$\frac{V_c}{V_z}=\frac{1}{\frac{z_0}{R_1}\,\mathrm{arctg}\left(\frac{z_0}{R_1}\right)}. \tag{13}$$

4. La distance focale des miroirs convergents ou divergents (figure 17.10)

$$\frac{f}{z_0}=\frac{0{,}72}{\sin\,[0{,}707\,\log_e\,(-x)+0{,}355]}, \tag{14}$$

5. Les éléments cardinaux des lentilles, où l'électrode centrale est positive et accélératrice (figure 17.11).

FIGURE 17.10.



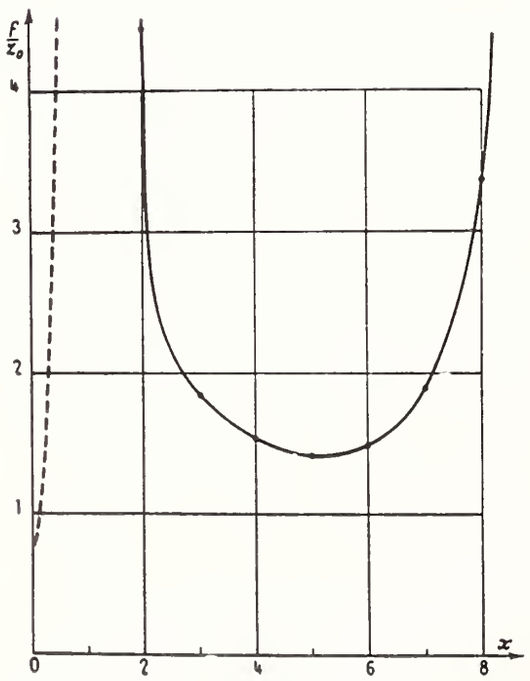


FIGURE 17.11. *Distance focale d'une lentille à une électrode centrale mince (électrode centrale positive ($x>1$)).*

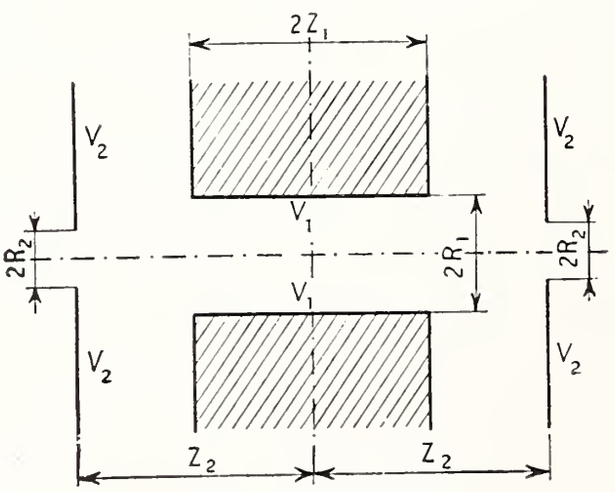


FIGURE 17.12. *Lentille électrostatique épaisse.*

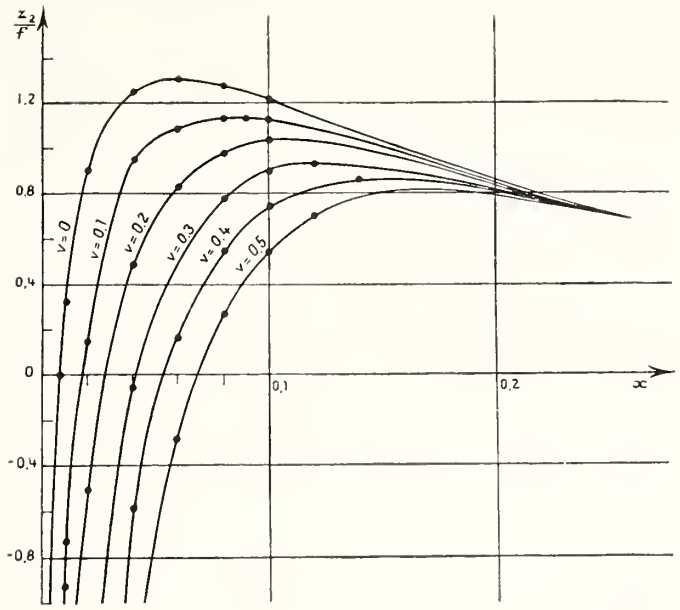


FIGURE 17.13. *Distance focale d'une lentille à une électrode centrale épaisse (lentilles fortes).*

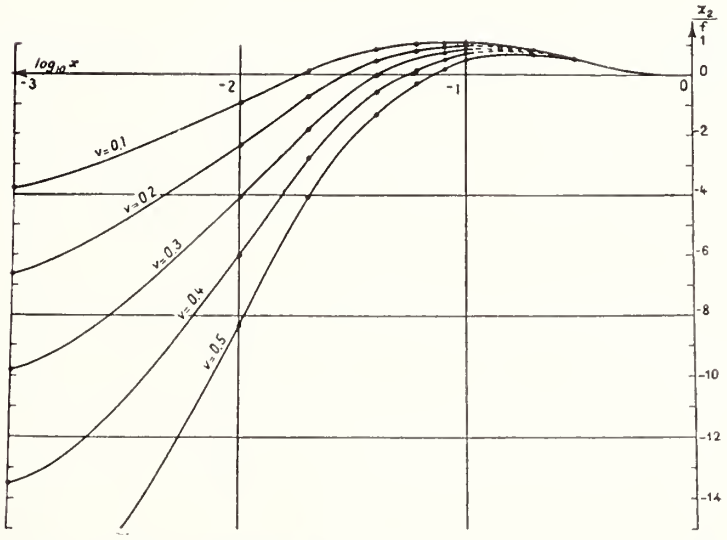


FIGURE 17.14. *Distance focale d'une lentille à une électrode centrale épaisse (lentiule faible).*

On dispose donc maintenant d'un arsenal de formules qui décrivent de manière extrêmement complète le comportement de cette famille de lentille. La confrontation de ces précisions théoriques avec le résultat de mesures faites au laboratoire et avec celles de Heise & Rang [12] montre un accord satisfaisant.

Enfin, le cas où l'électrode centrale est très épaisse, se traite sans nouvelle difficulté, mais mène à des formules plus complexes; dans ce cas la courbe du potentiel présente au centre de la lentille un plateau étendu; on le représente par un quatrième morceau de courbe : une droite horizontale et on obtient encore des résultats en bon accord avec l'expérience. La figure 17.12 donne le schéma de la lentille, les figures 17.13 et 17.14 les valeurs de f et de Z_F; le paramètre γ est défini par $\gamma = z_1/z_2$.

Aberrations—Astigmatisme d'Ellipticite

Les résultats précédents ont une importante conséquence dans le domaine des aberrations car depuis les recherches de Bruck [2] et de Heise [12] on savait que lorsque la distance focale est minima, les

aberrations de distorsion et de sphéricité passent par un minimum. On peut donc par les formules de Regenstreif déterminer facilement ces conditions qui correspondent à la valeur $x_m = 5{,}8.10^{-2}$ ($f_m = 0{,}7730$) lorsque l'électrode centrale est mince. Regenstreif a fait progresser récemment cette théorie d'un pas important en montrant que l'aberration un peu singulière mais qui en pratique est souvent la plus importante, l'astigmatisme d'ellipticité passait aussi par un minimum en même temps que la distance focale. On le comprend intuitivement en remarquant que les distances focales sont les mêmes pour deux lentilles rigoureusement rondes, qui auraient pour rayon du trou central, soit le grand axe, soit le petit axe de la lentille réelle. La proposition est plus difficile à démontrer; Regenstreif y réussit en faisant le calcul approché mais complet d'une lentille dont le trou central est faiblement elliptique [13].

Il est même parvenu récemment à traiter le cas où le trou central est fortement elliptique, ce qui pourra faciliter beaucoup l'usage des lentilles fortement elliptiques dont Scherzer [14] a montré tout l'intérêt pour corriger les abberrations sphériques et chromatiques. Nous renvoyons sur ce sujet aux deux communications de Regenstreif au présent Congrès.

Application au Microscope Électrostatique

Les résultats précédents se sont montrés très précieux pour nous guider dans le perfectionnement de notre microscope électrostatique:

1. Objectif

En particulier les formules mettent en évidence l'existence d'un type remarquable d'objectif:

(a) Son foyer affleure la face d'entrée à quelques dixièmes de millimètres de distance, si bien que sa "distance de travail" est minima.

(b) Simultanément sa distance focale est située au minimum de la courbe donnant f en fonction de x, si bien que:

L'objectif est extrêmement stable, il est tout à fait insensible aux petites variations de tension, telles celles, difficilement évitables, qui proviennent des petits courants de fuite,

Son aberration sphérique, et son astigmatisme d'ellipticité sont minima.

(c) Des objectifs de cette sorte peuvent être construits sur le type unipotentiel où le potentiel de l'électrode centrale est égal à celui de la cathode, l'électrode centrale étant alors assez épaisse.

(d) Mais il est encore préférable d'utiliser leur stabilité vis-à-vis des variations de tension, en utilisant une polarisation négative de l'électrode centrale (quelques milliers de volts obtenus avec une résistance d'autopolarisation) vis-à-vis de la cathode: dans ces conditions, et à distance focale égale, on diminue encore l'astigmatisme d'ellipticité, pour une précision de fabrication constante, parce que le trou central peut être choisi sensiblement plus gros.

2. Lentille de Projection

Les formules permettent de calculer une grande variété de lentilles de projection dénuées de distorsion: L'emploi d'une électrode centrale épaisse s'avère toujours précieuse mais on dispose maintenant d'une bien plus grande latitude dans le choix des dimensions.

Cette étude nous a amenés à réviser sensiblement la conception de notre microscope, où nous avons cherché à distribuer les fonctions importantes à des organes spécialisés: C'est pourquoi nous le munissons de quatre lentilles,

—l'objectif et la lentille de projection (première et quatrième lentilles) ont pour but de grossir,

—la deuxième lentille, immédiatement après l'objectif est faible, et sert à mettre au point, puisque maintenant l'objectif a une convergence rigoureusement fixe,

—la troisième lentille, de moyenne convergence sert à faire varier le grossissement,

L'appareil est actuellement à l'essai et donne des résultats encourageants.

IV. Un Nouveau Modèle Mathématique de Lentille

Lentille a Deux Cylindres

Il est connu depuis longtemps [15] que le potentiel le long de l'axe d'une lentille formée de deux cylindres de même axe et de même diamètre, séparées par une coupure infiniment mince est donné par la formule

$$\varphi(z)=\frac{\varphi_1+\varphi_2}{2}\left[1+\frac{1-\gamma}{1+\gamma}\,\mathrm{th}\;\omega z\right], \tag{15}$$

φ_1 et φ_2 étant les potentiels des deux cylindres de rayon unité, $\gamma=\varphi_1/\varphi_2$; $\omega=1{,}318$; mais on ne sait pas intégrer les équations des rayons correspondant à cette forme de $\varphi(z)$. Hutter [16] a remarqué que la courbe $\varphi(z)$ avait la même allure que

$$\psi(z)=\psi_0 e^{K\,\mathrm{arctg}\,(z/a)}, \tag{16}$$

qui mène à une équation (2) résoluble, celle de Glaser (cf. [1] p. 73) Mais il a essayé de représenter au mieux la fonction φ (z) par ψ (z) et n'a pu y parvenir que grossièrement, ce qui le mena à des résultats optiques peu précis. Au contraire, si comme Bernard et l'auteur [17] l'ont proposé, on renonce à représenter φ (z) avec une exactitude inutile, mais qu'on s'attache à trouver une cloche de Glaser $B(z)=B_0/[1+(z/a)^2]$ qui se rapproche le mieux de la courbe réelle $(\sqrt{3}/4)$ (φ'/φ), on obtient des résultats très précis car il se trouve que la cloche réelle est naturellement symétrique et fort voisine du modèle: Le lecteur se reportera à la communication de Bernard & Grivet sur ce point, et y verra aussi que la lentille à interstice de même forme, ainsi que la lentille formée de deux plans parallèles percés de trous égaux, peuvent être traitées de la même manière.

Utilité d'un Autre Modèle

Cette étude fait apparaître un défaut du modèle de Glaser [17] qui se manifeste également dans son application à la lentille magnétique, lorsque son blindage est loin de la saturation. La représentation est bonne au centre de la lentille, mais mauvaise sur ses deux bords: En

effet le champ de fuite réel décroît très vite dans ces deux cas, suivant une loi exponentielle comme le montre bien la formule (15) dans l'exemple électrostatique et comme on le constate aussi, sur les courbes expérimentales de lentilles magnétiques. Cette loi exponentielle paraît être caractéristique, de la décroissance du champ chaque fois qu'elle est produite par un effet de "blindage" marqué.

Au contraire, la courbe (17) a une décroissance très lente, beaucoup trop lente.

On rattrape naturellement largement l'écart entre la courbe (17) et la réalité, en choisissant habilement le paramètre a. Une méthode simple et efficace pour cela, consiste à définir a de manière que:

$$B_0 \int_{-\infty}^{+\infty} \frac{dr}{\left(1+\frac{z^2}{a^2}\right)^2} = \frac{3}{16} \int_{-\infty}^{+\infty} \left(\frac{\varphi'}{\varphi}\right)^2 dr, \tag{17}$$

ce qui revient à assurer la même valeur au premier terme, dans le développement de la distance focale, qu'à donné Picht [18]; ce développement étant très convergent, le procédé est bon. Mais il s'agit là d'une compensation dont l'exactitude est difficile (de l'ordre de quelques pour cent) à évaluer, et il vaudrait mieux l'éviter. C'est ce qui a conduit l'auteur à étudier un modèle plus satisfaisant sur ce point [19].

Le Nouveau Modèle

1. Le Nouveau Modèle

1. L'auteur a proposé le nouveau modèle

$$B(z) = B_0 . \operatorname{sech}(z/b), \tag{18}$$

pour lequel on a:

$$b = 0{,}7593a, \tag{19}$$

a étant la demi-largeur de la cloche à mi-hauteur.

2. Rayons de Gauss

En prenant b pour unité de longueur ($z/b = x$; $R = r/b$), l'équation des rayons gaussiens est

$$R'' ch^2(x) + h^2 R = 0 \tag{20}$$

avec $h = 0{,}7593\ K$, K = paramètre classique défini dans (1) par

$$K^2 = \frac{eB^2}{8m\varphi_0}\, a^2. \tag{21}$$

Le changement de variable: $u = th(x)$ ramène cette équation au type de Legendre

$$(1-u^2)\frac{d^2R}{du^2} - 2u\frac{dR}{du} + \nu(\nu+1)R = 0, \tag{22}$$

et l'expression du rayon général s'écrit

$$R=A.P_\nu[th\,(x)]+BP_\nu[-th\,(x)],$$

$P_\nu(\mathrm{u})$ étant la *fonction* de Legendre d'ordre ν avec $\nu(\nu+1)=h^2$. Les cas $\nu=n$ entier sont physiquement significatifs et mathématiquement singuliers; alors la *fonction* $P_\nu(\mathrm{u})$ devient le *polynome* $P_n(\mathrm{u})$ et $P_n(-\mathrm{u})$ doit être remplacé par $Q_n(\mathrm{u})$ *fonction* de deuxième espèce.

3. Valeurs Remarquables $\nu=n$ Entier

Que ν soit entier ou non, le rayon $r=P_\nu\mathrm{th(x)}$ est toujours un rayon principal, parallèle à l'axe à la distance 1, pour $x\to+\infty$ car on a toujours $P_\nu(1)=1$ et $P'_\nu(1)$ est fini; la dépendance $u=th(x)$ assure le comportement asypmtotique convenable. Mais si ν est un entier n, ce rayon principal émerge encore de la lentille parallèlement à l'axe pour $x\to-\infty$, car $P_n(-1)=1$ si n est entier: la lentille est alors afocale au sens classique du mot. A l'intérieur de la lentille le rayon coupe n fois l'axe, donnant lieu à n *foyers immergés* G_n symétriques par rapport au centre optique, qui est l'un d'eux lorsque n est impair; c'est le seul lorsque $n=1$,

$$r=A\ \mathrm{th}\,(x)+Bx\ \mathrm{th}\,(x). \tag{23}$$

Les valeurs $\nu=n$ pour lesquelles le système est classiquement afocal, séparent les valeurs de ν correspondant aux systèmes convergents ($0<\nu<1$; $3<\nu<4$; $5<\nu<6$; . . .) de celles donnant des systèmes divergents ($2<\nu<3$; $4<\nu<5$; . . .) Dans les instruments actuels on évite toujours les foyers multiples et nous pourrons limiter l'étude par la condition $0<\nu\leq 1$, $0<h^2\leq 2$.

4. Éléments Cardinaux Classiques

Ils définissent la correspondance homographique entre les deux asymptotes de chaque rayon. Or le développement de $P_\nu(u)$ au voisinage de $u=-1$ est connu (cf. [20] p. 224) et donne après quelques transformations:

$$P_\nu(\mathrm{th}\ x)\to\frac{\sin\nu\pi}{\pi}\left\{2x+2[\psi(\nu)+C]+\pi\cot(\pi\nu)\right\}, \tag{24}$$

C étant la constante d'Euler, $\psi(\nu)$ la dérivée logarithmique de $\nu!$. On obtient finalement pour la position du foyer z_F, la distance focale f, et la rotation de l'image φ:

$$(z_F/b)=\psi(\nu)+C+\frac{\pi}{2}\cot(\pi\nu) \tag{25}$$

$$f/b=\frac{\pi}{2\sin(\nu\pi)} \tag{26}$$

$$\varphi=\pi h. \tag{27}$$

Ces lois diffèrent peu de celles de Glaser quand les lentilles sont faibles, mais les écarts deviennent importants lorsque la convergence est

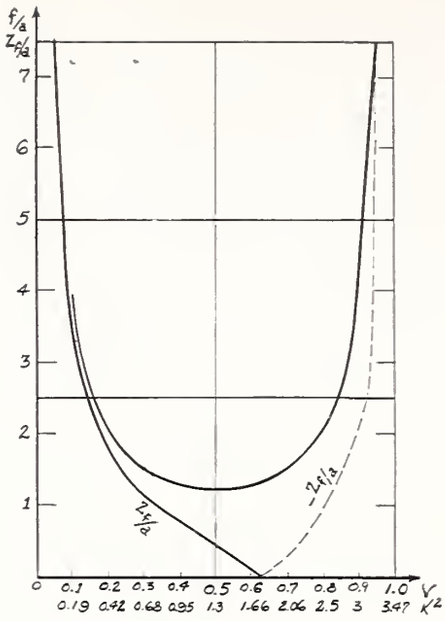


FIGURE 17.15. *Eléments cardinaux classiques.*

forte. On note en particulier que la lentille de projection la plus convergente est définie par $\nu=1/2$, $k=1,14$ $f=1,19a$ au lieu de $k=1$, $f=1,5a$.

La fonction $\psi(\nu)$ est bien tabulée en particulier dans les tables de Jahnke-Emde; elle ne varie d'ailleurs que de zéro à 1, lorsque ν passe de 0 à 1. La figure 17.15 montre l'allure des courbes z_F et f, en fonction de K.

5. Immersion

Lorsque l'objet ou l'image (ou l'un et l'autre) sont immergés dans le champ, il n'existe plus de correspondance homographique rigoureuse entre objet et image: le champ considéré ici n'est pas "newtonien" au sens de Glaser [21] et il n'existe plus d'éléments cardinaux valables pour *tous* les couples objet-image; il est facile de le vérifier dans le cas simple $\nu=1$. Mais les éléments cardinaux définis par Glaser [22] et que nous appellerons brièvement "immergés" restent précieux pour décrire la correspondance objet-image au voisinage du couple conjugué: Foyer-infini (cf. [21] p. 373). La position du foyer G est donnée par

$$P_\nu(u_\nu)=0 \qquad u_\nu=th(x_G) \qquad z_G=bx_G \tag{28}$$

u étant l'une des ν racines réelles de $P_\nu(u)=0$, racine unique si $0<\nu\leq 1$. La distance focale immergée g est alors

$$\frac{b}{g}=\frac{P'_\nu(u_\nu)}{ch^2(x_G)}=\frac{2\sin(\nu\pi)}{\pi P_\nu(-u_\nu)}, \tag{29}$$

la seconde expression étant commode lorsque ν est petit. La distance focale g diffère de f classique dans la mesure où $P_\nu(-u_\nu)$ s'écarte de 1 et l'immersion est négligeable tant que $\nu\leq 0{,}3$ $(P_\nu\geq 0{,}98)$; la position de G tend vers celle du foyer classique F quand ν est petit $(z_F=z_G=1/2\nu)$. La rotation de l'image dépend de la position de l'objet; entre le foyer et l'infini elle est:

$$\varphi=[(\pi/2)-\text{arctg}\,(\text{sh}\,x_G)]h. \tag{30}$$

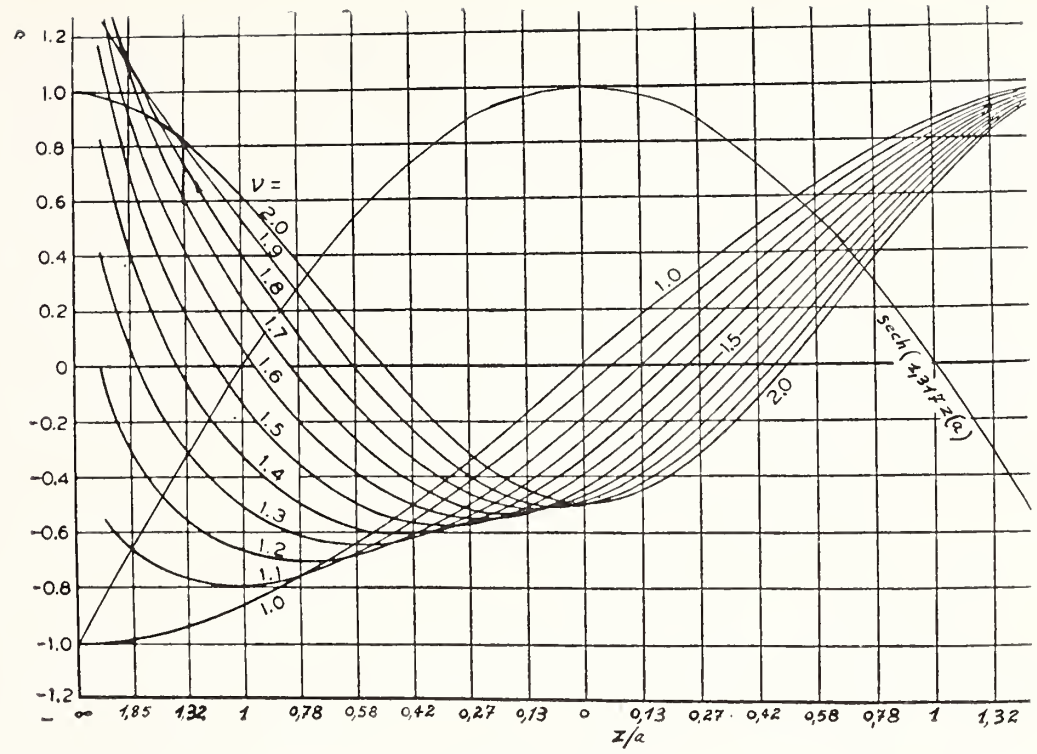


FIGURE 17.16.

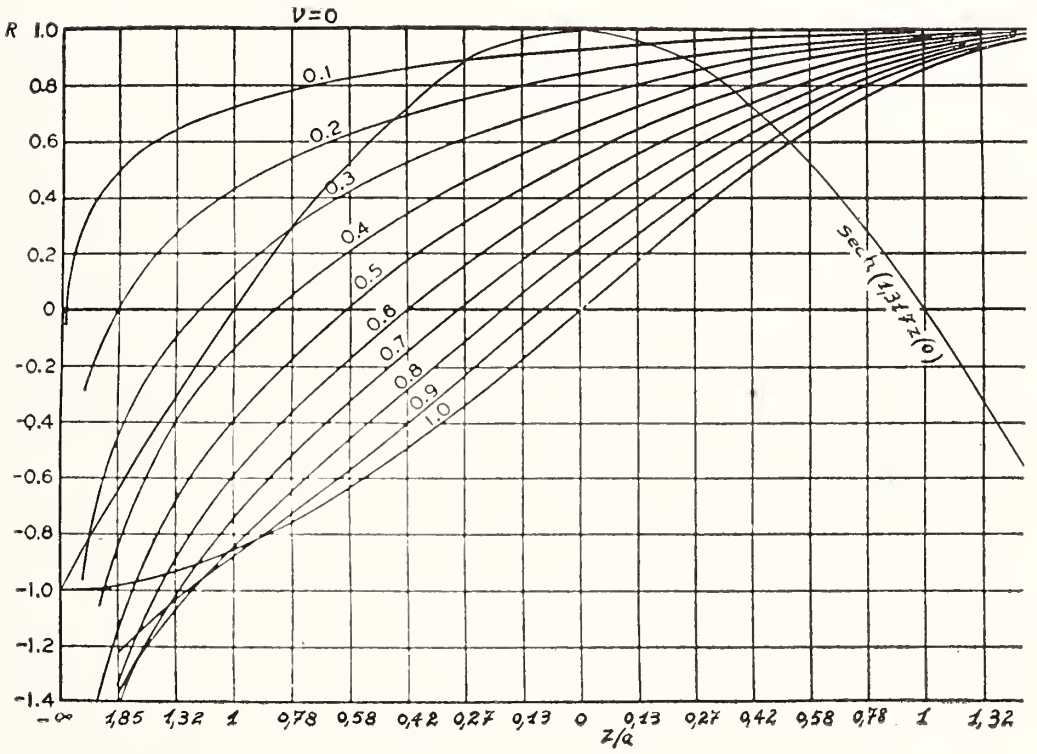


FIGURE 17.17.

La valeur minimum de g caractérise un type important d'objectif de microscope; elle vaut ici: $g=0{,}7593a$, pour $\nu=1$, $k^2=2$, valeurs sensiblement différentes de $g=a$, $k^2=3$ pour le modéle de Glaser et qui serrent de plus près la réalité en particulier la relation entre les ampères-tours (A_t), la demi-largeur a (millimètres) et le champ au centre de la lentille B (kilo-gauss) s'écrit ici:

$$B_0 a = 5{,}27 A_t, \tag{31}$$

alors que les valeurs expérimentales du coefficient numérique s'échelonnent entre 4,4 et 5,8 dans les mesures de Ruska [23], le modèle

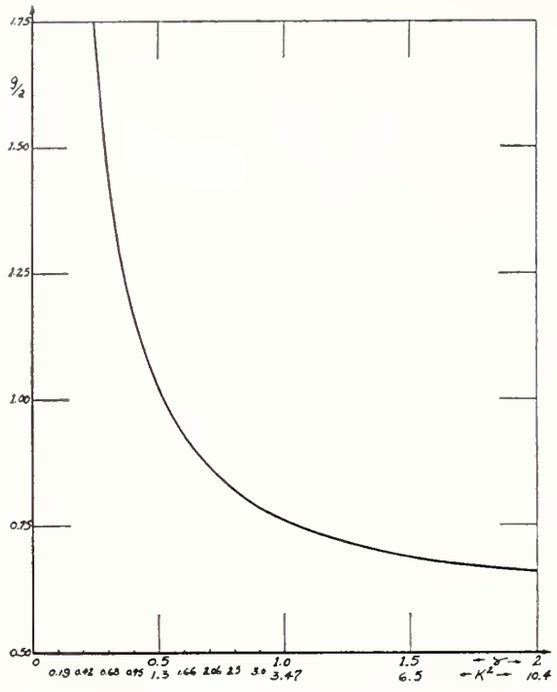


FIGURE 17.18a.

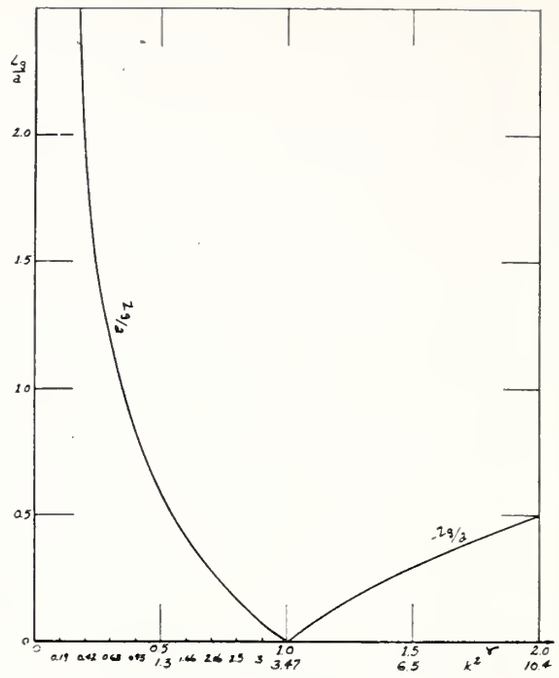


FIGURE 17.18b.

de Glaser donnant 4. Les valeurs de k qui rendent la lentille afocale sont faciles à mesurer; leur répartition fournit un test expérimental commode, permettant de choisir le modèle approprié: La suite de ces valeurs est ici 3,469; 10,404; 20,808 $n(n+1).1,7344$ au lieu de 3; 8; 15; $n'(n'+1)$.

Schelkunoff (cf. [24] p. 423–424) donne un réseau de courbes très précieuses représentant $P_\nu(\cos\theta)$ en fonction de θ, pour les valeurs de ν échelonnées de 0,2 entre 0 et 2; ce sont les *trajectoires principales* légèrement déformées à leurs extrémités par la transformation $\cos\theta = \text{th}\ z$; ces diagrammes sont reproduits figure 17.16 et 17.17 avec une échelle des abscisses graduée en x. Les courbes figure 17.18 donnent g et z_G; pour les calculer on peut s'aider des formules d'approximation de l'appendice II.

Le Nouveau Modèle et les Théories en Cours

Il est intéressant de situer notre modèle au milieu des théories en cours:

1. Modèle de Glaser

Nous pouvons subvenir au défaut de la cloche de Glaser de la manière suivante. Nous remarquons que l'équation différentielle dont l'intégration fait le succès de notre méthode, se réduit lorsque z est grand, au type:

$$\frac{d^2R}{dz^2} + 4h^2e^{-2x}R = 0 \tag{32}$$

dont la solution est:

$$R = AJ_0(2he^{-x}) + BN_0(2he^{-x}), \tag{33}$$

J_0 et N_0 étant les deux fonctions de Bessel, d'ordre 0; solution commode car les fonctions J_0 et N_0 sont bien tabulées. On peut donc remplacer

les deux ailes de la courbe $1/\left(1+\frac{z^2}{a^2}\right)$ par deux exponentielles et compléter la trajectoire classique de Glaser valable dans la région centrale, par deux morceaux de courbe (31) qu'on lui raccorde aux deux points où la caractéristique réelle commence à devenir exponentielle.

2. Méthodes de Minimum

Récemment Glaser [25] a cherché à intégrer l'équation des rayons par une méthode variationnelle; il se donne, à priori un point objet et le point image, et cherche à déterminer un paramètre de la courbe de champ, h par exemple ici, pour qu'il existe effectivement un faisceau de rayons joignant l'objet à l'image. On doit avoir alors (cf. référence [24], p. 224):

$$h^2=\frac{\int_0^i R'^2\,dx}{\int_0^i R^2\,\mathrm{sech}^2(x)dx}, \tag{34}$$

et comme le remarque d'abord Lord Rayleigh, on a une très bonne précision sur h, en calculant le deuxième membre à l'aide d'une fonction R grossièrement approchée.

Cette méthode est excellente pour déterminer la distance focale immergée g, car le choix du rayon principal approché est facile; il ne peut évoluer qu'entre des limites très étroites car il doit entrer dans la lentille parallèlement à l'axe à la distance 1, et couper l'axe au point $G(x_G)$ donné à l'avance et où il a une tangente d'inflexion.

Au contraire, si l'on désire F le foyer normal, on a une beaucoup plus grande latitude et la détermination d'une fonction approchée est bien plus délicate; la méthode est moins efficace. C'est ce que montre bien notre modèle: On peut pour chaque valeur de f et de g, calculer la valeur de l'ω de Glaser qui nous fournirait les valeurs exactes de f et g: on trouve deux fonctions $\omega_1(K)$ et $\omega_2(K)$ fort différentes pour f et pour g.

V. La Lentille Magnétique

Champ Axial

Le type le plus simple de lentille magnétique est représenté figure 17.19, et a été étudié expérimentalement par Le Poole & Ments [3].

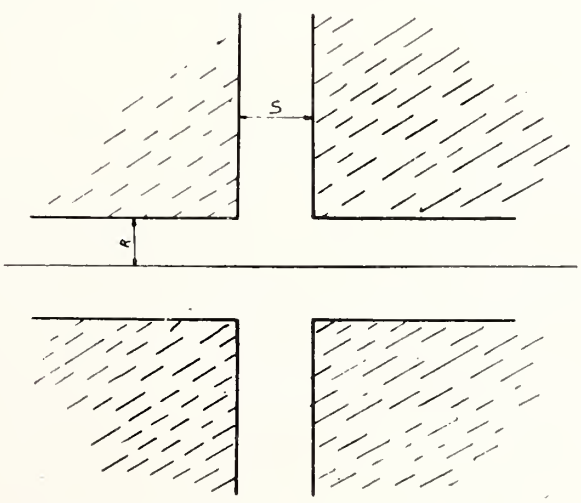


FIGURE 17.19.

Nous allons lui appliquer la théorie du paragraphe IV.

Nous nous limiterons au cas où la perméabilité du fer reste très grande en tout point des pièces polaires et nous la supposerons infinie. Nous pouvons alors essayer de représenter le champ par la formule donnée par Bertram [15]. Dans le cas du champ magnétique, la loi d'Ampère donne la différence de potentiel magnétique scalaire entre les deux pôles, $\phi_2 - \phi_1$ sous la forme:

$$\phi_2 - \phi_1 = nI = \int_{-\infty}^{+\infty} B dr \tag{35}$$

nI étant le nombre d'ampères-tours magnétisants et la formule de Bertram s'écrit:

$$B = \frac{\mu_0 nI}{2R} \frac{sh\,\omega\epsilon}{\epsilon ch\,\omega\left(z+\frac{\epsilon}{2}\right) ch\,\omega\left(z-\frac{\epsilon}{2}\right)} \tag{36}$$

avec $\epsilon = s/R$, s étant la valeur de l'entrefer, $\omega = 1{,}318$ formule qui s'écrit en utilisant la valeur maxima du champ sur l'axe, B_m, observée au centre de la lentille:

$$B = B_m \frac{ch^2 \frac{\omega}{2}\epsilon}{ch\,\omega\left(z+\frac{\epsilon}{2}\right) ch\,\omega\left(z-\frac{\epsilon}{2}\right)}, \tag{37}$$

car la valeur de B_m est

$$B_m = \frac{\mu_0 nI}{R} \frac{th \frac{\omega}{2}\epsilon}{\epsilon} \tag{38}$$

qu'il est commode de comparer à la valeur du champ dans l'entrefer B_0:

$$B_0 s = \mu_0 nI, \tag{39}$$

d'où

$$\frac{B_m}{B_0} = \text{th}\,\frac{\omega\epsilon}{2}. \tag{40}$$

La demi-largeur a de la courbe réelle à mi-hauteur, résulte de la formule (35) on obtient:

$$ch\left(2\omega\frac{a}{R}\right) = 2 + \text{ch}\left(\omega\frac{s}{R}\right) \tag{41}$$

les valeurs de a/R et de B_m/B_0 sont représentées figure 17.20.

On vérifie que la formule (39) donne des résultats en excellent accord avec le graphique expérimental de Le Poole (figure 17.20); la courbe théorique en est indiscernable tant que $s/R < 2$ et l'écart seulement de 5% environ pour $s/R = 4$ Le Poole n'a pas publié la courbe donnant B_0 en fonction de ϵ, mais nous vérifierons indirectement la formule (38) en calculant les propriétés optiques et en les comparant aux graphiques de Le Poole.

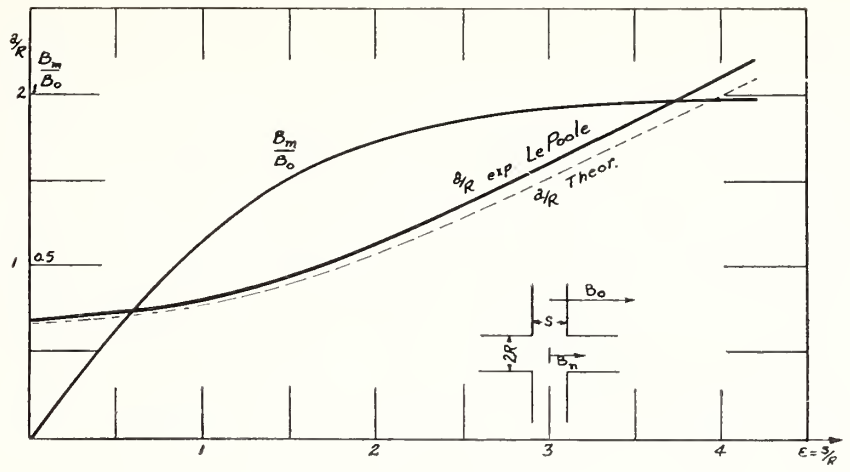


FIGURE 17.20.

Adaptation du Modèle

Pour calculer les éléments cardinaux, nous adopterons notre modèle du paragraphe IV à la courbe (18) en choisissant même valeur de B_0 et en calculant la largeur a′ du modèle par la condition

$$\int_{-\infty}^{+\infty} B_{\text{réel}}\, dr = \int_{-\infty}^{+\infty} B^2_{\text{modèle}}\, dr \tag{42}$$

ce qui donne:

$$\frac{a'}{R} = \frac{1}{2}\left(coth\ \frac{\omega\epsilon}{2}\right)\left(\frac{\omega\epsilon}{th\,\omega\epsilon} - 1\right). \tag{43}$$

L'évaluation de a′ montre que le modèle est bon, car a' se confond avec a (on a par exemple pour $s=0$, $a'_0=0{,}667$, $a_0=0{,}669$)

Propriétés Optiques

On vérifie sur les courbes de Le Poole que les formules données précédemment au chapitre IV donnent une prévision précise des valeurs observées pour f, Z_F, g, Z_G.

En particulier si 1′ on désigne avec Le Poole par K

$$K = \frac{(nI)^2}{U} \tag{44}$$

nI ampères-tours magnétisants, U tension d'accélération des électrons, on obtient:

$$K = 28{,}706\ \frac{\epsilon}{\text{th}\,\frac{\omega}{2}\,\epsilon}\ \frac{R^2}{a^2} \tag{45}$$

en bon accord avec l'expérience.

On vérifie aussi que cette formule donne bien la valeur de $K_{\min}$ correspondant à la valeur minimum de la distance focale d'une lentille projective: il suffit dans la formule précédente de faire $k^2_{\min} = 1{,}3$; on obtient un peu plus de précision en distinguant entre k' du

modèle et k de la réalité ce qui mène à

$$K_{\min}=115{,}184\ \mathrm{th}^2\frac{\omega}{2}\,\epsilon\,\frac{1.3}{[\omega\epsilon\ \coth\ (\omega\epsilon)-1]}. \tag{46}$$

Les valeurs prévues pour le minimum de f, $f_{\min}$ sont bonnes aussi. On peut faire un calcul analogue pour le foyer immergé et déterminer la valeur k_0 pour laquelle G vient au centre de la lentille et la valeur $g_{\min}$ correspondante; il y a encore accord satisfaisant avec l'expérience, comme on peut le vérifier en comparant le tableau suivant aux graphiques de Le Poole et il semble bien que les écarts soient attribuables à la saturation du fer, les écarts les plus grands s'observant aux valeurs élevées de l'induction.

Le tableau suivant donne quelques valeurs numériques ainsi calculées.

TABLEAU 17.1. *Cardinal elements in the new model.*

ν	k^2	z_G/a	g/a	f/a	z_F/a
0, 1	0, 1908	3, 787	3, 86	3, 850	3, 787
0, 2	0, 4163	1, 861	2, 025	2, 029	1, 860
0, 3	0, 6765	1, 190	1, 450	1, 474	1, 176
0, 4	0, 9539	0, 8302	1, 180	1, 254	0, 779
0, 5	1, 301	0, 5914	1, 0276	1, 193	0, 466
0, 6	1, 659	0, 4200	0, 932	1, 254	0, 1464
0, 7	2, 065	0, 2847	0, 864	1, 474	−0, 2700
0, 8	2, 496	0, 1745	0, 820	2, 029	−0, 987
0, 9	2, 965	0, 0810	0, 786	3, 860	−2, 962
0, 95	3, 211			7, 525	−6, 796
1	3, 469	0	0, 7593	∞	∞
1, 1	4, 007	0, 0717	0, 739		
1, 2	4, 579	0, 1359	0, 722		
1, 3	4, 186	0, 1359	0, 708	$T(z)=\mathrm{sech}\left(1,317\frac{z}{a}\right)$	
1, 4	5, 828	0, 2477	0, 697		
1, 5	6, 504	0, 2968	0, 688	$\frac{d^2P}{dz^2}+\frac{k^2}{a^2}T^2(z)P=0$	
2	10, 407	0, 500	0, 658		

TABLEAU 17.2. *Calculated element of a magnetic lens.*

s/P	a/P	K/k^2	$K_{\min}$	$f_{\min}/P$	K_0	g_0/R
0	0, 669	148, 5	193, 2	0, 798	515, 1	0, 508
0, 5	0, 698	146, 3	190, 35	0, 833	507, 6	0, 530
1	0, 783	141	183, 4	0, 934	489	0, 596
2	1, 104	126	164	1, 317	437	0, 838
3	1, 515	122	158, 7	1, 807	423	1, 151
4	2, 008	116, 7	152	2, 396	404, 8	1, 516
5	2, 5	115, 9	151	2, 982	402	1, 9

Appendice

1. Plan des Calculs

Les fonctions $\mathrm{P}_\nu(\mathrm{u})$ d'indice ν non entier et petit (ici pratiquement $0<\nu<1{,}5$) sont peu connues en physique. C'est pourquoi nous donnerons ici quelques formules qui en facilitent l'usage: elles nous ont servi à établir les courbes et les tableaux de cet article.

Bien que le rayon électronique soit donné par la formule $P_\nu(th\ x)$, il est commode de baser les calculs sur la fonction plus usuelle $\Theta_\nu(\theta)=$

$P_\nu(\cos\theta)$: le recours à la littérature mathématique et aux formulaires s'en trouve facilitée et la représentation des propriétés physiques n'est en rien obscurcie; en effet la correspondance $\text{th}\, x = \cos\theta$ est très simple, et les lois en θ, ont l'avantage de dilater la région la plus intéressante au point de vue de la formation des images; l'intérieur de la lentille où les rayons sont courbes, les espaces objets et images exterieurs à la lentille sont alors contractés à l'extrême mais ce défaut est de peu d'importance, car les rayons sont des droites bien représentées par les éléments cardinaux.

Les quantités qui nous intéressent ici sont:

Physique	Mathématique
Les éléments cardinaux classiques	Asymptotes de $P_\nu(\text{th}\, x)$ pour $x \to + \infty$
Position des foyers immergés	Valeur de la racine θ_G de $P_\nu(\cos\theta_G) = 0$
Distance focale immergeés	Valeur de $\dfrac{dP_\nu(\cos\theta)}{d\theta}$ pour $\theta = \theta_G$

2. Asymptotes des Rayons

Schelkunoff donne le développement de $P_\nu(\cos\theta)$ lorsque θ est voisin de π (formule 71, p.432 de la référence [16]). Limité au terme principal on a

$$P_\nu(\cos\theta) = \frac{2\sin\nu\pi}{\pi}\left[\log\cos\left(\frac{\theta}{2}\right) + \psi(\nu) - \psi'(0)\right] + \cos\nu\pi. \qquad (47)$$

On passe facilement aux x ($x \to - \infty$) par:

$$\cos\theta = \text{th}\, x = (-1 + 2e^{2x}).$$

On obtient

$$P_\nu(\text{th}\, x) = \frac{2\sin\nu\pi}{\pi}\left[x + \psi(\nu) + C + \frac{\pi}{2}\cot\pi\nu\right] \qquad (48)$$

avec

$$\psi(x) = \frac{d[\log(x!)]}{dx}$$

$$C = 0{,}577215$$

d'où les formules (25) et (26) du texte, d'où l'on déduit les valeurs représentées graphiquement en se rappelant que

$$\frac{z}{a} = 0{,}7593x. \qquad (49)$$

On a écrit ces formules en changeant le signe de x pour faire correspondre l'espace image à $x \to + \infty$.

3. Racines θ_G de P_ν (cos θ_G)=0

Il est commode de distinguer deux domaines de valeurs de ν

$$0<\nu<1/2$$

$$1/2<\nu<3/2$$

et les valeurs limites $\nu=1/2$, $\nu=1$, $\nu=3/2$, $\nu=2$.

3.1—Valeurs Limites $\nu=1$, $\nu=2$

Alors les P_ν ont une expression trés simple

$$P_1\ (\cos\theta)=\cos\theta \tag{50}$$

$$P_2\ (\cos\theta)=1/2\ (3\cos\theta-1)=1/4\ (3\cos 2\theta+1) \tag{51}$$

et l' on a $\nu=1$, $\theta_G=\pi/2$

$$\nu=1{,}5 \qquad \theta_G=\infty .$$

3.2—Valeurs Limites $\nu=1/2$, $\nu=3/2$

Les fonctions $P_{1/2}$ et $P_{3/2}$ s'expriment alors à l'aide des intégrales elliptiques complètes E et K de module $k=\sin\theta/2$. Il en va de même pour $P_{1/2}$ qui est utile comme auxiliaire dans les formules de récurrence à trois termes relatives à $P_{1/2}$ et $P_{3/2}$.

La formule de Mehler (référence [20] p. 27) donne

$$P_{1/2}(\cos\theta)=\frac{\sqrt{2}}{\pi}\int_0^\theta \frac{\cos\psi}{\sqrt{\cos\psi-\cos\theta}}\,d\psi. \tag{52}$$

En posant: $\sin\psi=\sin\varphi\,\sin\dfrac{\theta}{2}$ on ramène l'intégrale à

$$P_{1/2}\ (\cos\theta)=\frac{2}{\pi}\int_0^{\pi/2}\frac{1-2\sin^2\varphi\,\sin^2\frac{\theta}{2}}{\sqrt{1-\sin^2\varphi\,\sin^2\frac{\theta}{2}}}$$

$$=\frac{2}{\pi}\left[2\int_0^{\pi/2}\sqrt{1-\sin^2\varphi\,\sin^2\frac{\theta}{2}}\,d\varphi-\int_0^{\pi/2}\frac{dy}{\sqrt{1-\sin^2\varphi\,\sin^2\frac{\theta}{2}}}\right]$$

$$P_{1/2}\ (\cos\theta)=\frac{2}{\pi}\left[2E\left(\sin\frac{\theta}{2}\right)-k\left(\sin\frac{\theta}{2}\right)\right]. \tag{53}$$

avec des notations classiques que l'on trouvera décrites dans la référence [26] p. 170 et suivantes.

Par des raisonnements analogues on établit que

$$P_{-1/2}(\cos\theta)=\frac{2}{\pi}\,k\left(\sin\frac{\theta}{2}\right) \tag{54}$$

et les relations de récurrence des P (cf. référence [24] p. 433)

$$(2\nu+1)\cos\theta P_\nu-(\nu+1)\,P_{\nu+1}-\nu P_{\nu-1}=0, \tag{55}$$

donnent

$$P_{3/2}\,(\cos\theta)=\frac{2}{3\pi}\left[8\cos\theta E\left(\sin\frac{\theta}{2}\right)-(1+4\cos\theta)\,k\left(\sin\frac{\theta}{2}\right)\right]. \tag{56}$$

Les tables de fonctions elliptiques permettent alors de déterminer facilement, par interpolation les racines:

$\nu=\frac{1}{2}$, $\theta_G=132°42'5$
$\nu=\frac{3}{2}$, $\theta_G=68°9'$.

3.3—Faibles Valeurs de ν: $0<\nu<\frac{1}{2}$

Si ν est inférieur à 0,2 le foyer classique F et le foyer immergé G sont confondus à 1/1000 près. On a

$$\theta_G=\theta_F \tag{57}$$

et θ_F est fourni par l'asymptote, d'équation (48).

Lorsque ν est compris entre 0,2 et 0,6, cette formule du premier ordre n'est plus assez précise (erreur de 5% à l'extrémité de l'intervalle environ); il faut compléter l'équation (48) par un terme correctif qu'on obtient par la méthode des perturbations (cf. réf. [24] p. 196–199). En notant de manière abrégée l'équation de l'asymptote (48):

$$y_1=\alpha x+\beta, \tag{58}$$

on a pour la deuxième approximation y_2, x étant négatif:

$$y_2=\alpha x+\beta-h^2\int_{-\infty}^{x}\frac{(x-t)(\alpha t+\beta)}{ch^2(t)}\,dt \tag{59}$$

l'intégrale s'évalue aisément par intégration par parties, en utilisant successivement les fonctions 1+th(t), puis t th t-log ch t jusqu'à ce qu'on parvienne à: $\int_{-\infty}^{x}\log\text{ ch }(t)\,dt$ que l'on intègre par un développement en série de e^{2t}, e^{4t}, etc . . . On obtient

$$\left.\begin{aligned} &y_2=(\alpha x+\beta)[1-h^2\log\,(1+e^{2x})+\alpha h^2\left[e^{2x}-\frac{e^{4x}}{4}+\frac{e^{6x}}{9}\;\ldots\right],\\ &\alpha=\frac{2\sin\,\nu\pi}{\pi},\qquad \beta=\frac{2\sin\,\nu\pi}{\pi}\,[\psi(\nu)+c]+\cos\,\nu\pi,\\ &h^2=\nu(\nu+1), \end{aligned}\right\} \tag{60}$$

et en particulier on a très simplement la différence $x_G - x_F$

$$\left.\begin{aligned} x_G - x_F &= -\frac{\alpha h^2 \left[e^{2x} - \frac{e^{4x}}{4} + \frac{e^{6x}}{9} \cdots \right]}{1 - h^2 \log (1+e^{2x})}, \\ x &= x_G. \end{aligned}\right\} \tag{61}$$

3.4—Valeurs Moyennes $-1/2 < \nu < 3/2$

Au voisinage de $\theta = \pi/2$, $P_\nu (\cos \theta)$ peut être développé en série de puissances de cos θ (formule (72) de la référence [24], p. 432), mais les valeurs intéressantes de θ, s'étendent de $\pi/4$ à $3\pi/4$ et la série converge mal aux deux bords de cet intervalle. Nous nous en servirons seulement pour déterminer les éléments initiaux d'une représentation plus commode. En $\theta = \pi/2$ on a

$$P_\nu(0) = \frac{1}{\sqrt{\pi}} \frac{\left(\frac{\mu - 1}{2}\right)!}{\left(\frac{\nu}{2}\right)!} \tag{62}$$

$$\left.\frac{dP_\nu(\cos\theta)}{d\theta}\right|_{\theta = \pi/2} = -\frac{1}{\pi} \frac{\sin \nu\pi}{P_\nu(0)} \tag{63}$$

la fonction $x!$ est de variation très lente dans le domaine envisagé ici $0 < \nu \leq 1,5$, car l'argument dans les formules précédentes est toujours voisin de celui qui rend $x!$ minimum: $x_m = 0{,}46163$ $x_{m!} = 0{,}88560$. On peut toujours se ramener à des factorielles comprises entre 0,8856 et 1 en utilisant la définition $(x+1)! = (x+1).(x!)$ valable aussi pour x négatif ($\nu < 1$).

Il est facile d'interpoler la fonction $(x!)$ par une formule parabolique dans cette région mais c'est à peine nécessaire car elle est tres bien tabulée.

Nous utiliserons les formules (62) et (63) pour fixer les constantes φ et A dans la représentation approchée (cf. [24], p. 417 et suivantes)

$$P_\nu(\cos\theta) = \frac{A \sin(\gamma\theta - \varphi)}{\sqrt{\sin\theta}} \tag{64}$$

$$\gamma^2 = \left(\nu + \frac{1}{2}\right)^2 + \frac{1}{4}. \tag{65}$$

On obtient

$$\operatorname{tg}\left(\varphi - \frac{\pi}{2}\gamma\right) = \frac{\gamma}{2}\left[\frac{\left(\frac{\nu-1}{2}\right)!}{\left(\frac{\nu}{2}\right)!}\right]^2 \cot\frac{\pi\nu}{2} \tag{66}$$

$$\theta_G = \phi/\gamma. \tag{67}$$

La connaissance de A n'est pas nécessaire pour le calcul de θ_G, mais se montrera utile pour celui de $dP_\nu/d\theta$, on a

$$\gamma A = \frac{-2}{\sqrt{\pi}} \frac{\sin \frac{\pi\nu}{2}}{\cos\left(\gamma\frac{\pi}{2}-\varphi\right)} \frac{\left(\frac{\nu}{2}\right)!}{\left(\frac{\nu-1}{2}\right)!}. \tag{68}$$

Ces formules sont très précises et donnent θ_G avec une erreur relative bien inférieure à 0,001 dans tout l'intervalle utile.

4. Pente $dP_\nu(\cos\theta)/d\theta$

Nous distinguerons ici les mêmes domaines en ν, que pour le calcul de θ_G.

4.1—ν Entier

Alors les formules (50) et (51) donnent directement

$$\frac{dP_1(\cos\theta)}{d\theta} = -1, \qquad \frac{dP_2(\cos\theta)}{d\theta} = -1,5.$$

4.2—$\nu = 1/2$, $\nu = 3/2$

Les formules donnant la dérivée des intégrales elliptiques par rapport à leur module sont simples et connues (cf. réf. [26]), formules 789–1 et 789–2, p. 173). On obtient avec leur aide en dérivant notre formule (53)

$$\nu = \frac{1}{2}, \qquad \theta = (\theta_G)_{1/2}, \qquad \frac{dP_{1/2}}{d\theta} = -\frac{k\left(\sin\frac{\theta}{2}\right)}{\pi \sin\theta} = -0{,}9746. \tag{69}$$

Pour $\nu = 3/2$, il est d'ailleurs plus simple d'utiliser la relation de récurrence:

$$\frac{dP_\nu}{d\theta} = \frac{-\nu}{\sin\theta} P_{\nu-1}, \qquad \theta = (\theta_G)_\nu, \tag{70}$$

elle mène à:

$$\nu = \frac{3}{2}, \qquad \theta = (\theta_G)_{3/2}, \qquad \frac{dP_{3/2}}{d\theta} = \frac{-3}{\pi}\frac{(2E-k)}{\sin\theta} = -1{,}1884. \tag{71}$$

4.3—ν Petit $0 < \nu < 1/2$ (Bord de la Lentille)

Alors on utilise le Wronskien de l'équation de Legendre (cf. réf. p. 424) qui donne en $\theta = \theta_G$:

$$\frac{dP_\nu(\cos\theta)}{d\theta} = \frac{2\sin\gamma\pi}{\pi\sin\theta}\frac{1}{P_\nu(-\cos\theta)}, \tag{72}$$

où θ_G est assez voisin de π pour qu'on puisse utiliser le développement

en série classique

$$P_\nu(\cos\theta)=1-\nu(\nu+1)\sin^2\frac{\theta}{2}+\frac{(\nu-1)\nu(\nu+1)(\nu+2)}{4}\sin^4\frac{\theta}{2}$$

$$-(\nu-1)\theta_G\left(\sin\frac{\theta}{2}\right). \qquad (73)$$

4.4.—Intérieur de la Lentille $-\frac{1}{2}\leq\nu\leq\frac{3}{2}$

L'expression (64) est suffisamment exacte pour qu'on puisse la dériver ce qui donne.

$$\theta=\theta_G, \qquad \frac{dP_\nu(\cos\theta)}{d\theta}=\frac{\gamma A}{\sqrt{\sin\theta_G}}. \qquad (74)$$

Néanmoins aux bords de l'intervalle, l'erreur atteint plusieurs millièmes, mais on peut utiliser la deuxième approximation obtenue par la méthode des perturbations. On l'obtient le plus simplement en intégrant la formule classique

$$\frac{d}{d\theta}[\sin\theta P_\nu(\cos\theta)]=-\nu(\nu+1)\sin\theta P_\nu(\cos\theta). \qquad (75)$$

Si nous posons

$$m=\frac{dP_\nu(\cos\theta)}{d\theta} \qquad (76)$$

et que nous appelons, m_1 la valeur de m, en $\theta=\theta_G$ donnée par (74), m_2 la valeur améliorée et m_0 la pente en $\theta=\pi/2$, nous obtenons, en substituant (64) au deuxième membre de (75)

$$\left.\begin{aligned} m_2\cos\alpha&=m_0+h^2m_1\sqrt{\cos\alpha}\left[\frac{1-\cos\gamma\alpha}{\gamma^2}-\frac{\alpha^4}{48}\right],\\ \alpha&=\theta_G-\frac{\pi}{2}. \end{aligned}\right\} \qquad (77)$$

5.—Table Numérique des Valeurs de $P(-\cos\theta)$ pour les Valeurs Demi-Entières de n

Cette table nous a été aimablement communiquée par Miss Marion Gray and Mr. Schelkunoff, des Bell Telephone Laboratories, qui l'ont calculée dès 1940 pour résoudre un problème de rayonnement électromagnétique.

$P_n(-\cos\theta)$

θ	$n=.1$	.2	.3	.4	.5	.6	.7	.8	.9	1.0
5°	0.365201	−0.254581	−0.813813	−1.272544	−1.599553	−1.774742	−1.791031	−1.652930	−1.378654	−0.996444
10°	.501717	.005894	−.453932	−.847492	−1.150000	−1.343918	−1.420160	−1.378607	−1.227945	−.984808
20°	.638358	.268268	−.088558	−.411669	−.683193	−.888957	−1.019355	−1.069887	−1.041354	−.939693
30°	.718190	.423320	.130467	−.145688	−.391682	−.596024	−.749809	−.847187	−.885632	−.866025
40°	.774511	.534092	.289416	.0511663	−.170483	−.366364	−.528746	−.651698	−.734616	−.766044
50°	.817704	.620144	.414869	.209624	.0120125	−.170817	−.332468	−.467557	−.574395	−.642788
60°	.852374	.690081	.518406	.342882	.169084	.002434	−.151995	−.289652	−.406673	−.5
70°	.880955	.748422	.606031	.456324	.307261	.157754	.016269	−.116849	−.237227	−.342020
80°	.904886	.797813	.681210	.557670	.430035	.301038	.173516	.050203	−.066312	−.113648
90°	.925086	.839927	.746089	.645288	.539353	.430189	.319752	.209982	.102787	0
100°	.942171	.875872	.802069	.721889	.636309	.546730	.454374	.360536	.226528	.173648
110°	.956571	.906416	.850092	.788227	.721505	.650659	.576469	.499745	.421314	.342020
120°	.968597	.932102	.890814	.845072	.795249	.741748	.685007	.625480	.563646	.5
130°	.978471	.953322	.924694	.892752	.857676	.819665	.778935	.735715	.692387	.642788
140°	.986362	.970362	.952059	.931521	.908821	.884042	.857275	.828616	.798169	.766044
150°	.992387	.983428	.973140	.961544	.948665	.934528	.919163	.902601	.884877	.866025
160°	.996635	.992665	.988095	.982927	.977168	.970822	.963895	.956393	.948322	.939693
170°	.999161	.998171	.997028	.995735	.994289	.9926933	.990947	.989050	.987003	.984808
θ	$n=1.1$	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
5°	−0.542651	−0.058458	+0.413527	0.833237	1.165932	1.385435	1.476543	1.433916	1.265476	0.989351
10°	−.672324	−.318188	+.047601	.395130	.696918	.930007	1.077683	1.130627	1.087486	.954769
20°	−.775450	−.562976	−.319350	−.063159	+.186880	.412909	.599730	.735204	.811308	.824533
30°	−.792525	−.672251	−.514788	−.331564	−.135126	+.061648	.246321	.407651	.536224	.625000
40°	−.756377	−.705113	−.617004	−.498494	−.357352	−.202235	−.042220	.113682	.260689	.380236
50°	−.679240	−.680298	−.648801	−.587376	−.500235	−.392577	−.270327	−.139860	−.007413	.119764
60°	−.567487	−.607957	−.621232	−.608117	−.570350	−.510520	−.431952	−.338567	−.234713	−.125
70°	−.428902	−.496128	−.542575	−.567190	−.571540	−.555254	−.520788	−.468913	−.402560	−.324533
80°	−.269688	−.352628	−.421019	−.473792	−.510309	−.530315	−.533992	−.521924	−.495078	−.454769
90°	−.906662	−.185629	−.265506	−.335101	−.393447	−.439820	−.473745	−.495011	−.503662	−.5
100°	+.083160	−.003726	−.085870	−.162209	−.231828	−.293895	−.347694	−.392684	−.428461	−.454769
110°	.262707	+.184212	+.107352	+.032911	−.038356	−.105747	−.168608	−.226363	−.278485	−.324533
120°	.435048	.369302	.303277	.237487	+.172439	+.108625	+.046528	−.013395	−.070707	−.125
130°	.593260	.542960	.491146	.438457	.385144	.331534	.277906	+.224546	+.171992	.119764
140°	.732358	.697230	.660787	.623160	.584483	.544892	.504529	.463536	.422057	.380236
150°	.846085	.825099	.803106	.780153	.756288	.731556	.706009	.679698	.652678	.625
160°	.930510	.920782	.910521	.899731	.888427	.876616	.864311	.851520	.838256	.824533
170°	.982463	.979971	.977330	.974544	.971609	.968530	.965306	.961937	.958424	.954769
180°	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Bibliographie.

(1) P. Grivet, Advances in Electronics **2,** 47–100.
(2) H. Bruck, Cahiers phys. **24,** 1–14 (1944).
(3) J. B. Le Poole & M. J. Mentz, J. Appl. Sci. Research, Netherlands **1,** 3–17 (1947).
(4) C. Fert et P. Gautier, C. R. Acad. Paris **233,** 148–49 (1951).
(5) V. K. Zworykin and al., Electron optics, p. 477–480.
(6) G. Liebmann, Advances in Electronics 2, 115.
(7) M. Duchene, Le telescope électronique, thèse (Paris , 1948).
(8) L. Jacob, J. Appl. Phys. **21,**, 966–970 (1950).
(9) P. Grivet et Y. Rocard, Rev. Sci. **88,** p. 86–87 (1949).
(10) E. Regenstreif, Ann. Radioélectricité **6,** 62–94 (1951); 1–31 (1951).
(11) R. Rudenberg, J. Franklin Inst. **246,** 311–339 (1948); 377–408 (1948).
(12) F. Heise et O. Rang, Optik **5,** 201–216 (1949).
(13) E. Regenstreif, J. Phys. **12,** 760–761 (1951) et C. R. Acad. Paris **232,** 1818–20 (1951).
(14) O. Scherzer, Optik **2,** 114–132 (1947) et Communication n° 28 au Congrès international de Microscopie de Paris, septembre 1950.
(15) S. Bertram, Proc. Inst. Radio Engrs., **28,** 418–420 (1940) et J. Appl. Phys. **13,** 496–502 (1942).
(16) R. G. E. Hutter, J. Appl. Phys. **16,** 680–699 (1945).
(17) M. Bernard et P. Grivet, C. R. Ac. de Paris **233,** 788 (1951).
(18) J. Picht, Electronoptics, p. 115–121, Edwards 1944.
(19) P. Grivet, C. R. Ac. de Paris **233,** 971 (1951).
(20) E. W. Hobson, Spherical and ellipsoidal harmonics, Cambridge Univ. Press.
(21) W. Glaser et O. Bergmann, Z. angew. Math. u. Phys. **1,** 363–379 (1950).
(22) W. Glaser, Z. Phys. 285–315 (1941).
(23) E. Ruska, Arch. Elektrotechnik **38,** 102–130 (1944).
(24) Schelkunoff, Appl. mathematics (New-York, 1948).
(25) W. Glaser, Ann. Phys. **7,** 213–227 (1950).
(26) H. B. Dwight, Tables of integrals and other mathematical data (McMillan Book Co., New York, N. Y.).

18. A Property of the Paraxial Ray Equation and Some Consequences

By P. A. Clavier [1]

Introduction

Progress in the field of electron optics is hampered much too often by the fact that the paraxial ray equation is solvable in general only by iterative means. Graphical iteration will not give reliable results because it cannot be carried out with the required accuracy. If the path is determined by means of an algebraic computation, one type of error will be introduced that depends on the number of significant figures used in the calculations. Errors will also be introduced by the fact that the intervals in which the field plot is replaced by a known function are not infinitely small. Since both these types of error are present, it is difficult to estimate the total error. The purpose of this paper is to show that one can find means of integration that avoid inherent errors.

It is shown first that the characteristic equation of any linear differential equation of the second order can be transformed into a paraxial ray equation so that the proposed integration may be used for the solution of more general cases.

General Equation Into Standard Form

The characteristic equation of the general second-order linear equation

$$A(x)y''+B(x)y'+C(x)y=D(x) \tag{1}$$

is defined as

$$A(x)y''+B(x)y'+C(x)y=0. \tag{2}$$

If R is solution of the paraxial ray equation

$$\phi R''+\frac{1}{2}\phi' R'+\frac{1}{4}\phi'' R=0, \tag{3}$$

one can find ϕ such that the solution of eq (2) is

$$y=\alpha(x)R. \tag{4}$$

By substitution, one finds

$$\phi=\phi_0 \exp \pm \int_{x0}^{x} \frac{2}{3}\left[12\frac{C}{A}-6\left(\frac{B}{A}\right)'-3\left(\frac{B}{A}\right)^2\right]^{1/2}dx \tag{5}$$

[1] Sylvania Electric Products, Inc., Bayside, N. Y.

$$y=R\phi^{1/4}\exp\int_{x0}^{x}-\frac{1}{2}\frac{B}{A}dx. \tag{6}$$

A Property Derived from Equations (5) and (6)

Because of the + or — sign in eq (5), one sees that if ϕ_1 and ϕ_2 are two potentials

$$\phi_1\phi_2=\text{const.}, \tag{7}$$

the corresponding solutions R_1 and R_2 of the paraxial ray equations are

$$\frac{R_1}{R_2}=\left(\frac{\phi_2}{\phi_1}\right)^{1/4}. \tag{8}$$

This theorem can be extended to the following one:
If ϕ_1 and ϕ_2 are two potentials, α_1 α_2 two constants, R_1 one solution of the paraxial ray equation corresponding to ϕ_1 and ϕ_2 is such that

$$\phi_2=\frac{\alpha_1+\alpha_2\int\frac{\phi_1}{R_1^2}dx}{\phi_1}, \tag{9}$$

the solution R_2, corresponding to ϕ_2, is

$$R_2=R_1\phi_1^{1/2}\left[A_0+A_1\int\frac{dx}{\phi_2^{1/2}\phi_1^{1/2}R_1^2}\right] \tag{10}$$

in which A_0 and A_1 are integration constants. The same procedure may be used again replacing ϕ_1 by ϕ_2 and R_1 by any solution R_2, and thus a series of "integrable fields" may be obtained.

A Property of the Paraxial Ray Equation

The paraxial ray equation, eq (3), is a linear second-order equation in φ as well as in R. It permits one, when knowing a solution R_1, corresponding to a potential ϕ_1, to find the general form ϕ_2 of the potentials for which R_1 will be a solution

$$\phi_2=\phi_1\left[\alpha_1+\alpha_2\int\frac{dx}{\phi_1^2R_1^2}\right], \tag{11}$$

where α_1 and α_2 are constants. The solution R_2 corresponding to ϕ_2 is known

$$R_2=R_1\left[A_0+A_1\int\frac{dx}{R_1^2\phi_2^{1/2}}\right], \tag{12}$$

A_0 and A_1 being integration constants. Using now ϕ_2 for ϕ_1 and a solution R_2, which is linearly independent of R_1, for R_1, one can use the same method again and generate another series of "integrable fields". This series and the one found in the last paragraph may be combined,

meaning that at any time one can use one or the other form as the next step and thus increase considerably the number of "integrable fields."

Fringing Fields

The generation of a series of "integrable fields" would be of no great use, if such fields were physically not realizable. Restricting ourselves to electron optics, the potential $\phi(x)$ must be positive, never infinite, and continuous; the first derivative $\phi'(x)$ must be continuous. It is always possible to divide the axial coordinate into sections, the potential in each section representing a lens, and the total lens becoming the sum of the different arbitrarily created partial lenses. It is, of course, possible also to build a total lens by placing adjacent partial lenses along the axis. In such case the potential and its first derivative must be continuous at the boundary between any two adjacent lenses. One of the "integrable fields" can be used between two abscissae if the required boundary conditions are fulfilled. One way of fulfilling those conditions is to add on each side of the "integrable field" a "fringing field" for which the corresponding paraxial ray equation is integrable and which can fulfill the required boundary conditions. Such fields do not have to extend over large axial lengths nor show any rapid variations.

The simplest fringing field is given by

$$\phi = a_0 + a_2 x^2. \tag{13}$$

The corresponding paraxial ray equation yields

$$R = A_0 \exp \sqrt{\frac{a_2}{2}} \int \frac{dx}{\sqrt{a_0 + a_2 x^2}} + A_1 \exp - \sqrt{\frac{a_2}{2}} \int \frac{dx}{\sqrt{a_0 + a_2 x^2}}, \tag{14}$$

A_0 and A_1 being integration constants.

Double Riccati Transform

Let f and g be defined by:

$$\phi = e^{\int f dx}, \tag{15}$$

and

$$R = e^{\int g dx}. \tag{16}$$

One can replace ϕ by eq (15) and R by eq (16) in the paraxial ray equation, eq (3). This new equation is solvable in functions of S, which are defined by

$$S = g + \frac{f}{4}. \tag{17}$$

It yields

$$\phi = \phi_0 \exp \pm \frac{4}{3} \int \frac{\sqrt{-3S^2(S^2 + S')}}{S} dx, \tag{18}$$

$$R = \phi^{-1/4} e^{\int S dx} [A_0 + A_1 \int e^{-2\int S dx} dx], \tag{19}$$

A_0 and A_1 being integration constants. One can thus derive two "integrable fields" from any function S; this function may be chosen real or complex.

Equation (18) is the formal integration of

$$\frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2+S^2+S'=0, \tag{20}$$

which is of the form

$$S^2+S'+X=0. \tag{21}$$

In electron optics X is real and positive. When this method is used to solve the unidimensional Schrödinger equation, X can be kept real but may be negative. The integration of eq (21) which follows is restricted to the case X positive. The case of X negative can be solved also but by somewhat different means.

Integration of eq (21) when X is real and positive

If S_0 is one solution of eq (21), the general solution is

$$S=S_0+\frac{\exp -2\int_{x0}^{x} S_0 dx}{K_0+\int_{x0}^{x}\exp -2\int_{x0}^{x} S_0 dx dx}, \tag{22}$$

where K_0 is an integration constant. If S is complex,

$$S=Re+jI, \tag{23}$$

one gets from eq (21)

$$Re=-\frac{I'}{2I}. \tag{24}$$

If S_0 is real, and S is complex, K_0 must be complex:

$$Ko=a+\overline{b}_j, \tag{25}$$

and

$$I=-\frac{\overline{b}\exp -2\int_{x0}^{x} S_0 dx}{\left\{\left[a+\int_{x0}^{x}\exp -2\int_{x0}^{x} S_0 dx\right]2+\overline{b}^2\right\}}. \tag{26}$$

The use of a complex S simplifies the form of the solution of the corresponding paraxial ray equation. From eq (19),

$$R=\phi^{-\frac{1}{4}}(|I|)^{-\frac{1}{2}}\left[A_0\cos\int_{x0}^{x} I dx+A_1\sin\int_{x0}^{x} I dx\right], \tag{27}$$

where the vertical bars around I mean absolute value and A_0, A_1 are integration constants.

It is now necessary to find one particular real solution S_0. If the field extends from x_0 to x_1, one can particularize the solution by assuming

$$S_0(x_0)=0. \tag{28}$$

One can always divide the total lens into adjacent partial lenses and start afresh at each boundary between partial lenses. If one of those partial lenses extends from x_i to x_j, one can use between those abscissae an S_0 such that the limit of S_0 is zero if $x \to x_i$ and $x > x_i$.

Assuming the integration to extend in one step from x_0 to x_1 and assuming eq (28), one gets from eq (21)

$$S_0' < -x. \tag{29}$$

In the present case one can integrate both sides of eq (29); it yields:

$$S_0 < -\int_{x0}^{x} X dx. \tag{30}$$

Combining eqs (21) and (30), one gets:

$$S_0' < -X - \left[\int_{x0}^{x} X dx\right]^2, \tag{31}$$

both sides of which can be integrated as follows:

$$S_0 < -\int_{x0}^{x} X dx - \int_{x0}^{x}\left[\int_{x0}^{x} X dx\right]^2 dx. \tag{32}$$

The same procedure may be used again.

The method can be condensed in the following manner: Let

$$S_1 = -\int_{x0}^{x} X dx \tag{33}$$

$$S_2 = -\int_{x0}^{x} S_1^2 dx \tag{34}$$

$$S_3 = -\int_{x0}^{x} [S_1 + S_2]^2 dx \tag{35}$$

$$S_{n+1} = -\int_{x0}^{x} [S_1 + S_n]^2 dx. \tag{36}$$

One gets

$$S' < x - [S_1 + S_n]^2 \tag{37}$$

and

$$S < S_1 + S_{n+1}. \tag{38}$$

It is easy to prove that

$$S_{n+1} < S_n. \tag{39}$$

From eq (38) the S_{ns}' must have a limit if S is not infinite; this limit can be shown to be $S - S_1$.

If one writes:

$$S=S_1+S_{n+1}, \tag{40}$$

the maximum error will occur for $x=x_1$.

Example

For

$$X=1-x^2 \qquad (-1\leq x\leq +1), \tag{41}$$

the integration may be restricted because of the symmetry to

$$0\leq x\leq +1. \tag{42}$$

The solution S_0 is known; it is

$$S_0=x. \tag{43}$$

Using the method of the last paragraph, one can find successive approximations for S_0 and compute the maximum error existing for each approximation. The maximum errors are:

Approximation	Percent
1	33. 4
2	11. 7
3	3. 5
4	0. 77
5	. 14
6	. 02

Another Approximation

One may sometimes with advantage stop the approximation on S at an early stage, compute the potential ϕ_1 for which the approximate S is the true solution, find the solution R_1 of the corresponding paraxial ray equation and then use the following well known method.

One can write

$$\phi=\phi_1-\epsilon. \tag{44}$$

The difference ϵ must be small enough to assure the convergence of the series

$$\sum_{n=1}^{\infty}={}_1R+{}_2R+\ \ldots\ +{}_nR, \tag{45}$$

the terms of which are defined later.

Let

$$R_2=R_1+{}_1R \tag{46}$$

and

$${}_1R=A_1R_1. \tag{47}$$

In the paraxial ray equation, eq (3), one replaces ϕ by eq (44), R by R_2 in the term involving ϕ_1 and R by R_1 in the terms involving ϵ. eq (3) becomes

$$A_1''\phi_1 R_1 + A_1'\left(2\phi_1 R_1' + \frac{1}{2}R_1\phi_1'\right) = \epsilon R_1'' + \frac{1}{2}\epsilon' R_1' + \frac{1}{4}\epsilon'' R_1, \tag{48}$$

which yields

$${}_1R = R_1 \int_{x0}^{x} \frac{dx}{R_1^2\sqrt{\phi_1}} \int_{x0}^{x} \frac{R_1}{\sqrt{\phi_1}}\left[\epsilon R_1'' + \frac{1}{2}\epsilon' R_1' + \frac{1}{4}\epsilon'' R_1\right]dx. \tag{49}$$

Let now

$$R_3 = R_2 + {}_2R = R_1 + {}_1R + {}_2R \tag{50}$$

and

$${}_2R = A_2 R_2. \tag{51}$$

One replaces in eq (3) ϕ by eq (44), R by R_3 in the terms involving ϕ_1 and R by R_2 in the terms involving ϵ. Equation (3) becomes

$$A_2''\phi_1 R_2 + A_1'\left(\epsilon\phi_1 R_2' + \frac{1}{2}R_2\phi_1'\right) = \epsilon({}_1R)'' + \frac{1}{2}\epsilon'({}_1R)' + \frac{1}{4}\epsilon''({}_1R). \tag{52}$$

Equation (52) is of the form of eq. (48). It makes it possible to condense the successive approximations in the following manner.

The successive approximations for R are

$$R_2 = R_1 + {}_1R \tag{53}$$

$$R_3 = R_2 + {}_2R = R_1 + {}_1R + {}_2R \tag{54}$$

$$R_n = R_{n-1} + {}_{n-1}R = R_1 + {}_1R + {}_2R \ \ldots \ + {}_{n-1}R, \tag{55}$$

with:

$${}_1R = R_1 \int_{x_0}^{x} \frac{dx}{R_1^2\sqrt{\phi_1}} \int_{x_0}^{x} \frac{R_1}{\sqrt{\phi_1}}\left[\epsilon R_1'' + \frac{1}{2}\epsilon' R_1' + \frac{1}{4}\epsilon'' R_1\right]dx \tag{56}$$

$${}_2R = R_2 \int_{x_0}^{x} \frac{dx}{R_2^2\sqrt{\phi_1}} \int_{x_0}^{x} \frac{R_2}{\sqrt{\phi_1}}\left[\epsilon({}_1R)'' + \frac{1}{2}\epsilon'({}_1R)' + \frac{1}{4}\epsilon''({}_1R)\right]dx \tag{57}$$

$${}_nR = R_n \int_{x_0}^{x} \frac{dx}{R_n^2\sqrt{\phi_1}} \int_{x_0}^{x} \frac{R_n}{\sqrt{\phi_1}}\left[\epsilon({}_{n-1}R)'' + \frac{1}{2}\epsilon'({}_{n-1}R)' + \frac{1}{4}\epsilon''({}_{n-1}R)\right]dx. \tag{58}$$

Conclusion

A noniterative method of approximating the solution of the paraxial ray equation is discussed. This method requires fewer numerical computations than iterative methods to give results with the required degree of accuracy.

19. Etude Théorique de la Lentille Formée de Deux Cylindres Coaxiaux de Même Diamètre à l'Aide d'Une Représentation Approchée du Potentiel Axial

Par Pierre Grivet [1] et Michel Bernard [1]

Étude du Potentiel Exact

Les électrodes sont constituées par deux cylindres coaxiaux de même diamètre (figure 19.1); l'unité de longueur sera le rayon commun. Ils sont séparés par une coupure infiniment mince et portés respectivement aux potentiels Φ_1 et Φ_2; dans la suite on posera $\gamma=\Phi_1/\Phi_2$. Nous considérons les cylindres comme s'étendant indéfiniment de part et d'autre de la coupure.

L'équation de Laplace peut, dans ce cas particulier, s'intégrer rigoureusement. On sépare les variables, ce qui conduit à prendre la solution sous la forme

$$\Phi=\frac{\Phi_1+\Phi_2}{2}+\int_0^\infty A(k)\cdot J_0(kr)\,\sin kz\cdot dk,$$

et l'on détermine $A(k)$ en écrivant que pour $r=1$, la fonction est égale à Φ_1 lorsque z est négatif et à Φ_2 lorsque z est positif. Les théorèmes classiques de l'intégrale de Fourier conduisent à:

$$A(k)=\frac{2}{\pi}\,(\Phi_2-\Phi_1)\,\frac{1}{2kI_0(k)}$$

et finalement

$$\Phi=\frac{\Phi_2+\Phi_1}{2}\left[1+\frac{1-\gamma}{1+\gamma}\,H(z)\right]\qquad H(z)=\frac{2}{\pi}\int_0^\infty\frac{\sin kz}{kI_0(k)}\,dk.$$

Il est curieux de constater que H est très bien représentée par la fonction simple th ωz avec $\omega=1{,}318$. Ce résultat, signalé par Gray [1][2] nous permet d'écrire

$$\Phi=\frac{\Phi_1+\Phi_2}{2}\left[1+\frac{1-\gamma}{1+\gamma}\,\mathrm{th}\;\omega z\right],$$

expression simple et pourtant approchée à moins de 1%.

Mais il n'est quand même pas possible d'intégrer l'équation de Gauss avec cette expression du potentiel, sinon par les moyens de l'analyse numérique.

---

[1] Laboratoire de Radioélectricité de l'Ecole Normale Supérieure, Paris, France.

[2] Figures in brackets indicate the literature references on p. 211.

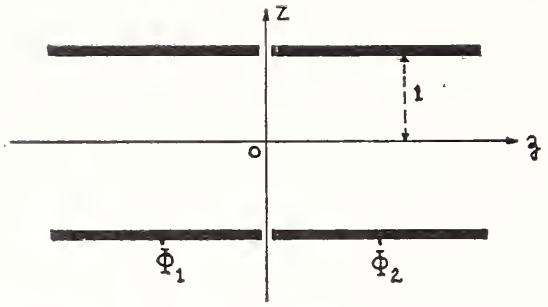


FIGURE 19.1. *Allure des electrodes.*

Nous utiliserons dans la suite, l'équation réduite de Picht [2]

$$R''+\frac{3}{16}T^2R=0, \qquad \text{avec } R=r\Phi^{\frac{1}{2}};\ T=\frac{1}{\Phi}\frac{d\Phi}{dz}.$$

Nous sommes donc conduits à étudier la fonction $T=\Phi'/\Phi$ qui sera la caractéristique de la lentille.

Nous posons $2\omega z=\log t$, ce qui nous donne

$$T=\frac{2\omega(1-\gamma)t}{(1+t)(\gamma+t)};$$

il apparaît alors un maximum pour $t_m=\sqrt{\gamma}$

$$z_m=\frac{1}{4\omega}\log\gamma, \qquad T_m=-2\omega \text{ th } \omega z_m.$$

Nous allons translater l'axe des ordonnées de façon à annuler l'abscisse du maximum, ce qui conduit à poser

$$Z=z-z_m, \qquad s=e^{2\omega Z}, \qquad T=yT_m,$$

après quelques transformations algébriques simples, on obtient une équation réduite donnant les deux valeurs de s qui correspondent à une valeur de y,

$$s^2-2s\left[\frac{(1-\sqrt{\gamma})^2}{2y\sqrt{\gamma}}-\frac{1+\gamma}{1-\sqrt{\gamma}}\right]+1=0,$$

ce qui donne, pour les abscisses z,

$$\text{ch } 2\omega Z=\frac{(1+\sqrt{\gamma})^2}{2y\sqrt{\gamma}}-\frac{1+\gamma}{2\sqrt{\gamma}}$$

et comme la fonction ch $2\omega Z$ est paire, cette formule montre que la courbe représentative de T est symétrique par rapport à l'ordonnée passant par le maximum.

Ainsi la dissymétrie de la lentille réelle a disparu dans la caractéristique. La courbe $T=\Phi'/\Phi$ possède encore un axe de symétrie mais celui-ci est décalé par rapport au plan de symétrie des cylindres. La figure 19.2 illustre ces résultats.

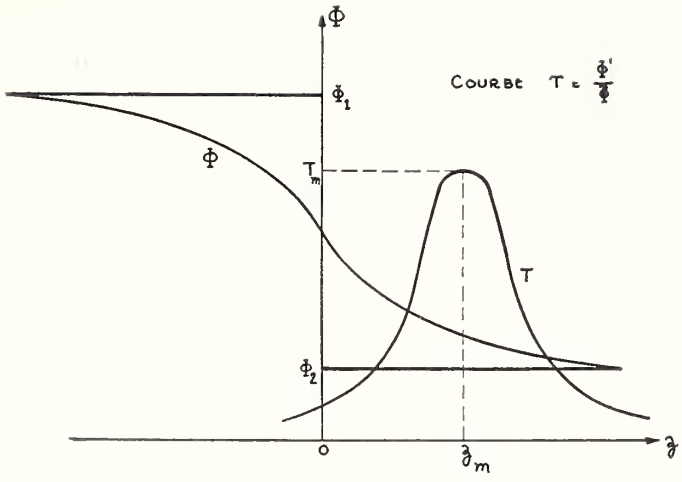


FIGURE 19.2. *Potentiel axial.*

Représentation Approchée du Potentiel

La courbe en T est une courbe en cloche; si nous comparons l'équation de Picht et l'équation de Gauss dans une lentille magnétique où le champ est celui de Glaser [3]

$$r''+\frac{e}{z_m\Phi_0}\frac{H_0^2}{\left[1+\left(\frac{z}{a}\right)^2\right]^2}r=0, \qquad R''+\frac{3}{16}T^2R=0,$$

on constate que si l'on représente la fonction $T=\Phi'/\Phi$ avec une fonction de la forme

$$T_a=\frac{T_0}{1+\left(\frac{z}{a}\right)^2},$$

on pourra représenter les trajectoires par des fonctions simples et conduire à son terme l'étude théorique de la lentille sans rencontrer de calculs trop complexes.

Il nous faut donc déterminer les deux paramètres T_0 et a, ainsi que l'origine des coordonnées de façon à représenter la fonction T le mieux possible.

Nous ferons tout d'abord coïncider le maximum de la courbe réelle et de la courbe approchée; nous aurons donc

$$T_0=-2\omega \text{ th } \gamma, \qquad \gamma=\tfrac{1}{4}\log\gamma,$$

l'origine des nouveaux axes étant à la distance $z_0=\frac{1}{4\omega}\log\gamma$ du plan de symétrie.

Quant à a, nous pourrons tout d'abord le choisir de façon que la demi-largeur de la courbe exacte et de la courbe approchée soit la même ce qui donne

$$y=\frac{1}{2}, \qquad \text{ch } 2\omega Z_1=2+\frac{1}{2}\left(\sqrt{\gamma}+\frac{1}{\sqrt{\gamma}}\right),$$

et par suite

$$a_1=\frac{1}{2\omega}\arg\text{ch}\left[2+\frac{1}{2}\left(\sqrt{\gamma}+\frac{1}{\sqrt{\gamma}}\right)\right].$$

Mais une meilleure approximation s'obtiendra avec une condition plus physique: On choisira a de façon que l'intégrale

$$\int_{-\infty}^{\infty} (T^2 - T_a^2) dz = 0$$

soit nulle. En effet, si la lentille est mince, la convergence est donnée par

$$\frac{1}{f} = \frac{3}{16} \int_{-\infty}^{+\infty} T^2 dz + \ldots, (\text{termes d'ordre supérieur}).$$

La série, utilisée par Goddard, est très rapidement convergente, de sorte qu'en choisissant les paramètres de façon à rendre égaux les premiers termes on égalise pratiquement les distances focales. La valeur de a ainsi calculée sera donc certainement bien supérieure à celle obtenue en faisant coïncider les courbes à mi-hauteur.

Nous avons sans difficulté, en posant $z = a \operatorname{ctg} \varphi$

$$\int_{-\infty}^{+\infty} T_a^2 dz = T_0^2 \int_{-\infty}^{+\infty} \frac{dz}{\left[1 + \left(\frac{z}{a}\right)^2\right]^2} = -T_0^2 a \int_{\eta}^{0} \sin^2 \varphi d\varphi = \frac{\pi}{2} T_0^2 a.$$

Le calcul de l'autre intégrale est plus compliqué mais Goddard [4] a donné le résultat. Nous l'avons vérifié en calculant par deux méthodes différentes: Tout d'abord en posant $\operatorname{th} \omega z = t$, ensuite en posant $e^{2\omega z} = u$. On trouve

$$\int_{-\infty}^{+\infty} T^2 dz = +\omega \left[\frac{\gamma + 1}{\gamma - 1} \log \gamma - 2\right];$$

finalement la seconde approximation donne, en remplaçant T_0 par sa valeur

$$a_2 = \frac{1}{\pi \omega} \left(\frac{1 + \sqrt{\gamma}}{1 - \sqrt{\gamma}}\right)^2 \left[\frac{\gamma + 1}{\gamma - 1} \log \gamma - 2\right].$$

La comparaison avec l'expérience permettra de choisir entre a_1 et a_2 et l'on verra plus tard que c'est la seconde hypothèse qui conduit au meilleur résultat, comme les considérations précédentes l'avaient laissé prévoir.

Élements du Premier Ordre

Nous pouvons intégrer l'équation de Picht en remplaçant T par sa valeur: On obtient une représentation simple pur la trajectoire réduite.

$$z = a \operatorname{ctg} \varphi, \qquad R = A \frac{\sin K(\varphi + B)}{\sin \varphi}, \qquad K = \sqrt{1 + \frac{3}{16} T_0^2 a^2},$$

si γ est voisin de 1, on peut poser $\sqrt{\gamma}=1+\epsilon$; il vient alors

$$K=1+0{,}0675\,\frac{\epsilon^2}{\sqrt{\gamma}}, \qquad \text{et} \qquad a=0{,}644\left(1+\frac{\epsilon^2}{10}\right).$$

Les distances focales sont données par les formules classiques

$$f_0=\frac{R(+\infty)}{R'(-\infty)}\left(\frac{\phi_2}{\phi_1}\right)^{-1/4}, \qquad f_i=\frac{R(-\infty)}{R'(+\infty)}\left(\frac{\phi_2}{\phi_1}\right)^{1/4},$$

qui vérifient bien la relation de Lagrange–Helmoltz.

$$f_i/f_0=-[\phi_2/\Phi_1]^{1/2}.$$

Avec les résultats précédents on a

$$f_i=-\frac{ak\gamma^{-1/4}}{\sin\,(K\pi)}, \qquad f_0=\frac{aK\gamma^{1/4}}{\sin\,(K\pi)}.$$

Si l'on peut négliger ϵ^4 on obtient les formules simples

$$f_i=3{,}036\,\frac{\gamma^{1/4}}{\epsilon^2}, \qquad f_0=-3{,}036\,\frac{\gamma^{-3/4}}{\epsilon^2}.$$

Pour obtenir les abscisses des foyers il faut faire un développement limité de la trajectoire réelle au voisinage de l'infini. Ce développement donnera l'équation de l'asymptote c'est-à-dire le "rayon" dans l'espace image.

Après quelques calculs on obtient

$$z_{F_i}=aK\ \mathrm{ctg}\ K\pi-\frac{T_m a^2}{4}+z_0;$$

en recommençant le même calcul du côté objet on obtient

$$z_{F_0}=-aK\ \mathrm{ctg}\ K\pi-\frac{T_m a^2}{4}+z_0.$$

Nous sommes donc capables de calculer tous les éléments du premier ordre de la lentille.

Rèsultats Numeriques

Nous avons calculé la distance focale image f_i et l'abscisse du plan principal image z_{F_1} pour des valeurs de γ allant de 1 à 10. Les courbes de la figure 19.3 illustrent ces résultats qui sont donnés par le tableau ci-joint. Les points expérimentaux de Spangenberg [5] montrent que les meilleurs valeurs de a sont celles données par la seconde hypothèse (a_2) que nous conserverons désormais.

Le tableau contient aussi les résultats calculés par Goddard [4]. L'accord est bon, pour les distances focales, il est moins correct pour les abscisses des plans principaux.

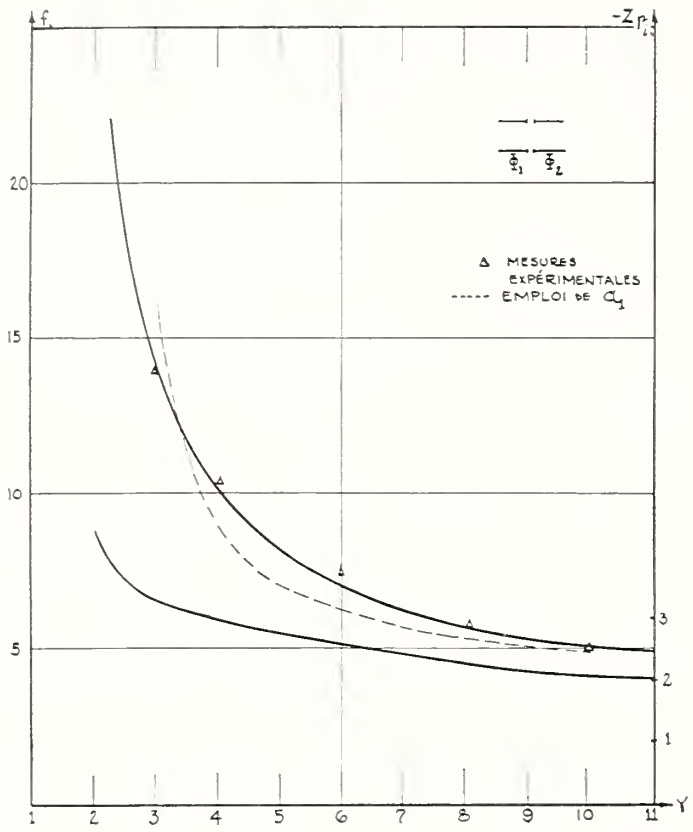


FIGURE 19.3.

TABLEAU 19.1. *Valeurs numériques de f_i et de Z_{P_i}*

γ	2	3	4	5	6
f_i:					
Bernard-Grivet [1]	30, 4	13, 9	9, 88	8, 13	6, 99
Goddard [2]	30, 5	13, 8	9, 57	7, 71	6, 65
Z_{P_i}:					
Bernard-Grivet	−4, 4	−3, 3	−2, 95	−2, 79	−2, 67
Goddard	−5, 0	−3, 55	−3, 11	−2, 93	−2, 83
	7	8	9	10	11
f_i:					
Bernard-Grivet	6, 20	5, 81	5, 46	5, 09	4, 89
Goddard	5, 98	5, 53	5, 20	4, 95	4, 75
Z_{P_i}:					
Bernard-Grivet	−2, 45	−2, 30	−2, 18	−2, 11	−2, 07
Goddard	−2, 75	−2, 70	−2, 65	−2, 62	

[1] Correspondant aux résultats de Bernard-Grivet calculés par la méthode de l'article.
[2] Correspondant aux résultats de Goddard.

Lentille à Coupure Large et Lentille à Deux Diaphragmes

En supposant que le potentiel varie linéairement le long de la coupure, depuis Φ_1 jusqu'à Φ_2 on peut trouver le potentiel axial par une méthode analogue à celle du texte. Après quelques calculs, on obtient [8]

$$\Phi = \frac{\Phi_1 + \Phi_2}{2}\left[1 + \frac{1-\gamma}{1+\gamma}\log\frac{\operatorname{ch}\,\omega(z+d)}{\operatorname{ch}\,\omega(z-d)}\right]$$

où d est l'épaisseur de la coupure.

Tant que d est inférieur à 0,5, cette fonction compliquée est bien représentée par

$$\Phi=\frac{\Phi_1+\Phi_2}{2}\left[1+\frac{1-\gamma}{1+\gamma}\,\mathrm{th}\;\omega' z\right]$$

avec $\omega'=\mathrm{th}\;\omega d/d$ et $\omega=1{,}318$ la précision est de 2% environ. Les formules du texte sont donc valables, il suffit de remplacer ω par ω'.

Si d est plus grand que 0,5, l'approximation précédente n'est plus bonne. Il faut refaire le calcul de T avec la valeur exacte du potentiel. Les formules sont plus compliquées mais l'allure générale de la courbe est conservée. L'analyse détaillée de ce cas sera publiée incessamment.

Enfin la lentille constituée par deux diaphragmes plans, distants de d, portés aux potentiels Φ_1 et Φ_2, percés de trous circulaires de rayon unité relève encore de la même méthode, car le potentiel est encore représenté par

$$\phi=\frac{\Phi_1+\Phi_2}{2}\left[1+\frac{1-\gamma}{1+\gamma}\log\frac{\mathrm{ch}\;\omega(z+d)}{\mathrm{ch}\;\omega(z-d)}\right].$$

La fidélité de cette représentation est un peu moins bonne que dans le cas précédent et sera discutée dans la publication détaillée des résultats.

Conclusion

Nous sommes donc arrivés par des calculs beaucoup plus simples que ceux de Goddard à des résultats au moins aussi précis sous forme d'expressions très maniables.

D'autre part, l'obtention d'une représentation paramétrique simple pour le rayon, rend possible le calcul des abberrations en utilisant les formules nouvelles de Sturrock [6].

Les résultats de Glaser et de Glaser & Lammel [7] peuvent s'étendre aux lentilles électriques et rendre des services en tous points comparables.

Enfin, l'exactitude des résultats peut être considérablement augmentée en représentant $T(z)$ par sech (z/b) ainsi qu'il est exposé dans la communication de P. Grivet.

Bibliographie

[1] T. Gray, Bell Syst. Tech. J. **28,** 25 (1939).
[2] T. Picht, Ann. Phys. Lpz **15,** 926 (1932).
[3] W. Glaser, Z. Phys. **117,** 285–315 (1941).
[4] S. Goddard, Proc. Phys. Soc. **61,** 372–396 (1944).
[5] K. Spangenberg, Elec. Comm. **21,** 194 (1943).
[6] P. A. Sturrock, Compt. Rend. **233,** 146–147, 243–245 (1951).
[7] W. Glaser et Lammel, Arch. Elec. **37,** 347–356 (1943).
[8] S. Bertram, J. Appl. Phys. **13,** 496–502 (1942).

20. Etude Théorique des Lentilles à Grille à l'Aide d'Une Représentation Approchée du Potentiel

Par Michel Bernard [1]

1. Introduction

11—Les lentilles électroniques classiques présentent la propriété bien connue d'être à la fois convergentes en certains endroits et divergentes en certains autres; cela est dû au fait que la composante radiale de la force qui agit sur la particule, proportionnelle à Φ'', change de signe à chaque point d'inflexion de la courbe $\Phi(z)$.

Cela crée certains inconvénients dont les deux principaux sont: (1) La difficulté d'obtenir de très fortes convergences puisque il y a obligatoirement des parties divergentes dans le montage, et (2) l'impossibilité de réaliser des lentilles divergentes avec des potentiels moyens, car on démontre que les parties convergentes sont prépondérantes.

Il est donc intéressant de chercher à réaliser des lentilles qui n'auraient pas ces inconvénients; elles seront caractérisées par une courbe de potentiel dont la concavité est toujours tournée du même côté.

12—La plus simple de ces lentilles est constituée par un simple diaphragme, au potentiel Φ_0 séparant deux régions de *champ uniforme*. Elle présente un intérêt historique puisqu'elle fut la première étudiée par Davisson et Germer [2].

Ces lentilles sont intéressantes théoriquement car elles donnent un exemple simple de lentilles à rayons courbes; en effet les "rayons" de l'espace objet et de l'espace image sont des paraboles. Mais elles n'ont pas trouvé d'application jusqu'à présent, sans doute parce qu'elles utilisent mal le potentiel et ne sont que faibles. Aussi nous nous limiterons aux lentilles pour lesquelles l'espace objet et l'espace image sont des régions où le potentiel est constant.

13—La courbe de potentiel doit donc avoir une concavité toujours de même sens, et posséder des asymptotes horizontales; il est aisé de se rendre compte que ces conditions sont incompatibles si la courbe a une pente continue, il doit nécessairement y avoir des points anguleux.

La manière la plus simple de réaliser un champ discontinu sera de placer au milieu de la lentille, une grille assez transparente aux particules, portée à un potentiel plus haut ou plus bas que les potentiels extérieurs. Nous arrivons donc aux lentilles à grille mises en valeur par Cartan [3] dans son spectrographe de masse et étudiées à sa suite par Knoll & Weichart. [4] Le texte qui suit contient une étude

[1] Laboratoire de Radioélectricité de l'Ecole Normale Supérieure, Paris, France.
[2] Davisson et Germer, Phys. Rev. **38**, 585 (1931).
[3] Cartan, Jour. Phys. Rad. **8**, 111 (1937).
[4] Knoll et Weichart, Z. Phys. **110**, 233–236 (1938).

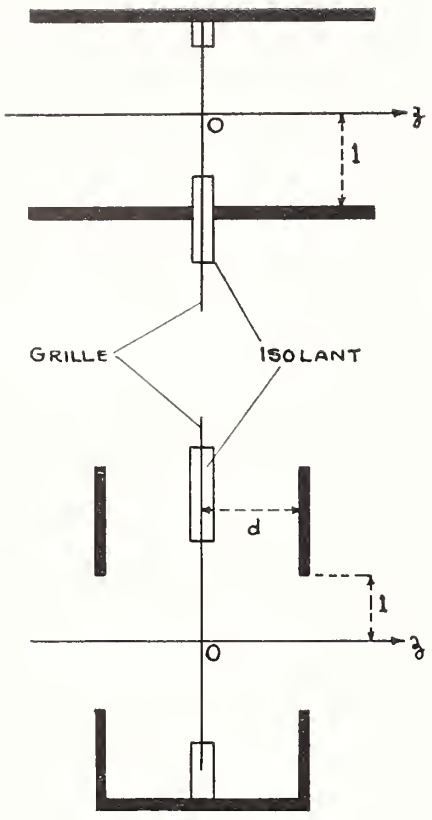


FIGURE 20.1. *Modèle de lentilles à grille.*

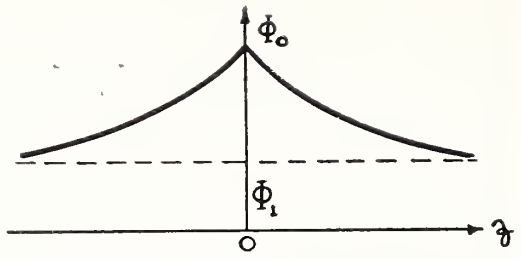


FIGURE 20.2. *Allure du potentiel* Φ [z].

précise de ces lentilles, menée jusqu'au bout grâce à un modèle simple de potentiel qui a permis l'intégration de l'équation des trajectoires.

14—Nous étudierons deux types de lentilles dont la figure 20.1 indique la forme des électrodes et la figure 20.2, l'allure du potentiel axial:

Lentilles à cylindres. Deux cylindres, de rayon unité, coaxiaux, sont portés au potentiel Φ_1. Ils sont séparés par la grille au potentiel Φ_0.

Lentilles à diaphragme. Deux électrodes planes, distantes de 2d sont portées au potentiel Φ_1; elles sont percées d'ouvertures circulaires de rayon unité. Entre les deux, à mi-distance se trouve la grille au potentiel Φ_0.

Le premier modèle facile à construire est moins souple que le second qui contient un paramètre supplémentaire. Nous verrons par la suite, qu'au point de vue électrique, ce n'est qu'un cas particulier. Comme la grille constitue un écran électrostatique nous n'étudierons que le potentiel pour $z>0$. La partie négative se déduira par symétrie.

2. Étude du Potentiel sur l'Axe

21—Lentille à cylindres

Nous intégrons l'équation de Laplace en séparant les variables; nous prenons la solution sous la forme

$$\Phi=\Phi_0+\int_0^\infty A(k)\,I_0(kr)\,\sin kz\cdot dk,$$

et nous déterminons $A(k)$ par la condition aux limites imposée, à savoir $\Phi=\Phi_1$ pour $r=1$ quel que soit z. Le calcul classique de l'inté-

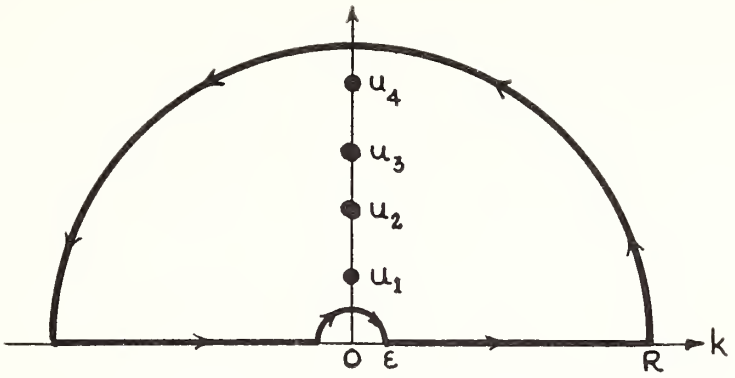


FIGURE 20.3. *Contour d'intégration.*

grale de Fourier conduit à

$$A(k)=\frac{2}{\pi}\,(\Phi_1-\Phi_0)\,\frac{1}{kI_0(k)}.$$

Nous poserons dorénavant $\sigma=\frac{\Phi_1}{\Phi_0}$; $\varphi=\frac{\Phi}{\Phi_0}$. Le potentiel s'écrira alors:

$$\varphi=1+(\sigma-1)F(z), \qquad F(x)=\frac{2}{\pi}\int_0^{\infty}\frac{\sin kz}{kI_0(k)}\,dk.$$

La fonction $F(z)$ est assez simple à calculer si l'on prend la précaution de la transformer auparavant en une série en intégrant par la méthode des résidus. Le contour d'intégration est indiqué par la figure 20.3. On intègre la fonction

$$\frac{2}{\pi}\,\frac{e^{jkz}}{kI_0(k)}$$

de $-\infty$ à $+\infty$ le long de l'axe réel, encontournant l'origine par un petit cercle de rayon ϵ, et l'on termine en intégrant le long du cercle de l'infini. La figure 20.3 donne la forme du contour d'intégration, ce qui conduit à

$$2\pi j\Sigma R=\frac{2}{\pi}\int_{-\infty}^{+\infty}\frac{e^{jkz}}{kI_0(k)}\,dk+\oint_{\epsilon}+\oint_{R}=2jF(z)+\int_{\epsilon\curvearrowleft}+\int_{R\curvearrowleft}$$

les pôles sont situés sur l'axe imaginaire et correspondent aux racines de $J_0(u)=0$. Les résidus s'obtiennent par l'emploi des formules classiques; enfin, on calcule l'intégrale le long du cercle de rayon ϵ en posant $k=\epsilon e^{j\varphi}$, on trouve $\int_{\epsilon\curvearrowleft}=-2j$, et celle du cercle de rayon R en posant $k=Re^{j\varphi}$, on trouve qu'elle tend vers 0 si R tend vers l'infini. Finalement

$$F(x)=1-2\sum_{n=1}^{n=\infty}\frac{e^{-\mu n^2}}{\mu_n J_1(\mu_n)}, \qquad J_0(\mu_n)=0.$$

Nous avons calculé les valeurs de cette fonction pour les principales valeurs de z (tableau 20.1) il est intéressant de remarquer que la fonction th ωz coïncide pratiquement avec F si l' on prend $\omega=1.315$. L'erreur ainsi commise n'atteint pas 1%.

Nous poserons donc dorénavant $\varphi=1+(\sigma-1)$ th ωz.

TABLEAU 20.1. *Valeurs numériques des fonctions $F(z)$ et th 1,315z*

Z	$F(z)$	th 1,315z
0	0	0
0, 25	0, 319	0, 318
0, 50	0, 575	0, 573
0, 75	0, 752	0, 756
1	0, 859	0, 866
1, 25	0, 922	0, 929
1, 50	0, 957	0, 963
1, 75	0, 977	0, 981
2	0, 988	0, 991
2, 25	0, 994	0, 996
2, 50	0, 997	0, 998

22—Lentille à diaphragmes

Nous faisons le même raisonnement qu'au paragraphe précédent mais les conditions aux limites sont plus difficiles à exploiter. Aussi ferons nous la même hypothèse que Bertram[5] dans un cas semblable: Nous posons que, pour $r=1$, le potentiel varie linéairement de Φ_0 à Φ_1, lorsque z varie de 0 à d, puis reste ensuite constant et égal à Φ_1,

$$\Phi=\Phi_0+(\Phi_1-\Phi_0)\frac{z}{d}, \qquad 0\leqq z\leqq d, \qquad \Phi=\Phi_1, \qquad z\geqq d.$$

Les résultats qui s'en déduisent ont été bien vérifiés à la cuve rhéographique dans un cas analogue. Nous pouvons alors calculer $A(k)$ comme au paragraphe 21; on trouve

$$A(k)=\frac{2}{\pi}(\phi_1-\phi_0)\frac{\sin kd}{dk^2I_0(k)}$$

de sorte que le potentiel sur l'axe va s'écrire, avec les mêmes notations réduites

$$\varphi=1+(\sigma-1)G(z,d), \qquad G(z,d)=\frac{2}{\pi}\int_0^\infty \frac{\sin kd.\sin kz}{dk^2I_0(k)}\,dk.$$

La fonction $G(z,d)$ ne va pas nécessiter de nouveaux calculs car il est facile de la transformer en introduisant la fonction $F(z)$. On a

$$G(z,d)=\frac{2}{\pi d}\int_0^\infty \frac{\cos k(z-d)}{2k^2I_0(k)}\,dk-\frac{2}{\pi d}\int_0^\infty \frac{\cos k(z+d)}{2k^2I_0(k)}\,dk,$$

et si l'on pose

$$\psi(\alpha)=\frac{2}{\pi}\int -\frac{\cos k\alpha}{k^2I_0(k)}\,dk,$$

on est conduit à

$$G(z,d)=\frac{1}{2}[\psi(\alpha)]_{z-d}^{z+d},$$

mais il est évident que la fonction Ψ n'est qu'une primitive de F que

[5] S. Bertram, Proc. Inst. Radio Engrs. **28**, 418–420 (1940).

nous garderons par la suite

$$G(z,d)=\frac{1}{2\omega d}\log\frac{\operatorname{ch}\omega(z+d)}{\operatorname{ch}\omega(z-d)}.$$

Si d tend vers 0, on montrera facilement que G tend vers F si bien que la lentille à cylindres n'est qu'un cas particulier de la lentille à diaphragmes, et à partir de maintenant, c'est sur cette dernière que nous raisonnerons.

23—Étude de la courbe $T=\Phi'/\Phi$

Nous prendrons l'équation de Gauss sous la forme de Picht [6] il nous faut donc étudier la fonction caractéristique Φ'/Φ. Il y a deux cas à distinguer suivant que la lentille est convergente ($\sigma<1$) ou divergente ($\sigma>1$):

Si la lentille est convergente, la courbe possède un extremum dans sa partie utile, c'est-à-dire pour z positif, dont l'abscisse est donnée par la racine de l'équation

$$\frac{1}{2\omega d}\log\frac{\operatorname{ch}\omega(z+d)}{\operatorname{ch}\omega(z-d)}=\frac{1}{1-\sigma}-\frac{\operatorname{sh}2\omega d}{2\omega d}\cdot\frac{1}{\operatorname{sh}2\omega z}$$

qui peut être résolue graphiquement.

L'ordonnée du maximum est

$$T_m=-\omega\frac{\operatorname{sh}2\omega z_m}{\operatorname{ch}\omega(z_m+d).\operatorname{ch}\omega(z_m-d)}$$

et enfin l'ordonnée à l'origine veut

$$T_1=(\sigma-1)\frac{\operatorname{th}\omega d}{d}$$

Quant à la partie de la courbe valable pour z négatif il est évident qu'elle est symétrique de la précédente par rapport à 0. La figure 20.4

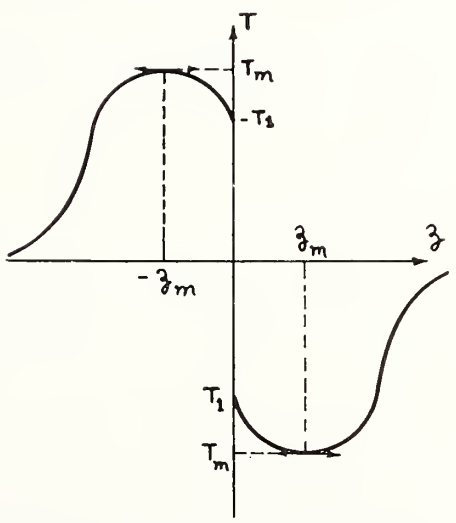


FIGURE 20.4. *Caracteristiques des lentilles* $T=\Phi'/\Phi$
Lentille convergente.

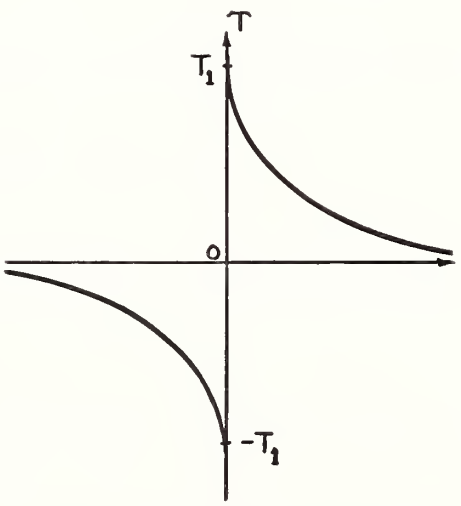


FIGURE 20.5. *Caracterisiiques des lentilles* $T=\Phi'/\Phi$.
Lentille divergente.

[6] J. Picht, Ann. Phys. Leipzig **15**, 926 (1932).

indique l'aspect de cette courbe.

Si la lentille est divergente, il n'y a plus d'extremum dans la partie utile de la courbe, celle-ci est constamment décroissante. L'ordonnée à l'origine vaut toujour $T_1=(\sigma-1)\,th\,\omega d/d$ et l'aire limitée par la courbe et l'axe des z est égale à: Log σ. L'autre moitié de la courbe s'obtient toujours par symétrie. La figure 20.5 donne l'allure de la courbe T dans le cas de la lentille divergente.

Signalons pour terminer que la lentille à cylindres est régie par des équations plus simples, obtenues en faisant tendre d vers 0 dans les relations précédentes. On trouve les valeurs

$$z_m=\frac{1}{4\omega}\log\left(\frac{2-\sigma}{\sigma}\right),\qquad T_m=-2\omega\ \mathrm{th}\ \omega z_n,\qquad T_1=(\sigma-1)\omega.$$

3. Représentation Approchée de T Par Une Courbe de Glaser

31—Nécessité d'une représentation du potentiel.—Aussi simples que soient les expressions précédentes elles sont cependant trop complexes pour permettre de tenter l'intégration de l'équation de Gauss avec quelques chances de succès. Aussi allons-nous représenter chacune des moitiés de la courbe par une cloche de Glaser,

$$T=T_0\left[1+\left(\frac{z-z_0}{2}\right)^2\right]^{-1},$$

dont une partie seulement sera valable. Une telle fonction permet d'intégrer l'équation de Gauss, et la forme de sa courbe représentative est très analogue à celle de la fonction préalablement étudiée. Il nous faut déterminer les paramètres T_0, z_0, a, de façon à obtenir le moins d'écart possible.

32—Lentilles convergentes

Nous faisons coïncider le maximum et l'ordonnée à l'origine. Ces conditions nous donnent: $T_0=T_m$, $z_0=z_m$, $a=z_0(T_0/T_1-1)^{-1}$, pour la moitié positive et $T_0'=-T_0$, $z_0'=-z_0$, $a'=a$. pour la moitié négative. Pour chaque valeur de σ plus petite que 1 il est ainsi possible de trouver une représentation de T qui soit assez proche de la valeur réelle et qui permette l'intégration par des fonctions simples.

33—Lentilles Divergentes

Ici les conditions à remplir sont moins impératives, car il n'y a plus d'extremum. Nous faisons toujours coïncider l'ordonnée à l'origine et nous choisissons les autres paramètres de façon que l'aire limitée par la courbe et l'axe des z soit égale à celle limitée par la courbe réelle.

Cela ne fait que deux équations pour trois paramètres; il nous faut prendre une troisième condition la plus simple possible; nous prendrons $z_0=0$ ce qui conduit à

$$z_0=0,\qquad T_0=(\sigma-1)\,\frac{\mathrm{th}\ \omega d}{d},\qquad a=\frac{2}{\pi T_0}\log\sigma.$$

Il est évident que la troisième condition est artificielle et que l'on pourrait la remplacer par d'autres plus physiques; mais il ne semble pas que la précision de la représentation soit beaucoup augmentée, tandis que les calculs sont passablement plus complexes. La partie négative de la courbe s'obtient toujours par symétrie.

4. Éléments du Premier Ordre

41—Les rayons de l'espace objet et de l'espace image sont les asymptotes aux trajectoires. En effet, on ne peut rien immerger dans un champ électrostatique, sous peine de modifier le potentiel et par suite de perturber la lentille. Tant que la trajectoire a un rayon de courbure non négligeable, nous sommes à "l'interieur" de la lentille, espace où il n'est pas possible de pénétrer.

Les objets et les images seront définis par les points de rencontre des asymptotes. Ils seront réels si ils sont à l'extérieur de la lentille car trajectoire et asymptote sont confondues; ils seront virtuels dans le cas contraire.

42—L'équation de Gauss s'intègre dans chacune des moitiés de la lentille et fournit pour la trajectoire la représentation:

$$\left.\begin{aligned} R&=A'\frac{\sin K(\phi+B')}{\sin\phi}\\ z-z_0'&=a\ \mathrm{ctg}\ \phi\end{aligned}\right\}-\infty<z<0 \qquad \left.\begin{aligned} R&=A\frac{\sin K(\phi+B)}{\sin\phi}\\ z-z_0&=a\ \mathrm{ctg}\ \phi\end{aligned}\right\}0<z<+\infty$$

K est donné par $K=\sqrt{1+(3/16)T_0^2a^2}$. A' et B' sont déterminées par les conditions initiales de la particule. Quant aux constantes A et B elles seront déterminées par les conditions qui régleront la traversée de la grille par le corpuscule. Au premier ordre, nous admettrons que la trajectoire ne subit pas de déviation sensible lorsqu'elle traverse la grille. Nous avons alors

$$\frac{1}{R}\frac{dR}{dz}-\frac{1}{4}T=\text{constante}, \qquad R=\text{constante}$$

de part et d'autre de la grille.

Les conditions permettent de calculer la totalité de la trajectoire. Dans le cas particulier où nous sommes, elles s'écrivent

$$A'\sin K(\pi-\alpha+B')=A\sin K(\alpha+B)$$

$$-K\ \mathrm{ctg}\ K(\alpha+B)=-K\ \mathrm{ctg}\ K(\pi-\alpha+B')-2\ \mathrm{ctg}\ \alpha+\frac{T_0a}{2}$$

est défini par $\mathrm{ctg}\ \alpha=-z_0/a$.

43—Les distances focales sont ici égales et de signe contraire. On les obtient par les formules classiques

$$\frac{1}{f_i}=-\frac{1}{f_0}=-\frac{R'(+\infty)}{R(-\infty)},$$

qui conduisent après calculs à

$$\frac{1}{f_i}=-\frac{1}{f_0}=-\frac{\sin^2 K\alpha}{K^2 a}\left[\frac{T_0 a}{2}+2(K\operatorname{ctg} K\alpha-\operatorname{ctg}\alpha)\right].$$

Pour avoir les abscisses des foyers, il faut faire un développement limité de la trajectoire au voisinage de l'infini. Si l'on a

$$r=\alpha z+\beta+\frac{\alpha}{z}+\ \ldots,$$

on en déduira

$$z_{F_i}=-\beta/\alpha.$$

Après calculs on obtient

$$z_{F_i}=-z_{F_0}=z_0-\frac{T_0 a^2}{4}+$$
$$f_i \sin^2 k\alpha\left[1+\frac{2}{K}\operatorname{ctg} K\alpha\cdot\operatorname{ctg}\ \alpha-\operatorname{ctg}^2 K\alpha-\frac{T_0 a}{2K}\operatorname{ctg} K\alpha\right].$$

Nous avons conduit les calculs numériques jusqu'au bout pour une lentille particulière, correspondant à $d=1$, pour laquelle on possède des valeurs expérimentales de f_i mesurée par Knoll et Weichart.[7] La figure 20.6, donne les deux courbes de f_i^{-1} et de Z_{Fi} en fonction de σ pour les lentilles convergentes, calculées par nos formules. On voit que les points expérimentaux se placent dessus. Le tableau 20.2 qui contient les valeurs calculées et les valeurs mesurées permettra de rendre compte de la précision de la méthode. Le graphique cor-

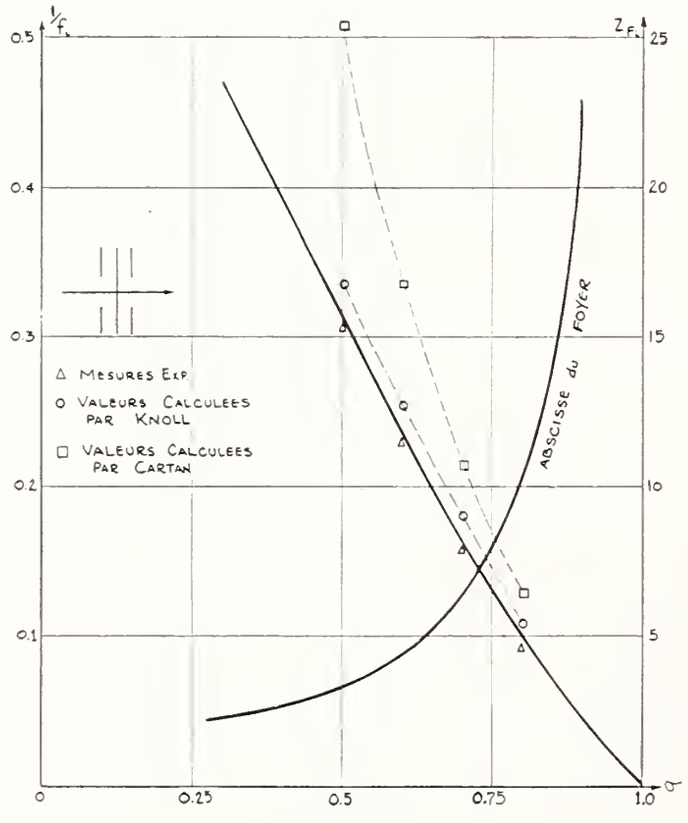


Figure 20.6.

[7] See footnote 4.

TABLEAU 20.2. *Valeurs numériques des éléments du premier ordre d'une lentille à grille avec d=1*

Les valeurs expérimentales proviennent de mesures faites par Knoll et Weichart; elles ne s'étendent que sur l'intervalle 0,8 à 0,5

σ	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2
$1/f_{cal}$	0, 044	0, 097	0, 162	0, 233	0, 311	0, 392	0, 470	0, 515
$1/f_{mes}$		0, 095	0, 162	0, 233	0, 310			
z_{F_i}	22, 8	10, 3	6, 23	4, 38	3, 30	2, 66	2, 23	2, 05

tient aussi les valeurs calculées par Knoll et celles calculées par Cartan;[8] l'accord est moins bon, mais il ne faut pas oublier que leurs calculs sont moins longs et leurs formules plus maniables.

44—Nous n'avons considéré que des lentilles à électrodes symétriques par rapport à la grille. La théorie précédente s'étend sans difficulté au cas des électrodes quelconques; seulement les formules sont beaucoup plus compliquées puisque les coefficients sont entièrement différents de part et d'autre de la grille.

Nous ne donnerons pas les formules correspondantes car il est impossible d'en tirer rapidement des résultats numériques relatifs à une classe de lentilles. Elles ne présentent d'intérêt que dans l'étude d'un cas particulier.

Par contre, revenant au cas de la symétrie, nous étudierons rapidement les simplifications qui peuvent être apportées aux formules si K est voisin de 1 (lentille faible).

En négligeant le produit $T_0'^2 a^2$ on a $K=1$ et la distance focale devient

$$\frac{1}{f}=\frac{\phi_1-\phi_0}{\phi_0}\,\frac{\operatorname{th}\omega\, d}{2d}$$

formule qui est analogue à celle proposée par Cartan. Si l'on veut tenir compte de $T_0'^2 a^2$ mais si l'on néglige $T_0'^4 a^4$ en posant $\sqrt{1+\epsilon}=1+\frac{\epsilon}{2}$ on obtient des formules aussi complexes que les formules primitives.

5. Conclusion

Nous avons donc ainsi obtenu des formules précises et cependant assez maniables, pour calculer à priori les éléments du premier ordre. Indépendamment de l'intérêt spécial des lentilles à grille, il est intéressant de remarquer que le modèle de Glaser est aussi utile en électrostatique qu'en magnétisme et que grâce à l'équation réduite de Picht il joue un rôle exactement symétrique.

[8] See footnote 3.

21. Note sur la Théorie de la Lentille Electrostatique à Electrode Centrale Elliptique

Par Edouard Regenstreif [1]

1. Introduction

L'influence de l'ovalisation du champ magnétique ou électrique sur le pouvoir résolvant de la lentille électronique a été mise en évidence pour la première fois par les expériences de Hillier [1][2]. Glaser [2] a attiré l'attention des théoriciens sur ce problème et les travaux de Bertein [3] ont fourni les bases d'une compréhension solide des phénomènes en jeu en même temps que des propositions pratiques sur les moyens extérieurs à mettre en oeuvre pour combattre les effets néfastes des dissymétries axiales.

Cependant, on n'était pas encore parvenu à établir des formules finies et explicites donnant directement les propriétés optiques fondamentales de la lentille elliptique en fonction de sa structure géométrique et électrique. D'un autre côté on ne s'était occupé que très peu des possibilités offertes par les relations entre les paramètres internes de la lentille pour agir sur son fonctionnement et amener celur-ci à l'optimum compatible avec une ellipticité donnée.

La présente Note résume un travail en cours de publication et dont le but est de fournir d'abord des relations explicites pour les caractéristiques optiques de la lentille elliptique et de préciser ensuite les conditions optima de fonctionnement.

2. Bases du Calcul

Considérons d'abord une lentille électrostatique indépendante, formée de trois électrodes planes, minces, parallèles et percées de trous rigoureusement circulaires et coaxiaux.

Les paramètres mécaniques de la lentille sont z_0, distance entre les électrodes extérieures et l'électrode centrale et R_1 et R_2, rayons des trous des électrodes. Les paramètres électriques sont V_1 et V_2, tensions appliquées aux électrodes centrale et extérieures.

La théorie de la lentille ronde, établie antérieurement [4] fournit alors pour la trajectoire de Gauss d'un rayon initialement parallèle à l'axe, des expressions de la forme:

$$\left.\begin{aligned} r &= r_0 \operatorname{ch}\left[\frac{2}{\sqrt{2}} \operatorname{arc\,sin}\, \beta(z+z_0)\right] && \text{dans la région d'incidence} \\ r &= A \cos\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sh}\, \alpha z - \varphi\right] && \text{dans la région centrale} \\ r &= B \operatorname{sh}^{\operatorname{ch}}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\, \beta(z-z_0) - \psi\right] && \text{dans la région d'émergence} \end{aligned}\right\} \quad (1)$$

[1] École Normale Supérieure, Paris, France.
[2] Figures in brackets indicate the literature references on p. 230.

avec

$$\alpha=\frac{1}{z_0}\sqrt{\frac{1}{k}\,\frac{1-x}{x}},\quad \beta=\frac{1}{z_0}\sqrt{\frac{1}{1-k}\,(1-x)},\quad x=\frac{\Phi(0)}{\Phi(z_0)}. \tag{2}$$

Dans ces relations k est un facteur numérique, déterminé par la position du point d'inflexion sur la courbe de potentiel (d'une manière usuelle $k\cong\frac{2}{3}$) et $\Phi(0)$ et $\Phi(z_0)$ sont les potentiels au centre de la lentille et à l'aplomb des électrodes extérieures. $\Phi(0)$ et $\Phi(z_0)$ sont calculables [4] à partir de z_0. R_1R_2 V_1 et V_2.

La manipulation mathématique des équations (1) permet de déterminer les caractéristiques optiques de la lentille, savoir la distance focale,

$$\left.\begin{aligned}&\frac{1}{f}=\frac{\beta}{\lambda\sqrt{2}}\,(\mathrm{ch}^2\,\beta_0+\lambda^2\,\mathrm{sh}^2\,\beta_0)\,\sin 2(\alpha_0-\gamma_0),\\ &\text{et l'abscisse du foyer gaussien,}\\ &\frac{z_F}{z_0}=1+\frac{f}{2\lambda z_0}\,4\lambda^2+(\lambda^2+1)\,\mathrm{sh}^2\,2\beta_0\,\sin 2(\alpha_0+\delta_0)\end{aligned}\right\} \tag{3}$$

avec

$$\alpha_0=\frac{1}{\sqrt{2}}\,\text{arc sh}\,\alpha k z_0,\qquad \beta_0=\frac{1}{\sqrt{2}}\,\text{arc sin}\,\beta(1-k)z_0,$$

$$\mathrm{tg}\,\gamma_0=\lambda\,\mathrm{th}\,\beta_0,\quad \mathrm{tg}\,2\delta_0=\frac{2\lambda}{\lambda^2-1}\,\frac{1}{\mathrm{th}\,2\beta_0}$$

et

$$\lambda=\sqrt{\frac{k}{1-k}}. \tag{4}$$

Considérons maintenant une lentille elliptique, d'ellipticité relative $\eta=\Delta R_1/R_1$. La distribution du potentiel s'écrit alors [5]

$$\phi(z,r,\theta)=\phi(z)-\frac{r^2}{4}\,\phi''(z)[1-\epsilon(z)\,\cos 2\theta]. \tag{5}$$

La fonction $\epsilon\,(z)$ a été étudiée notamment par Bertein [3] qui a trouvé que ϵ acquiert des valeurs importantes seulement dans la région centrale de la lentille où $\epsilon(0)\cong\eta$, et tend vers zéro au voisinage des électrodes extérieures. *L'hypothèse fondamentale* que nous faisons dans ce travail est de supposer ϵ constant dans la région centrale de la lentille et nul dans les régions d'incidence et d'émergence. Cette hypothèse peut se justifier aisément en pratique. En effet, la marche des trajectoires dans les régions extrêmes de la lentille est très peu affectée par modification de sa structure; par contre, il est indispensable de représenter correctement $\epsilon(z)$ au voisinage de $z=0$, où la trajectoire est très sensible à toute variation des paramètres géométriques et électriques de la lentille.

3. Équations Différentielles des Trajectoires Gaussiennes dans la Lentille Elliptique

En partant des bases ainsi exposées, les équations générales du mouvement de l'électron

$$\left.\begin{aligned} \frac{m}{e}(\ddot{r}-r\dot{\theta}^2)&=\frac{\partial\phi}{\partial r}\\ \frac{m}{e}\frac{d}{dt}(r^2\dot{\theta})&=\frac{\partial\phi}{\partial\theta}\\ \frac{m}{e}\ddot{z}&=\frac{\partial\phi}{\partial z}\end{aligned}\right\} \tag{6}$$

fournissent pour le régime gaussien de la lentille elliptique les relations

$$\left.\begin{aligned} \phi(z)r''+\frac{1}{2}\phi'(z)r'+\frac{1}{4}\phi''(z)r&=\left[\frac{\epsilon}{4}\phi''(z)\cos 2\theta+\phi(z)(\theta')^2\right]\\ r\Phi(z)\theta''+\left[2r'\phi(z)+\frac{r}{2}\phi'(z)\right]\theta'+\frac{\epsilon}{4}r\phi''(z)\sin 2\theta&=0\end{aligned}\right\} \tag{7}$$

qu'on peut écrire sous forme cartésienne

$$\left.\begin{aligned} \phi(z)r_x''+\tfrac{1}{2}\phi'(z)r_x'+\tfrac{1}{4}\phi''(z)(1-\epsilon)r_x&=0\\ \phi(z)r_y''+\tfrac{1}{2}\phi'(z)r_y'+\tfrac{1}{4}\phi''(z)(1+\epsilon)r_y&=0.\end{aligned}\right\} \tag{8}$$

Le système (7) permet de se faire immédiatement une idée qualitative de la marche du rayon de Gauss dans la lentille elliptique.

Si l'on considère un rayon initialement parallèle à l'axe; il cheminera dans la région d'incidence exactement comme dans une lentille ronde puisque ϵ est, par hypothèse, nul dans cette région. En particulier, l'électron restera constamment dans son méridien d'entrée θ_0. Dans la région centrale par contre, ϵ possède une valeur finie. Le rayon de Gauss subira alors une rotation autour de l'axe et quittera par conséquent son méridien initial. Dans la région d'émergence ϵ est à nouveau nul, mais à l'entrée dans cette région la vitesse de l'électron n'est plus situé dans un plan méridien. On aura donc à faire à une trajectoire gauche dans une lentille de révolution.

Toutes ces questions peuvent être précisées mathématiquement.

4. Trajectoires dans les Plans de Symetrie

Le système (7) montre que les solutions de la forme $\theta=$Cte sont données par $\sin\theta=0$, soit $\theta=\theta_0=K\frac{\pi}{2}$. En d'autres termes les trajectoires situées initialement dans les plans de symétries, et ces trajectoires seulement, restent planes.

En résolvant le système (7) ou (8) pour $\theta=0$ et $\theta=\pi/2$ on trouve pour les équations explicites des trajectoires situées dans les plans de symétrie de la lentille elliptique.

région d'incidence:

$$r(0)=r\left(\frac{\pi}{2}\right)=r_0 \operatorname{ch}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\ \beta(z+z_0)\right]$$

région centrale:

$$r(0)=A(0)\cos\left[\frac{\sqrt{1-\epsilon}}{2} \operatorname{arc\,sh}\ \alpha z-\varphi_x\right]; r\left(\frac{\pi}{2}\right)$$

$$=\left(A\frac{\pi}{2}\right)\cos\left[\frac{\sqrt{1+\epsilon}}{2} \operatorname{arc\,sh}\ \alpha z-\varphi_y\right]$$

région d'émergence:

$$r(0)=B(0)^{\operatorname{ch}}_{\operatorname{sh}}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\ \beta(z-z_0)-\psi_x\right]; r\left(\frac{\pi}{2}\right)$$

$$=B\left(\frac{\pi}{2}\right)^{\operatorname{ch}}_{\operatorname{sh}}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\ \beta(z-z_0)-\psi_y\right]. \qquad (9)$$

Les coefficients A à ψ peuvent être déterminés par des considérations de continuité.

5. Étude des Propriétés Focales

Le traitement mathématique des équations (9) permet de déterminer les caractéristiques optiques de la lentille elliptique.

On trouve pour les distances focales dans les deux plans de symétrie des formules généralisant (3_1)

$$\frac{1}{f_x}=\frac{\beta}{\lambda_x\sqrt{2}}(\operatorname{ch}^2\beta_0+\lambda_x^2\operatorname{sh}^2\beta_0)\sin 2(\alpha_{0X}-\gamma_{0X})$$

$$\frac{1}{f_y}=\frac{\beta}{\lambda_y\sqrt{2}}(\operatorname{ch}^2\beta_0+\lambda_y^2\operatorname{sh}^2\beta_0)\sin 2(\alpha_{0Y}-\gamma_{0Y}). \qquad (10)$$

Le calcul fournit pour la différence $\Delta f=f_X-f_Y$

$$\frac{\Delta_f}{f^2}=\frac{\beta\operatorname{ch}^2\beta_0}{\lambda\sqrt{2}\cos^2\gamma_0}[(2\alpha_0+\sin^2\gamma_0)\cos 2(\alpha_0-\gamma_0)$$

$$+\cos 2\gamma_0 \sin 2(\alpha_0-\gamma_0)]\epsilon+\text{termes en } \epsilon^2.$$

La figure 21.1 montre l'allure de la variation de Δf en fonction de x. Aux termes en ϵ^2, pres Δf est nul pour

$$tg\,2(\alpha_0-\gamma_0)=-\frac{2\alpha_0+\sin 2\gamma_0}{\cos 2\gamma_0} \qquad (11)$$

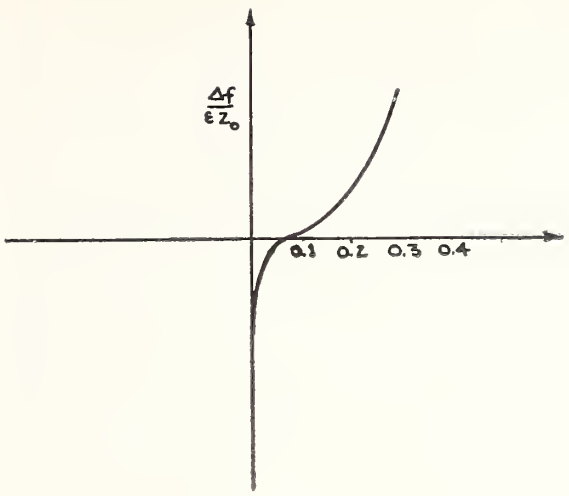


FIGURE 21.1.

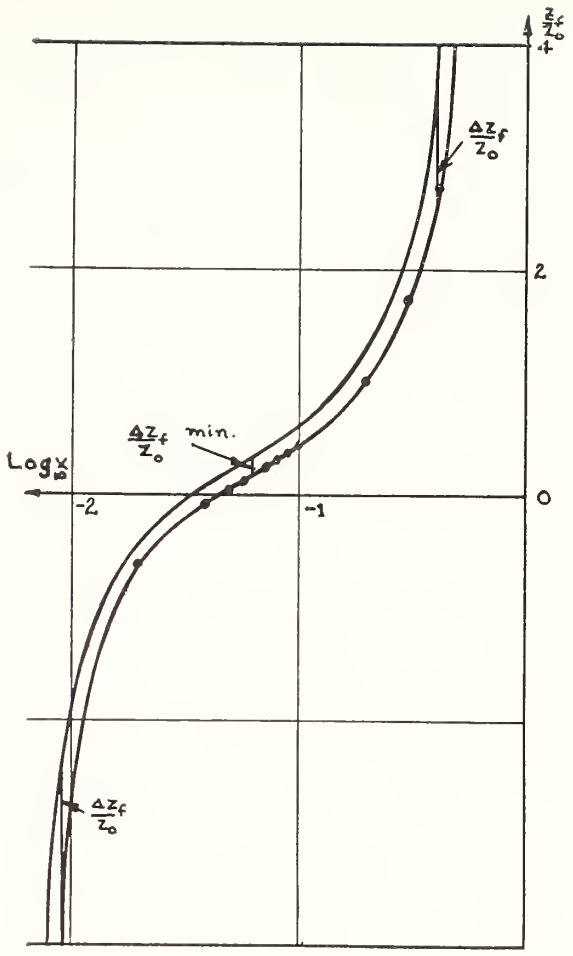


FIGURE 21.2.

ce qui fournit $x \simeq 5{,}8.10^{-2}$ *c'est la valeur même qui rend minimum la distance focale* [4] *de la lentille non perturbée.*

Le fait que Δf peut être nul en dépit d'une valeur finie de la différence entre les deux axes du trou central est dû à la forme bi-branche de la courbe focale $f=f(x)$ (réf [4]). Dans le fonctionnement habituel de la lentille le point figuratif est situé sur le branche droite de la courbe et la distance focale la plus grande correspond au rayon le plus fort. Par contre, sur la branche gauche de la courbe, la distance focale sera d'autant plus forte que le rayon du trou sera plus faible. On conçoit donc que pour une certaine valeur de x, il puisse y avoir égalité entre les deux distances focales. De l'égalité $\Delta f=0$, il ne faut pas conclure à la nullité des aberrations d'ellipticité. Elle montre seulement que pour une certaine valeur de x, la différence entre l'abscisse du foyer et celle du plan principal ($f=z_F-z_P$) est la même pour les deux plans de symétrie. Par contre, si la différence $\Delta z_P=z_{FX}-z_{FY}$, passe par un minimum, les aberrations d'ellipticité passeront bien par un minimum. La figure 21.2 montre que la différence Δz_F passe effectivement par un minimum dû essentiellement au caractère constamment décroissant de la fonction z_F. La question peut être traitée par le calcul; on trouve en généralisant (3_2) au cas elliptique et en explicitant la différence Δz_F, une expression de la forme $\Delta z_F=[A(x)f+B(x)\Delta f]\,\epsilon+$termes en ϵ^2. Δz_F est minimum pour $\Delta f=0$; il en résulte que le *minimum de la courbe focale* $f=f(x)$ constitue un point de fonctionnement remarquable qui rend effectivement *minima les aberrations d'ellipticite.*

On peut généraliser ce raisonnement et montrer que toutes les aberrations sont minima en ce point. Cette démonstration théorique a été corroborée par des preuves expérimentales dans le cas de la distorsion [6] et des aberrations de sphéricité [7].

6. Trajectoires Situées dans un Méridien Initial Quelconque

En explicitant le système [8] dans les trois régions de la lentille, on trouve pour les trajectoires de Gauss les équations explicites

Région d'incidence:

$$r_x = r_y = r_0 \operatorname{ch}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\, \beta(z+z_0)\right]$$

Région centrale:

$$\left.\begin{aligned} r_x &= A_x \cos\left[\sqrt{\frac{1-\epsilon}{2}} \operatorname{arc\,sh}\, \alpha z - \varphi_x\right]; \\ r_y &= A_Y \cos\left[\sqrt{\frac{1+\epsilon}{2}} \operatorname{arc\,sh}\, \alpha z - \varphi_y\right] \end{aligned}\right\} \qquad (12)$$

Région d'émergence:

$$\left.\begin{aligned} r_x &= B_x \overset{\operatorname{ch}}{\underset{\operatorname{sh}}{}}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\, \beta(z-z_0) - \psi_x\right]; \\ r_y &= B_y \overset{\operatorname{ch}}{\underset{\operatorname{sh}}{}}\left[\frac{1}{\sqrt{2}} \operatorname{arc\,sin}\, \beta(z-z_0)\psi_y\right]. \end{aligned}\right\}$$

Les coefficients A et ψ peuvent encore être déterminés par des considérations de continuité et les trajectoires peuvent ensuite être calculées numériquement en fonction de l'ellipticité relative ϵ, du paramètre x qui caractérise la structure de la lentille non perturbée et des conditions initiales r_0, θ_0.

En développant les équations (12) en fonction de ϵ, on peut mettre le rayon gaussien de la lentille ellipitique sous la forme

$$\left.\begin{aligned} r_x &= r_r(x,\ z)[1+\epsilon g(x,\ z)] \cos\theta_0 \\ r_y &= r_r(x,\ z)[1-\epsilon g(x,\ z)] \sin\theta_0 \end{aligned}\right\} \qquad (13)$$

où $r_r(x,\ z)$ représente le rayon gaussien de la lentille ronde calculé une fois pour toutes [4] et $g(x,\ z)$ est calculable numériquement à partir des caractéristiques de la lentille non perturbée.

Pour $r_r=0$ c'est-à-dire si l'on considère la région extérieure de la lentille au foyer gaussien de celle-ci, la quantité $G(x,z)=r_r(x,z)\ g(x,z)$ reste finie et $r(r_r=0)=G(x,z)$ représente le cercle de moindre confusion ou l'aberration transversale. Le rayon de ce cercle dépend de x et est calculable pour une lentille donnée; il est minimum en même temps que l'aberration longitudinale.

Les considérations qui précèdent montrent qu'il est possible de prédéterminer par le calcul le point de fonctionnement optimum c'est-à-dire celui où les aberrations d'ellipticité sont minima.

La vérification expérimentale de cette théorie est actuellement en cours au laboratoire du Professeur Pierre Grivet.

7. Quelques Considérations sur les Images Gaussiennes Fournies par une Lentille Elliptique Éclairée en Lumière Parallèle

Un faisceau parallèle, s'appuyant à l'entrée de la lentille sur un cercle de rayon r_0, possèdera encore une section circulaire dans la région d'incidence (en conformité avec notre hypothèse fondamentale) mais présentera une section elliptique dans toutes les autres régions. Les axes de cette ellipse sont entièrement déterminés par les relations que nous avons établies et calculables numériquement en fonction de la structure de la lentille. Les équations données montrent également que l'ellipse peut dégénérer en une droite focale $X=0$ ou en une droite focale $Y=0$ et les valeurs de z pour lesquelles ces dégénérescences ont lieu sont très voisines et situées de part et d'autre de la valeur de z qui correspond à l'intersection avec l'axe du rayon r_r. Les calculs que nous avons effectués permettent de déterminer numériquement la position des focales et la distance qui les sépare en fonction de x et de ϵ. Pour $r_r=0$, l'ellipse dégénère en cercle de moindre confusion dont nous avons calculé le rayon.

Si l'on considère maintenant une droite coupant l'axe Oz, son image sera en régime elliptique gaussien encore droite. Mais cette droite aura subi une rotation donnée par $\mathrm{tg}\theta=(1-2\epsilon g)\mathrm{tg}\theta_0$ et calculable numériquement.

Si la droite est parallèle à OX ou OY son image sera encore parallèle à OX et OY.

Un carré centré sur Oz et dont les côtés sont parallèles aux axes OX et OY et égaux à $2a$, aura pour image un rectangle dont les côtés sont donnés par des expressions de la forme $2ah(x,z)[1+\epsilon g(x,z)]$ et $2ah(x,z)[1-\epsilon g(x,z)]$ calculables numériquement.

Enfin, une droite ne coupant pas l'axe de la lentille et ayant pour équation $r_0=d/[\cos(\gamma_0+\theta_0)]$ aura pour image une courbe d'équations $\mathrm{tg}\ \theta=(1-2\epsilon g)\ \mathrm{tg}\ \theta_0$ et

$$r=\frac{dh(x,z)}{\cos(\gamma+\theta_0)}(1+\epsilon g\cos 2\theta_0).$$

Cette courbe diffèrera d'autant plus d'une droite que ϵ sera plus fort.

8. Le Miroir Elliptique Gaussien

En posant $\alpha'=i\alpha(i=\sqrt{-1})$ et en procédant à quelques légères retouches, la théorie précédemment établie peut être étendue au miroir elliptique gaussien. Ses propriétés optiques peuvent alors être étudiées comme dans le cas de la lentille, c'est-à-dire en poussant jusqu'aux valeurs numériques.

9. Symétries d'Ordre Supérieur

La théorie peut être généralisée aux symétries de répétition d'ordre m. On peut montrer dans ce cas que les trajectoires situées initialement dans un plan de symétrie restent planes et admettent des solutions analytiques. Pour les autres trajectoires les solutions sont analytiques seulement pour $m=2$. Le problème peut néanmoins être résolu par des intégrations numériques.

10. Lentille Décentrée

La théorie peut être appliquée au cas $m=1$, à condition d'y regarder ϵ comme le décentrage relatif $\epsilon=e/R_1$, e étant la distance entre l'axe des électrodes extérieures, supposées alignées, et l'axe de l'électrode centrale. La lentille aura un seul plan de symétrie, le plan XOz passant par les deux axes mentionnés. On peut montrer qu'à la sortie de la lentille, les trajectoires cheminant dans ce plan de symétrie sont de la forme

$$r(0)=r_c+\epsilon g_c(x,z)$$

$$r(\pi)=r_c-\epsilon g_c(x,z)$$

où r_c représente la trajectoire de Gauss dans la lentille centrée et g_c un terme correctif, imputable au décentrage de l'électrode centrale. Si l'on fait subir maintenant à l'axe Oz une translation d'amplitude ϵg_c dans le plan XOz, $r(0)$ et $r(\pi)$ deviendront égaux, aux termes en ϵ^2 près, par rapport à ce nouvel axe de symétrie. On peut montrer alors que dans un azimut quelconque la distance de la trajectoire de Gauss au nouvel axe de symétrie est $r_c+\epsilon^2 g_{c2}$. En particulier au foyer gaussien ($r_c=0$) elle sera $\epsilon^2 g_{c2}$. Par conséquent, le pouvoir séparateur de la lentille, qui était proportionnel à ϵ dans le cas de la lentille elliptique est proportionnel à ϵ^2 dans le cas de la lentille décentrée. L'effet sur la qualité de la lentille peut donc être négligé. Le seul résultat pratique est un déplacement de l'image d'une quantité proportionnelle à ϵ.

11. Références

[1] J. Hillier, J. Appl. Phys. **17,** 411–419 (1946).
[2] W. Glaser, Z. Phys. **120,** 1–15 (1943).
[3] F. Bertein, Ann. Radio **2,** 379–408 (1947); **3,** 49–62 (1948).
[4] E. Regenstreif, Ann. Radio **6,** 51–83, 114–155 (1951).
[5] P. Grivet, Advances in electronics, t. II, (Academic Press Inc., New-York, 1950).
[6] O. Rang, Optik 2/3, 251–257 (1948).
[7] F. Heise, Optik 8/9, 479–489 (1949).

22. Note sur la Théorie de la Lentille Électrostatique Elliptique Fonctionnant en Regime Transgaussien

Par Edouard Regenstreif [1]

1. Introduction

Le régime transgaussien de la lentille électrostatique elliptique a fait jusqu'à présent l'objet d'un nombre extrêmement réduit de publications et se présente par conséquent comme un domaine à peu près inexploré de l'optique électronique. Les raisons de cet état de choses sont imputables premièrement aux difficultés mathématiques du problème: le régime gaussien de la lentille elliptique s'avère déjà très rebelle au calcul et l'impression générale était qu'on n'arriverait pas à résoudre les équations du régime transgaussien. Deuxièmement, du côté expérimental la recherche a très peu avancé, probablement à cause du fait qu'on s'était limité aux lentilles perturbées, c'est-à-dire aux faibles valeurs de l'ellipticité ϵ; le phénomène elliptique transgaussien est ainsi complètement masqué ou très peu observable.

Actuellement, un programme de recherches est en cours au laboratoire du Professeur P. Grivet au sujet de la lentille elliptique. La première partie de ce programme comporte l'étude théorique poussée de la lentille elliptique, non seulement en tant que système perturbé mais aussi en tant que phénomène principal. La deuxième partie du programme est consacrée à l'étude expérimentale des lentilles de forte ellipticité (lentilles cylindriques au sens optique du terme) dont le but est de réaliser sur l'écran un phénomène "macroscopique" ou tout au moins observable. L'objectif final de ce programme est la connaissance approfondie des éléments cylindriques constituant les systèmes corrigés d'aberrations sphériques proposés par Scherzer.[2]

La présente Note résume un travail en cours de publication sur la théorie de la lentille électrostatique elliptique fonctionnant en régime transgaussien. Un autre travail montrera les résultats de recherches expérimentales.

2. La Lentille Elliptique Transgaussienne: Hypothèses et Définitions

Les calculs qui suivent sont articulés sur la théorie du régime transgaussien de la lentille ronde [3] et sur l'étude du régime gaussien de la lentille elliptique (objet de la communication No. 21).

Nous garderons l'hypothèse fondamentale suivant laquelle la lentille peut être décomposée en trois morceaux suivant l'axe, les

[1] École Normale Supérieure, Paris, France.
[2] O. Scherzer, Optik, p. 114–132 (Juillet 1947).
[3] E. Regenstreif, Ann. Radio **6**, 51–83, 114–155 (1951).

fractions extrêmes étant des éléments de révolution et la région centrale étant affectée d'une ellipticité constante η.

Le régime transgaussien sera caractérisé par le fait que la distance à l'axe r_0 du rayon incident initialement parallèle à l'axe n'est plus considérée comme négligeable par rapport aux dimensions z_0 de la lentille, et qu'en outre les quantités r^2 et $(r')^2$ ne sont pas considérées comme négligeables par rapport à r et r'. Dans ces conditions, les équations de Gauss cessent d'être valables et pour établir la théorie des rayons ainsi définis il est nécessaire de recourir aux équations générales de l'optique électronique. Toutefois, le procédé qu'on utilise fréquemment et qui consiste à poser $r=r_{\text{gauss}}+\Delta r$ et à calculer l'aberration Δr ainsi définie n'est pas légitime lorsqu'on approche des zones où les trajectoires oscillent autour de l'axe et à fortiori lorsqu'on passe au miroir. En effet, l'hypothèse $\Delta r \ll r_{\text{gauss}}$ est erronée dans ces cas et le seul procédé exact est de résoudre jusqu'au bout les équations générales des trajectoires. Nous allons montrer qu'une telle solution est possible, même dans le cas de la lentille elliptique fonctionnant en régime transgaussien.

3. Équations des Trajectoires Transgaussiennes dans la Lentille Elliptique

En désignant par r_X et r_Y les projections de la trajectoire sur les deux plans de symétrie, les équations générales du mouvement de l'électron s'écrivent:

$$\left.\begin{aligned} \frac{m}{e}\,\ddot{r}_X &= \frac{\partial \Phi}{\partial X} \\ \frac{m}{e}\,\ddot{r}_Y &= \frac{\partial \Phi}{\partial Y} \\ \frac{m}{e}\,\ddot{z} &= \frac{\partial \Phi}{\partial z} \end{aligned}\right\}. \tag{1}$$

Des transformations appropriées conduisent alors aux équations

$$\begin{aligned} &\{-4\Phi(z)+[r_x^2+r_y^2-\epsilon(r_x^2-r_y^2)]\Phi''(z)\}r_y'' \\ &\qquad =(1+(r')_x^2+(r')_y^2)[(1-\epsilon)r_x\Phi''(z)+2r_x'\Phi'(z)] \\ &\{-4\Phi(z)+[r_x^2+r_y^2-\epsilon(r_x^2-r_y^2)]\Phi''(z)\}r_x'' \\ &\qquad =(1+(r')_x^2+(r')_y^2[(1+\epsilon)r_y\Phi''(z)+2r_y'\Phi'(z)] \end{aligned} \tag{2}$$

Nous n'avons pas à nous occuper des régions d'incidence et d'émergence où, par hypothèse ϵ est nul. Par contre, dans la région centrale (2) devient:

$$\left[-\frac{2}{\alpha^2}2z^2+r_x^2+r_y^2-\epsilon(r_x^2-r_y^2)\right]r_x''=(1+(r')_x^2+(r')_y^2)[(1-\epsilon)r_x+2zr_x']$$

$$\left[-\frac{2}{\alpha^2}2z^2+r_x^2+r_y^2-\epsilon(r_x^2-r_y^2)\right]r_y''=(1+(r')_x^2+(r')_y^2)[(1+\epsilon)r_y+2zr_y']. \tag{3}$$

C'est un système différentiel non linéaire, du deuxième ordre, à variables non séparables. A priori il ne semble pas possible de trouver des solutions à ce système. Nous chercherons cependant—en généralisant les formules obtenues dans l'étude du régime transgaussien de la lentille ronde et du régime gaussien de la lentille elliptique—des solutions de la forme:

$$r_x = A_x \cos\left[\sqrt{\frac{1-\epsilon}{2}}\ \text{arc sh}\ \gamma z - \varphi_x\right]$$

$$r_y = A_y \cos\left[\sqrt{\frac{1+\epsilon}{2}}\ \text{arc sh}\ \gamma z - \varphi_y\right]. \tag{4}$$

Le calcul conduit alors à une solution univoque qui s'écrit:

Dans la région d'incidence:

$$r_x = r_y = r_0\ ch\left[\frac{1}{\sqrt{2}}\ \text{arc sin}\ (z+z_0)\sqrt{\frac{2\beta^2}{2+r_0^2\beta^2}}\right],$$

Dans la région centrale:

$$r_{x,\,y} = A_{x,\,y} \cos\left[\sqrt{\frac{1\pm\epsilon}{2}}\ \text{arc sh}\ z\sqrt{\frac{2\alpha^2}{2-A_x^2\alpha^2(1-\epsilon)-A_y^2\alpha^2(1+\epsilon)}} - \varphi_{x,\,y}\right],$$

Dans la région d'émergence:

$$r_{x,\,y} = B_{x,\,y}\ \begin{matrix}\text{ch}\\ \text{sh}\end{matrix}\left[\frac{1}{\sqrt{2}}\ \text{arc sin}\ (z-z_0)\sqrt{\frac{2\beta^2}{2+B_x^2\beta^2+B_y^2\beta^2}} - \psi_{x,\,y}\right]. \tag{5}$$

Telles sont les équations des trajectoires transgaussiennes dans la lentille elliptique. Les coefficients A à ψ peuvent être déterminées par des considérations de continuité.

4. Caractères Physiques des Trajectoires Transgaussiennes dans la Lentille Elliptique

L'étude des équations (5) conduit aux conséquences suivantes:

(a) Les trajectoires situées initialement dans un plan de symétrie et ces trajectoires seulement restent planes.

(b) Les trajectoires qui ne sont pas situées initialement dans un plan de symétrie subissent une rotation dans la région centrale de la lentille et de ce fait se comportent dans la région d'émergence comme des trajectoires gauches dans un système transgaussien de révolution.

(3) Considérons une lentille électrostatique dont l'électrode centrale est légèrement négative par rapport à la cathode. Les rayons cheminant au voisinage de l'axe seront entièrement décrits par la théorie des rayons paraxiaux de la lentille elliptique. Par contre lorsque l'incidence r_0/z_0 croît, le rayon rencontrera dans la région centrale de la lentille des zones où le potentiel est de plus en plus négatif. La quantité

$2-A_x^2\alpha^2(1-\epsilon)-B_y^2\alpha^2(1+\epsilon)$ sera alors de plus en plus voisine de zéro et la trajectoire effectuera un nombre de plus en plus grand d'oscillations.

(d) Pour une incidence supérieure à une certaine valeur limité r_{0l}, le système fonctionnera en miroir, le nombre des oscillations décroissant lorsque l'incidence augmente.

(e) En raison de la rotation générale des trajectoires dans une lentille elliptique, l'on pourra trouver, et surtout pour ϵ assez fort, des zones où le système fonctionne comme lentille pour une certaine incidence r_0/z_0 et un certain azimut θ_0 et des zones où, pour la même incidence r_0/z_0 et un azimut différent, le système fonctionne en miroir.

Les équations des trajectoires transgaussiennes dans les zones fonctionnant en miroir et dans le miroir lui-même peuvent être obtenues à partir de (5) en tenant compte du caractère imaginaire du radical dans (5_2).

Remarquons que dans la théorie précédente nous n'avons fait aucune hypothèse sur la valeur relative de ϵ. Cette théorie pourra donc servir de base à l'étude détaillée des lentilles cylindriques (au sens optique du terme) et des assemblages de lentilles cylindriques proposés pour la correction des aberrations de sphéricité.

5. Étude Particulière de la Lentille Perturbée

Cette fois on a à faire à une lentille ronde dont ϵ constitue seulement un paramètre perturbant. On peut admettre que ϵ est une quantité très faible et procéder par conséquent à un développement en série des expressions (5). On trouve alors en négligeant les termes d'ordre supérieur en ϵ, des expressions de la forme

$$\left.\begin{aligned} r_x &= r_r\left(x,\frac{r_0}{z_0},z\right)\left\{1+\epsilon\left[g\left(x,\frac{r_0}{z_0},z\right)+h\left(x,\frac{r_0}{z_0},z\right)\cos 2\theta_0\right]\right\}\cos\theta_0 \\ r_y &= r_r\left(x,\frac{r_0}{z_0},z\right)\left\{1-\epsilon\left[g\left(x,\frac{r_0}{z_0},z\right)-h\left(x,\frac{r_0}{z_0},z\right)\cos 2\theta_0\right]\right\}\sin 2\theta_0 \end{aligned}\right\}. \quad (6)$$

Les projections de la trajectoire transgaussienne elliptique sont ainsi exprimées en fonction de r_r, trajectoire transgaussienne de la lentille ronde, et de deux fonctions g et h dont la deuxième est due essentiellement au régime transgaussien, la première existant même dans le fonctionnement gaussien de la lentille elliptique sous la forme $g(x.z)$.

Le calcul fournit des formules explicites pour g et h et les trajectoires puevent ainsi être calculées numériquement et jusqu'au bout.

Pour $r_r=0$, c'est-à-dire au point d'intersection avec l'axe du rayon transgaussien de la lentille ronde, les quantités

$$r_r\left(x,\frac{r_0}{z_0},z\right)xg\left(x,\frac{r_0}{z_0},z\right)=G \quad \text{et} \quad r_r\left(x,\frac{r_0}{z_0},z\right)\times h\left(x,\frac{r_0}{z_0},z\right)=H$$

restent finies et les expressions (6) deviennent

$$\left.\begin{aligned} r_x &= \epsilon(G+H\cos 2\theta_0)\cos\theta_0 \\ r_y &= -\epsilon(G-H\cos 2\theta_0)\sin\theta_0 \end{aligned}\right\}. \quad (7)$$

Ce sont les équations d'une hypocyclöide à 4 rebroussements. La "figure de moindre confusion," généralisation au régime transgaussien du "cercle de moindre confusion" est ainsi une hypocycloïde à 4 rebroussements pour une lentille elliptique, en bon accord avec l'expérience.[4]

6. Formation des Images

Les formules données permettent d'étudier en détail le mécanisme de la formation des images en régime transgaussien elliptique. Pour donner un seul exemple l'image de deux fils croisés $\times$ est un huit 8 en bon accord avec l'expérience.

Enfin des propriétés focales peuvent être définies dans les deux plans de symétrie. Cette fois les distances focales dépendront non seulement du paramètre fondamental x mais aussi de l'incidence r_0/z_0 et de l'ellipticité ϵ.

Disons pour finir que l'étude du régime transgaussien de la lentille électrostatique elliptique est en plein travail et que les résultats déjà obtenus permettent de prévoir un développement assez considérable de la question.

[4] P. Grivet, F. Bertein et E. Regenstreif, The use of marginal rays for the study of assymetry in electrostatie lenses (Proceedings of the Conferenee on Electron Microseopy, Delft, Netherlands, juillet 1949).

23. Progress in Microscopy by Reconstructed Wavefronts

By D. Gabor [1]

Microscopy by Reconstructed Wavefronts, or "Diffraction Microscopy" is an attempt to overcome the limitations of the electron microscope in a roundabout way. As it is impossible to take a perfect image with imperfect lenses, we produce first an imperfect one, in which the information is present, though in a distorted form. This is "decoded" in a second step, the reconstruction, which is carried out by optical means, and which restores the original, as if it had been imaged with a perfect optical system.

The information is never completely destroyed if the first electron-optical image is taken with a *coherent* electron beam; it appears, in a not easily recognizable form, in the diffraction fringes. If a strong *coherent background* is superimposed, the diffraction pattern will not only contain a maximum of information on the object, but will allow its reconstruction by a very simple method: If one illuminates the photograph by a wave that has suffered distortions corresponding to those of the background in the original electron-optical system, a correct image of the object will appear in the beam. In addition there will appear also a second or "conjugate" image, but this can be separated to some extent by focusing sharply on the correct image.

This principle has been demonstrated by light-optical experiments three years ago, and has been described in detail in several publications.[2] The realization of the electron-optical scheme has been, for the past two years, in the hands of Mr. M. E. Haine, and his collaborators, who will report on its progress, and on its difficulties. Most of these difficulties would have to be overcome in any scheme that aims to penetrate into the region of 5 A and below, and they will not concern us here. There are others, chief of which is the question of obtaining sufficient intensities with coherent electron beams to avoid excessively long exposures, which are peculiar to diffraction microscopy, but which will have to be overcome by electronic improvements. We will talk here only of optical improvements in the reconstruction process, and of certain new applications of the basic idea of diffraction microscopy.

The reconstructions that have been obtained so far from diffraction patterns produced in the electron microscope appear decidedly inferior to those that had been previously produced in optical experiments. This may be due in small part to differences in the optical and photographic techniques, but making all allowances for these, there remains a deficiency which is evidently due to the difference in *objects*. In the optical experiments the objects were of the type that obscure only a small fraction of the field, and allow a strong

[1] Imperial College, London, England.
[2] D. Gabor, Proc. Roy. Soc. [A] **197,** 454 (1949); Proc. Phys. Soc. [B] **64,** 449 (1951).

background intensity to fall on the photographic plate. The theory predicts the best results for objects of this type.

We thought until recently that in electron optics the reconstruction principle could be applied only to self-supporting objects, as the thin supporting membranes used in electron microscopy were believed to destroy the coherence of the transmitted electron waves almost completely. Consequently the objects used for the tests were mostly self-supporting zinc oxide crystals. Apart from the fact that such objects tend to grow under the electron beam, rather rapidly, by contamination, they are also sensitive to vibrations of the apparatus, and ultimately the recognition of fine details might be frustrated by the appreciable amplitude of their thermal vibrations, due to their small strength. It has now been realized that sufficiently thin membranes do not destroy the coherence, though they somewhat reduce the contrast of the diffraction fringes, and appreciable progress may be expected in the near future by using the right type of thin objects, on sufficiently thin supports.

It has been recently realized that in light optics it is possible to obtain good reconstructions even with the wrong type of objects, which obscure a great part of the field and thus cut out most of the coherent background, by adding a background wave that has not gone through the object but *around* it. This will be discussed later. To some extent this artifice could be used also in the "projection method" of diffraction microscopy, as originally proposed. This, however, had many disadvantages, and has now been replaced by the "transmission method" of Haine and Dyson.[3] which consists simply in taking a *defocused* image of the microscopic object, using coherent light. Let us see whether improvements in the reconstruction process are possible if we take a hologram obtained by this method as the starting point.

The spurious part of the reconstructed image has always been interpreted up to now as caused by the presence of a "conjugate" image. There is, however, an alternative interpretation: The reconstruction is imperfect, because two wave fields appear in it, superposed. One is due to the "amplitude contrast," the other to the "phase contrast" of the original, *plane* object. This is by no means an obvious result, and we will explain it, assuming for simplicity that there were no geometric-optical errors present in the taking of the hologram.

If we illuminate such a hologram by a plane wave, as shown in figure 23.1, the two conjugate objects appear at both sides of it, at equal distances z_0. Let us place a lens of focal length f in the emerging beam, and consider its focal plane, which can be called the "Fourier-plane." If the transmission of an object (in any plane) is a function $t(x,y)$ of its transversal coordinates, the amplitudes in the Fourier plane will be, apart from a constant factor

$$T(\xi, \eta) = \iint t(x, y) e^{-2\pi i (x\xi + y\eta)} \, dx \, dy \tag{1}$$

that is to say the Fourier transform of t. ξ,η are the "Fourier coordinates," which are connected with the transversal coordinates x_F,y_F

[3] M. E. Haine and J. Dyson, Nature **166**, 315 (1950).

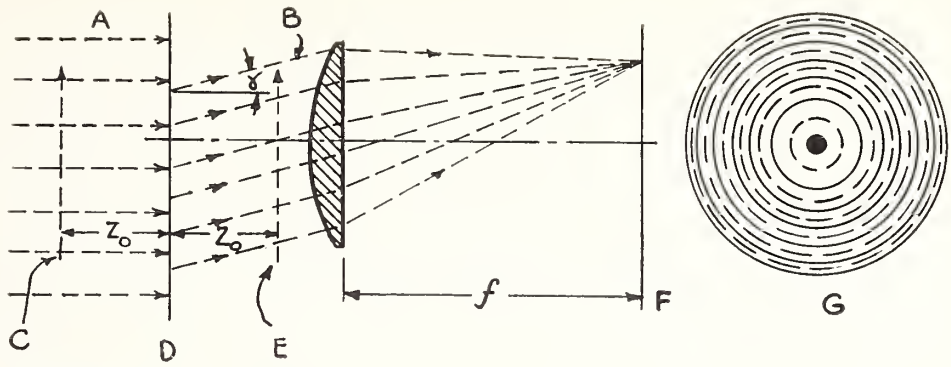


FIGURE 23.1. *Principle of full reconstruction by interpolation.*

A, Parallel illuminating beam; B, beam diffraeted in a certain direetion; C, first conjugate image; D, plane of hologram; E, second eonjugate image; F, focal plane of exploring lens "Fourier plane"; G, zones in Fourier plane. Solid line, amplitude contrast; broken line, phase eontrast.

in the Fourier plane by

$$x_F = f\lambda\xi, \qquad y_F = f\lambda\eta. \tag{2}$$

We should obtain the same wave as in eq (1), if instead of a photograph we had a "complex hologram" in the H-plane, whose transmission function is

$$\tau(X,Y) = \iint T(\xi,\eta) e^{-\pi i\mu\rho^2} e^{2\pi i(X\xi+Y\eta)}\, d\xi d\eta. \tag{3}$$

X and Y are the transversal coordinates in the H-plane. For proof see Gabor (1951). $\mu = z_0\lambda$ is the characteristic parameter of the transformation. $\rho = (\xi^2+\eta^2)^{\frac{1}{2}}$ is the radius in the Fourier plane.

We do not, however, possess the function τ, but only its *real* part, which is the amplitude transmission of the hologram

$$\begin{aligned}\frac{1}{2}(\tau+\tau^*) &= \frac{1}{2}\iint \left[T e^{-\pi i\mu\rho^2} e^{2\pi i(X\xi+Y\eta)} + T^* e^{\pi i\mu\rho^2} e^{-2\pi i(X\xi+Y\eta)}\right] d\xi d\eta \\ &= \frac{1}{2}\iint \left[T(\xi,\eta) e^{-\pi i\mu\rho^2} + T^*(-\xi,-\eta) e^{\pi i\mu\rho^2}\right] e^{2\pi i(X\xi+Y\eta)}\, d\xi d\eta.\end{aligned} \tag{4}$$

In the second line we have written this in the form of a standard Fourier transform, which can be compared with eq (3).

Let us now decompose the object transmission into an "amplitude contrast" and a "phase contrast" part

$$t = t_1 + it_2, \tag{5}$$

where both t_1 and t_2 are real functions. Let T_1 be the Fourier transform of t_1, T_2 the transform of it_2, so that $T = T_1 + T_2$, where T_1 and T_2 are not of course, necessarily real. They have the symmetry properties

$$T_1(\xi,\eta) = T_1^*(-\xi,-\eta), \qquad T_2(\xi,\eta) = -T_2^*(-\xi,-\eta). \tag{6}$$

Substituting $T = T_1 + iT_2$ into eq (4), and using the relations (6), we

obtain

$$\frac{1}{2}(\tau+\tau^*)=\int\int[T_1\cos\pi\mu\rho^2-iT_2\sin\pi\mu\rho^2]e^{2\pi i(X\xi+Y\eta)}d\xi d\eta. \quad (7)$$

The expression in square brackets under the integral sign gives the amplitude in the Fourier plane, and it is clear that this differs from the correct function T. We notice, however, that on all those radii ρ in the Fourier plane for which

$$\mu\rho^2=n \qquad (n \text{ integer}) \quad (8)$$

this amplitude is the same, apart from a factor $(-1)^n$ as if the original object were present, but with amplitude contrast alone. Similarly on the radii

$$\mu\rho^2=n+\frac{1}{2}, \quad (9)$$

T_1 is suppressed, and the amplitude is as if the object had only phase contrast, again apart from a factor $(-1)^n$. In short, *we possess the correct values of T, but only on certain circles*, which are spaced like Newton fringes, and contain information on amplitude contrast and on phase contrast alternatively.

In order to utilize this fact we could think first of masking the whole Fourier plane, except on the circles where the wave amplitude has its correct value. That is, we admit only the circles as given by eq (8) if we want the amplitude contrast image, or the circles given by eq (9), if we want phase contrast, in both cases with n an even number. Next, we could also utilize the circles with n odd, if we reversed the sign of the amplitude in these zones, by advancing or retarding the wavefront by half a wave. This could be done, for example, by a glass plate into which a groove of suitable depth is etched at every second circle, but it can be done much more simply by backing the mask with a lens of suitable strength, positive or negative.

This way we obtain narrow annuli, cut out of the correct wavefront corresponding either to the amplitude contrast or to the phase contrast in the object. The idea naturally suggests itself that we might obtain a further improvement by interpolating the wavefront between these circles.

Before discussing means for realizing this interpolation, let us estimate the error in the imaging of a point. Let this be a "Gaussian point", at a distance x_0,y_0 from the axis of the Fourier lens, with a transmission

$$t(x,y)\alpha\exp[-\pi\sigma^2[(x-x_0)^2+(y-y_0)^2]], \quad (10)$$

whose Fourier transform is

$$T(\xi,\eta)\alpha\exp(-\pi\rho^2/\sigma^2)e^{2\pi i(x_0\xi+y_0\eta)}. \quad (11)$$

We can now go over to a sharp point, by making $\sigma\to\infty$. Introducing polar coordinates ρ,θ in the Fourier plane this is

$$T(\rho,\theta)=\exp[2\pi i\rho r_0\cos(\theta-\theta_0)], \quad (12)$$

where θ_0 is the azimuth of x_0,y_0. Using a well known formula, we can convert this into a Fourier-Bessel series

$$T(\rho,\theta)=J_0(2\pi r_0\rho)+2i\sum_1^\infty J_k(2\pi r_0\rho)e^{ik(\theta-\theta_0)}. \tag{13}$$

We possess the value of this for any θ, but only on the circles $\pi\rho^2=n$. The question is therefore how well we can approximate the Bessel functions J_k by interpolating at these points.

This question can be decided without much calculation. It is known that the "half period," i. e., the distance between consecutive zeros of Bessel functions is always larger than π, which is their limit for large values of the argument. Thus we can expect good interpolation only if we have at least one interpolation point per interval π of the argument $2\pi r_0\rho$. This gives

$$r_0(\rho_{n+1}-\rho_n)=r_0\Delta\rho<\frac{1}{2}. \tag{14}$$

On the other hand either eq (8) or eq (9) give approximately

$$2\mu\rho_n\Delta\rho=1. \tag{15}$$

It is sufficient to consider only $\rho_n=\rho_{max}=\gamma_{max}/\lambda$, where γ_{max} is the aperture angle, because it is well known that a narrow annular aperture gives practically the same resolution as a full one. Combining eq (14) and (15) we now obtain the necessary criterion for good interpolation

$$r_0\leq\mu\rho_{max}=z_0\gamma_{max}. \tag{16}$$

This means that we can simultaneously reconstruct only a small part of the image; not more than the area that has contributed to one point of the hologram in the transmission method. This area is traced out by the aperture cone whose apex is a point of the H-plane. Even better results can be obtained if the object area is smaller. This means that, in order to obtain good results by the interpolation method, the object must be gradually reconstituted *by scanning*.

An optical method of interpolation is sketched out in figure 23.2. From the mask the light enters a number of fine, transparent laminae, with boundary layers of slightly different refractive index. By multiple reflection it appears as if the light were coming, not from a point, but from a patch with approximately Gaussian distribution of light amplitude, but the same phase. Such an interpolator may be perhaps realized by means of a tightly wound reel of film behind the mask. It would be premature to enter into the discussion of details at this stage.

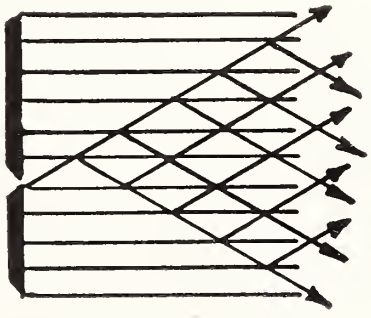


FIGURE 23.2. *Optical interpolation.*

The conclusion is that, at least mathematically, it is possible to eliminate the "conjugate image" in the reconstruction, even if we have only one ordinary hologram at our disposal. This conclusion is somewhat surprising because of our instinctive belief that we must pay in some way for the loss of information by the suppression of phases. The reason that the reconstruction nevertheless succeeds (at least in theory) is that we have not dealt with *any* kind of objects but only with *plane* objects. Our mathematical process presupposed the knowledge of the object distance z_0, and in principle one would have to use a different mask for every z_0. It may be hoped that in practice it will be possible to *search* for the right object position as we are now searching for the errors of the electron-optical system. The right distance is found if the object shows the sharpest details.

Our best hope is, however, that it may not be necessary to go to the complications of the interpolation method, as the masking method, especially in its perfected form with phase reversal in every second zone, may give us sufficient improvement in the reconstruction of our more difficult holograms.[4]

When the idea of wavefront reconstruction first emerged, there seemed to be little advantage in applying it to light microscopy, where we have lenses at our disposal that are nearly perfect. It was only recently that the author became aware of a distinct advantage of the reconstruction method: In diffraction microscopy it is not necessary to illuminate the object as strongly as in ordinary microscopy. Paradoxically, the illumination need be only a small fraction of what is required for seeing or for photographing the object.

The principle is illustrated in figure 23.3. The instrument is essentially a Jamin-Mach interferometer, with a special beam-splitting prism at one end, which will be explained later. Otherwise the apparatus differs from an ordinary interferometer mainly in that its two branches carry very unequal light intensities. The object is introduced in the branch in which the intensity is very low, for example, 0.01 to 0.001 of the total. The rest goes through the other branch, and constitutes the coherent background, which carries no information on the object, and serves only to *amplify* the weak image carrying beam, with which it is united only at the photographic plate. In order to explain this amplifying effect let us consider first only one photograph. At one point of this plate let A be the large background amplitude, a the weak image-carrying amplitude. The photographic plate records the resulting amplitude, whose square is

$$A^2+a^2+2Aa\cos\phi \tag{17}$$

where ϕ is the phase angle between the vectors A and a. If, as we have assumed, A is large compared with a, we can neglect the small term a^2 in eq (17). The third term will be relatively large. For instance, if the intensities A^2 and a^2 are in the ratio 100:1 in the two branches, A and a are only as 10:1, and the contrast between the maxima and the minima in the interference fringes will be $(1.1/0.9)^2=1.50$, which is more than sufficient for photography. It can be expected that it will be possible to take photographs even if only 1/1000 of the light that reaches the plate has gone through the object. This

[4] *Note added in the proof, Feb. 2, 1952.* In the meantime I have found that it is, in fact, possible to construct zone filters that reduce the disturbing effect of the conjugate image to a small fraction, without having to go to the complications of the interpolation and without using the necessarily very slow scanning process.

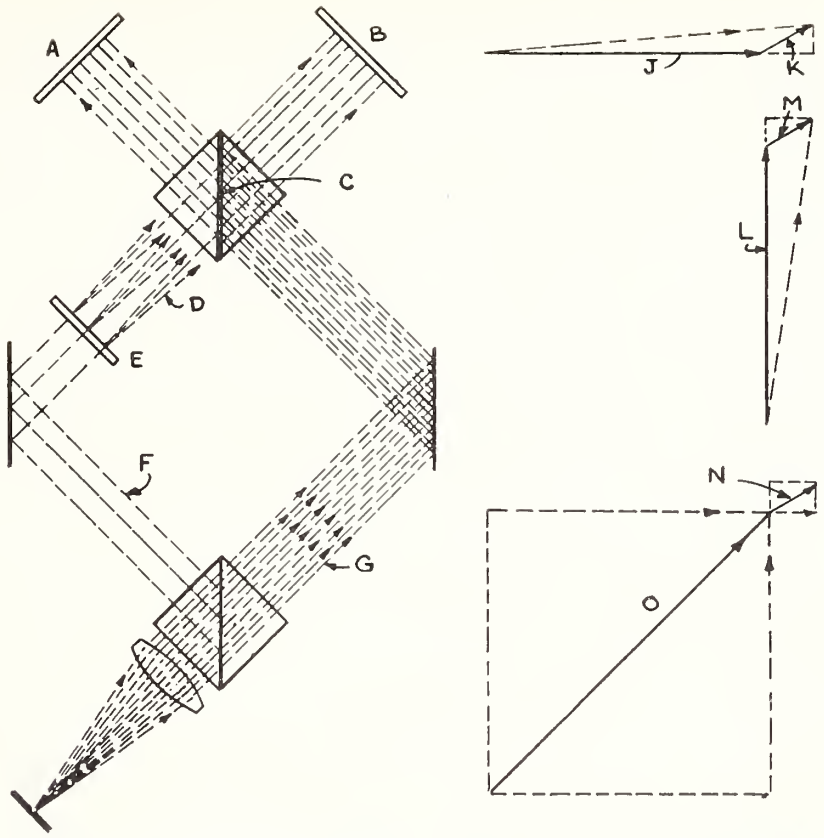


FIGURE 23.3. *The principle of full reconstruction from two complementary holograms.*

The two holograms record two components of the image wave amplitude at right angles to one another. A, Hologram I; B, hologram II; C, quadrature prism (produces 90° phase shift between the two backgrounds relative to the image waves); D, diffracted waves; E, object; F, weak illuminating wave; G, strong background wave; J, background I; K, amplitude of image wave; L, background; M, amplitude of image wave; N, illuminating the two holograms with waves in quadrature reproduces the original image amplitude; O, resulting background.

may enable us to take photographs of objects that are far too light sensitive for ordinary microphotography.

The strong background assures in itself that the reconstructions will be of good quality, but the scheme is capable of a further improvement, by which the conjugate image is *completely* eliminated, and the object becomes visible in *three dimensions*.

As seen from eq (17), and also from the vector diagram in the top right corner of figure 23.3 in a photograph taken with a strong coherent background only that component of the image-carrying amplitude will be recorded that is in phase with the background, the component in quadrature will be surpressed. Assume now that we take a second photograph, in which we produce a 90° phase shift between the amplitudes A and a. This photograph will record exactly that part of the wave which had been suppressed in the other. If now we illuminate the two photographs, with two background waves again at right angles to one another, we obtain the original wavefront, added to a uniform background. This is *full reconstruction* (in which no reference whatever is made to the position of the object) by means of *two "complementary" holograms*.

Figure 23.4 is a sketch of the experimental instrument, under construction. One of the two branches contains the object and the objective. A dummy object plate and objective are introduced into the other, to make the light paths approximately equal. The most

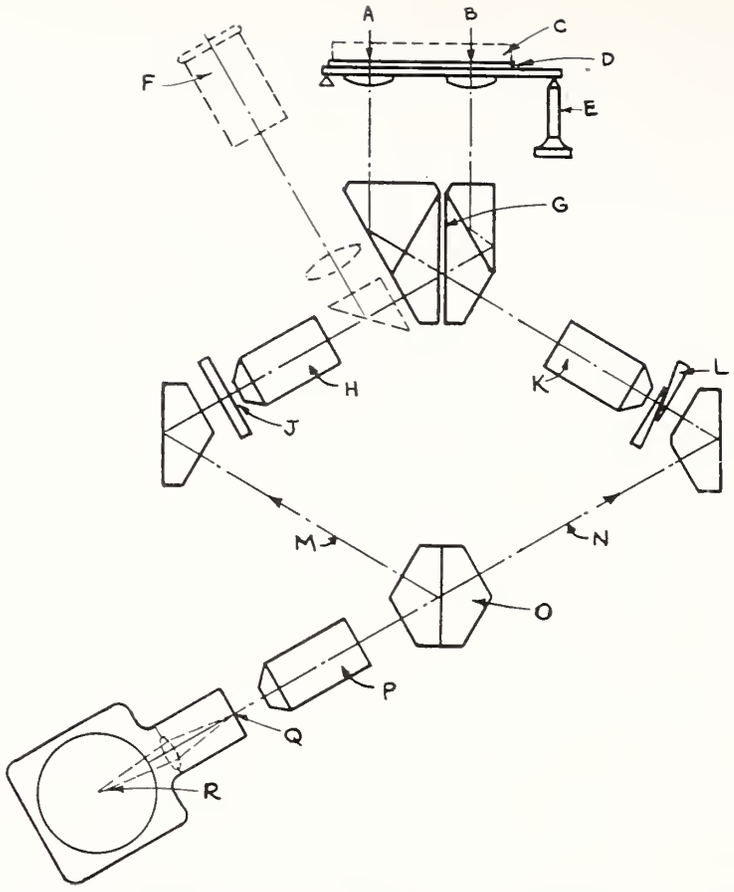


FIGURE 23.4 *Interference microscope utilizing the reconstruction principle.*

The elements shown in interrupted lines are used in the reconstruction only. A, Hologram I; B, hologram II; C, mirror, backing the plate, used in the reconstruction; D, photographic plate; E, tilting screw used in the reconstruction; F, eyepiece for observing the reconstructed image; G, sandwich layer for producing a quarter-wave phase shift between the two backgrounds; H, objective; J, object; K, dummy objective for symmetrizing the two branches; L, compensator; M, illuminating wave containing 1 percent or less of the energy; N, background wave containing 99 percent or more of the energy; O, beam-splitting prism with very small reflection coefficient; P, condenser; Q, pinhole; R, light source.

important element of the instrument is the "quadrature prism", which is to produce the 90° phase shift between the two photographs. This may be obtained, according to the theory which we will not detail here, by a "sandwich" reflecting layer, consisting of two semi-reflecting coatings, separated by transparent material of suitable thickness. Two further prisms are attached to the quadrature prism, so designed that the two complementary photographs appear side by side on the same plate.

In the reconstruction the object is taken out, and the double hologram is backed by a mirror. The plate and the mirror are now tilted until there is a 90° phase difference between the two backgrounds, and the resulting wave is observed by means of the viewing device, shown in hatched lines. The correct phase shift reveals itself by the vanishing of the conjugate image and its accompanying fringes.

It is hoped to obtain three advantages over ordinary microscopy:

1. Once the double photograph is taken, it preserves the image of the object in three dimensions. Moreover, this image can still be viewed in bright-field, dark field, or phase contrast. We must, however, mention the disadvantage, that coherent illumination is not very suitable for deep objects, as every object leaves a "wake" of interference fringes behind it.

2. Only a small fraction of the light required for the exposure of the photographic plate need go through the object.

3. Moreover, it is not necessary to destroy light-sensitive objects by the illumination required for focusing, as there is no need for focusing, and a blind shot can be taken.

Though this new interference microscope exists as yet only on paper,[5] and some instrumental difficulties may have to be overcome, the electron microscopist might well feel a pang of envy at the thought of what can be done with light, and not with electrons. But is it really impossible to realize at least the second advantage, the small illumination? Two years ago I thought that this must be answered in the negative, as I believed that efficient electron-beam splitters are impossible. I am not so sure now, after some experiments with electron reflection on graphite, which we have carried out recently at the Imperial College. But at this point I must break off, as already I may have ventured too far into the realm of speculation and conjecture.

Discussion

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: Which face was used on the graphite reflector, and what were the reflection angles?

Dr. Gabor: It was microcrystalline graphite, polished in a certain way, so that the cleavage faces were alined with the surface. The angles were of the order of 1 to 2 degrees.

Dr. J. Hillier, RCA Laboratories Division, Princeton, N. J. In regard to using a supporting membrane and having a large amount of coherent radiation coming through—if you are talking about it coming through in the sense that it has not lost any energy and in the sense that nothing much has happened to the diffraction fringes, I will agree with you, but their contrast will be appreciably reduced with supports of perhaps 100 angstroms thickness. Besides, you will have local phase shifts of the order of half a wavelength. I wonder if that is not a serious disturbance.

Dr. Gabor: Originally we feared that coherence would be quite destroyed by such local phase shifts as Dr. Hillier mentioned, but now it appears that coherence, defined as the number of discernible interference fringes, is not much altered by the supporting membrane. The local phase shifts introduced by the support do not really matter very much. If this structure does give an extra set of fringes, it merely means that when you come to reconstruct them by the phase contrast method the variable optical thickness of the support will appear superimposed on the object. On the other hand, of course, it is too much to hope that one will be able to get away with a supporting film if we come down to resolutions of one or two angstrom units because the atomic structure of the support may become visible and obscure the structure of the object.

[5] This work is now in progress at the Imperial College, London, with the assistance of the Paul Instrument Fund of the Royal Society.

24. Problems in the Realization of Diffraction Microscopy with Electrons

By M. E. Haine [1] and T. Mulvey [1]

Many of the practical problems connected with the realization of diffraction microscopy with electrons are essentially common with or extensions of those associated with high resolution electron microscopy by classical methods. There are, however, certain additional difficulties which arise as a result of the high coherence of illumination required for the new method. Basically these result only when the new method aims at a higher resolution, though even for comparable resolution practical necessities result in some increased difficulties. It is the purpose of this paper to discuss some of the practical aspects and difficulties of the method as we see them at the present time.

At the Paris Conference last year we showed that the projection method of diffraction microscopy originally suggested by Gabor was severely limited by practical difficulties introduced by chromatic effects. A new method, the transmission method, previously proposed, was shown to be free from these effects. The new method comprises the recording of the Fresnel diffraction pattern produced under conditions of high coherence suitably magnified to enable the fine high order fringes not to be lost in the record as a result of the finite photographic plate resolution. Optical reconstruction of these transmission diffraction images, or holograms, are possible by methods such as suggested by Dyson and by Gabor. The reconstruction process has, in fact, certain advantages over the old method.

The apparatus used for the transmission method is essentially identical with the transmission electron microscope. The main difference is that the illuminating system must be arranged to give a highly coherent or parallel beam which results in lower illumination densities and therefore longer exposure times. The diffraction process only compensates for geometric aberrations and all other disturbing factors such as variations in lens strength, movement of object or microscope column and stray field effects must be reduced by the amount which would be required for the same improvement in resolution of the classical instrument. The elimination of such disturbances is actually rendered more difficult by the increased exposure times. Fortunately, the diffraction process itself provides very powerful means for testing the instrument for the presence of these disturbing factors. The diffraction test methods were also described in Paris last year.

The important fundamental conditions necessary for the recording of the diffraction pattern fringes out to an angle α corresponding to a resolution d, defined by $d=\lambda/2\alpha$, can be derived from a geometrical path difference equation. The derived expressions are shown in table 24.1 compared with the corresponding requirements for classical

[1] The Associated Electrical Industries Research Laboratory, Aldermaston, Berkshire, England.

electron microscopy where dV/V and dI/I are the permissible fractional variation in voltage and current during exposure; f, the objective lens focal length; λ, the wavelength; α_c, the maximum angle of illumination; γ, the angle subtended by the effective source at the object; Z, the axial off-focus distance; C_s, the spherical constant; q, the ratio of the maximum to the actual angle of illumination; and Ω, a factor depending on the source brightness (β) the photographic plate resolution Ψ and sensitivity η given by

$$\Omega=\frac{n\Psi^2}{\beta}.$$

TABLE 24.1.

	Classical	Diffraction
Longitudinal coherence	$\frac{dV}{V}<4d^2/f\lambda$; $\frac{dI}{I}<2d^2/f\lambda$	Same as column 2.
Transverse coherence	$\alpha_c<\lambda/2d$	$\gamma<\frac{2d}{Z-C_s\alpha^2}$
Exposure time	$t_c=\frac{4q^2\Omega}{\pi\lambda^2}$	$t_d=\frac{Z^2}{4\pi d^4}\Omega$

Practically, the important factor is the exposure time. The transverse coherence is a condition which must be met, while the exposure time is an inevitable result which must be tolerated. It is seen that the exposure increases rapidly with reduction in resolution instead of being independent of resolution as in the classical instrument. On the other hand, some control of exposure is possible through the value of Z which is, within limits, under control of the operator.

In the classical instrument, the exposure time worked out from the expression shown for a value of $q=5$ and for typical values of the other parameters is 0.05 sec. In practice, of course, this is inconveniently short and lower intensities are used with the corresponding reduction in damage to the object, also full use is not normally made of the photographic plate resolution.

Diffraction microscope exposures (in seconds) for various values of d and Z are shown in table 24.2.

TABLE 24.2.

$Z(\mu)$ \ d(A)	1	2	4	8
1	130	32	2	0.125
10	13,000	3,200	200	12.5
100	1,300,000	320,000	20,000	1,250

These figures assume $\psi=20\mu$, $\eta=4\times10^{-11}$ coulombs and $\beta=100{,}000$ amp/cm^2 ster. It is seen that in order to keep exposures to a reasonable value, the off-focus distance must be kept small.

Our experience so far has shown that much is to be gained by reduc-

ing exposure times to a minimum to overcome the disadvantageous effects of slow drifts in object position, etc. The logical conclusion would seem to be to work on focus. In fact, although the exposure time required is a minimum for best focus, incidental disadvantages result. In the first case the difficulty of obtaining accurate focus for object detail of the type that might be usefully observed by diffraction microscopy would be very great because of the very small contrast. Also, within the aberration caustic of the objective lens some peculiar secondary effects complicate the diffraction fringe pattern by folding it over on itself. It would seem practicable to use a value of Z down to 3 or 4 μ which already gives reasonably small exposure times (e. g., a few minutes for $d=2$ A).

The part played by the photographic plate is important. The plate must record the diffraction fringes in such a way that reconstruction is possible. The photographic resolution is determined largely by the path of the electrons in the emulsion, a minimum value of 10 to 20 μ is obtained for reasonable voltages. The effect of the random distribution of grains in the plates is of interest. For a given area of plate the randomness of the recorded signal will be at least as large as the randomness in the number of electrons falling on that area. We have made measurements of this randomness, or noise, and the measurements show the noise to be approximately equal to that which would be expected from the random distribution in the electron beam. Consider an object consisting of a circular opaque disc of diameter equal to the resolution limit, calculation shows that the fringe contrast in the hologram of such an object would be such that for resolutions of 1 to 2 A the noise over the area of one fringe would exceed the fringe contrast. This would be very much more so for a practical object of much less than 100-percent contrast. Thus, the hologram would appear to the eye or to a microdensitometer of infinite sensitivity and resolution as a completely random distribution of photographic grains. The question arises as to whether such a hologram would reconstruct. Irradiation of a completely random distribution of scattering points with coherent radiation will produce on a subsequent screen no contrast variation. However, any degree of order in the distribution will produce a contrast variation or image. The diffraction pattern, although constituting a negligible departure from randomness over small areas, will, when treated as a whole, produce a pattern. This pattern will under the correct focusing conditions be the reconstructed image of the original object. The contrast and noise over the area of the reconstructed image will be the same as would be expected in a directly formed image. In practice, it is very desirable to stop out the background wave in the reconstruction apparatus and so obtain a dark field image, since particles only a few angstrom units in size will otherwise give a contrast too small for observation.

In our experiments, by paying careful attention to mechanical stage design and general instrument stability both mechanical and electrical, we have so far been able to obtain a resolution of 5 to 6 A in the diffraction image.

Reconstruction of such images is rendered difficult by the unsuitability of the type of object that it has so far been possible to use. Diffraction microscopy does not take kindly to an object that occupies too much of the total area of the field. The ideal object is very small, as, for example, a small collection of single scattering atoms or atom

clusters. On the other hand, the object should be freely supported without the usual film which would tend to destroy the coherence of background wave. The objects so far used have been metal oxide crystals or carbon black. In fact, whatever objects have been used rapidly contaminate. In theory, contamination should not necessarily be any worse than in the classical instrument, as the same coulomb loading is required for an exposure. However, in seeking a higher resolution, a reduction of contamination is important. Also, since the current density at the object is less, the object operates at a lower temperature with a resulting increase in contamination.

The final object never departs significantly from straight edges which are far from ideal as test objects. Further progress depends very much on eliminating contamination and devising suitable means of mounting specimens. Supporting films need not be entirely ruled out as experiments have shown the loss of coherence may not be as serious as might be expected.

Dr. Hillier: You said, and I quite agree with you, that the noise in the electron beam is a fundamental limitation—That also agrees with some work we have been doing recently. However, you say that your mechanism integrates out the noise within the range, but is that not defeating your purpose? Is it not the noise from the diffraction range that constitutes the information that is going to give the high resolution?

Mr. Haine: No. At one time we thought actually that it was the noise, particularly in the last ring, that determined whether we could get the resolution. More recently we have come to realize that you integrate the noise over the whole hologram. If you carry out this integration, you find that the signal gives a contrast that is exactly the same as the contrast in a transmission microscope, and the noise again is exactly the same. The signal noise ratio is then identical.

Dr. Hillier. Yes, but you are referring to a single point.

Dr. Gabor: The hologram for each point overlaps the whole plate.

25. On the Theory of the Magnetron

By L. Brillouin [1]

A rather complete report on the theory of the one anode magnetron was recently published in "Advances in Electronics", vol. III (1951), and the reader will find in this report [2] the detailed calculations. Since experiments do not seem to check too well with the theory, it may be useful to specify that the assumptions made in the discussion were the usual ones: I, electrons leaving the cathode with zero initial velocity; II, space-charge limited current, when the electric field is zero on the cathode, or saturation current, when the field is not zero on the cathode; III, steady conditions with a well defined electric potential at all points between cathode and anode; IV, The space charge modifies the electric potential distribution between cathode and anode, and it is assumed that direct collisions or interactions between electrons need not be introduced. The average potential is supposed to represent the only important effect of electron interactions, through space-charge.

Condition III implies that electrons may strike back on the cathode, but in doing so they come back to the same potential from which they started—hence they must have a back velocity equal to their initial velocity, which is zero (assumption I). Back bombardment by electrons with large velocities is thus forbidden. We must of course admit that assumption I is only an approximation. There may be small initial velocities. In such a case the back bombardment velocities will also be small.

Two different types of solutions were proved to be theoretically possible: A, double stream solutions with part of the electrons moving away from the cathode with a velocity $+V_r$, at a certain distance r, and part of the electrons moving back with a velocity $-V_r$; B, single stream solutions with no electrons moving backwards.

The discussion was first carried out for a plane magnetron, and this was found to represent a good approximation whenever the cathode-anode distance was small compared to the cathode radius. Practically, the plane solution is a good first approximation when

$$R=\frac{r_a}{r_c}<1.4$$

r_a is anode radius, and r_c, cathode radius. If the ratio R becomes larger, the cylindrical model must be discussed, and gives very different results when $R>2.3$.

This point should be noted in connection with some recent experiments performed by L. Marton and D. L. Reverdin at the National

[1] IBM Watson Scientific Computing Laboratory, New York, N. Y.
[2] L. Brillouin, Plane Magnetron, p. 85–144; L. Brillouin and F. Bloch, Cylindrical Magnetron, p. 145–181.

257899—54——17

Bureau of Standards,[3] where magnetrons with a large value of R were used in the investigation. Two different kinds of space-charge distributions, apparently corresponding to solutions A or B, were observed, and these results check with theoretical predictions. For large R values, it appears that both solutions may take place. For low R values, when the plane approximation can be used, the conditions for excitation of either A or B are clearly defined.[4] If the current never reached saturation since the time when operation of the tube was started, the single stream B solution should obtain. If saturation has been reached at some time or other, the double stream A solution has great probability of being established.

In addition to these results, the conditions for negative resistance were computed, and the peculiarities of the corresponding trajectories discussed. The stability or instability of trajectories was discussed only for motions maintaining the cylindrical symmetry, with a space-charge density ρ (r) independent of the angle θ. Experiments[5] clearly indicate some results which are not accounted for by the theory: I, large noise level; II, anode current below cut-off; and III, large back bombardment on the cathode.

The third result clearly indicates that the motion is not actually a steady motion, and that our previous assumption III cannot be maintained. There must be large oscillations in the space-charge with an average potential V depending on r and maybe also on θ. The mechanism for the excitation of such internal oscillations may be found in the interaction between electron beams, discovered by A. V. Haeff.[6] This interaction may greatly amplify the initial small perturbations in electronic emission, and explain the three experimental results just described.

It might be interesting to check whether similar experimental anomalies could not be discovered in a triode with positive grid and low potential anode, which may also exhibit either single-stream or double-stream solutions.[7]

[3] J. Applied Phys. **21,** 42 (1950) and **22** (1951).
[4] L. Brillouin, Phys. Rev. **70,** 187 (1946). Elec. Commun., **22,** 110 (1944); **23,** 458 (1946.)
[5] R. L. Jepsen and M. W. Muller, J., Applied Phys. **22,** 1196 (1951); D. Gabor and G. D. Sims, NBS Circ. 527, p. 253.
[6] A. V. Haeff, Proc. Inst. Radio Engrs. **37,** 4 (1949); Phys. Rev., **74,** 1532 (1948).
[7] C. E. Fay, A. L. Samuel and W. Shockley, Bell System Tech., J. **17,** 49 (1938).

26. On the Theory of the Magnetron

By D. Gabor [1] and G. D. Sims [2]

The most elementary fact about a cylindrical magnetron in dc-operation is the absence of a sharp cut-off. It is also the most difficult to explain. All theories about the "steady state" of the magnetron agree in the point that there must exist a maximum orbit, determined by the energy theorem and by the momentum integral. If this orbit does not reach the anode, there should be no current, and the cut-off ought to have a width of the order of the emission velocities, i. e., a small fraction of a volt. In fact, one finds a tail, which a steady-state theory could explain only by an "electron temperature" of the order of 10–20 volts.

It has been suspected for a long time that the tailing-off of the cut-off curve is in fact due to internal oscillations of the magnetron. Since the tail appears also if the a-c impedance of the outer circuit is zero, i. e., at short circuit, measurements by ordinary methods are impossible. One can only approach the short-circuit case by leaving a small impedance sufficient for measurement and extrapolate these observations. These show a "noise" in the region of the operating frequencies of the magnetron, not sharply defined. It is also known that the noise increases sharply at certain critical values of the magnetic field. Another observation is the self-heating of the cathode in the cut-off state, where there is only a d-c current flowing to the anode. These are essentially all the experimental data at our disposal.

On the theoretical side the situation was long obscured by our ignorance on several fundamental points. Perhaps the most important is the question whether we are at all justified in applying "self-consistent," i. e., "single-electron" theories to the magnetron? This was assumed from the start by Brillouin and by Hartree independently. Hartree made first a rough estimate of the effect of "electron-collisions," using the formulae of L. H. Thomas, and justified it a posteriori by showing that the average time spent by an electron in the cloud is not sufficient for the electron-electron interaction to become important. This procedure appeared also justified by the good agreement with experiments of the theory of the oscillating magnetron.

In view of the evident clash of theory and experiment in the case of the nonoscillating, almost cut-off magnetron one must ask two questions: 1. Is Thomas' formula correct? 2. Even if the interaction is as small as given by the Thomas formula, can we neglect it in the cut-off condition, in the region where, according to Brillouin's theory, electrons describe extremely long, almost circular trajectories around the cathode?

The first question can now be answered. Mr. E. A. Ash, at Imperial College, London, has recently measured the diffusion of a weak electron

[1] Imperial College, London, England.
[2] General Electric Co., Wembley, England.

beam in a plasma, and found that it agreed closely enough with Thomas' formula.

The second question is more difficult to answer, but it appears that we can evade it, because electron interaction is not the only thing which has been neglected in the single-stream theory. The *initial velocities* have been also neglected, and, though this appears a small matter at first sight, it has serious consequences.

One of us [3] has investigated the theory of the steady state with initial velocities, but neglecting electron-electron interaction. The results are rather surprising. They depend quite essentially on how many degrees of freedom one allows at the start for the initial velocities. If one allows none, one obtains the Brillouin single-stream solution. If one allows two, one obtains a double stream, at all anode-cathode ratios (not only where, for example, the Page-Adams theory postulates them). If one allows three degrees of freedom, and if one makes in addition the assumption that the electron trajectories are so long that "ergodic disorder" can be postulated, one obtains again a solution rather similar to the Brillouin steady state in appearance, though the space charge density falls off towards the anode with a somewhat different law from Brillouin's. This is a somewhat simplified representation of the results, valid for small initial velocities. At large initial energies one can obtain a confusing variety of equilibrium states of the electron cloud.

It is very striking that finite initial velocities, *however small*, give results so strongly dependent on the degree of freedom. An unexpected feature of magnetron theories with zero initial velocity is revealed: They relate to a *degenerate* case. In order to make them physical one must use criteria which could not be even suspected. One of us (G. D. S.) has found recently that at a cathode temperature of 1,000° K and a magnetic field of 1,700 gauss the emission current must exceed 1.2×10^6 amps/cm^2 for a single stream state to be possible.

Before trying to clear up all the questions that arise in connection with the questions of initial velocity and electron interaction, we have made an attempt at an investigation of the *stability of the two-stream steady state*. For reasons such as the above mentioned, we expect this to apply to a wide variety of magnetrons, possibly to all which are of practical importance.

It is to be understood that this steady state will exist only so long as the electron cloud keeps at a safe distance from the anode, and is undisturbed, because once disturbances of a certain type are produced, they will grow and maintain themselves automatically in a state of more or less steady oscillation.

We investigated first the possibility of radially symmetrical *pulsations* of such a cloud, but we found that these will always be damped out if there is an outer short circuit. But when we tried azimuthally periodic disturbances the result was quite different (fig. 26.1). It was found that *beyond a certain azimuthal number* (number of periods round the circle) there was always a tendency for the cloud to start *self-oscillations*. The mechanism is roughly this: The electrons that reach the anode will have energies *smaller* than corresponding to the voltage between the cathode and the anode.

Such an electron leaves a part of its energy behind, in electromagnetic form. This in turn is passed on to one or several electrons

[3] D. Gabor, Stationary electron swarms in electromagnetic fields, Proc. Roy. Soc. [A] **183**, 436 (1945).

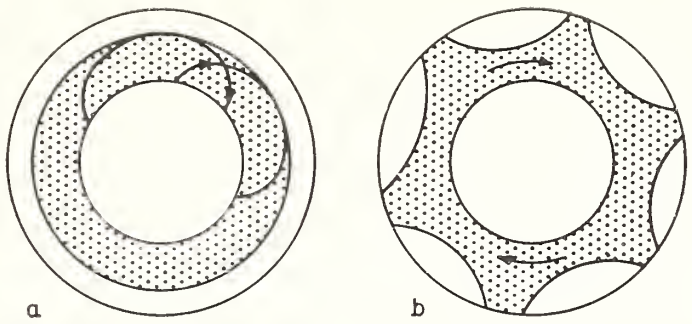


FIGURE 26.1. *a, Steady state; b, preoscillating state with* $n=5$.

that return to the cathode with more than their initial velocity. As the anode potential is completely steady, no oscillating energy appears in the outer circuit. The result is merely a deficiency in anode heat, and an equal excess of cathode heat.

One can imagine such a mode as a wheel with n spokes, rotating in the same direction as the electrons. The spokes oscillate a little radially, and they touch the anode always near the instant when the electrons have minimum velocity.

The frequencies are very nearly integer multiples of the inverse transit time of an electron, from the cathode back to cathode. For any such frequency any number of oscillating modes are possible, provided that the spoke number n is large enough. This, of course, is only a first approximation; in reality, one must expect the frequencies somewhat dependent on n. The noise-like character of the pre-oscillation is explained by this near-degeneracy, as in general no mode can be excited in its pure form for any length of time, and the energy probably pendulates between them, in a complicated manner.

One can expect these oscillations to grow particularly easily if the time taken by an electron for an excursion to the anode and back resonates with the cyclotron frequency or one of its multiples. One must therefore expect certain critical combinations of potential and magnetic field, for which the noise is maximum. This seems to be borne out by observations.

This is a short preview of our results, and necessarily of a provisional nature. Further progress is expected to be difficult, because while the critical frequencies and the stability criteria can be, at least approximately, obtained from a linear theory, the coupling between these modes requires a nonlinear treatment, which will have to be also statistical, in order to account for the width of the noise-spectrum. The most difficult problem is the calculation of the cut-off curve, resulting from these oscillations, which shows that this most elementary of observations on magnetrons relates in fact to the most intricate phenomena. It is likely that the theory will have to stop somewhere short of a complete quantitative explanation, until new experiments enable us to make useful simplifying assumptions.

27. A Point X-ray Focus for Shadow Microscopy

By V. E. Cosslett [1] and W. C. Nixon [1]

Introduction

A number of schemes have been proposed for high resolution microscopy with X-radiation. The wave-length being of the order of 1 A it is in principle possible to obtain a much higher resolving power than that of the best optical microscope, which gives about 1000 A when ultra-violet radiation is employed. The simplest arrangement is that in which X-rays are allowed to illuminate a thin specimen placed in close contact with a photographic plate at some distance from the source (fig. 27.1a). The resulting negative is then enlarged photographically for observation, so that the limiting resolution is essentially that of the grain of the recording emulsion. Engström [1] [2] in particular has refined this method and obtained resolutions of the order of 1 μ, as also have Ehrenberg and Spear [2]. It appears to be impossible to equal optical resolving power by this means, however. The second system is that proposed by Kirkpatrick and Baez [3], and since investigated by several other workers, in which curved crystals are used to focus X-rays incident at glancing angles (fig. 27.1b); this is directly analogous to reflecting optical microscopes. Its practical realization encounters mechanical difficulties, as pointed out by Ehrenberg [4], and it is doubtful again whether the optical resolution limit can be approached.

The remaining possibility is to use X-radiation simply to project shadowgraphs of objects, magnification being obtained partly geometrically, by recording at a distance, and partly by subsequent photographic enlargement. The system was originally proposed by Sievert [5], with a conventional X-ray source (fig. 27.1c). Later von Ardenne [6] suggested the use of an electron lens to reduce the electron beam to a minute spot on the target, since the obtainable resolution depends in the first place on the size of the X-ray source. An alternative system [7, 8] is to form the electron spot on a Lenard window as target, utilizing the transmitted radiation (fig. 27.1d); this allows the specimen to be brought very close to the source, giving a large geometrical magnification with short camera length and minimising diffraction effects in the specimen. Such an arrangement has the advantage over electron microscopy that the object may be kept in air, so that biological specimens can be examined without being dried and exposed to vacuum, which normally kills them. It is this last system that has now been investigated in some detail, and from which a resolution of optical order has been already obtained from a variety of specimens. The work was begun at Oxford in 1939 by one of us (V. E. C., with the aid of a Research Fellowship from the Uni-

[1] Cavendish Laboratory, University of Cambridge, Cambridge, England.
[2] Figures in brackets indicate the literature references on p. 264.

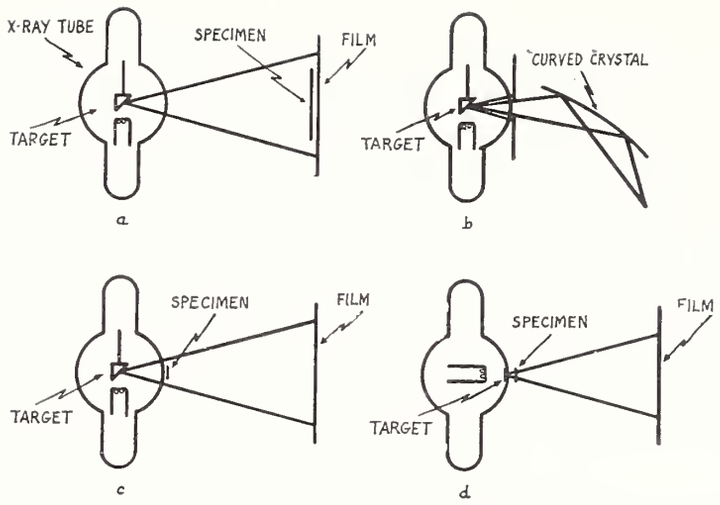


FIGURE 27.1. *Methods of obtaining X-ray micrographs.*

a, Contact photography on fine-grain film; b, true focusing by reflecting surface; c, shadow projection from normal target; d, shadow projection by X-rays transmitted through thin target.

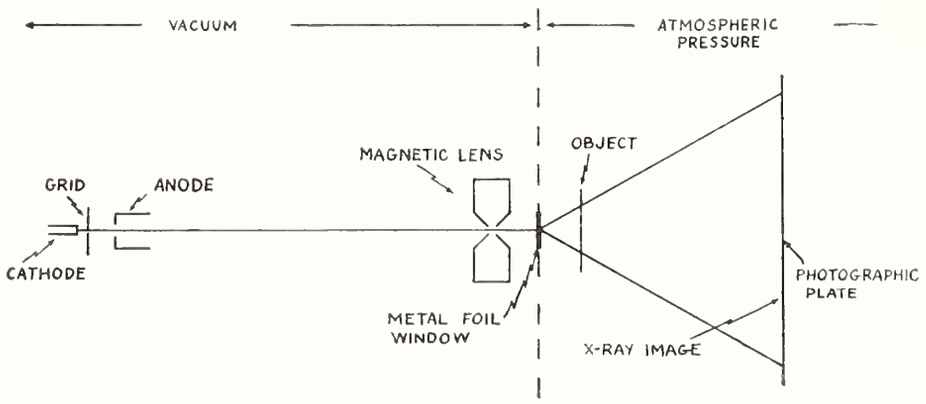


FIGURE 27.2. *Schematic arrangement of X-ray shadow microscope.*

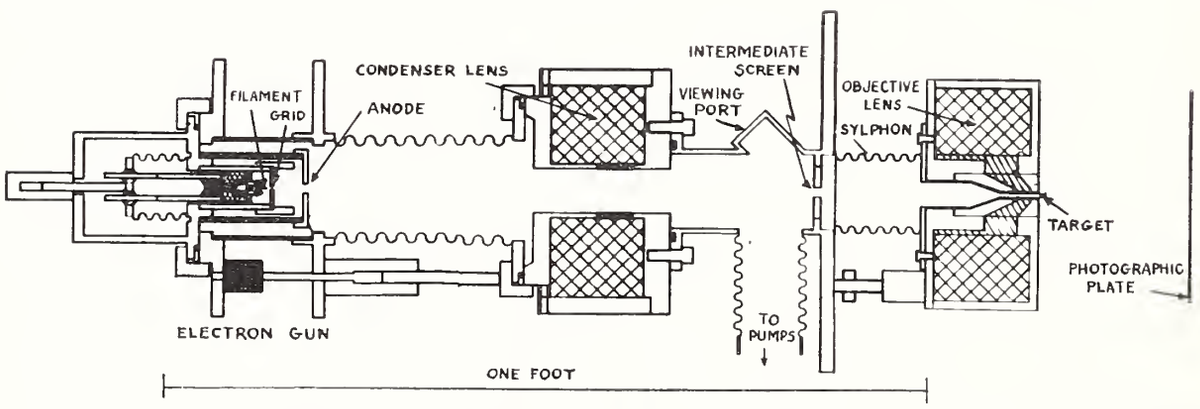


FIGURE 27.3. *Constructional details of X-ray shadow microscope (diametrical section).*

versity of London), and was taken up again after the war; the main development has taken place subsequently in Cambridge.

Experimental Apparatus

The experimental arrangement is shown in outline in figure 27.2, and in detailed cross section in figure 27.3. A beam of electrons is obtained from a triode gun of normal construction. A tungsten hairpin filament is situated at the centre of an aperture (diameter,

1.0 mm) in the surrounding grid cylinder, which carries a negative bias potential. The emitted electrons are accelerated towards the anode, by a potential difference that is variable from 3 to 15 kv, and pass through a hole in its surface of 3.5 mm diameter. The electron beam thus formed passes down the axis of the evacuated metal tube into the pole-pieces of an electron lens, which focuses it into a small spot on the metal foil that closes the end of the tube. It is found convenient to employ a first lens as condenser, to collimate the beam; the cross section of the beam is then visible on an intermediate viewing screen covered with fluorescent material. The second electron lens is of high power and narrow bore (6 mm). The magnification of such a system is a minimum, in the absence of intermediate image formation, when the beam forms a parallel bundle between the two lenses, and is then given by the ratio of their focal lengths. The focal length of the first lens was 15 cm and that of the second was 2.5 mm at minimum; the minimum magnification, M, was thus 1/60. The electron gun gives a beam having a minimum cross section, somewhere between cathode and anode, of about 0.05 mm diameter. The minimum spot size to be expected would then be 0.8 μ.

To ensure that the focused spot remains stationary during exposures, the applied potential and the current in the lenses must be constant to a high degree. The former is obtained from a high frequency rectifying set of conventional design, and the latter is supplied from 6 v accumulator batteries. The sharpness of the photographs shown indicates that the requisite stability is obtained. The several components of the apparatus are provided with transverse adjustments working about flexible metal bellows (fig. 27.3), so that accurate alinement can be carried out and maximum brightness of the focused beam be assured. During this procedure a small fluorescent screen is placed at the end of the tube instead of the metal foil. With the target in position, the emitted X-rays are still powerful enough at 10 kv for the shadowgraph of an object to be visible on a screen placed at the position of the recording plate.

The size of the electron spot is determined by the aberrations of the final electron lens as well as by the size of the source. It is therefore necessary to limit its aperture by means of a stop of small diameter; a platinum or copper aperture of 0.25 mm diameter was used in most of the work. At a focal length of 2.5 mm and spherical aberration coefficient (C_s) 1.5 mm (cf. van Ments and Le Poole [9]; Liebmann [10]), this would limit the spot in the paraxial plane to a diameter 0.4 μ, or to about one-third of this in the plane of least confusion, well below the minimum spot size as fixed by the diameter of the effective cathode.

The target used for most of the work was a tungsten foil of about 1 μ thickness, supplied by the Philips Laboratories, Eindhoven; the unsupported area had a diameter of 0.25 mm. Careful search with an optical microscope enables pieces of foil to be chosen that have no visible holes and that hold vacuum over long periods. The comparison made by Oosterkamp [11] of the relative merits of various metals as X-ray targets showed that tungsten has a distinct advantage over all others, including copper, so far as thermal properties and efficiency of X-ray production are concerned. Some advantage would accrue from the use of copper, in the form of its strong characteristic radiation, if voltages greater than 10 kv were to be employed. However, longer wavelength X-rays were desirable because of their stronger absorption

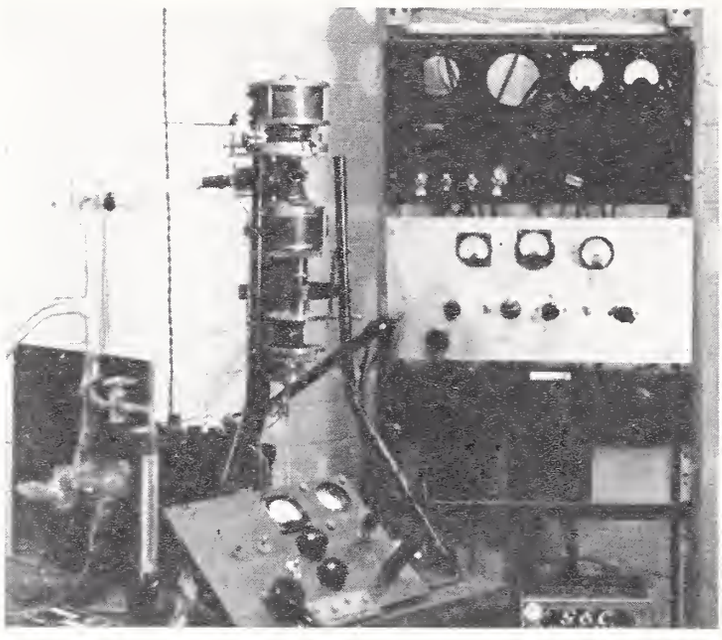


FIGURE 27.4. *Experimental X-ray shadow microscope.*

in biological material. Wavelengths of around 1 A were used in most of this exploratory work, for reasons of experimental convenience, but photographs have also been taken with longer wavelengths, up to 4 A. Beyond this wavelength, the absorption of the radiation in the air path of the camera becomes appreciable.

The camera is a small light-tight box, normally of length 2.5 cm for an image 1.5 cm in diameter, but extending to 8 cm in length for a larger image of 4 cm diameter. The specimen to be examined is mounted close to the target in such a way that it can be moved across the axis in two directions at right angles, and also along the axis for varying the magnification. With an accelerating voltage of 10 kv, the exposure time is of the order of 1 min on slow, fine-grain X-ray emulsions, with a beam current in the order of 10 μa on a spot of about 1 μ diameter. The complete apparatus is shown in figure 27.4, being mounted vertically on a tripod support, with the electrical gear at the rear.

Preliminary Results

The initial experiments have thrown light on the possible range of application of the method, and its probable ultimate resolving limit. Figure 27.5 shows test grids, the finest being 1,500 meshes per inch, along with a 100 μ thick section of frog kidney impregnated with phosphotungstic acid; the fissures are drying artefacts. Clearly the contrast in biological tissues is so poor at a wavelength of 1 A that thicknesses considerably greater than 10 μ are needed to give an adequate picture, unless dense structures or heavy elements are present. Figure 27.6 and 27.7 are a comparison optical and X-ray pair of a small insect (Cyclops). The resolution is rather better in parts of the optical picture, but the depth of focus is limited. The X-rays render visible certain of the internal structures and also show the whole specimen in focus at the same time. Figure 27.8 shows greater detail in the rat flea (Xenopsylla cheopis). In both figure 27.7 and 8 the reference grid is composed of silver bars of width 2.5 μ and 3.5 μ in the respective directions. The resolution in the shadowgraph thus may be estimated as approaching 1 μ.

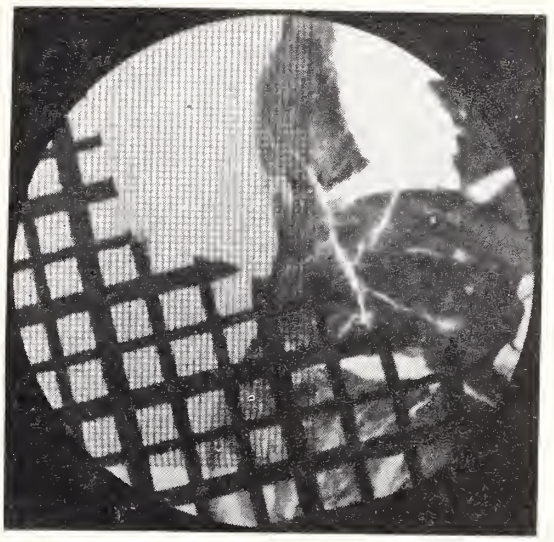


FIGURE 27.5. *X-ray micrograph of section of frog kidney, 100 μ thick and stained with phosphotungstic acid.*

Also 200-mesh copper grid and 1,500-mesh silver grid (x40).

FIGURE 27.6. *X-ray micrograph of cyclops specimen with 1,500-mesh silver grid (x55).*

FIGURE 27.7 *Light optical micrograph of cyclops specimen with 1,500-mesh silver grid (x55).*

FIGURE 27.8. *X-ray micrograph of rat flea (Xenopsylla cheopis), and 1,500-mesh silver grid (x30).*

The present exposure time is too long for successive pictures of an insect to be taken without immobilising it temporarily. However, it is quite practicable to anaesthetize it from time to time for photography, then to permit it to revive and continue development between exposures. Figure 27.9, 10 and 11 are a series thus obtained of a red mite, the second photograph being taken four days, and the third seven days after the first. Considerable change has visibly taken place in the interim, but it is probable that it is mostly due to inadequate feeding and partial desiccation in this first experiment. The effect of the X-radiation on the insect needs also to be investigated; the dose rate is of the order of 50 roentgen per minute. All these matters are now under active examination.

It is readily possible to take stereo-micrographs of an object, by

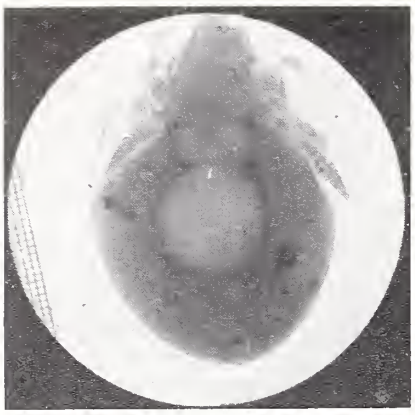


FIGURE 27.9. *X-ray micrograph of anesthetized red mite, and 1,500-mesh silver grid (x40).*

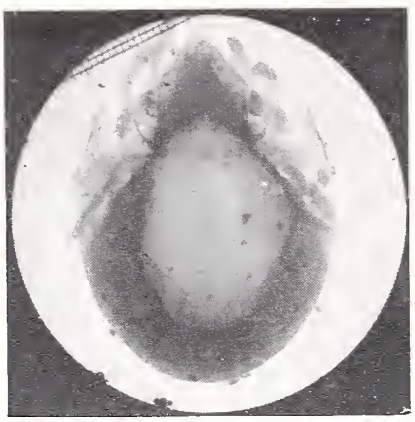


FIGURE 27.10. *Same red mite 4 days later (x40).*

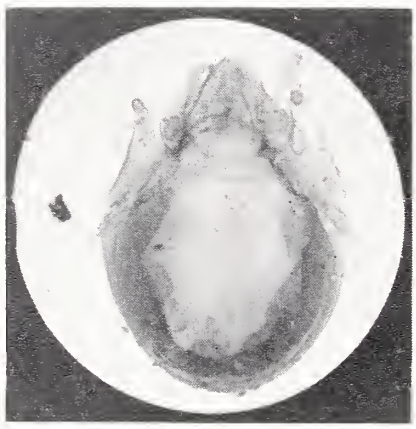


FIGURE 27.11. *Same red mite 7 days later (x40).*

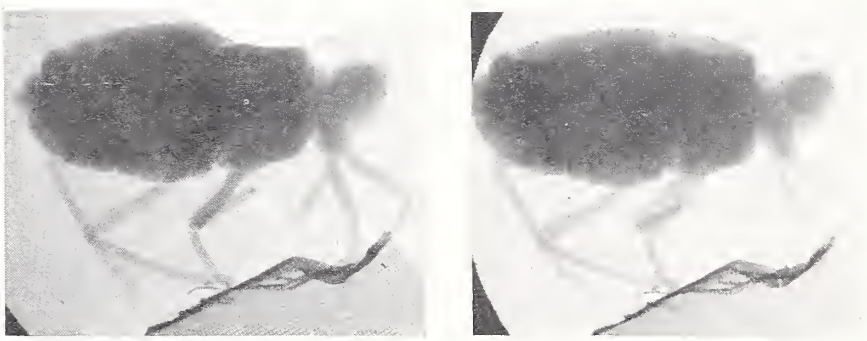


FIGURE 27.12. *Stereographic X-ray micrographs of Aphis Fabae, and 1,500-mesh silver grid (x28).*

moving it slightly between two exposures or by tilting its support or that of the recording plate. The first method has been used on several subjects, and a typical pair of pictures is shown in figure 27.12. It is thus possible to get an accurate three-dimensional view of a specimen, which is particularly important in the case of insects, the total volume of which is large compared with the size of detail to be seen.

Resolution Limit

The electron spot in the work here described was designed to be 1 μ in diameter, and the results show that a resolution of this order can be obtained. In the first place the resolution limit is set by the size of the X-ray source, taking into account the spread of the incident

electrons within the target. In tungsten the range of 10 kv electrons is appreciably less than 1 μ; in copper it would approach this figure. Consideration of the practicable limit of resolution for this type of microscope rests finally on the tolerable length of exposure. By using two stages of demagnification, an electron spot of the order of 100 A diameter can be obtained without great difficulty (Hillier and Baker [12]). The intensity then becomes so small that the exposure might be impracticably long, especially for biological specimens. It becomes a question of what intensity can be drawn from an electron gun, and whether the incident energy can be dissipated from the spot without melting the target metal.

For a spot diameter, small compared with the area and thickness of the target, Oosterkamp [13] has shown that the permissible loading (energy per unit area) is inversely proportional to the diameter; the energy input is then directly proportional to the diameter. Although correction is necessary when the spot is of similar dimensions to the thickness of the target, the relation still proves to be nearly linear in the region of interest. A more exact consideration, taking into account the penetration of the beam into the target, does not change the result appreciably (Cosslett [14]).

The electron beam has to be delivered to the target through an aperture, the size of which must be reduced with the spot diameter so as to ensure that the spherical aberration is always smaller than the latter. If the reduction in spot size is effected by increasing the lens strength, rather than by reducing the cathode size or distance, it fortunately appears [14] that the aberration coefficient of practical lenses falls in such a way as to ensure that the energy delivered through the aperture is approximately proportional to spot size, and not to some higher power.

The question remains whether the energy delivered by the X-ray beam can be dissipated in the target without melting it. Application of Oosterkamp's figures shows at once that electron beams of normal intensity would not raise its temperature appreciably when the spot size is of the order of 1 μ, provided that the target is not thinner than this. In order to approach the melting point of tungsten, the beam intensity must approach a value of 10^6 amp cm^{-2} steradian^{-1}, corresponding to a filament temperature of 3,000° K. With $\beta = 10^5$ amp cm^{-2} ster^{-1}, the load on the target is of the order of 0.1w, or 10^7 w/cm^2, the current density being approximately 10 amp/cm^2 at the filament and 10 μ amp/μ^2 at the target. Intensities of this order have been drawn from the gun in the present work, to keep the exposure time small. Further reduction in spot size would lead to more favorable conditions still in the target, so that forced cooling is unnecessary in this apparatus provided that the spot size is never allowed to become larger than a few microns. At the same time, there is little point in overheating the filament still further, since its life would soon fall below the required exposure time at high resolution, as mentioned above.

It thus appears that a further reduction of 100-fold in spot size is entirely feasible, giving an equivalent resolution limit for X-ray microscopy, which is some ten times beyond the limit with ultraviolet light. The exposure time, however, would then be 100 times greater and hence in the order of hours. Apart from the long immobilization of biological specimens, the exposure time now becomes of

the same order as the life of the filament. There may well be specimens, especially those impregnated with heavy elements and thus already killed, for which such long exposures could be tolerated. It would appear more useful, however, to aim in the next stage of development at a resolution in the same order as that of ultraviolet microscopy; with a spot of 0.1 μ, exposures would be of the order of 10 min. For optimum efficiency the target would need to be about 0.25 μ thick, but it could be strengthened to withstand atmospheric pressure by a layer of beryllium or other light metal, which would not seriously weaken the X-ray beam.

During the course of this work one of us (V. E. C.) has been supported by a Keddey Fletcher-Warr Research Studentship of the University of London (at the Electrical Laboratory, Oxford, 1939–41) and by an I. C. I. Research Fellowship (at Cambridge, 1946–49), and the other (W. C. N.) by a Scholarship of the Research Council of Ontario. Our thanks are due also to D. A. Taylor and Major S. S. D. Jones, who assisted in the preliminary experiments. We are indebted to Electrical and Musical Industries, Ltd., for fine silver grids, to the Philips Laboratories for thin tungsten foil, and to Metropolitan-Vickers for the magnetic electron lens used in the early work.

References

[1] A. Engström, Progr. Biophys. **1,** 164–196 (1950).
[2] W. Ehrenberg and W. E. Spear, Nature **168,** 513 (1951).
[3] P. Kirkpatrick and A. V. Baez, J. Opt. Soc. Am. **38,** 766 (1948).
[4] W. Ehrenberg, J. Opt. Soc. Am. **39,** 741 (1949).
[5] R. Sievert, Acta. Radiol. **17,** 299 (1936).
[6] M. von Ardenne, Naturwiss. **27,** 485 (1939).
[7] M. von Ardenne, Elektronen–Uebermikroskopie, p. 73 (Springer, Berlin, 1940).
[8] V. E. Cosslett, The Electron Microscope, p. 72, 116 (Sigma Books, Ltd., London, 1947).
[9] M. van Ments and J. B. LePoole, Appl. Sci. Res. **B1,** 1 (1947).
[10] G. Liebmann, Proc. Phys. Soc. **64B,** 972 (1951).
[11] W. J. Oosterkamp, Philips Res. Rep. **3,** 58 (1948).
[12] J. Hillier and R. F. Baker, J. Appl. Phys. **15,** 633 (1944).
[13] W. J. Oosterkamp, Philips Res. Rep. **3,** 313 (1948).
[14] V. E. Cosslett Proc. Phys. Soc. (London) **B65,** 682 (1952).

28. X-ray Shadow Microscope

By L. Marton [1]

Several attempts at microscopy by X-rays have been reported in the literature. One of the latest is that of Burgers, who employed a method originally proposed by Börsch and later employed by Bragg. The Börsch-Bragg-Burgers method is a two-step method giving the highest resolution of all possible attempts at X-ray microscopy. Its essential limitation, however, is that it can be applied only to such periodic structures as crystal lattices.

Another attempt by Kirkpatrick and collaborators is based on grazing reflection on curved surfaces. Up to the present its resolution has been rather poor, not even attaining that of light microscopes.

A third approach to X-ray microscopy consists in producing an extremely small source of X-rays, from which a shadow image is formed by central projection. The original idea was conceived about twelve years ago. At that time I had calculated the feasibility of such an instrument. My own manuscript on the subject, however, was held back from publication because of the publication of the same idea at the same time by von Ardenne. With the technical means at our disposal at that time, however, the available intensity was extremely low, and exposure times of several years—even as long as one hundred years—could be expected. For this reason both von Ardenne and I dropped the idea at that time. Furthermore, no similar research was performed elsewhere.

One of the chief limitations of the whole project was the low current density emitted by the electron source. Since that time considerable progress has been made in the development of emitters of high current density. The chief reason why the project, which seemed utterly impractical twelve years ago, is today a sound one is the availability of field emitters producing a current density as high as 10^8 amp/cm^2, instead of 0.1 amp/cm^2 as in the old project. It is this tremendous increase in emission, by the factor of 10^9, which makes it advantageous to revise completely the entire calculations and, indeed, an instrument of this kind is today a perfectly feasible one.

The proposed X-ray microscope, in its present conception, consists of a field emitter, an electron lens, and a suitable target. The field emitter is of the tungsten point type, which has been very thoroughly investigated by a group of workers of whom particularly E. W. Muller in Germany, J. A. Becker at Bell Telephone Laboratories, and W. P. Dyke at Linfield College, Oregon, should be mentioned. Figure 28.1 summarizes some of the data on the total emission from a point, as a function of the reciprocal voltage (taken from Dyke and Trolan's ONR Report, Contract N8onr–72400, dated June, 1950). Part of the total current passes through the aperture of an electrostatic lens, which forms, on the target, a one-to-one image of the point. Figure 28.2

[1] National Bureau of Standards, Washington, D. C.

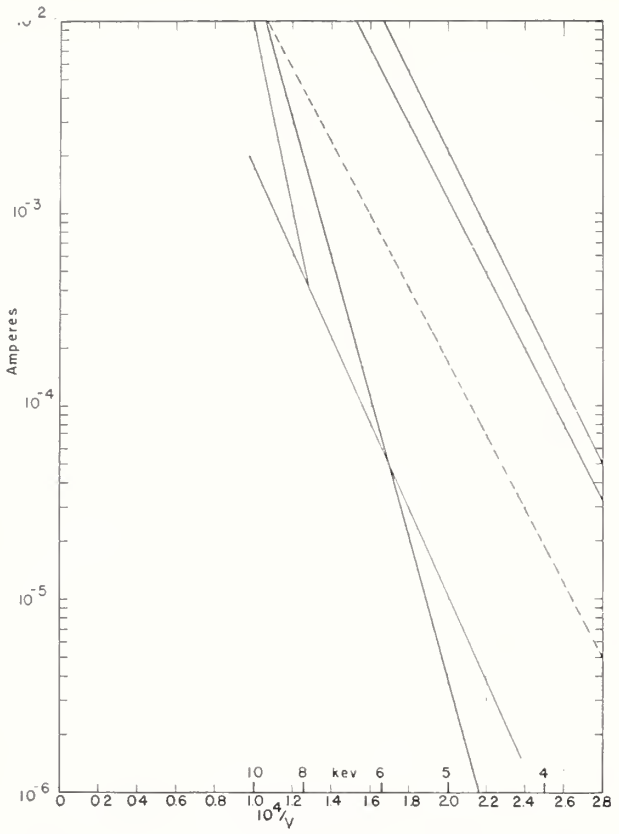


FIGURE 28.1. *Field emission from Dyke & Trolan, ONR-Report, June 1950.*

——, Measured by Dyke & Trolan; ——, assumed averaged.

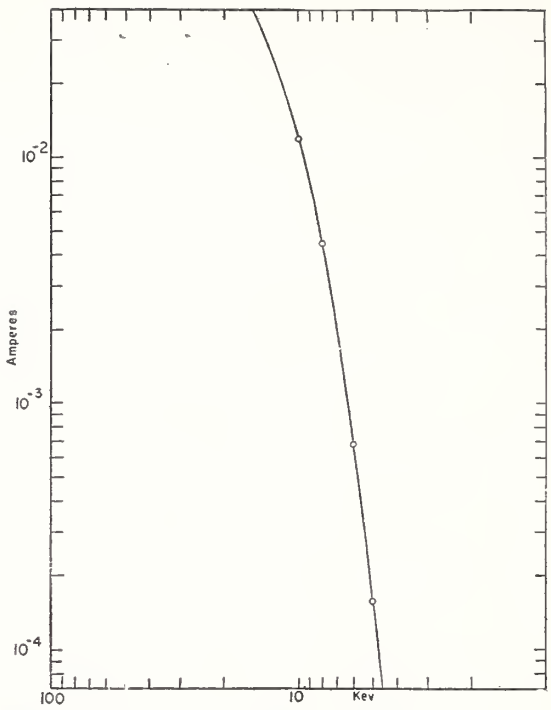


FIGURE 28.2. *Assumed target current.*

gives the total target current as a function of applied voltage. This target current is concentrated into a spot whose diameter, in absence of aberrations, can be computed to be of the order of 5 to 10 A. Spherical aberration will enlarge this image considerably, and it can be assumed that the total diameter of the image, as spread by spherical aberration, will be of the order of 50 A.

The graphs of figures 28.1 and 2 are plotted only for relatively low accelerating potentials, because the X-rays are emitted not only from the area of the target hit by the electrons, but from the whole volume in which the electrons loss energy. It follows that the next step in computing the X-ray emitting area is to determine the effective range of penetration of the electrons into the target material. Figure 28.3 indicates the dependence of the range on beam energy for different target materials. As can be seen from that graph, the penetration is considerably larger than the computed dimensions, not only of the Gaussian image, but even of the image as enlarged by aberrations.

In the preceding paragraph no mention was made of the spreading of the beam by space charge effects. The reason for the omission is that a reasonable estimate of space charge effects has been made, and the total widening of the beam due to space charge has been found to be a secondary effect as compared to the spreading of the X-ray source due to scattering of the electrons in the target material. For a similar reason, furthermore, the effect of thermal distribution of the initial energies of the electrons has been unmentioned, being also entirely negligible as compared to the scattering effect.

Essentially then, the size of the X-ray source being determined by the range of penetration of the electrons, we can estimate the resolving power of such an instrument for two types of operation. One type of

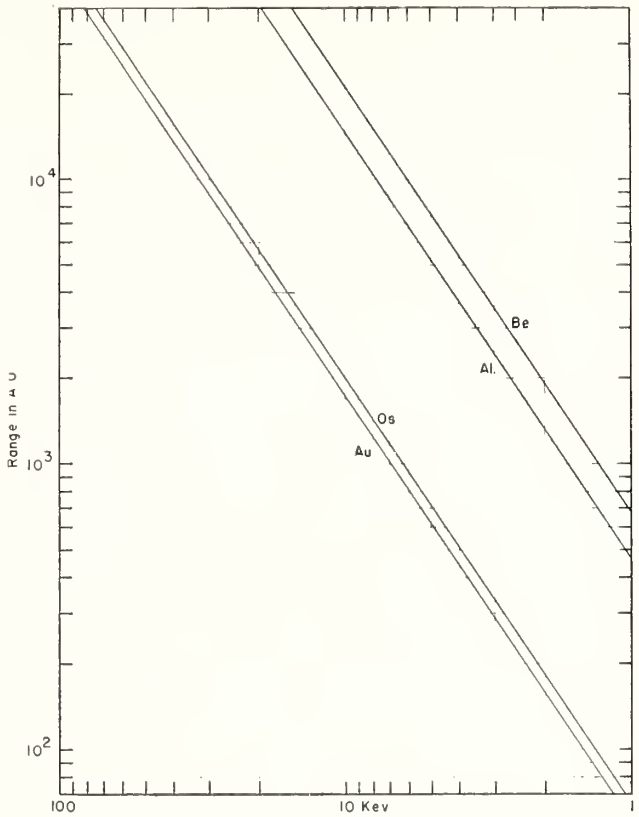
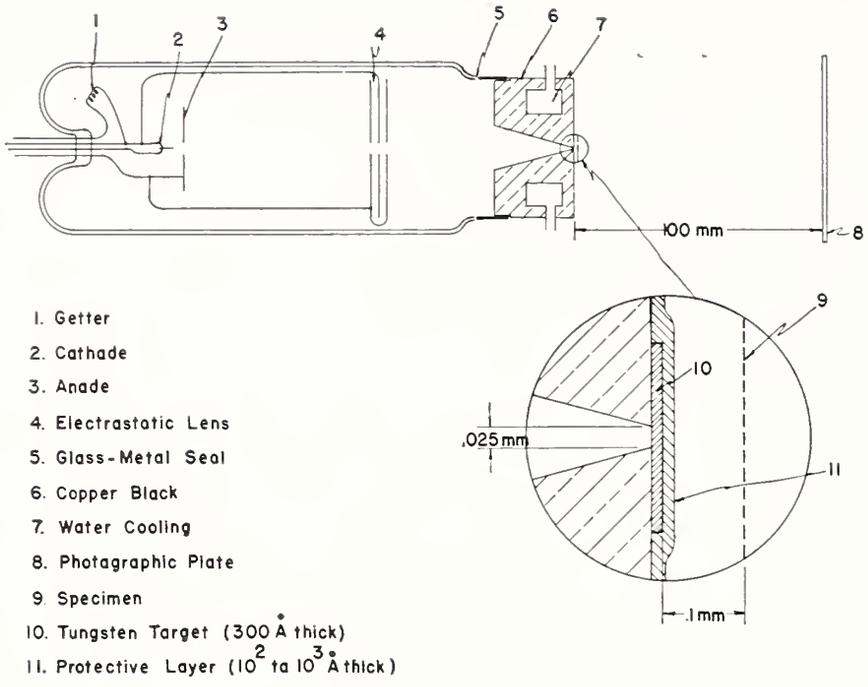


FIGURE 28.3. *Electron range extrapolated from data by Bleuler and Zunti.*

operation consists in having the beam strike the target obliquely, and placing the specimen on the same side of the target as the incident beam but outside of the incident beam itself. In that case, the least resolvable distance can be estimated to be, to first approximation, twice the range of penetration. For practical purposes, the specimen must be brought as close as possible to the X-ray-emitting surface, and hence the incident beam should make a relatively small angle with the emitting surface. If one takes into account, also, that the specimen must be in the same vacuum as the electron emitter and electron lens, it turns out that this type of operation is rather disadvantageous, particularly in view of the extremely high vacuum requirements.[2]

A better way of operation consists in producing a thin target, which is bombarded by electrons from one side and the X-ray emission observed from the other side. Such a thin target could consist of a thin evaporated layer of tungsten or of gold. The thickness of this layer should be approximately equal to the range of penetration of the electrons at the required beam potential. Thus, for example, for 4-kev beam energy, a tungsten target will have to have a thickness of approximately 300 A. To maintain the required vacuum, a sealed-off tube should be used, and the question arises whether or not it is feasible to have a 300-A–thick layer of tungsten sufficiently vacuum-tight to maintain such a good vacuum for long periods. The answer is twofold. First of all, the total window area need not exceed the dimensions of the geometrical image. While the feasibility of such an extremely small hole in a supporting structure of the window seems to be out of the question, it appears reasonable to assume that a

[2] The required vacuum is about 10^{-14} mm of Hg. This extremely high vacuum is necessary for a stable emission of the tungsten point in absence of any absorbed oxygen layers. Vacuum estimates are based on table IX of Ashworth's paper in volume III of Advances in Electronics, p. 35.



1. Getter
2. Cathode
3. Anode
4. Electrastatic Lens
5. Glass-Metal Seal
6. Copper Black
7. Water Cooling
8. Photagraphic Plate
9. Specimen
10. Tungsten Target (300 Å thick)
11. Protective Layer (10^2 ta 10^3 Å thick)

FIGURE 28.4.

tapered hole of the kind indicated in figure 28.4, ending in an opening of approximately 0.001 inch, is well within the practical limits. For that reason it is worth trying to see whether or not such a thin layer alone will be vacuum-tight. If not, it is reasonable to expect that covering it with a layer of low–X-ray–scatering material such as a few hundred angstroms of beryllium, or 50 to 100 Å of SiO^2, could make the window sufficiently vacuum-tight for the purpose of the X-ray tube.

Figure 28.4 then, illustrates the over-all features of the X-ray tube of the microscope. The sealed-off tube with a thin window target can be used in conjunction with a specimen and a photographic plate placed outside of the vacuum. By assuming a distance of 0.1 mm between the X-ray source and the specimen and 100 mm between the X-ray source and the photographic plate, a linear magnification of 1,000 is easily obtained. Further enlargement of that negative will give magnifications up to 10,000 to 30,000 diameters.

The least resolvable distance in that type of operation will again be roughly equal to twice the range of penetration. This figure will have to be modified by taking into account the absorption of X-rays in the target. Due to the hemispherical shape of the emitting volume (see fig. 28.5) this absorption will have a tendency to reduce the effective

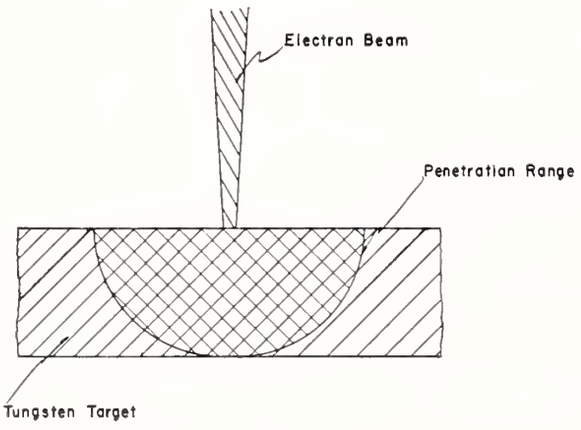


FIGURE 28.5.

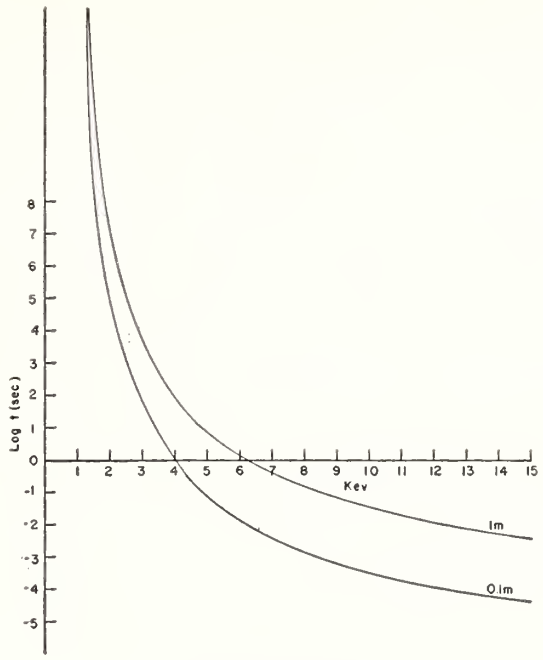


FIGURE 28.6.

diameter of the X-ray source and thus to improve the resolving power. That effect may be counterbalanced, however, by the added scattering of the X-rays in the target and in the added protective layers for vacuum-tight operation.

For a calculation of the exposure times, the first step is to determine the efficiency of X-ray generation as a function of electron beam energy. The efficiency can be assumed to be equal to

$$\epsilon = 2.5 \times 10^{-4} A \beta^2$$

where A is the atomic weight of the target material and $\beta = \frac{v}{c}$ is the ratio of the beam velocity to the velocity of light. By using this equation we can calculate the total number of quanta emitted and determine the number of quanta per unit area reaching the photographic plate. The quantum efficiency of different photographic materials has been determined as a function of X-ray energy (see, for instance, H. Hoerlin, J. Opt. Soc. Am. **39,** 891–97, 1949). Figure 28.6 gives the calculated exposure time as a function of electron beam energy for two different distances of the photographic plate (that is, two different linear magnifications).

Two more effects remain to be discussed that may limit the usefulness of the instrument. One is a possible spreading of the bombarded area of the target due to charging up of that area. The electrical conductivity of the target being finite, it is conceivable that, at the very high current densities employed, a potential barrier can be built up in front of the bombarded area and spread the electron beam. Estimates were made, using the data employed in calculating the exposure times, and it was found that potential barriers of the order of 1 v may be expected, which can be considered negligible as compared to the several kilovolts beam energy.

A much more serious effort is due to the limited heat conductivity

of the target material. This can be overcome by intermittent rather than continuous operation. If one assumes that the target loading increases inversely as the focal spot diameter below the 40 μ size which permits 11 kw/mm^2,[3] the maximum average power dissipation for a 50 A spot is 8.8×10^4 kw/mm^2. Suppose that a potential of 6 kv is used with a target current of 0.75 amp giving a target loading of 2.3 by 10^8 kw/mm^2. This means that the target current can flow for 0.04 percent of the time. From figure 28.6 the resultant exposure time is 25 and 2,500 sec for the two magnifications.

[3] Ehrenberg and Spear, Proc. Phys. Soc. *64***B**, 67 (January 1, 1951).

29. Marton's Schlieren Method and Weak Lenses

By F. Bertein [1]

The basis of the paper is the method devised by L. Marton to determine electric and magnetic fields by means of a grid and electron beam. We study whether it can be approached from a systematic point of view as the action of weak lenses on monokinetic electrons.

We shall consider only weak fields, that is, those having a short range, and the potential function $\varphi(xyz)$, which is small compared to the particle accelerating voltage V; we shall put $V=1$ in order to simplify the formulas.

Then according to a first approximation, rays may be considered as straight, but undergoing a refraction in the midplane of the field.

Supposing, as is the usual case, that the beam direction is not very different from Oz, the ray passing through the field at the point $M(xy)$ exhibits an angular deviation δ (xy). This angular deviation has two components along the transverse axes (Ox, Oy) and can be written as a vector or complex quantity (fig. 29.1). The problem is thus to determine this function δ. This determination is as follows: First in the case of the electric field putting

$$\Phi(xy)=\int_{-\infty}^{+\infty}\varphi(xyz)dz \tag{1}$$

(function φ integrated along a parallel to Oz),

$$\delta=\frac{1}{2}\,\mathrm{grad}\,\Phi. \tag{2}$$

Φ is a harmonic two dimensional function, from which some simple features result.

Let us consider an incident beam parallel to Oz; the equation

$$z=\frac{1}{2}\,\Phi(xy) \tag{3}$$

is the one of a wave surface S for the emerging rays; S has a mean curvature equal to zero. Now it is clear, and may be accurately stated, that this curvature, at each point M(xy), appears as a mean convergence of the emerging beam around M (fig. 29.2). Thus the field gives a null mean convergence to all the pencils crossing it, assuming that these pencils do not surround singular points, that is, free charges.

For a magnetic field, formula (2) differs only by the grad Φ coefficient; this coefficient is here purely imaginary. The same observations are valid concerning the mean convergence. Besides, it can be

[1] Ecole Normale Supérieure, Paris, France.

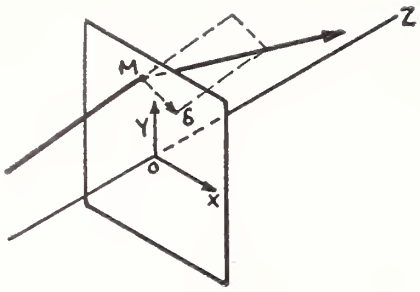


FIGURE 29.1.

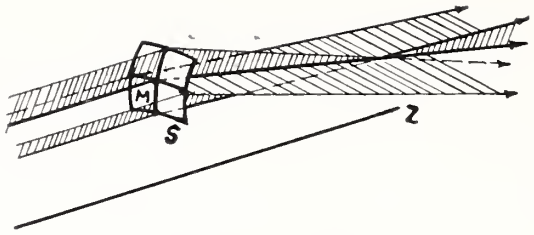


FIGURE 29.2.

observed that the structure of the emerging beams are similar in both cases, electric and magnetic, if the area does not contain charges or poles. That is, the shadow diagram cannot point out, by itself, the kind of field involved.

On the other hand, the neighborhoods of simple charges or poles are different. The shadow has a dilatation (or contraction) in the first case, a rotation in the second. It can then be shown that, putting $s=x+iy$, $\delta=a/\bar{s}$, with a real if charges are present, purely imaginary if poles are present.

But on the other hand, electric and magnetic dipoles show no difference $\delta=b/\bar{s}^2$, where b is complex.

Now it is possible to extend our point of view to take into account devices in which the potential distribution $\varphi(xyz)$ is not alone, but superposed on uniform fields. Let us consider, for instance, the particular case of a flat metallic plate, thin or thick, perpendicular to Oz and separating two uniform fields, E_1 and E_2, parallel to Oz. One or more apertures, of any shape, in this plate will result in a potential $\varphi(xyz)$ superposed on fields E_1 and E_2 (fig. 29.3). The same holds with several parallel plates near each other.

Such devices are electron lenses. It must be pointed out that rays on both sides are parabolas because of the uniform fields, but if, as we assumed, the field $\varphi(xyz)$ is weak, then both parts of a ray are again connected by a deviation $\delta(xy)$, and formulas (1) and (2) are valid. However, $\Phi(xy)$ is no longer an harmonic function; it is determined by $\Delta\Phi=E_2-E_1$.

The surface S is again a wave surface in the condition previously mentioned: S is now a surface having constant mean curvature, equal to $(E_2-E_1)/2$.

Thus one may say that the lens gives a mean convergence having the same value at every point in its aperture. This value depends only on (E_2-E_1) and equals zero if $E_2-E_1=0$ (Marton's method).

Let us consider a particular case. The simplest one is that of a long slit directed along Oy; S is a part of a symmetric cylinder that gives the focusing properties. The beam converges on a focal line, the cylinder axis. Supposing φ to be symmetric about the plane $x=0$

$$\delta=\frac{E_2-E_1}{2}\,x, \tag{4}$$

that is, a focal length

$$f=\frac{2}{E_2-E_1}.$$

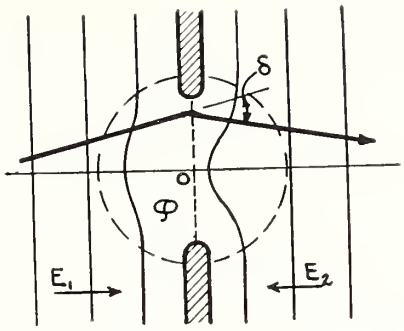


FIGURE 29.3.

This is a well-known result. We even see that the focusing is valid up to the extreme rays at the aperture boundary, in spite of singularities possibly existing in the field.

The other important case is an axially symmetric aperture. S is part of a sphere, and the beam converges on a point, the center of the sphere. The focal length is $f=4/E_2-E_1$.

We see why the convergence is smaller than in the first case; (E_2-E_1) being the same, the mean convergence is the same in both lenses. However, in the first one, it concerned only the x coordinate, instead of being shared between the two coordinates as in the second case.

The case of a slit applies for instance, to ion sources in mass spectroscopy. We see that in the calculations of first approximation, the slit being totally filled by the beam, the focusing holds.

Now it is important to check the validity of this approximation by computing a better one that will give us the deviation from the focusing, that is, the aberrations, in the above cases.

Such an approximation of higher order can be obtained by closer resolution of equations of the electron other than the aberration calculus of strong electron lenses, because we have now to consider the whole aperture, not only the paraxial region. The simplification comes from this fact that lenses are supposed weak; thus here the calculus takes into account developments in term of powers of $\varphi(xyz)$.

One can get formulas without supposing symmetry of any kind in the lenses; but we only shall consider one interesting case, the case of one or more plane electrodes with long slits parallel to Oy, being symmetric about the plane $x=0$.

The first approximation formula (4) must be replaced by the following

$$\delta=\frac{(E_2-E_1)}{2}x-\frac{1}{2}\int_{-\infty}^{+\infty}w\,d\psi,$$

where w and ψ are functions related to φ.

The integration is over z and can be computed in the following form

$$\delta=\left(\frac{E_2-E_1}{2}+a\right)x+a_3x^3+a_5x^5+\ldots,$$

The first term gives exact focusing; the other gives aberrations.

1. If $E_1=E_2=0$ (for example, in a weak microscope lens), focusing

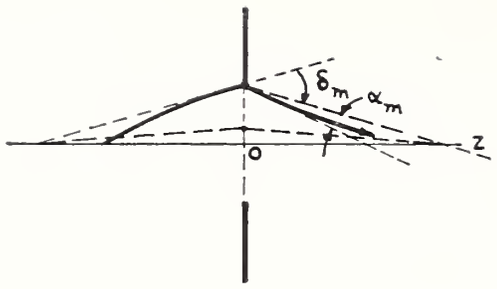


FIGURE 29.4.

depends only on a, that is, terms of second approximation. Consequently aberrations of marginal rays are of the same order of magnitude. The lens can be used only with paraxial rays.

2. If $E_1 - E_2 \neq 0$, focusing appears with the first order; aberrations always are of second order. The lens can be used with all its rays.

For instance, on the axis of a slit in this plate, the extreme angular aberration is $\alpha_m = \delta_m^2/2$ (fig. 29.4).

However, one must take into account the fact that the rays to be connected by δ here are parabolas. That adds a supplementary aberration, but not of a larger order of magnitude. One can then see the focusing properties of slits and its limitation, which will be taken into account in questions involving the thinness of image, when properties like resolving power are concerned.

30. Achievement of Accuracy in Electron-Optical Shadowgraphs of Electromagnetic Fields

By J. Arol Simpson [1]

The electron-optical shadow method of investigating magnetic fields developed in this laboratory under the direction of Dr. Marton consists of measuring the deflections of an electron beam as a function of position in the plane of the object magnetic field. This is done by a method similar to the Ronchi test method used in light optics. The optics of the system and the instrumentation have been described in the literature [2] and are again reviewed in Dr. Marton's introduction to this Symposium. From this measured deflection versus distance curve, or a series of them, the field may be calculated. The question of obtaining precise quantitative measurement of magnetic fields resolves into three parts: (1) Does the shadowgraph give an accurate representation of $K\int H\,dx$ through the object field? (2) Can K be precisely determined from instrumental constants? (3) Can the field be synthesized, given the curves of $\int H\,dx$? I will speak of the first and second problems. Mr. Lachenbruch will discuss the third at some length.

The fact that the electron-optical shadow method is most applicable to fields beyond the reach of conventional field measurements causes no little difficulty in investigating the first two questions. It was hoped at first to make use of a small electromagnet with pointed pole pieces to give a field of small extent and containing sharp gradients. Such a magnet was constructed and Mr. VanBronkhorst and others of this laboratory attempted to measure the field with a small rotating coil magnetometer. After considerable time and labor were expended on it, it was found that the calibration could not be made exact enough and a great deal of difficulty was experienced due to specimen-lens interaction because of the iron of the magnet. We then turned to a small Helmholtz coil of 1 mm radius wound on a copper form with 100 turns per coil of 0.05 mm wire. This was by no means an ideal test object since the field is too large and neither the field integral, amounting to about 0.5 gauss cm, nor its gradient was strong enough to be a severe test of the method. But it contains no iron to cause difficulty and no calibration is needed.

The field in the meridian plane was calculated using well-known formulas. The field was plotted to large scale and the values of $\int H\,dx$ for various values of y were obtained by planimeter integration. This work was done by Mr. Mendlowitz and Mr. Stewart of our laboratory.

The coil was mounted in our modified electron microscope and a series of shadowgraphs taken.

---

[1] Electron Physics Section, National Bureau of Standards, Washington. D. C.

[2] L. Marton and S. H. Lachenbruch, J. Appl. Phys. **20,** 1171 (1949); J. Arol Simpson and A. VanBronkhorst, Rev. Sci. Instr. **21,** 669 (1950).

The usual technique at this point is to plot the position of the shadow measured from an arbitrary base line, which we call a, versus arbitrary mesh numbers. In the absence of a field this will be a straight line whose slope is a measure of the magnification and the constant of the mesh. The ordinate of a point on this straight line for a given mesh number is called A_0. The curve of $(A-A_0)$ versus A where both A and A_0 are taken at the same mesh is then the measure of the beam deflection versus position in the object plane that we sought. In this case, however, the field was of so large extent that several shadowgraphs were necessary to cover it and there was no undistorted mesh visable to give us A_0. Hence we used current-no-current pairs and obtained the distortion caused by the field, $(A-A_0)$, by subtraction. A set of fiducial wires was set axially in the Helmholtz coil, and these allowed us to overlap successive plates. This tedious work was carried out by Messrs. McCraw and Suddeth.

It was in this overlapping process that we discovered the first effect of lens aberrations. As the exposures were made the fiducial wires fell alternately in the center and the two extreme edges of the field of view. Figure 30.1, show the effects of this use of different zones of the lens at different angular apertures obtained by varying the condensor current. Note that both the slope of the lines and the divergence due to the two types of zoning vary with the aperture.

The change of slope with illuminating aperture is of no great importance since it is possible to write valid formulas giving deflection as a function of field integral that is independent of source position, which we at first assumed to be at infinity. The divergence due to zoning on the other hand can lead to grievous errors if care is not taken to work with beams of small aperture.

The question of the effect of inexact focus was made the subject of a separate investigation by Mr. Moreland. In this case the Helmholtz coil was mounted in the electron-optical bench and a through focus series of exposures was made. Figure 30.2 shows the results of this work. The best visual focus, as determined when there was no visual movement of the fiducial wire when the current was switched, corresponded to an objective current of 52 ma ± 5. This corresponds to the position of minimum $A-A_0$ as shown on the graph. A tedious but positive way to determine exact focus would be the

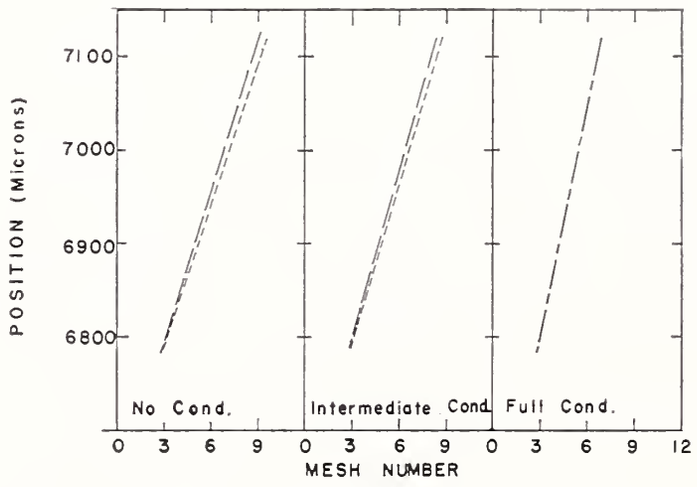


FIGURE 30.1.

——, Wire; — — —, wire 1 and 2.

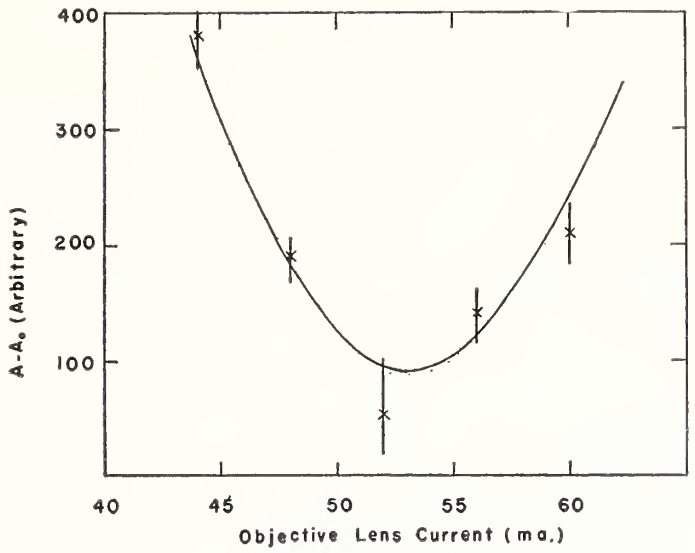


FIGURE 30.2.

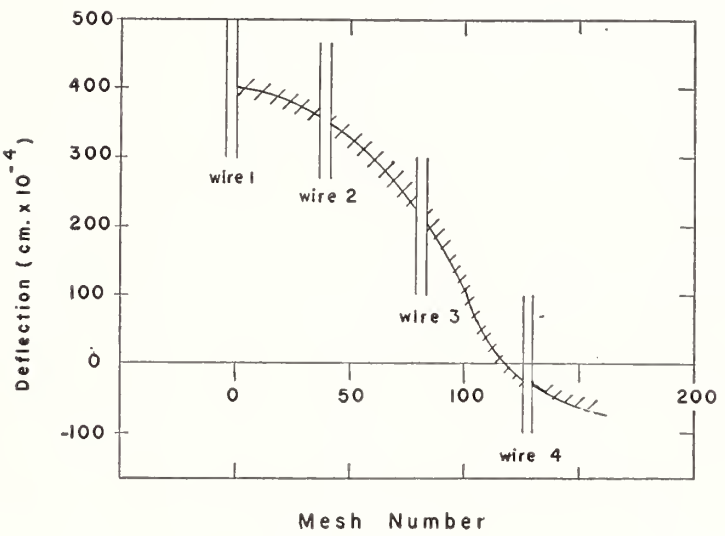


FIGURE 30.3.

determination of this minimum. The minimum is flat enough, however, to focus visually for all but the most exacting work.

The final results of the determination of $K\int Hdx$ is shown in figure 30.3. The agreement with theory is shown by the extent of the shaded portion. The maximum error is 20 μ where the experimental and theoretical curves were matched at only one point. The resetting error in the travelling microscope used to measure the plates is about 5 μ. The systematic deviation at the end of the curve near wire 4 is due to the finite extent of the winding that was not corrected for in the calculation of the field of the coil.

In the case of the second problem, the determining of the scale factor K, the formula we use to determine the value of the constant K involves determination of the magnification of the objective, the distance from object to principal plane of the objective, and the total magnifications of object and the mesh.

All the formulas are inexact to the extent that electron lenses are not thin and hence their principal planes are not easy to define. The formulas require, moreover, that the magnification of the objective must be determined in the object plane of the projector. We

accomplish this in a way made possible by the construction of the RCA EMB microscope by moving the objective and everything above it a measured distance, measured on a dial indicator, by use of the alinement adjustment, and measuring the resulting movement on the final plate by use of a travelling microscope. This gives the projector magnification that divided into the total magnification gives the objective magnification. If care is taken with these measurements (the distortion of the projector is very bothersome), it is possible to obtain agreement between the calculated and measured value of K within the uncertainty of the position of the principal planes. We have determined K to within an error of 8 percent.

We have thus shown that when proper precautions are taken, the electron-optical shadow method will give the value of $K\int Hdx$ along a straight line to within 5 percent and the value of $\int Hdx$ to within 10 percent for fields totally inaccessible to other methods of measurement. The greatest error is the lack of precision of measurement of instrumental constants.

31. Interpretation of Electron-Optical Shadowgraphs

By S. H. Lachenbruch [1]

Mr. Simpson has given a description of the experimental portion of the shadow method. He has shown how the deflection of electron paths by an unknown magnetic field is determined using an electron-optical system. There remains only the problem of deriving from these results the actual field distribution.

Figure 31.1 summarizes this theoretical problem. The coordinate system is attached to the magnetic field, which is directed normal to the plane of the figure (just as it would be if the figure were an equatorial plane). The relative orientation of field and optical system may be varied experimentally. The curve L represents a typical electron trajectory through the field. It may be considered as approximately straight, since the deflection angle ϕ is necessarily small—otherwise the method, which involved geometrical optics, would not have been applicable.

Thus for each of a certain set of directed lines L in the plane, each identified by coordinates (p,β), we have an experimentally determined deflection angle ϕ, and hence a value of

$$I(p,\beta)=\int_{L(p,\beta)} H(r,\theta)ds=\sqrt{\frac{e}{2mV}}\,\phi, \tag{1}$$

the integral of H along that line. We want to derive from this the magnitude and sign of the unknown H at an arbitrary point Q with polar coordinates (r,θ).

For the frequent case where the field is known to have axial symmetry, we can choose the origin as the center of symmetry. Then I is independent of β, and also H is independent of θ, so the problem is greatly simplified.

The methods that have been used or considered in this laboratory might be divided into two general categories.

The first category consists of those methods in which we assume at the start that H may be expressed in some specific analytical form

$$H(r,\theta)=F(r,\theta,a_1,a_2,\ \ldots\ ,a_n), \tag{2}$$

involving one or more unknown parameters or constants a_j. This form is generally chosen from known or predicted properties of the field.

Integrating the specified function F in advance gives a predicted analytical form for I—as a definite function of $(n+2)$ quantities:

$$I(p,\beta)=\int_{L(p,\beta)} F(r,\theta,a_1,\ \ldots\ ,a_n)ds=G(p,\beta,a_1,\ \ldots\ ,a_n). \tag{3}$$

[1] National Bureau of Standards, Washington, D. C.

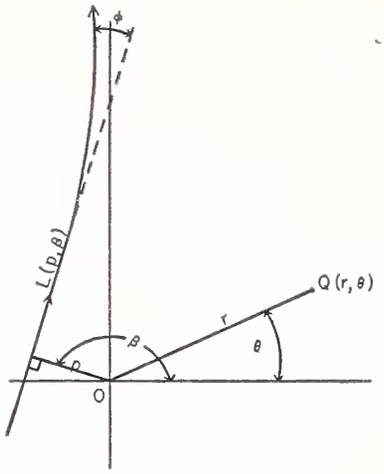


FIGURE 31.1.

We wish to solve for the n quantities a_j by matching this function against the values of the I-function obtained experimentally for different values of p and β. One way to do this is statistical, by applying least squares to the experimental results, as indicated by the equation

$$\sum_{k=1}^{m}[I_k - G(p_k,\beta_m,a_1, \ldots, a_n)]^2 = \text{minimum}. \tag{4}$$

Another way which is sometimes possible is to choose n electron paths in such a way that the resulting equations

$$G(p_k,\beta_k,a_1, \ldots, a_n) = I_k, \qquad k=1,2, \ldots, n \tag{5}$$

can be solved for the a's.

In the statistical method, the closeness of fit can usually be used as a partial check on the validity of the functional form F chosen. But *how good* a check is not known until we consider the question of uniqueness, which will be mentioned later.

The following are examples of the widely different techniques included under the category just described: (In each case the resulting I-function, as obtained by integration, is also given.)

(1) H may be taken to be a simple continuous function involving two parameters, for example.

$$H(r) = a_1/(r^2+a_2^2)^{3/2} \tag{6}$$

$$I(p) = 2a_1/(p^2+a^2_2). \tag{7}$$

(2) H may be expressed as a series, the unknown a's being the coefficients:

$$H(r) = a_1/r^2 + a_2/r^3 + \ldots, \tag{8}$$

$$I(p) = \pi a_1/p + 2a_2/p^2 + \ldots. \tag{9}$$

(3) The plane may be somehow divided up into regions R_k, and H taken to be a step-function constant in each region, the a's being these

constant values:

$$H(r,\theta)=a_k \text{ where } (r,\theta) \text{ is within } R_k;\ k=1,2,3,\ \ldots,\ n \tag{10}$$

$$I(p,\beta)=\sum_{k=1}^{n} a_k \lambda_k(p,\beta). \tag{11}$$

In eq (11), λ_k refers to the length of the segment of intersection of a trajectory with the region R_k.

We shall return later to these examples. Meanwhile we turn to an entirely different approach, of an analytical nature. Here the unknown function H is obtained directly, theoretically at least, by mathematical operations on the experimentally obtained function I, with no a priori assumptions regarding the functional form of H, and no unknown constants for which to solve.

We take first the case of axial symmetry, so that H is a function of r alone. By definition, I is given in terms of H by

$$I(p)=2\int_{|p|}^{\infty} H(r)r\,dr/(r^2-p^2)^{1/2}=T(H(r)). \tag{12}$$

But this can be reduced to a standard integral equation, Abel's equation. From the known solution of that equation, we find that the only continuous function H satisfying this equation is

$$H(r)=\frac{1}{\pi}\int_{r}^{\infty} I'(p)\,dp/(p^2-r^2)^{1/2}=T^{-1}(I(p)). \tag{13}$$

So we have, at the start, an explicit expression for the desired field function in terms of the function I which is known from experiment. The only remaining problem is the practical one of devising simple methods of evaluating the integral of this empirical function for different values of r.

Note that we obtained H essentially by inverting a *linear integral operator* T in function-space. Also, the questions of existence and uniqueness are answered at once.

Finally we turn to the general case, removing the symmetry condition. For any fixed r, the field H must be a periodic function of θ (with period 2π), and so it has a unique Fourier series expansion with functions of r as coefficients:

$$H(r,\theta)=\sum_{n=-\infty}^{\infty} H_n(r)\ \exp\ (in\theta). \tag{14}$$

Similarly, I has a unique Fourier expansion with functions of p as coefficients:

$$I(p,\beta)=\sum_{n=-\infty}^{\infty} I_n(p)\ \exp\ (in\beta). \tag{15}$$

If we substitute (14) and (15) into the definition of I, and equate

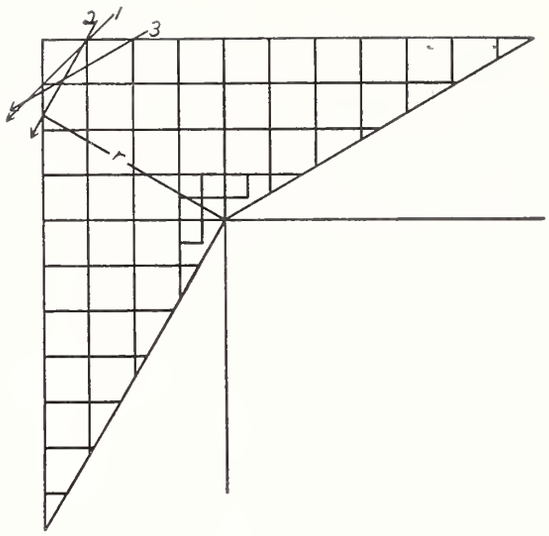


FIGURE 31.2.

coefficients of like powers of exp $(i\beta)$, we get the integral equations

$$I_n(p)=2i^{-n}\int_{|p|}^{\infty} H_{-n}(r)\cos\left(n\cos^{-1}p/r\right)\frac{rdr}{(r^2-p^2)^{1/2}}. \qquad (16)$$

We find that these have unique solutions and may be solved in turn for the H_n. If, then, we have enough experimental data to define the function $I\ (p,\beta)$, and therefore its Fourier coefficients, we will have the desired function $H\ (r,\theta)$ in the form of a Fourier series.

As a byproduct we have again settled the question of existence and uniqueness, this time for the general case. It is to be observed that in this general case we must use what amounts to a 2-parameter family of rays; otherwise the solution obtained for H by *any* method would not be unique.

We now return to the first category, for further comment on the examples (6), (8), and (10). The first method is most appropriate when our knowledge of the field-producing object tends to suggest a certain analytical form. The particular form given by (6) and (7) is for the field in the equatorial plane between two point poles an unknown distance apart. When $a_2=0$, it reduces to the dipole field.

The second method, while it has a certain amount of generality, must be used with some caution. A series of form (8) could never be used in the neighborhood of $r=0$, for example.

In the third method (10), with any mode of subdivision, each experimental measurement gives rise to a *linear* relation (11) among the a_i. Figure 31.2, prepared by Mr. Simpson and Mr. Ford of our laboratory, shows how one may go about choosing a set of n trajectories such that the resulting simultaneous linear equations have a triangular matrix and so can be solved readily, one by one.

Although the choice of subdivision is not in general critical, some times known properties of the field suggest a special one. If the field is known to have axial symmetry, for example, it is constant along any circle about the center of symmetry, and so the most appropriate type of subdivision would be into rings bounded by such circles. The field surrounding a magnetized wire has been computed by this method, and shows good agreement with the result obtained by the first method, in which we assumed that the field was that of a pair of point poles.

32. Characteristics of Symmetrical Magnetic Electron Lenses [1]

By G. Liebmann [2]

At the Electron Microscopy Conference in Delft in 1949, the author described improved numerical and experimental methods for the investigation of electron lenses. The field distribution is measured to a high degree of accuracy (1 to 2 parts in 10,000) with a resistance network analog, figure 32.1.

The electron trajectories and lens aberrations are then found by step-by-step methods. In this way, accurate information can be obtained on any given electron lens used in practice, without appeal to mathematical idealizations or the use of approximations.

These methods have meanwhile been applied, partly in collaboration with Miss E. M. Grad, to the investigation of a series of symmetrical magnetic lenses of varying relative pole piece separation S/D (fig. 32.2); this figure also shows the axial field distributions. The main objects of this study were the collection of working data and design rules for electron lenses used in electron microscopy, and the investigation of questions of the best possible practical electron microscope objective and the best projector lens.

There is no difference in the lens constants if a lens is used as an objective or as a projector in electron microscopy, as long as the lens excitation is weak or only moderately strong. At high values of lens excitation, however, one has to distinguish between the focal length, etc., of the lens used as projector (called f_0) and used as objective (called f_1), as shown in figure 32.3.

The excitation parameter used here is

$$k^2=0.022H_0^2R^2V_r^{-1}, \tag{1}$$

where V_r is the relativistically corrected accelerating voltage, H_0 is measured in gauss and the lens radius R in centimeters. One can also write

$$k^2=\beta(NI)^2V_r^{-1} \tag{1a}$$

where β is a known function of the ratio S/D.

A comparison of the four lenses $S/D=0.2$ to $S/D=2$, used as *objectives*, is given by the next two figures in which f_1/R, z_1/R (fig. 32.4), and C_s/R (fig. 32.5), are plotted as function of k^2. It is seen that in all cases the stronger the lens the shorter its focal length and the smaller its spherical aberration constant C_s/R. Most practical electron microscope objectives are worked with excitation parameters around $k^2\cong 0.5$ to $k^2\cong 1$. Near this excitation value, the lens

[1] The full results of this study are being published in the Proceedings of the Physical Society, London.
[2] Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berkshire, England.

257899—54——19

$S/D=1$ happens to have a spherical aberration constant C_s/R which is very nearly the same as that calculated by Glaser for his assumed bell-shaped field distribution if the half-width value for this lens, $a=1.13R$, is inserted: Hence, the many earlier discussions of possible electron microscope performance based on Glaser's "bell-shaped" field distribution (with assumed numerical values of the half width a) are confirmed by our evaluation of the lens constants of "real" lenses.

Projector lenses are mostly operated near the point of minimum focal length; a comparison of four lenses $S/D=0.2$ to 2 is given in figure 32.6. If for each lens the focal length is expressed relatively, as the ratio $f/f_{\min}$, and plotted against the relative excitation parameter $k^2/k^2_{\min}$, one finds that all these four lenses, as well as the longer lens $S/D=4$, can be represented by one single curve. This "general f-curve" is shown in figure 32.7.

The shortest focal length is given by

$$f_{\min}=BV_r^{\frac{1}{2}}H_{p\,\max}^{-1}, \tag{2}$$

where the factor B depends only on S/D, and $H_{p\,\max}$ is the highest field strength in the pole-piece gap that iron saturation will allow. In figure 32.8, curve 1 gives a plot of the factor B, and we see that the higher the ratio S/D the shorter the obtainable absolute focal length. Curve 2 shown in figure 32.8 gives the optimum lens radius, which is uniquely determined for each value of S/D. As numerical example, take $V=60$ kv and $H_p=26{,}000$ gauss. Then $f_{\min}=0.075$ cm, which is exactly Ruska's measured value for these excitation data.

An important fault in projector lenses is distortion. The radial distortion is proportional to the square of the distance of the image point from the axis and to the distortion constant C_d. If the distortion constant is plotted relatively, as the ratio C_d/C_{do}, where C_{do} is the distortion constant for weak excitation, one obtains the set of curves shown in figure 32.9. The significant point is that for an excitation value around $k^2_{\min}$, C_d changes its sign, i. e., the usual pincushion distortion changes to barrel distortion. Any projector lens can therefore be operated such as to be free from radial distortion (provided that iron saturation does not limit the maximum value of k^2 that can be used).

We return now briefly to the question of the *objective lens*, and the conditions for the lens of lowest (absolute) spherical aberration and lowest chromatic aberration. The earlier figure shows that C_s/R falls rapidly at first with increasing value of k^2, but only slowly for very high values of k^2. As $k^2 \propto R^2$, C_s (in absolute units) will at first drop as the lens radius is increased, go through a minimum (for $R=R_0$) and then rise again. The value of $k^2=k_0^2$ for which the minimum of C_s is reached depends on S/D; this is shown by table 32.1 We find then (in a similar way as for the projector lens of shortest focal length) that there is an expression

$$C_{s\,\min}=bV_r^{\frac{1}{2}}H_p^{-1} \tag{3}$$

where the constant b depends on S/D. We see that the longer lenses are more favorable, but the number of ampere turns increases rapidly for $S/D>1$, and the immersion of the specimen in the lens field be-

comes rather deep, which is undersirable as then a large part of the objective lens field becomes part of the illuminating system. A practical compromise would be to use a lens geometry $S/D \simeq 1$, and figure 32.10 shows the value of $C_{s\ \min}$ which could be reached in a practical electron microscope objective. We see that values of $C_s = 0.04$ cm to $C_s = 0.06$ cm would be practicable. These are values smaller by a factor of 10 to 20 than have been used hitherto. The combination of spherical aberration error and diffraction error, setting a fundamental limit to the resolving power of electron microscopes, would lead to a resolution limit of 3.5 to 4 A, for accelerating voltages of 60 to 90 kv. (Similar values had been predicted by Cosslett and by Glaser, on the basis of the "bell shaped" field, but in our estimate we adhered strictly to *measured* lens field distributions.)

Table 32.1.

S/D	H_o/H_p	k_0^2 (approximate)	C_s/R	b	NI (relative)
0.2	0.263	9	0.170	13.1	0.85
0.6	0.648	6.5	0.220	5.8	0.88
1	0.834	5	0.267	4.8	1
2	0.964	4.5	0.300	4.5	1.60

The old problem of the objective of shortest focal length, which would give low chromatic aberration as $C_c \leq f$, was taken up recently by Le Poole and van Dorsten; our lens data can be applied to this question, and in a similar way as in the case of the lens of lowest spherical aberration one can show that there is for each value of S/D a value of k^2 that yields the shortest possible (absolute) focal length, and another slightly higher value of k^2 which gives the lens of smallest (absolute) chromatic aberration constant C_c. The result is shown in figure 32.11. If we insert the following numerical values: $V = 80$ kv ($V_r = 86$ kv), $H_p = 24{,}000$ gauss, we find $f_{\min} = 0.084$ cm, in excellent agreement with van Dorsten's experimental value of 0.085 cm. Figure 32.11 also shows that the spherical aberration constant C_s of a short focal length objective is fairly small.

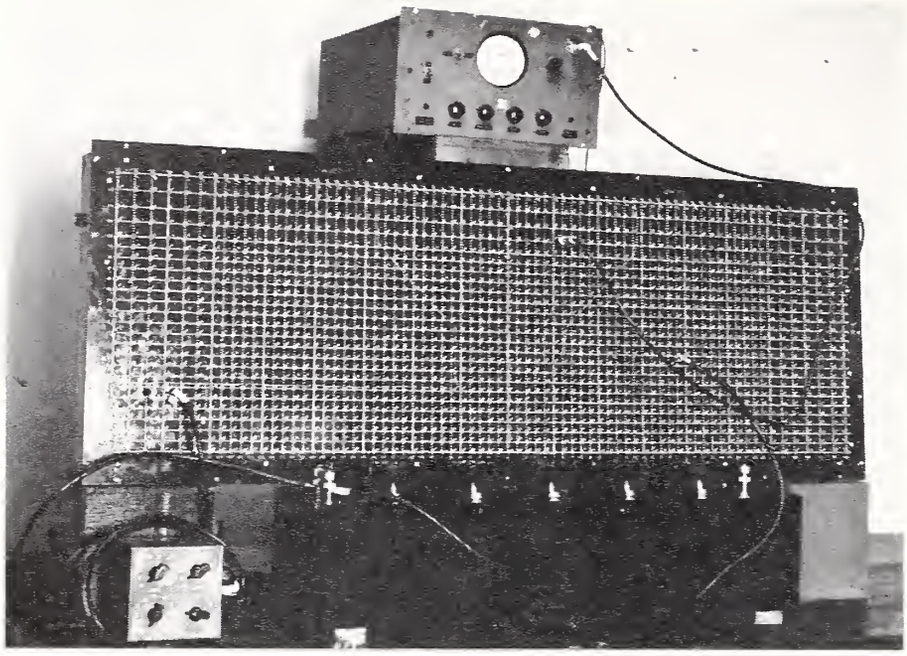


FIGURE 32.1.

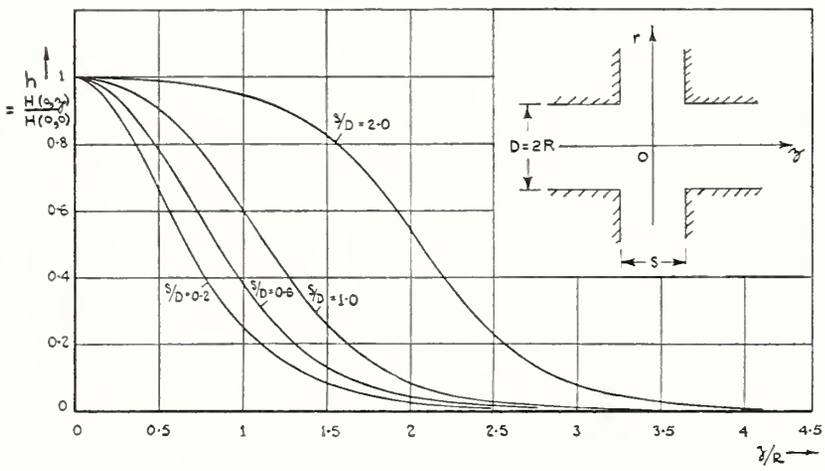


FIGURE 32.2.

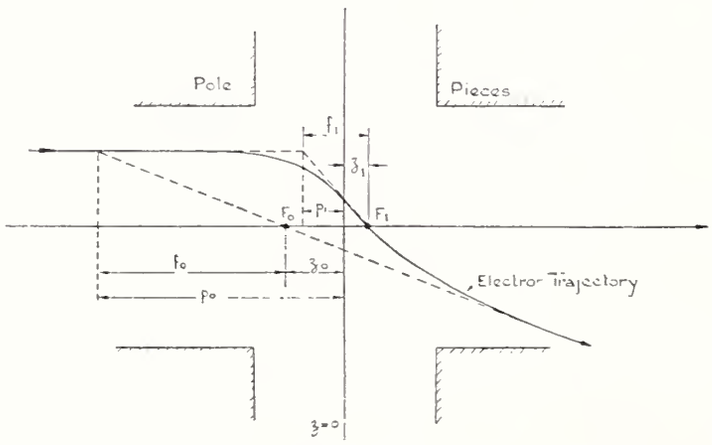


FIGURE 32.3.

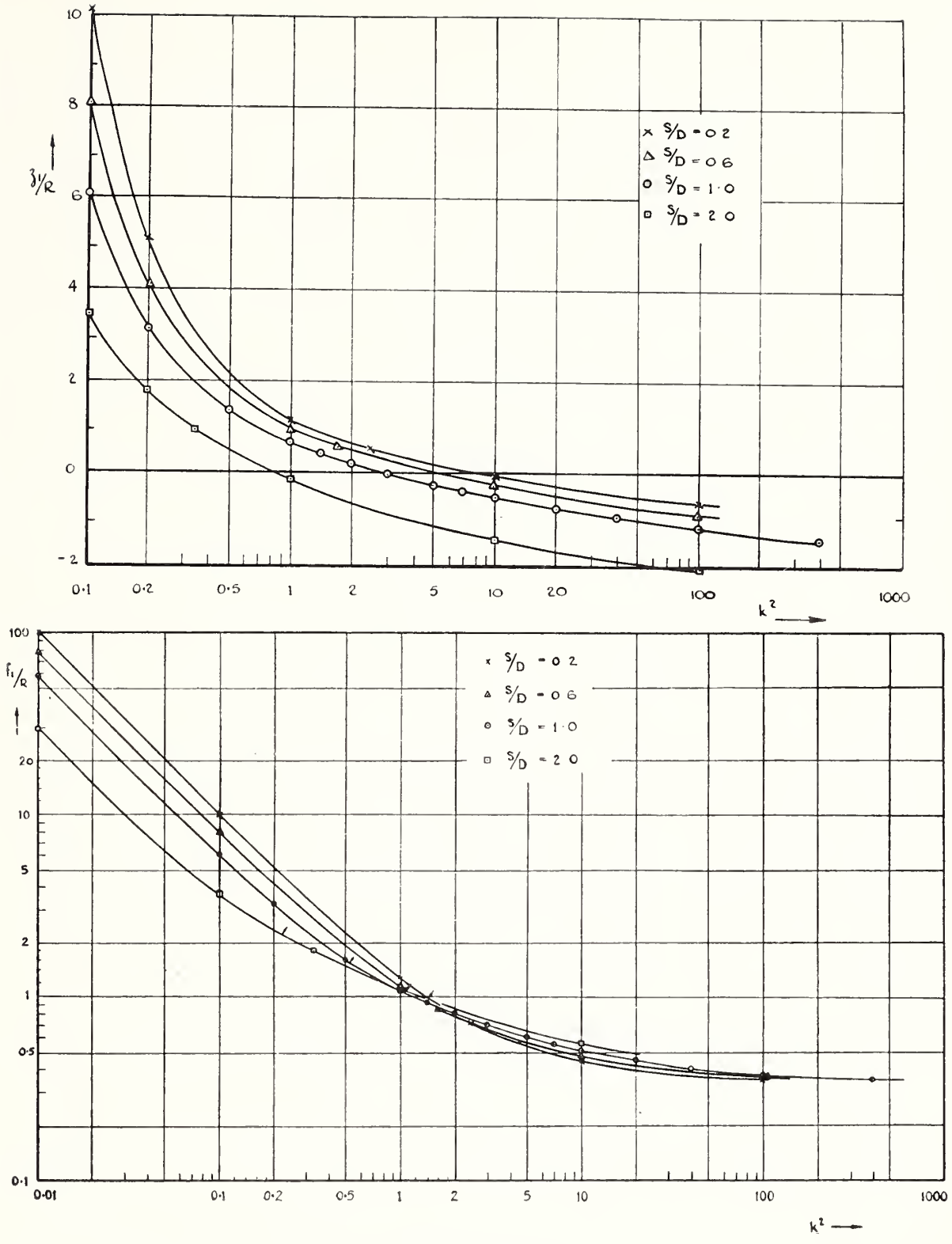

10
8
6
4
2
0
-2
β_1/R
× S/D = 0·2
△ S/D = 0·6
○ S/D = 1·0
□ S/D = 2·0
0·1
0·2
0·5
1
2
5
10
20
100
1000
k^2
100
20
10
5
2
1
0·5
0·1
f_1/R
0·01

FIGURE 32.4.

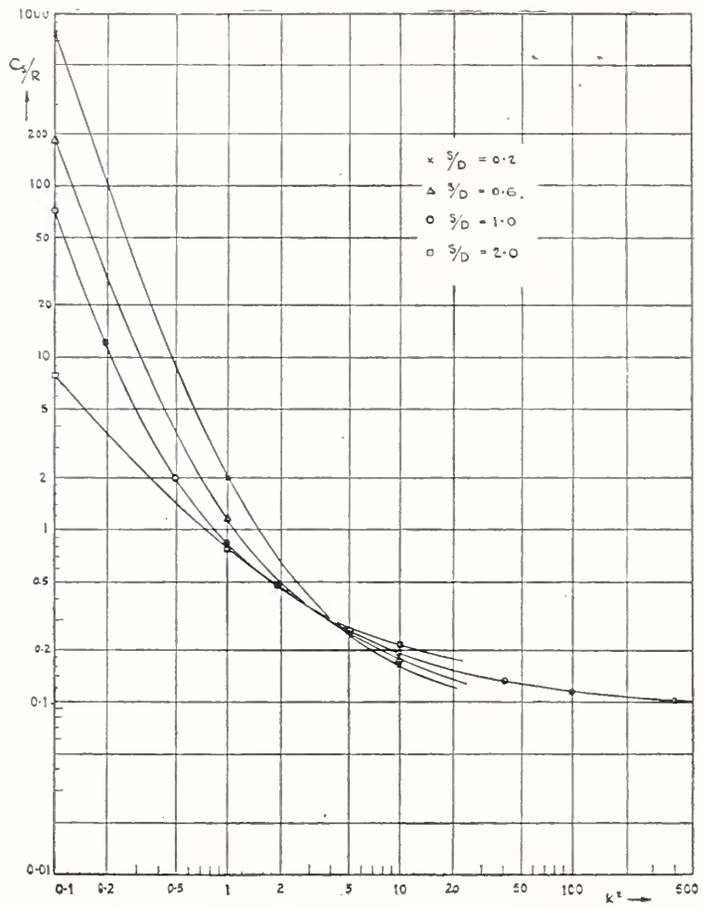


FIGURE 32.5.

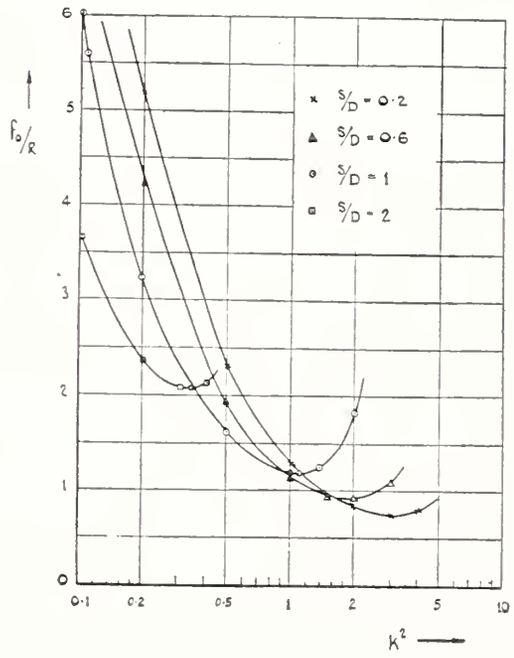


FIGURE 32.6.

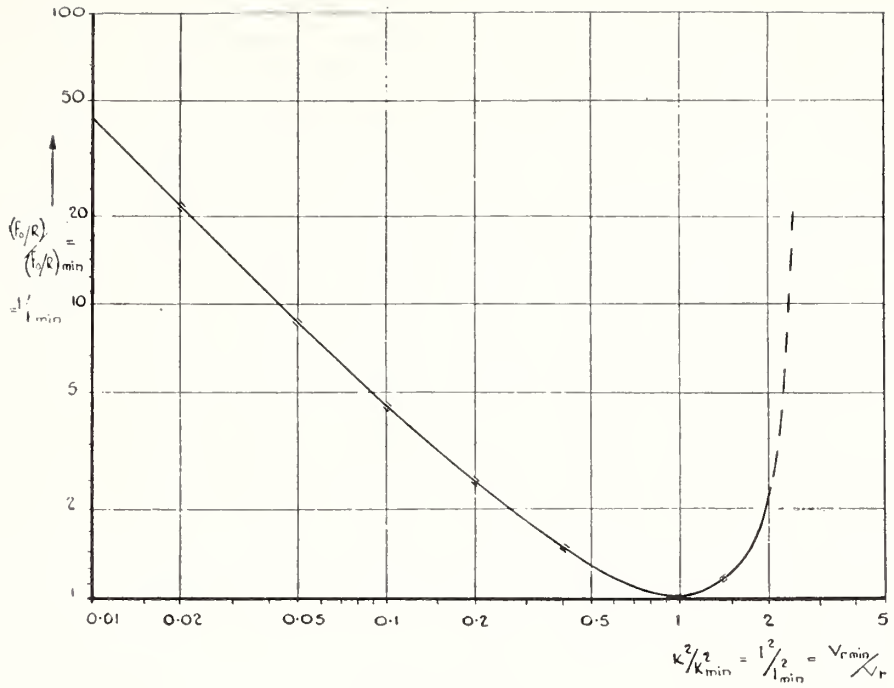
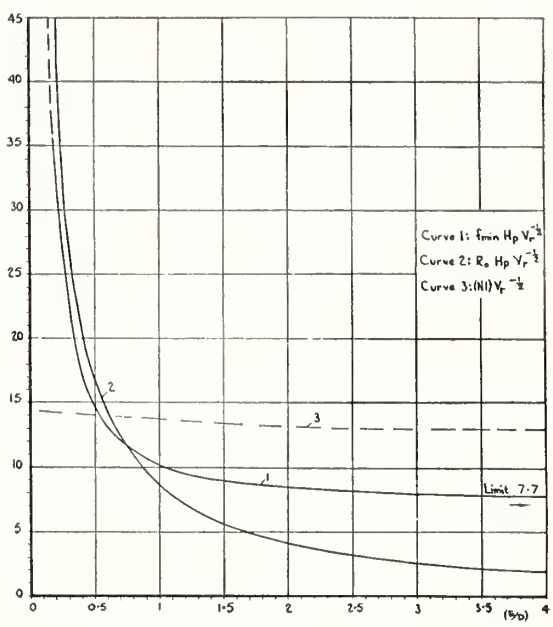


FIGURE 32.7.

FIGURE 32.8. *Minimum possible focal length of EM projector lenses for given H_p and V_r.*



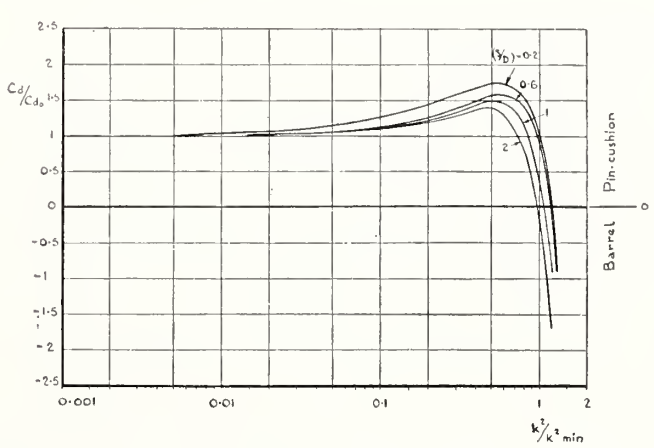


FIGURE 32.9. *Relative scale distortion constant.*

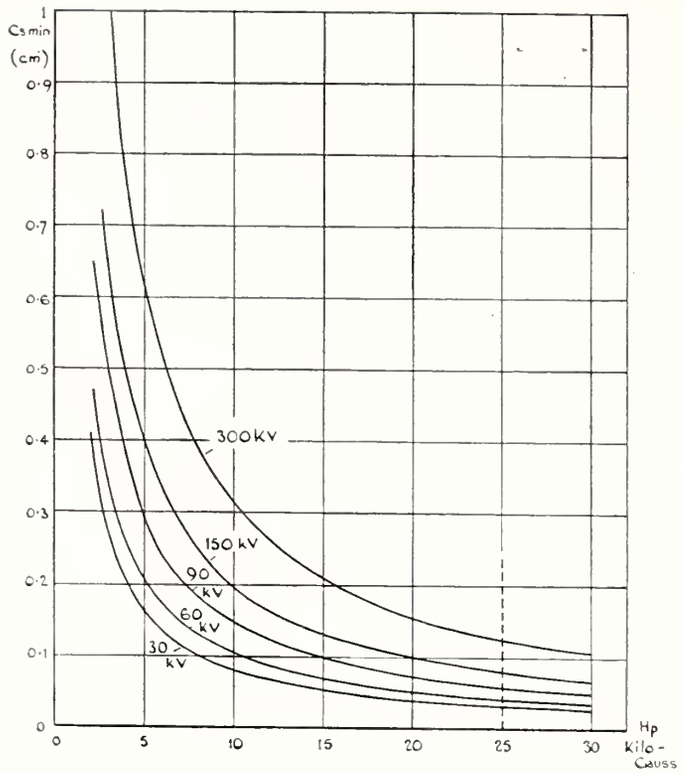


FIGURE 32.10.

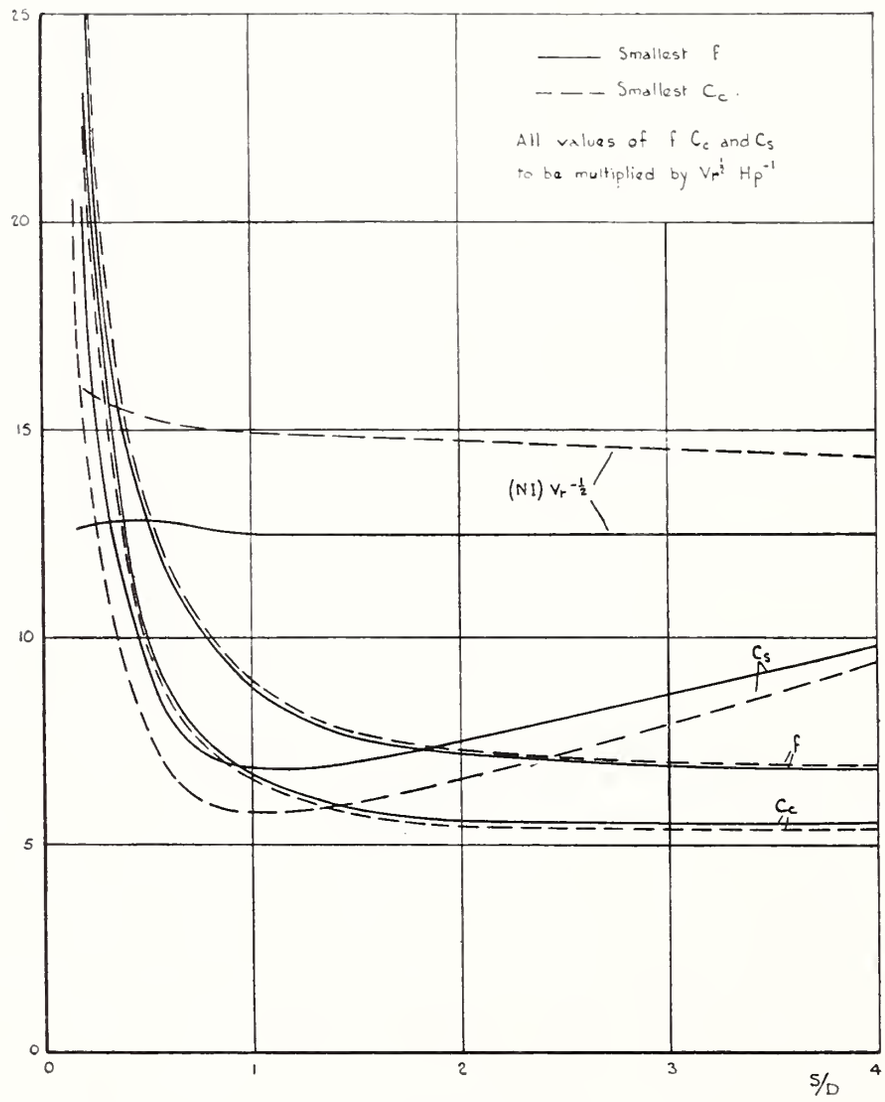


FIGURE. 32.11.

33. Present Trends in Electron Microscopy

By V. E. Cosslett [1]

Introduction

My intention is to discuss the directions in which electron microscopy is moving at present, and to try to forecast how far it may progress along them. It is always a little difficult when one is in the middle of developments to see directions clearly, but I feel that we have begun to recognize, in the past year or so, the lines along which the most fruitful developments are likely to take place. The field is large, and the abstract of this contribution promised more than I can possibly cover. I shall restrict myself, in tune with the general character of the conference, mainly to the electron optical aspects of the future of electron microscopy, with some reference to instrumental factors. Indeed, since most of those actively advancing the subject are assembled here, my paper inevitably becomes largely a survey and commentary on many of the contributions appearing on the program. It may help to put them in perspective, and in some cases to provide an elementary introduction to new ideas.

Limits of Present Methods

In the first place let us consider how far present techniques will carry us. It seems that we find difficulty in getting beyond a practical resolution limit of somewhere between 10 and 20 Å—the exact figure depends on one's degree of optimism and perhaps also on nationality. It is difficult to estimate precisely what resolution is obtained in this region of size, but there is no doubt that 20 A has been exceeded by several workers. The immediate question is what limitations stand in the way of achieving better resolving power, and indeed whether existing instruments can be expected to do better?

It is necessary to be rather more precise as to what is meant by resolution. Professor Glaser's paper in a previous session has clarified considerably the distinction between the limit of detection and the true resolution according to the accepted optical definition. A single opaque object will be imaged as a diffuse blur, the radius of which will be greater than that of the true image; the disparity in radius will increase as the size of the object is made smaller and smaller. But the presence of some sort of object will be detectable so long as the contrast of this blur against the general background in the image is above the discriminating power of the eye or of a photometer. On the other hand, the classical method of defining resolution is to consider two neighbouring object points, and to assume them to be separately distinguishable so long as the drop in intensity between their

[1] Cavendish Laboratory, Cambridge, England.

two images is greater than 25 percent of the peak intensity. The separation of the two peaks when this limit is reached is said to be the linear resolution. Clearly it will be considerably greater than the size of the smallest detectable isolated object, as Glaser has pointed out; he gave a factor of three for a circular object. However, the individual peaks of the pair are themselves much broader than the object points which formed them, so that it will remain true that when we have a resolution limit of (say) 12 A, we may be distinguishing particles as small as 4 A, or they may be larger particles which are only semi-opaque. A resolution of atomic dimensions would thus call for test objects of sub-atomic size; fortunately the nucleus, opaque to electrons, fulfills this demand.

It is thus clear that we must be careful about relating an observed linear resolution to the true size of the particles resolved. There is the further question as to whether this problem has any real meaning where atoms are concerned. It is very doubtful if individual atoms can ever be imaged, even if one succeeded in laying them down in isolation on a supporting film. It is not easy to think of suitable test objects for such a high resolution. A two-dimensional lattice might be more readily obtained, but resolution would not then have the same meaning owing to the special nature of the interaction of electrons with a periodic structure [1].[2] The best prospect for utilizing an atomic resolution seems to be for spotting a heavy impurity atom in a lattice or amorphous group of lighter atoms. This would involve the limit of detection, rather than the true resolution. It is another question whether the isolated atom could be at all closely localized, owing to its disturbance by impact of the imaging beam, as de Broglie has discussed [2].

Leaving this problem to the future, let us now consider some of the efforts made to discover how high a performance is to be expected of present lenses. Glaser's work on the bell-shaped field is well known, which gave the aberrations in detail and the focal length in different operating conditions [3]. Towards the end of the war I used his data to calculate prospective resolutions [4], and found one could expect a limiting value as small as 5 A. Glaser has shown in this meeting that a more refined treatment gives 4 A. However, such estimates were open to the objection that the bell-shaped field is not that of an actual lens, being taken for reasons of mathematical convenience. It is satisfactory that results are now available for the sort of magnetic lens used in practical electron microscopes.

Liebmann has lately made a detailed investigation [5] of the focusing properties of a typical magnetic lens in which the pole-piece separation is of the same order as the bore. Careful measurement of the field by the resistance network, followed by computation, gives the focal length and aberrations. Fig. 33.1 shows how the focal length, plotted against the lens power parameter k^2, falls continuously without reaching a minimum value as in a bell-shaped field. Different ratios of gap to diameter give curves differing only in gradient. The practical limit to the focal length is about $R/3$, where R is the internal radius of the polepiece. Fig. 33.2 shows similarly the spherical aberration coefficient, which again shows no minimum such as appears for a bell-shaped field. The limiting value of the coefficient proves to be slightly smaller than that deduced by Glaser. The chromatic aberration curves are of similar shape.

[2] Figures in brackets indicate the literature references on p. 303.

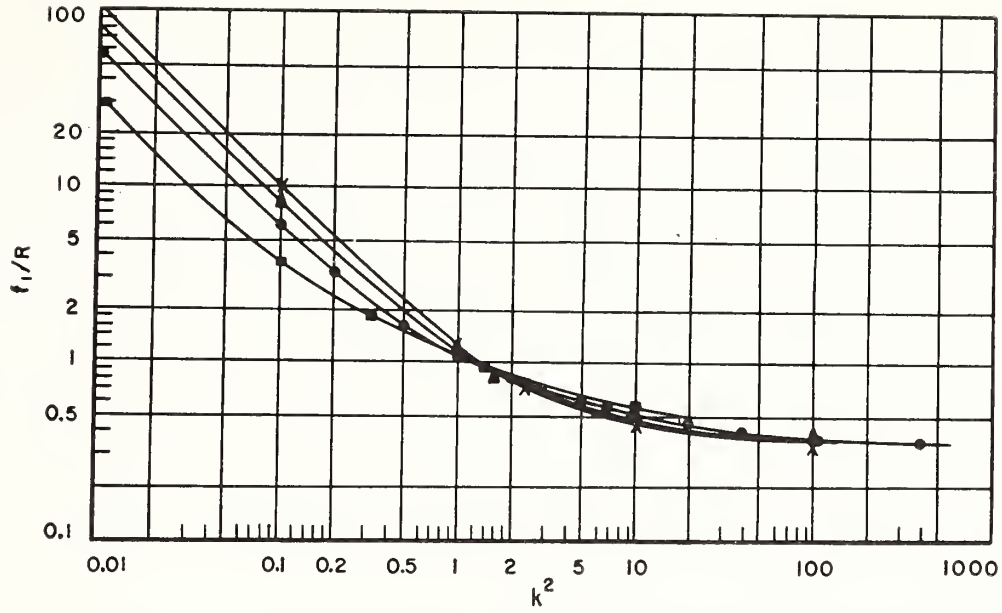


FIGURE 33.1. *Variation of focal lengths (divided by lens radius R) with lens power in magnetic lens, for different ratios of gap width S to bore D.*

Liebmann and Grad, Proc. Phys. Soc. London [B] **64,** 956, (1951), figure 15. ×, $S/D=0.2$; ▲, $S/D=0.6$; ●, $S/D=1.0$; ■, $S/D=2.0$.

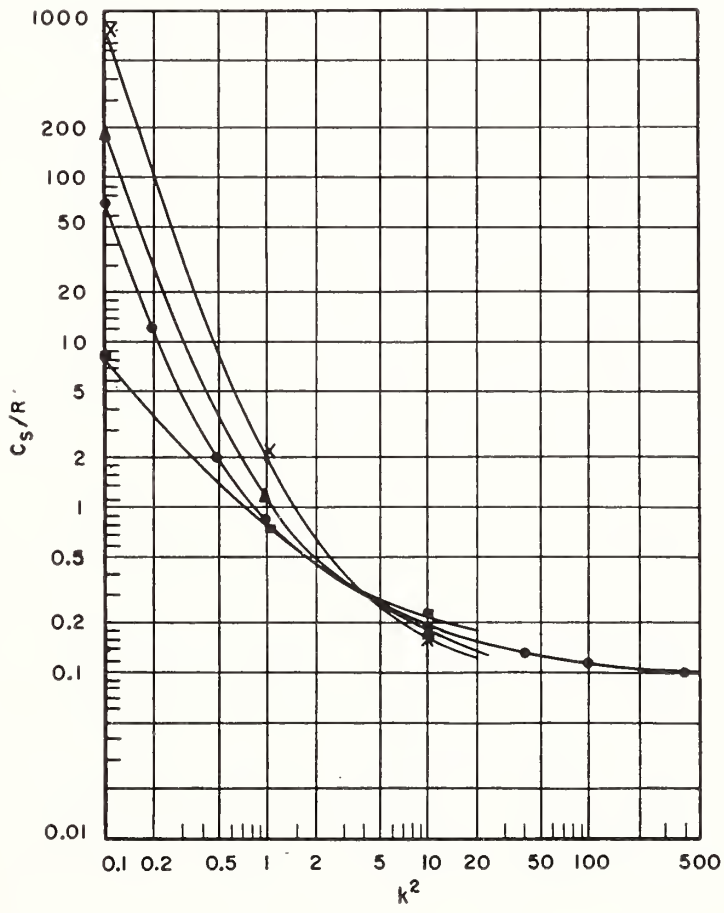


FIGURE 33.2. *Variation of spherical aberration (divided by R) with lens power in magnetic lens for different ratios of gap width S to bore D.*

Liebmann and Grad. Proc. Phys. Soc. London [B] **64,** 956 (1951), figure 20. ×, $S/D=0.2$; ▲, $S/D=0.6$; ●, $S/D=1.0$; ■, $S/D=2.0$.

From these data, Liebmann finds a limiting resolution of 3 A when the optimum compromise between spherical aberration and diffraction is adopted. We are thus justified in the belief that even with the imperfect lenses now available we should be able to reach a resolution three or four times better than the best so far reported. The residual astigmatism can be corrected, and the chromatic error made almost negligible by attention to the stabilization of the electrical supplies, so that we must enquire further as to what are the limiting factors in practice. It seems probable, as Haine will show in a later contribution, that the main difficulty is mechanical and thermal instability of the specimen. There are also some residual electrical troubles, since a stabilization of the order of 1 in 10^5 is called for over a period of a minute. These matters are now receiving detailed experimental investigation, but we have to recognize that they may prove intractable, so that we may not be able to utilize a resolving power that has been made potentially available.

Correction of Spherical Aberration

These difficulties will be discussed in a later session, so that we may proceed to enquire what prospects exist for eliminating spherical aberration altogether, so as to make possible a resolution approaching the limit set by the wave-length of the electron beam (1/20 A for 60kv electrons). As was shown by Scherzer [6], spherical aberration cannot be reduced to zero in a system that has rotational symmetry, contains no space charge and to which static fields only are applied. Many workers have reviewed the consequences of relaxing one or more of these conditions, and foremost Scherzer himself [7]. It is typical of the present stage of development that the possibilities that have been discussed theoretically since 1947 are now in process of experimental trial. One possibility has been investigated by Marton,—the introduction of a controlled space charge. If the charge density varies radially in the right way, the marginal rays will be repelled further from the axis, so that the greater refractive power of the outer zones of the lens (which is responsible for spherical aberration) may be neutralized. Such a charge distribution occurs in a magnetron, but the practical difficulties of controlling it with precision are great. However, it is encouraging to learn from Gabor that recent work by Ash and himself has shown that such a cloud of charge would not have the diffusing effect on the image, similar to the action of ground glass, that was indicated by the results of Marton and Reverdin [8]. But it still remains to be shown that practical use can be made of such a system.

The second possibility is even more difficult to realize in practice—to apply rapidly varying fields to the lens. Looking upon the beam now as a spherical wave front approaching the lens, it is in principle conceivable that the strength of the lens could be varied in the short time between the arrival of the central and outer parts of the front. If the outer zones of the lens were made progressively weaker as the wave reached them, the spherical aberration could then be compensated. However, not only are the technical requirements very severe in the millimeter wavelength region that would be involved, but it has been shown that the magnitude of the effect would not be sufficient to correct completely the degree of spherical aberration encountered in practical lenses [9].

The remaining alternative is the abandonment of rotational symmetry in the lens system, which may take two different forms. It may either mean making the apertures of the lens elements of non-circular shape, in effect introducing cylindrical lenses in the optical sense of the term, or it may involve curving the lens axis instead of it being straight as in normal optical devices. The latter procedure involves new methods of mathematical treatment, and it is encouraging that such attempts have recently been made by Gabor [10] and by Sturrock [11]. I shall not pursue this in detail, except to say that reliable methods are now available for discussing the motion of electrons in arbitrary electromagnetic fields. It will be enough to explain the physical principles of the proposed correction devices, and to indicate how far experimental work has progressed.

The most important work on the use of noncentered systems has been done by Scherzer himself. He has shown that it is possible to introduce astigmatic components into an imaging system in such a way as to correct the spherical aberration, and then to remove the astigmatism again. The essence of the method is to introduce a set of 8 radial electrodes bearing alternately positive and negative potentials (fig. 33.3). In the planes of the positive electrodes the electron beam will be attracted outwards from the axis, and it can be shown that the radial variation in this force is just that required to overcome spherical aberration. But, so long as the beam is of symmetrical form in cross section (i. e., is cylindrical in this region), the effect of the negative electrodes will be such as to double the spherical aberration in the diagonal planes. Scherzer's brilliant idea was to make the beam highly asymmetrical, by making the objective astigmatic and placing the correcting electrodes at one of the line foci thus formed (fig. 33.4). Correction is then obtained in one of the positive planes, and the effect on the beam is small in all other planes. A second set of electrodes at the other astigmatic focus will correct the aberration in the second positive plane. Scherzer showed that such an arrangement also ensures correction in all other planes, except for a small residual increase in aberration in the diagonal planes, which is removed by inserting a third set of correcting electrodes. The whole train of lenses is shown in figure 33.5, in horizontal and vertical section so as

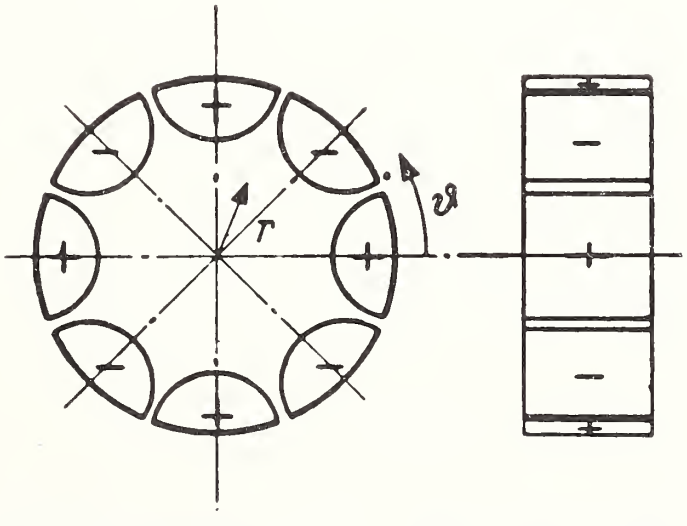


FIGURE 33.3. *Correction element with radial electrodes.*

Seeliger, Optik **5**, 490 (1949), figure 1.

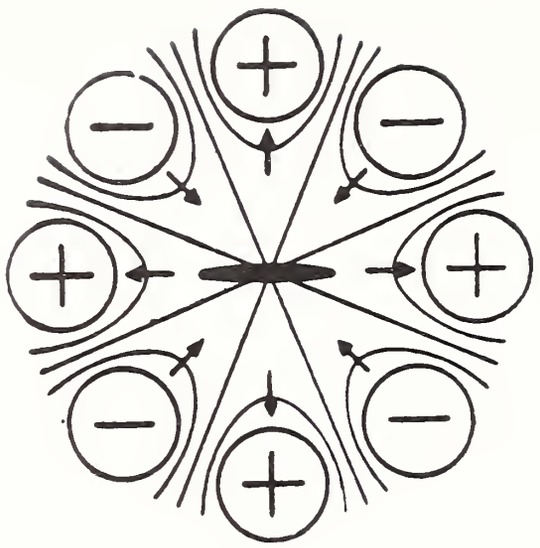


FIGURE 33.4. *Forces on flat electron beam passing through correction element.*

Scherzer, Optik **5**, 497 (1949), figure 1.

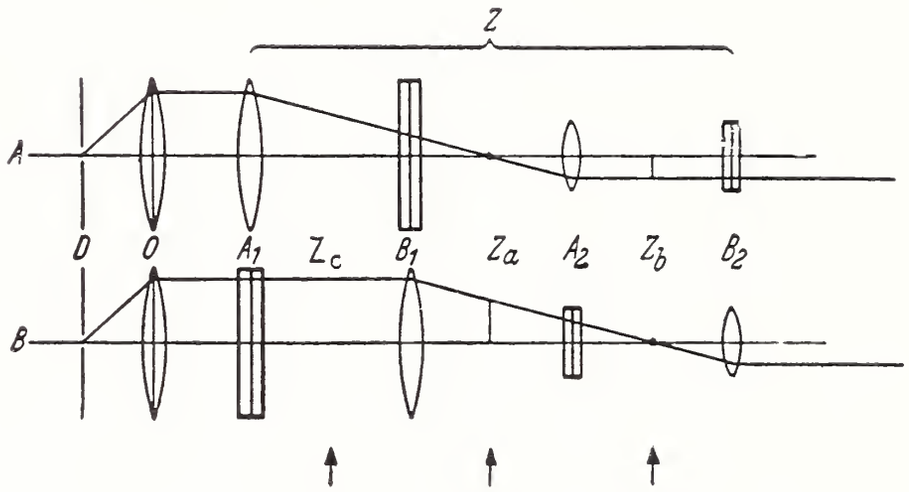


FIGURE 33.5. *Components of complete correction system, in horizontal and vertical section, with ray paths.*

Seeliger, Optik **5**, 490 (1949), figure 4.

to distinguish clearly the cylindrical components. The objective O is followed by two cylindrical lenses, A_1 and B_1, which produce longitudinally separated line foci at Z_a and Z_b respectively, at which are placed the rings of correcting elements, indicated by arrows. The succeeding cylindrical lenses A_2 and B_2 are needed to remove the astigmatism introduced by A_1 and B_1. The third correcting ring is placed at Z_c to correct the diagonal planes.

It will be seen that the objective with its correcting system consists of 8 lenses in all, which makes it complicated to construct and aline, apart from finding the conditions for correction. An electron microscope embodying such lenses has in fact been built by Scherzer and his collaborators at Mosbach-Baden [12], and was described at the International Congress on Electron Microscopy in Paris in 1950 [13] (Fig. 33.6). He showed convincing proof that it was possible to introduce a pronounced astigmatism into the imaging beam, and then to remove it so as to give a stigmatic final image. He also claimed that the lenses and correcting elements had been adjusted so as visibly to reduce the spherical aberration, but he confessed that the difficulty was to maintain everything constant for a long enough period to focus the image with precision and obtain a photograph. It seems

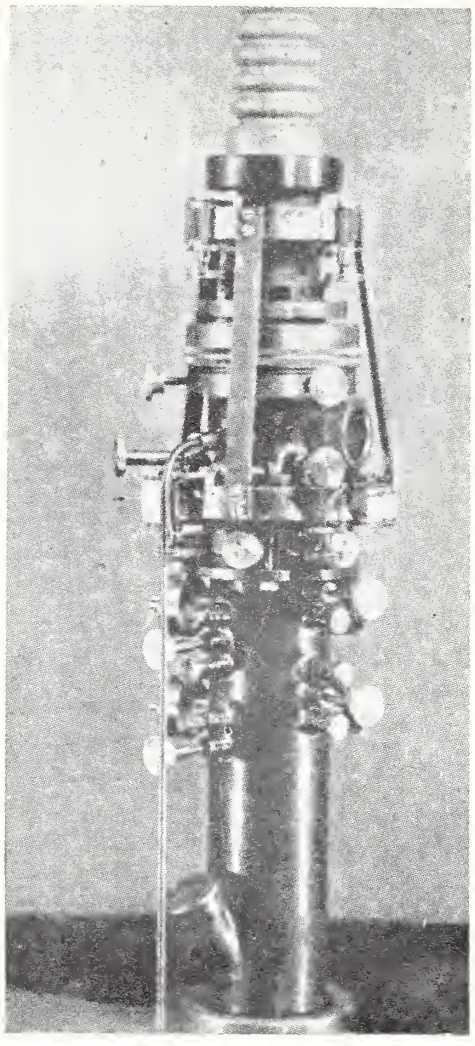


FIGURE 33.6. *Electrostatic microscope with correcting components.*

Scherzer, in Report of Paris Congress by Grivet, Biochim. et Biophys. Acta **7**, 1 (1951), figure 11.

possible that Scherzer has, indeed, attempted too much: He has broken down the problem into its ultimate elements and employed separate lenses for each required function. This allows in principle the correction of any degree of spherical aberration at any magnification and accelerating voltage, but so many degrees of freedom inevitably bring in complications. It might be more practicable to aim at correction only within a restricted range of working conditions, or even just in a single prescribed state. With this in mind, Burfoot in my laboratory has been investigating how far it is possible to combine the functions of some of the lenses of figure 33.5, for instance, by taking a unipotential electrostatic lens and making the apertures in its electrodes noncircular. The investigations of Regenstreif [14] bear on the same problem. There may be some hope that a system can be found of only 3 or 4 lenses, which would correct spherical aberration in one set of working conditions. In any case, it seems that the investigation of these noncentered systems offers the best hope of attaining such an end, which must be accounted the chief task of practical electron optics in the immediate future.

Lenses With Curved Axes

The other direction in which a relaxation of symmetry is likely to lead to practical results is that of using systems with curved instead of straight axes, so that rectilinear propagation of the beam is abandoned. The design of an objective system, in which the beam is curved

and focused by a specially shaped magnetic field, is now being investigated by Le Poole at Delft, but it is in a stage too early for more to be said than that such systems are now seriously looked on as practical possibilities. Curved beam paths have often been used, of course, in beta-ray and mass spectrometers. I wish to mention here a simpler system, a lens with helical beam path which is designed to correct chromatic aberration [10].

It has long been known that a beam moving in a circular orbit in the annular space between two concentric electrodes, appropriately charged, will be focused at intervals of 127° in the plane of its motion. It was shown by Wendt that focusing can also be produced in the perpendicular direction if the electrodes are shaped as hyperboloids of revolution (fig. 33.7a); focusing then occurs every 180° (fig. 33.7b). Gabor has since shown that this system can be used either as a velocity spectrograph or as a lens with negative chromatic aberration, depending on the plane of recording. Electrons with energy in excess of the standard value will travel in an orbit of larger radius (fig. 33.7c), so that at the opposite end of the diameter through the source, electrons of varying velocities will be focused at different radial positions (such as P′, P″, fig. 33.7c). But, at the completion of approximately one revolution, it is found that the points of focus fall behind each other circumferentially, as at P, P‴ (fig. 33.7d). This is equivalent to a negative chromatic aberration, since the faster electrons are focused before the slower. As all ordinary electron lenses have positive chromatic aberration, with a longer focal length for fast than slow electrons, the possibility exists of using lenses of these two types to form an achromatic combination. Such a system would reduce the need for costly and complicated electrical stabilizing circuits, and ultimately give the prospect of reducing the chromatic aberration below 1A or whatever other limiting resolution would be attainable if spherical aberration were corrected.

The difficulty remains of designing a practical form of the Gabor lens, since, as it stands, the focus is in an inaccessible position behind the source or object. Gabor has proposed to apply a longitudinal field at the same time, so that the beam path is drawn out into a helix. The necessary form of the electrodes and the axial trajectory for a correcting lens are shown in figure 33.8a; figure 33.8b shows the use of the system as a velocity spectrograph. The focusing properties of these new lenses requires more detailed investigation, so as to establish what residual aberrations (such as astigmatism) may be present. Some

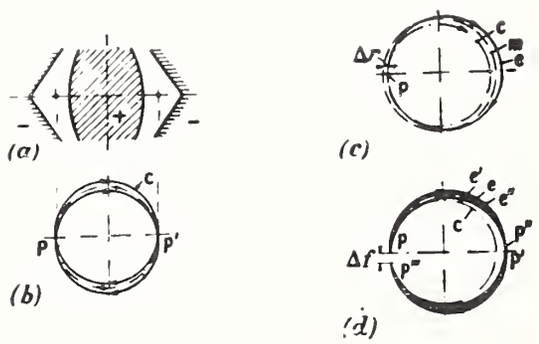


FIGURE 33.7. *Principle of helical lens.*

(a), Electrode system of Wendt (1943); (b), focusing of electrons with standard energy (c is the circular axis); (c), the axis e for electrons with excess energy; (d), focusing of electrons with excess energy. Gabor, Proc. Phys. Soc. London [B] **64,** 244 (1951), figure 1.

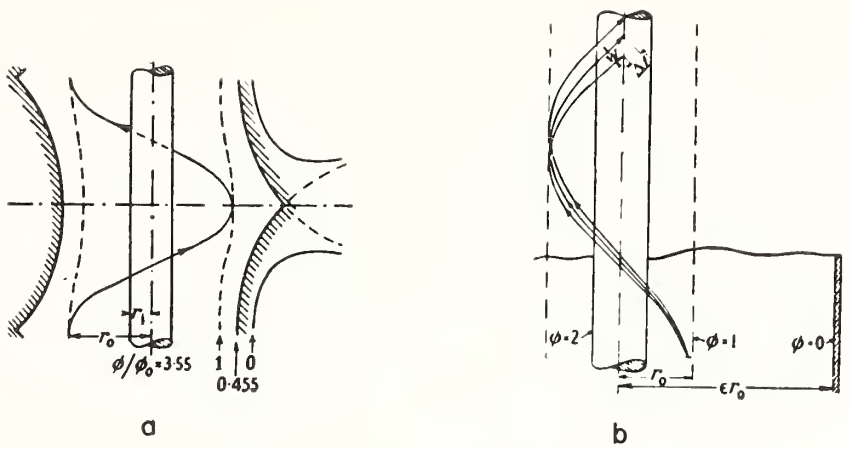


FIGURE 33.8. *Electrode design for (a) a helical lens, (b) a helical velocity spectrograph.*

Gabor, Proc. Phys. Soc. London [B] **64,** 244 (1951), figures 4 and 5.

work of Sturrock suggests that Gabor's treatment is not sufficiently rigorous, but does not invalidate the main conclusions. Sturrock has also been investigating the aberrations of more general systems with curved axes, [11, 16] such as the double-focusing spectrometer of Svartholm and Siegbahn [15], rather like a betatron magnetic field. Here focusing occurs at intervals of 255°, and in both planes. Sturrock has shown that the second-order errors can be eliminated, and perhaps also those of third order, by introducing an intermediate focus, rather like the astigmatic lines in Scherzer's system. Again it is a far cry to anything like a practicable imaging system, especially with prospective difficulties of alinement, but it is worth noting that these unorthodox investigations are on foot.

Diffraction Microscopy

It remains to say something about another scheme of Gabor's that differs from those so far discussed in that it seeks to bypass rather than to correct the spherical aberration of electron lenses. Diffraction images are formed that contain the effect of the aberration, but these are then "reconstructed" in an optical system that has exactly identical aberrations; the result should be cancellation of their disturbing effects, and the production of a perfect image. The principle of this "diffraction microscopy" is shown in figure 33.9, in the form originally proposed by Gabor [17]. In the electronic stage, a point focus of electrons is obtained by use of a strong condenser system. If an opaque object is placed close to this point, a shadow of it surrounded by diffraction fringes will be formed on a photographic plate placed at a distance. This picture, or "hologram", is printed as a positive and placed in an optical reconstruction device, or synthesizer, which contains a set of lenses identical with those of the electronic apparatus, except that the lens parameters are all scaled up in the ratio of the wavelength of light to that of the initial illuminating beam of electrons. If light is projected through the synthesizer as shown, a virtual reconstruction of the object should be formed in the corresponding position close to the point source of light. It is convenient to provide a reconstructing lens for throwing an enlarged image of it on to a photographic plate.

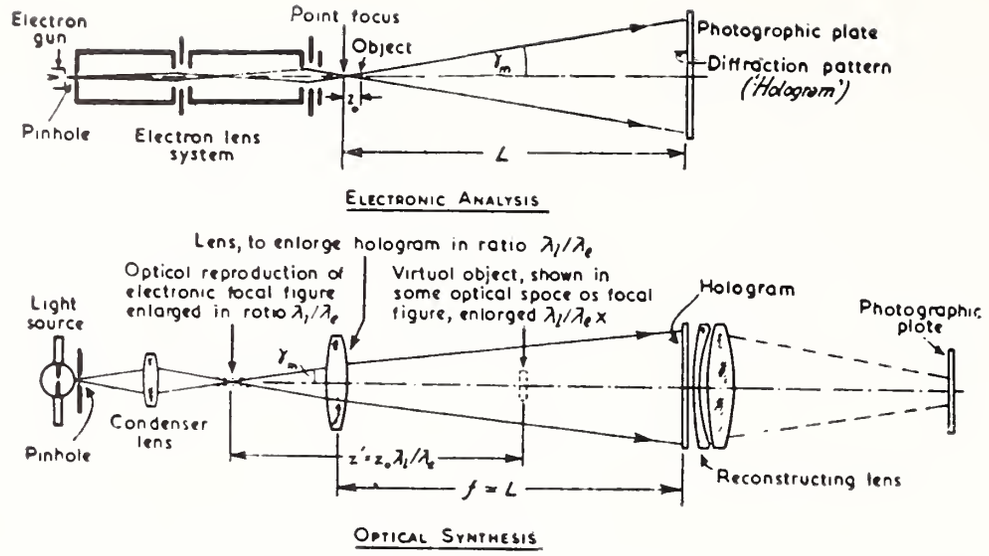


FIGURE 33.9. *Principle of electron microscopy by reconstructed wave fronts.*

Gabor, Proc. Roy. Soc. London [A] **197,** 454 (1949), figure 1.

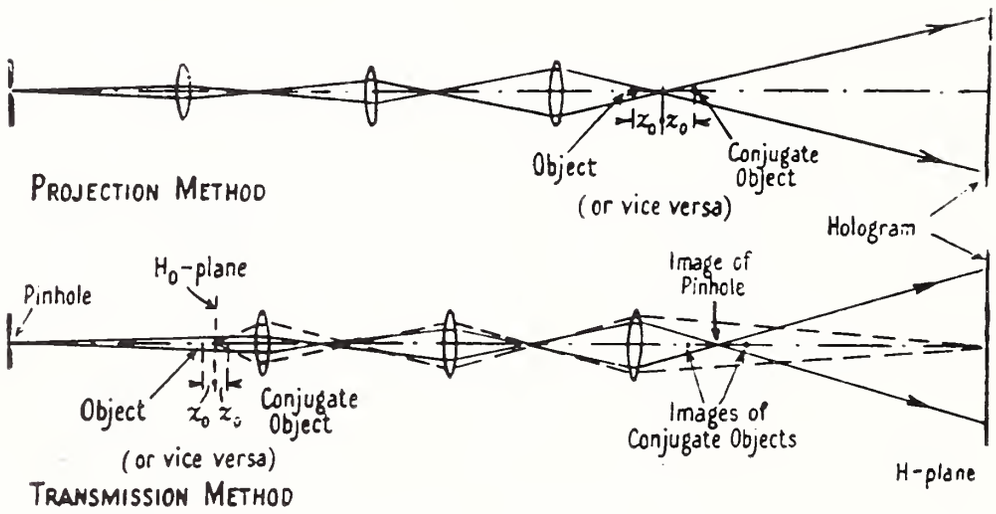
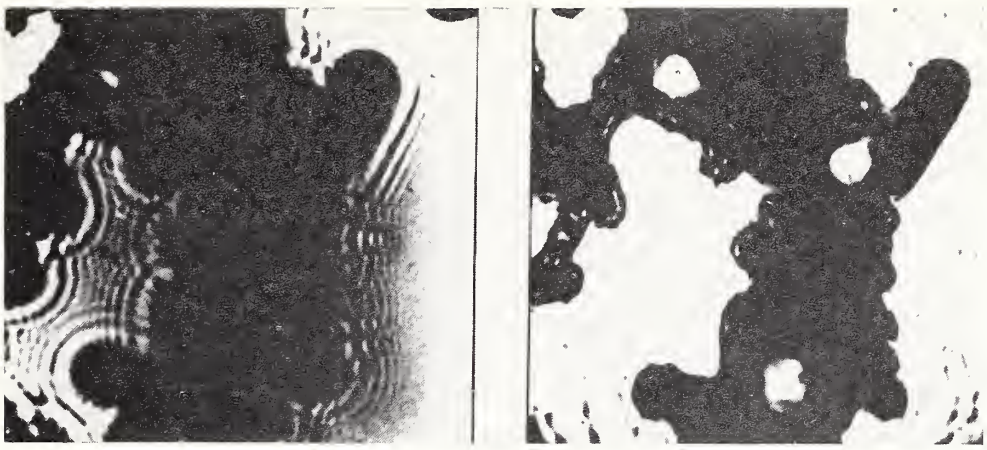


FIGURE 33.10. *Electron optical systems for the two methods of producing a hologram.*

Gabor, Proc. Phys. Soc. London [B] **64,** 449 (1951), figure 5.

Gabor showed that the proposals were sound by forming the hologram optically and then reconstructing it in the same system. It was soon realized, however, that it was impossible to do the same with electrons, owing to the difficulty of holding stable such a very small electron focus for the long time necessary to record a hologram (about a half hour). Haine and Dyson [18] then suggested that a practicable scheme would be to take an out-of-focus image in the normal type of electron microscope, so that the necessary information would be recorded in Fresnel fringes. Figure 33.10 shows this transmission system in comparison with the original projection method; the broken lines in the lower diagram show the paths of rays from the plane that is in focus to the final screen, the objective lens being over-focused with respect to the actual object. This system has the advantage of requiring a much shorter exposure (order of 30 sec.) than the originally proposed scheme, and also of needing no special apparatus, being carried out in a tuned-up electron microscope of normal design.



a b

FIGURE 33.11. *Zinc oxide smoke.*

a, Highly defocused electron micrograph (or hologram); b, optical reconstruction from hologram. (Haine, unpublished.)

Haine will be reporting in a later session on the progress so far made, so that it is unnecessary to go into details here. Figure 33.11 shows the type of result, with the hologram of a zinc oxide specimen (a) and its reconstruction (b). It is claimed that a resolution of about 10 A has been obtained already, the limitation at the moment being as much in the reconstruction procedure as in the actual electron microscopy. The latter is again affected by mechanical and electrical instabilities, as in normal operation of the machine. Haine will describe the attempts to track down and eliminate these troubles; he has succeeded to the point of being able to record diffraction patterns equivalent to a resolution of 5 or 6 A. The optical problem is partly one of introducing continuously variable spherical aberration and astigmatism in the reconstruction system, and partly that of judging the precise focus for reconstruction, since one does not know what the object really should look like. It comes down again to taking a through focal series and choosing the most convincing picture.

It is encouraging that this new method has already reached the same level of resolution as straight electron microscopy, but it appears that it is liable to be limited in the end by much the same factors: The instabilities of the instrument, rather than the aberrations of the lenses. But if these mechanical and electrical problems can be solved, it holds out a clear prospect of circumventing spherical aberration and so allowing a resolution of atomic order to be achieved.

Observation of Living Matter

Brief mention must be made of the other main problem before electron microscopists: The need to devise means of viewing organisms in the living state, instead of in the dead and desiccated condition imposed by present methods. It has been realized for some time that the best chance of doing so is to employ voltages of the order of 1 million. Sufficient penetration would then be available to use a double-walled cell in which the specimen could be kept in a liquid medium. The lethal effect of the electron beam is reduced as the voltage rises, but so also is the efficiency of the photographic plate. The complex of problems thus raised is being approached in several labora-

tories, and in the first place by gaining experience at voltages between those now common and those which will finally be needed. The original project of Marton [19] for a 1.4 mv electron microscope has encountered serious difficulties owing to the limitations of the existing source of high voltage. Power supplies and microscopes for around half a million volts are now being constructed in the Philips laboratories at Eindhoven [20], in those of Metropolitan-Vickers and at King's College, London. They are naturally all of magnetic type, the difficulties of employing electrostatic lenses at these voltages being considerable. We may therefore hope that accumulation of experience will enable us to judge more accurately the difficulties and advantages of operating at even higher voltages. Meanwhile, there are strong indications from the effects of beta-rays that it would in any case be difficult to maintain an object in a living condition under such an electron beam for long enough to focus it and take a single micrograph, let alone a series for following its living processes [21]. However, we may still learn a great deal about these processes from watching their breakdown,—indeed, possibly more than from seeing them in action. There is also a great deal to be learned from seeing such specimens in their natural wet state. The various projects in hand will be followed with interest for these reasons, and we wish them a successful outcome.

An alternative to using high voltage electron beams is to employ X-rays for forming shadow images. Electron optical means can be devised for making the source very small, and the resolution is then limited primarily by the size of the electron spot. If this falls on a target in the form of a thin foil sealing the tube, the specimen may be placed close to it, in air, and an X-ray shadowgraph formed on a plate at a distance. The primary magnification can be made large by making the ratio of these two distances large, and also by subsequent enlargement. I shall be describing such a system in a separate contribution and only mention it here as being a more immediately practicable means of viewing living specimens than by a million-volt microscope. The present resolution approaches that of the optical microscope, being around 1 micron, but it is limited in principle only by the exposure which can be tolerated, since an extremely fine electron probe can be obtained. The exposure time at present is about 1 minute. In the full communication [22] are shown typical microradiographs of insects, with portions of a 1,500 mesh silver grid, in which the wires are 3μ wide in one direction and 2μ in the other. The practical limitations will again be the lethal dosage for the organism, as in the case of the high voltage microscope, and the poor contrast obtainable when X-rays are used of sufficient energy to penetrate specimens in the wet state.

Discussion

Dr. Hillier agreed that the limiting factors on resolution at present are largely the mechanical and thermal disturbances, but felt that even better results than 10 A had, in fact, been achieved. He had obtained particles well separated at this distance, whereas, the classical definition allowed them to overlap considerably. He thought that the poor contrast was also an important limiting factor.

Dr. Cosslett agreed, and regretted that he had not been able to deal with contrast problems, on which important work was now being done.

There had been several attempts to deduce the minimum size of particle, possibly atoms, which could be detected on grounds of contrast, but these had relied on theoretical expressions for electron scattering. The investigations now being made by Hall [23] end by Ellis [24], on scattering in the experimental conditions of electron microscopy, at last gave promise of estimates based on solid facts. Unfortunately, the indications were that only the heavier atoms would scatter sufficiently to be visible.

In reply to Dr. Clavier, Dr. Cosslett said that a voltage of 10 kv was being used for most of the X-ray microscope work, as difficulty was experienced at lower voltages. Higher voltages, such as those of conventional X-ray tubes, would give very little contrast with biological specimens.

Mr. Haine criticized the usual treatment of resolution, since there is almost always enough coherence in the electron beam for interference to occur between the rays forming the images of two neighbouring points; they should not be treated separately, as Airy disks. He also emphasized the contrast difficulty, which seemed likely to prevent increased resolution being obtained, even if spherical aberration could be corrected. In these circumstances the best prospect was to employ phase contrast, as suggested several years ago by Gabor. As regards difficulty of focusing in reconstructing an object from the hologram, exactly the same problem arose in focusing an electron microscope at high resolution for normal microscopy. There was a slight difference, but in the method that Gabor would describe, the process was absolutely identical.

Dr. Gabor then discussed the possibilities of diffraction microscopy.

References

[1] H. Boersch, Z. Naturforsch. **2a,** 615 (1947).
[2] L. de Broglie, Rev. opt. **26,** 397 (1947): Compt. Rend. **224,** 1743 (1947).
[3] W. Glaser, Z. Phys. **117,** 285 (1941); W. Glaser and E. Lammel, Arch. Elektrotech. **37,** 347 (1943).
[4] V. E. Cosslett, J. Sci. Instr. **22,** 170 (1945); Proc. Phys. Soc. London **58,** 443 (1946).
[5] G. Liebmann, Proc. Phys. Soc. London B. **64,** 972 (1951); G. Liebmann and E. M. Grad, Proc. Phys. Soc. London **64,** 956 (1951).
[6] O. Scherzer, Z. Phys. **101,** 593 (1936).
[7] O. Scherzer, Optik **2,** 114 (1947); **5,** 497 (1949); J. Appl. Phys. **20,** 20 (1949).
[8] L. Marton and D. L. Reverdin, J. Appl. Phys. **21,** 842 (1950).
[9] R. Kompfner, Phil. Mag. **32,** 410 (1941). D. Gabor, The Electron Microscope, p. 70. (Hulton Press, London, 1946).
[10] D. Gabor, Proc. Phys. Soc. London [B] **64,** 244 (1951).
[11] P. A. Sturrock, thesis, Cambridge University, 1951 (in publication).
[12] R. Seeliger, Optik **5,** 490 (1949).
[13] O. Scherzer, Proc. Intern. Congr. Electron Microscopy, p. 191 (Paris, 1950).
[14] E. Regenstreif, thesis (Paris, 1951).
[15] N. Svartholm and K. Siegbahn, Ark. Math. Astr. Fys. **33a,** No. 21 (1947).
[16] P. A. Sturrock, Proc. Roy. Soc. **210A,** 269 (1951).
[17] D. Gabor, Proc. Roy. Soc. **197A,** 454 (1949); Proc. Phys. Soc. London [B] **64,** 449 (1951).
[18] M. E. Haine and J. Dyson, Nature **166,** 315 (1950).
[19] L. Marton, Electron Microscopy Conf. (Cambridge, England, 1948).
[20] A. C. van Dorsten, W. J. Oosterkamp, and J. B. Le Poole, Philips Tech. Rev. **9,** 193 (1947).
[21] V. E. Cosslett, Proc. Intern. Congr. Electron Microscopy, p. 555 (Paris, 1950).
[22] V. E. Cosslett, and W. C. Nixon, Paper No. 27, this symposium.
[23] C. E. Hall, J. Appl. Phys. **22,** 655 (1951).
[24] S. G. Ellis, Paper No. 42, this symposium.

34. Microanalysis by Means of an Electron-Probe Principle and Corrections

By R. Castaing [1]

The electron probe microanalyser developed in the Laboratories of the ONERA has already been described at the International Conference on Electron Microscopy, held in Paris in 1950. Its principle may be summarized as follows:

A selected spot on the surface of a sample, about one micron in diameter, is bombarded by an electron probe. The X-ray spectrum that is emitted by this spot contains the characteristic radiation of all the elements present at this spot. The operation consists of determining the chemical constitution of the very small volume of matter irradiated (about one cubic micron) by analysing the X-ray spectrum emitted from the sample.

Figure 34.1 represents a diagram of the apparatus. The probe is obtained by forming a reduced image of the cathode by means of two electrostatic lenses. The accelerating voltage is 30 kv. The electron beam current with a 1 μ probe is 1.5×10^{-8} amperes. Although the incident power on the sample is about 600 w/mm^2, yet the local increase in temperature of the sample is only a few degrees since the bombarded area is very small; so that the analysis can be performed at room temperature.

The spectrometer includes a curved quartz crystal used in connection with a Geiger-Müller counter. The reflected intensity of the characteristic radiation of the main elements that constitute the sample usually measures several hundred pulses per second, whereas the continium does not exceed one or two pulses per second. Consequently, an accurate analysis is possible. The purpose of this communication is to describe the experimental process of such an operation.

To start with, the various elements contained in the irradiated volume are easily detected by the presence of their characteristic radiation in the emitted spectrum. We will be concerned now with quantitative analysis.

Let I_A be the intensity of the characteristic radiation of the element A, emitted by the sample. Let I(A) be the intensity of the same radiation $K\alpha_1$ emitted by a sample of pure A under the same conditions. An approximate calculation shows that the ratio of I_A/I(A) is equal, to the first order, to the mass concentration c_A of the element A in the analyzed volume. This has been experimentally verified for large samples. One can get a rough idea of this property by noticing that the different elements have approximately the same mass absorption coefficient for electrons of the same energy. Thus the mass of alloy irradiated by the electrons is approximately independent of its composition, and the number of atoms of A excited by the electrons

[1] Office National d'Etudes et de Recherches Aéronautiques, Chatillon, France.

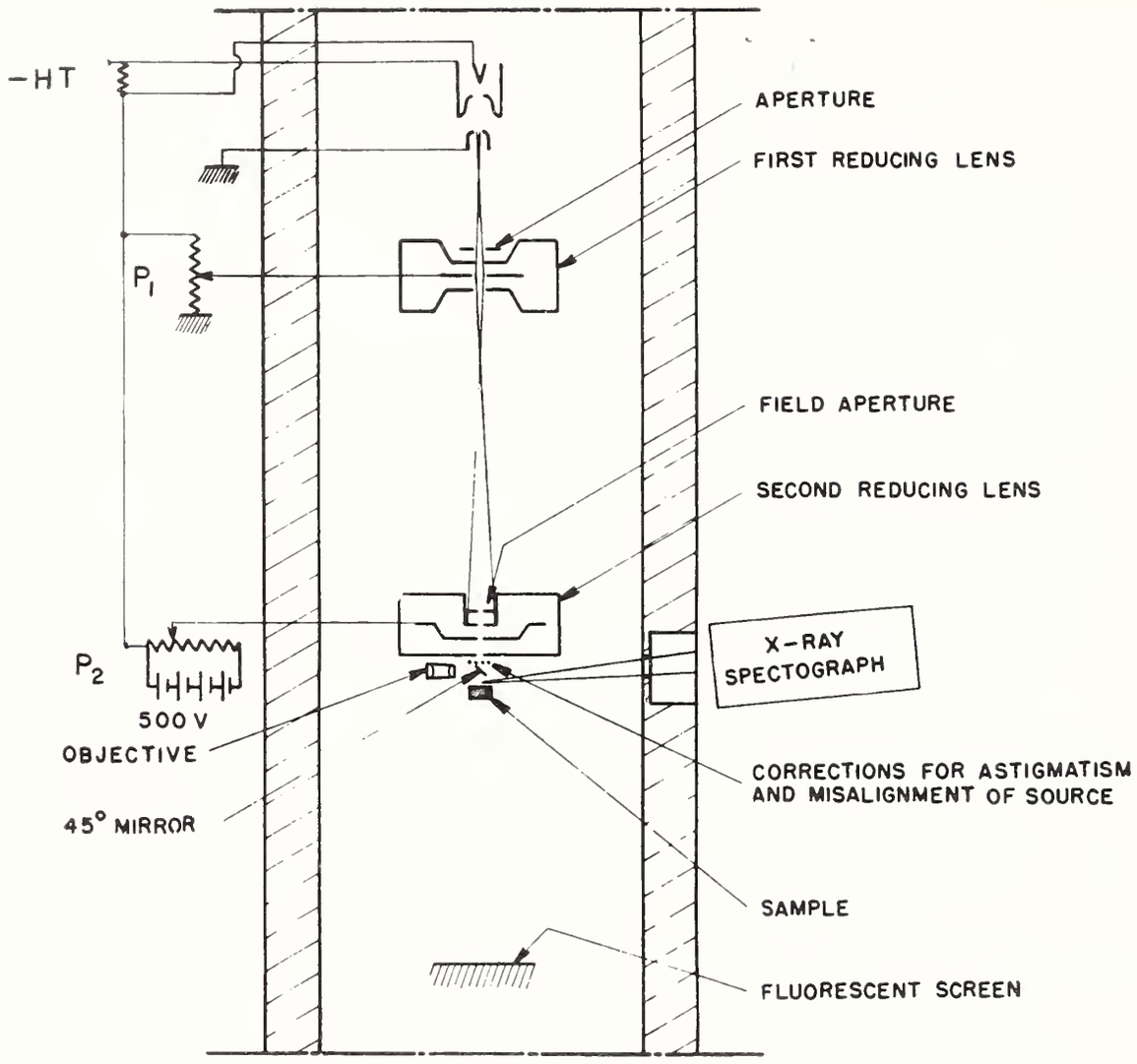


FIGURE 34.1.

is proportional to the mass concentration of the element A in the alloy.

A better approximation is given by the relation

$$\frac{I_A}{I(A)} = \frac{\alpha_A c_A}{\Sigma_i \alpha_i c_i}$$

where α_1 is the electron mass absorption coefficient of the exponential Lenard law corresponding to each element. In fact, these coefficients are empirically determined once and for all by an analysis of an alloy of known composition.

Table 34.1 represents the results of four analyses performed on copper-aluminum alloys with copper concentrations 1, 4, 53, and 88 percent. It gives an idea of the accuracy of our approximations. The first approximation gives the copper concentration within a 2.6 percent maximum error. For the second approximation the results

TABLE 34.1.

Alloy	True concentration	First approximation	Second approximation
A_1	0.01	0.0099	0.01075
A_2	0.04	0.0373	0.0404
A_3	0.53	0.504	0.525
A_4	0.88	0.867	0.876

are still better, since the error can never exceed 0.5 percent. Yet aluminum and copper are far from each other in the periodical table. For closer elements the first approximation gives currently an accuracy of 1 percent, and we have seen that this approximation does not require any previous calibration.

The main advantage of this method is that the only emitted intensities that we have to compare to one another are always those of radiations that have the same wavelength, so that the variations in sensitivity of the counter and reflecting power of the crystals for various wavelengths do not have to be taken into account.

The problem is not always so simple, however, and some corrections are sometimes necessary. In the first place, when the absorption coefficient of the sample for the characteristic radiation that is being measured is large, the intensity that has been absorbed in the sample itself must be computed and this correction must be applied. Let us consider that radiation which is emitted from an elementary layer located at the depth z under the surface of the sample. The corresponding absorption is equal to $e^{-\mu z \operatorname{cosec} \theta}$, where μ is the linear absorption coefficient of the alloy for this wavelength, and θ the angle between the analyzed X-ray beam and the surface of the sample. It can be shown by a calculation, which will not be given here, that the ratio $I(\theta)/I_0$ of the total intensity which is measured to the intensity which is actually emitted is a function of the form $f\left(\frac{\mu}{\rho} \operatorname{cosec} \theta\right)$ where $\frac{\mu}{\rho}$ is the mass absorption coefficient of the alloy for that wavelength, and f is, to the first order, a universal function, valid for any alloy, any characteristic radiation and any angle θ. If we let the angle θ vary, for a single given sample, we can determine the function f. Figure 34.2 shows the curve which represents $\log f$ versus the quantity (μ/ρ) cosec θ. Such a curve is valid only for a given accelerating voltage. In the present case, the voltage is 27.5 kv. Figure 34.3 shows the

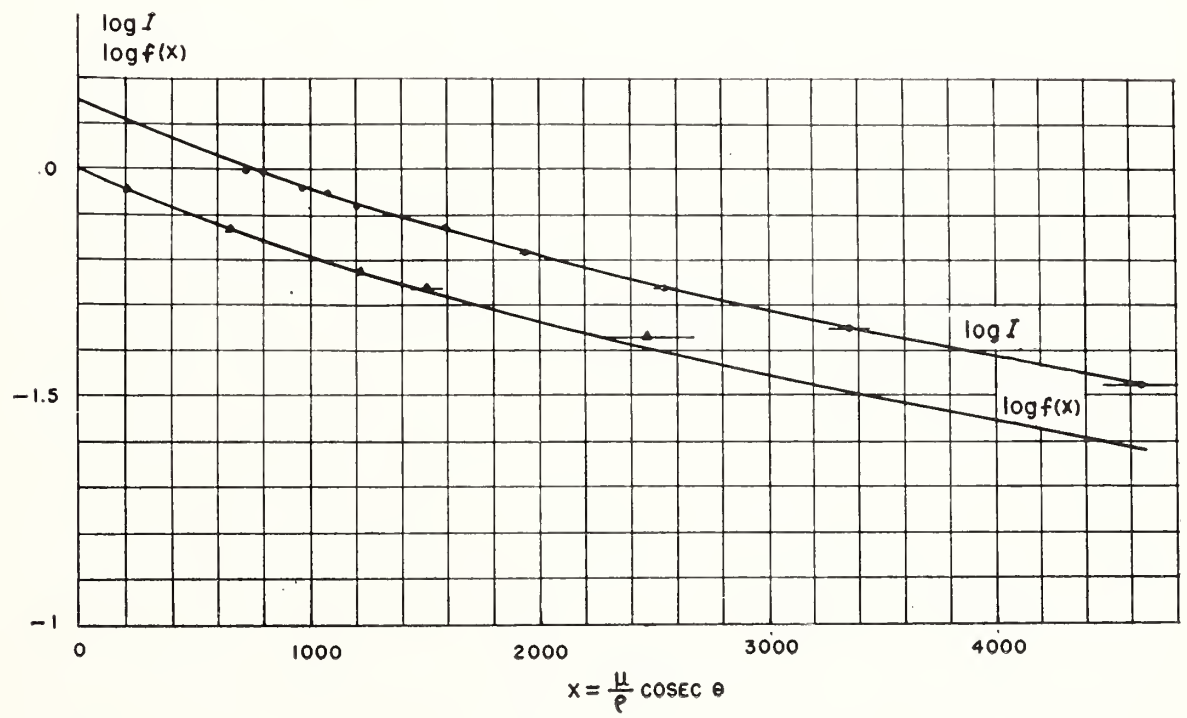


FIGURE 34.2.

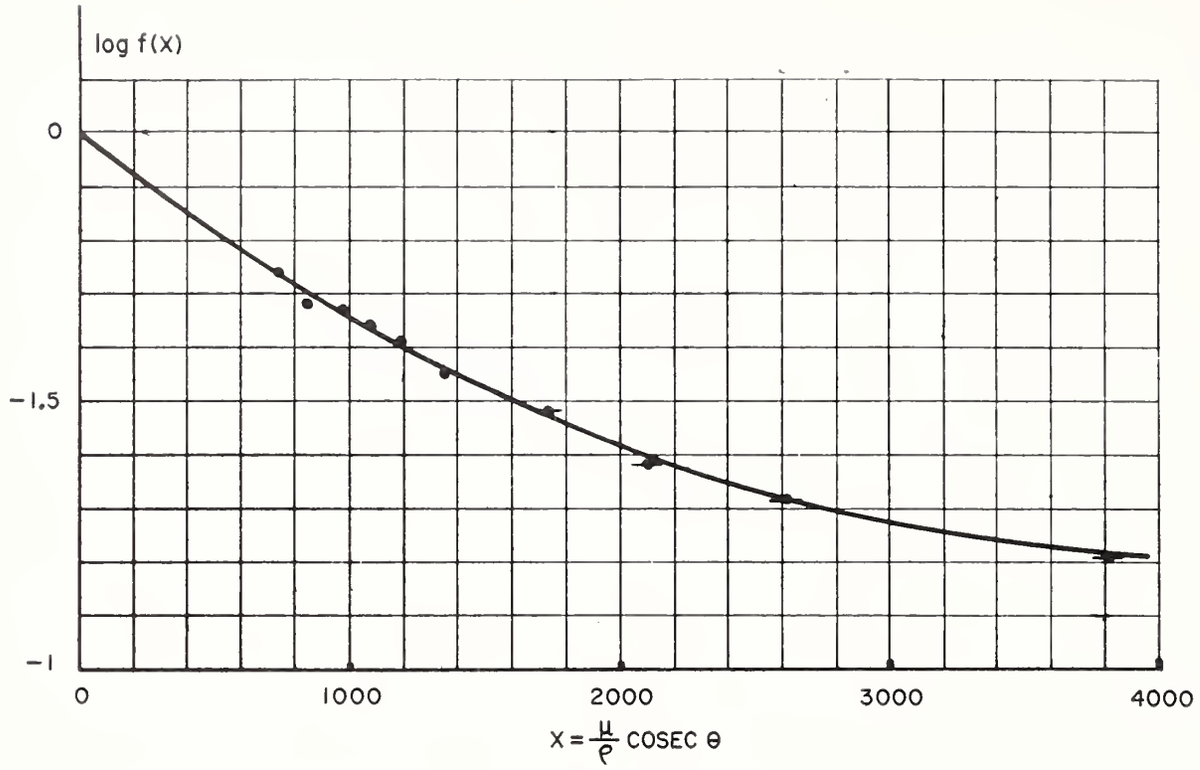


FIGURE 34.3.

correction curve for a 38 kv. The correction is larger for the same value of (μ/ρ) cosec θ because the electrons go deeper into the sample and consequently the X-ray beam goes through a thicker mean layer of matter before emerging.

It so happens that the function f is the Laplace transform of the function that represents the decrease in intensity of the emitted X-rays versus the depth of emission in the anticathode. Thus, information about the law of decrease of this intensity may possibly be obtained by performing the transformation. It is found that the most important factor in the phenomenon is the exponential Lenard absorption, and not the slowing down of the electrons, as it was usually presumed. The absorption coefficients measured by Lenard are very consistent with the results obtained by this process.

The concentrations of the various elements in the alloy can be determined directly from the absorption correction curve by a graphic method or by repeated approximations.

Let us say a few words now about a second correction, the correction for fluorescence. We have only considered so far the X-rays produced by atoms which had been directly ionized by the electron beam. But other ionizations may be produced in the sample by the X-rays themselves. This entails a secondary X-ray emission, which contains only the characteristic radiation of the excited atom, and which is known as the fluorescence emission. This secondary emission may affect the result of the measurement considerably and it is necessary to find a way of estimating its value. This tedious calculation will not be expanded here. We have verified experimentally the formula in the case of an iron-chromium alloy and the results were very satisfactory indeed.

As a conclusion, we may point out that the corrections which may be found necessary in microanalysis by means of an electron probe are simple enough and leave the convenience and the accuracy of this method practically untouched.

35. Applications of the Electron Probe Microanalyzer

By R. Castaing [1]

There are many possible applications of the electron probe microanalyzer. We were mainly concerned with its metallurgical applications. Let us mention in this field the determination of curves to represent the diffusion of metals into one another, the study of local variations of the concentrations in alloys, and more especially the analysis of precipitates or of inclusions of an unknown or incompletely known nature.

Figure 35.1 is the micrograph of a copper-zinc diffusion sample that was obtained by heating a plating of the two metals at 400° C. for one hour. Two intermediate phases are visible. The whole diffusion area is about 0.1 mm wide. It is possible to perform a rapid quantitative analysis of the sample at any point, that is to determine the copper-zinc diffusion curve by simply moving the probe across the diffusion area.

Figure 35.2 shows the curve obtained by plotting the zinc-concentration across the sample, that is by adjusting the spectrometer on the radiation Zn $K\alpha_1$. The abrupt variations in concentrations that correspond to the change from one intermediate phase to the next, and the steady variation inside a given phase are very noticeable. The equilibrium diagram of the two metals is also given. Phase β does not show in this diffusion sample. Further experiments showed that phase β was actually present in the sample, but the layer was less than 1 micron thick. The intersection of the curve and the straight line $T=400°$ should give the limit concentrations that correspond to the phase boundaries. The agreement is good, except at the limit of the phase γ, on the side of high copper concentrations. This disagreement could be explained by assuming a very steep decrease of the zinc concentration in phase γ, close to this limit.

The diffusion curve can also be drawn by analysing quantitatively the copper present in the sample, that is by adjusting the spectrometer on the radiation Cu $K\alpha_1$. Figure 35.3 shows the curve obtained. It is easy to check that the sum of the ordinates of the two curves is at any point equal to one within 1 percent. This is an excellent verification of the basic relations between concentrations and intensities of the characteristic radiation.

We will consider now an example of an analysis of precipitates. Figure 35.4 represents a micrograph of a copper-tin-antimony alloy which, in addition to big blocks of antimony about 0.1 mm wide, contains needles with a dark central area and a brighter rim. The problem was to determine the composition of these needles. A qualitative analysis shows immediately that they contain some copper; the absorption correction being taken into account, a quantitative analysis shows that the central part is constituted by the Cu_3Sn phase

[1] Office National d'Études et de Recherches Aéronautiques, Chatillon, France.

whereas the bright rim corresponds to the formula Cu_5Sn_4. But the rim may contain some antimony. This will be easily checked as soon as the frequency band of our spectrometer is a little broadened so as to include the characteristic radiation of tin and antimony.

The next problem is to estimate the resolving power of this method, that is to determine the smallest diameter of the precipitates that can be quantitatively analysed. Figure 35.5 shows a copper-aluminum alloy that contains precipitates of Al_2Cu about 1 to 15 μ in diameter. If the probe is moved on precipitates of decreasing sizes, the intensity emitted of radiation Cu $K\alpha_1$ remains constant as long as the diameter of the precipitate is larger than 2 or 3 microns. For smaller sizes, there is a loss of intensity that is due to the diffusion of electrons outside the precipitate. We may derive from this experiment that for a very tiny probe the minimum diameter of the precipitates that are accessible to an accurate quantitative analysis would be about 1 μ. On the other hand, a qualitative analysis can often be performed on much more minute precipitates, close to the resolving power of a conventional light microscope.

I conclude with a few words about the possibilities of punctual crystallographic analysis with the same apparatus. When an electron beam hits a single crystal that acts as an anticathode, the emitted X-rays are diffracted in the anticathode itself, and what is known as Kossel pattern originates as in figure 35.6. The characteristic rays which propagate from the striking point S meet in the anticathode lattice the family of reticular planes P. All rays which have the Bragg angle with planes P are reflected according to the so-called selective reflection. Therefore, the reflected rays are located on cones of revolution, the axes of which are normal to planes P and the half-angle of which is equal to 90 degrees minus the corresponding Bragg angle. If a photographic film is placed above the anticathode, conic sections are printed on the negative. They are the intersections of the plane of the film with the peak intensity cones; consequently they are black on the negative. The operation can be performed by transmission through a thin anticathode; a film placed below it registers conic sections, which show in white. They are the intersections of the plane of the film with the cones of X-rays weakened by extinction. These conic sections are usually hyperboles and are known as Kossel lines. They characterize the crystallographic nature and orientation of the involved region of the anticathode. Thus the phenomenon can be used to perform a rapid crystallographic analysis of a sample by means of the Kossel lines it produces under the excitation of an electron probe. It could be feared that the low rate of the total intensity of the X-rays would entail very long exposure times if a 1-μ probe was used; it is not so, for the film can be placed very close to the anticathode without injuring the very good definition of the image because the lines on the negative are extremely thin. Figure 35.7 shows the Kossel pattern that has been obtained through a 0.1 mm thick aluminum leaf which had been covered by a 1-μ copper layer by vaporization in vacuum. The film lay 2 mm from the sample and the original picture is only 5 mm in diameter. Despite the large maginificaton, the hyperboles remain very thin and allow an accuracy of 2 min in the determinations of orientation.

Furthermore, after a method used by K. Lonsdale, the tiny curvilinear triangles formed by some of the Kossel lines can be used to determine the crystallographic parameters of the area which is studied,

with a 10^{-3} accuracy. This result is very interesting since the method does not require any precision in the experimental setting. The fact that a 3 min exposure time was sufficient to obtain this pattern is very remarkable, too. The same exposure time is necessary to obtain a conventional Kossel pattern with an X-ray tube when the electron current intensity is 10^5 times higher.

Discussion

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: I was very much interested in your experiments because at one time we had contemplated something somewhat similar, but the total energy required seemed to be excessive, and I see that, in fact, you need 60 kw/cm^2. How far did you observe any modification of the samples?

Dr. Castaing: For an incident energy of 60 kw/cm^2 on a 1 micron probe, the calculated local rise in temperature is only 0.3° for a copper sample and less than 10° for other metallic samples. I think it would be possible to use fairly high intensities, as some experimenters have obtained with magnetic lenses, without being disturbed by thermal limitations. In fact, we have never observed any modification of the samples.

Dr. D. Gabor, Imperial College, London, England: Is it carbonaceous contamination?

Dr. Castaing: I have seen layers of contamination, but the rate of contamination is not so fast; the layer is visible under the light microscope after irradiating the same spot for about 5 min; it is generally possible to perform the whole analysis before the production of any visible layer.

Figure 35.1.
Magnification ≈ X700

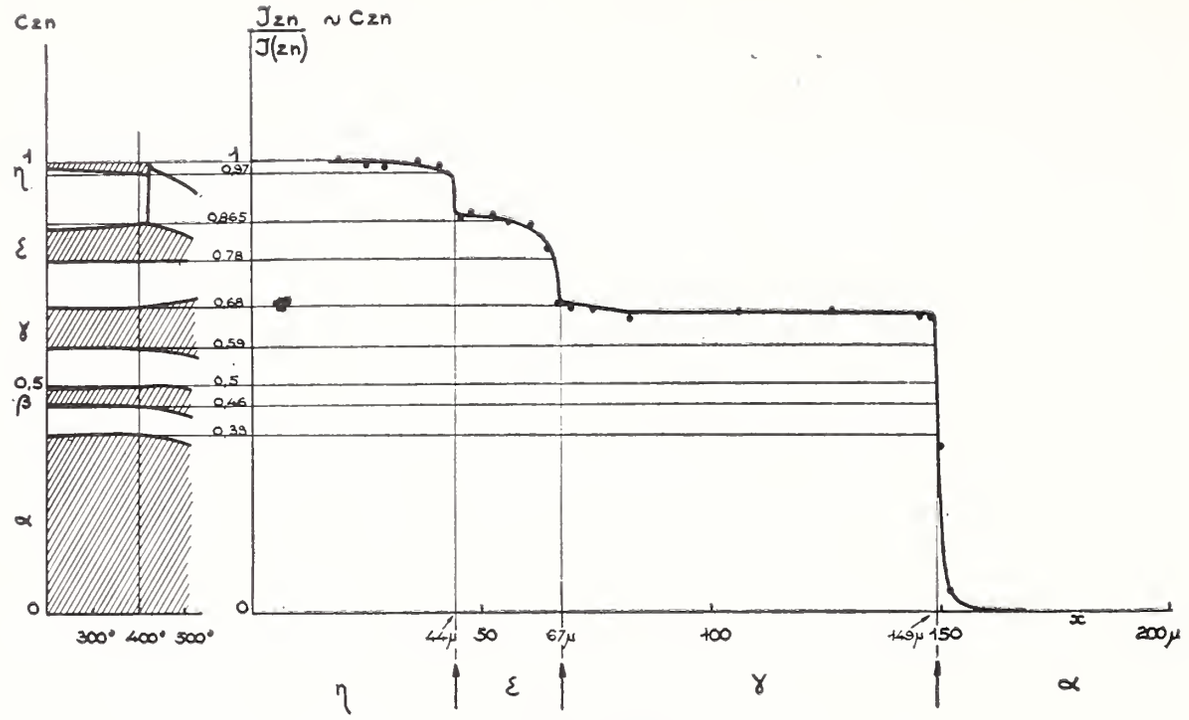


FIGURE 35.2.

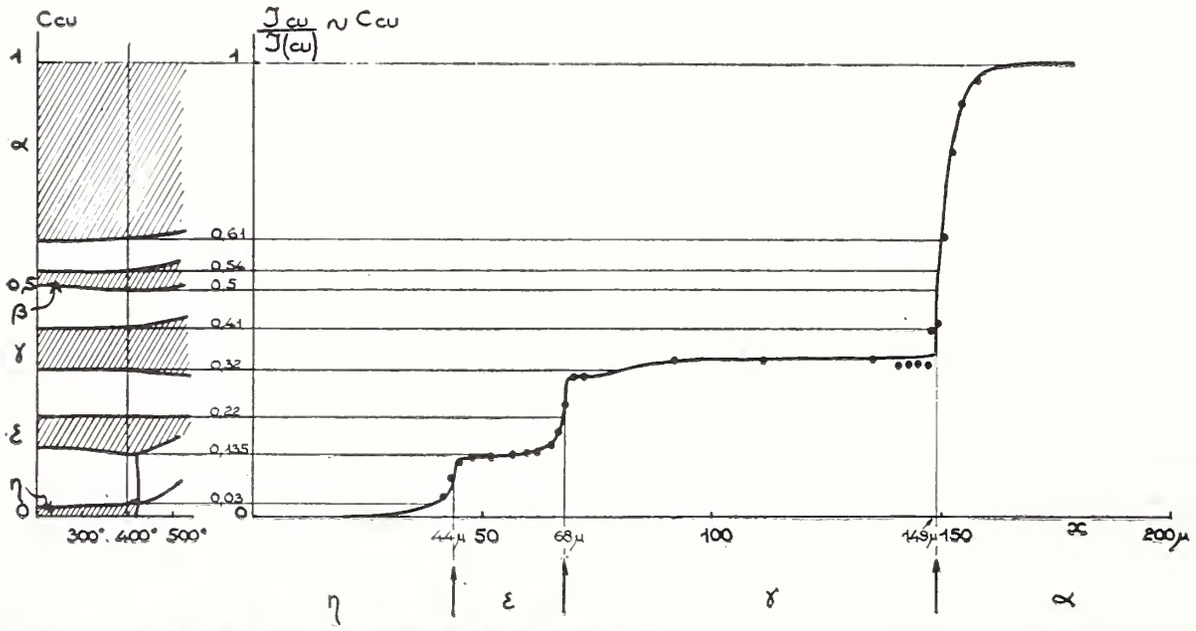
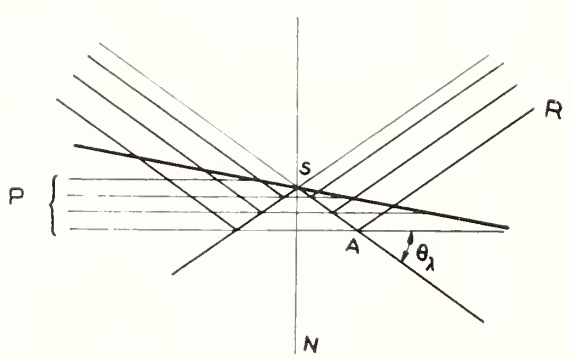


FIGURE 35.3.

FIGURE 35.4.
Magnification ≈ X250

FIGURE 35.5.
Magnification ≈ X300

FIGURE 35.6.



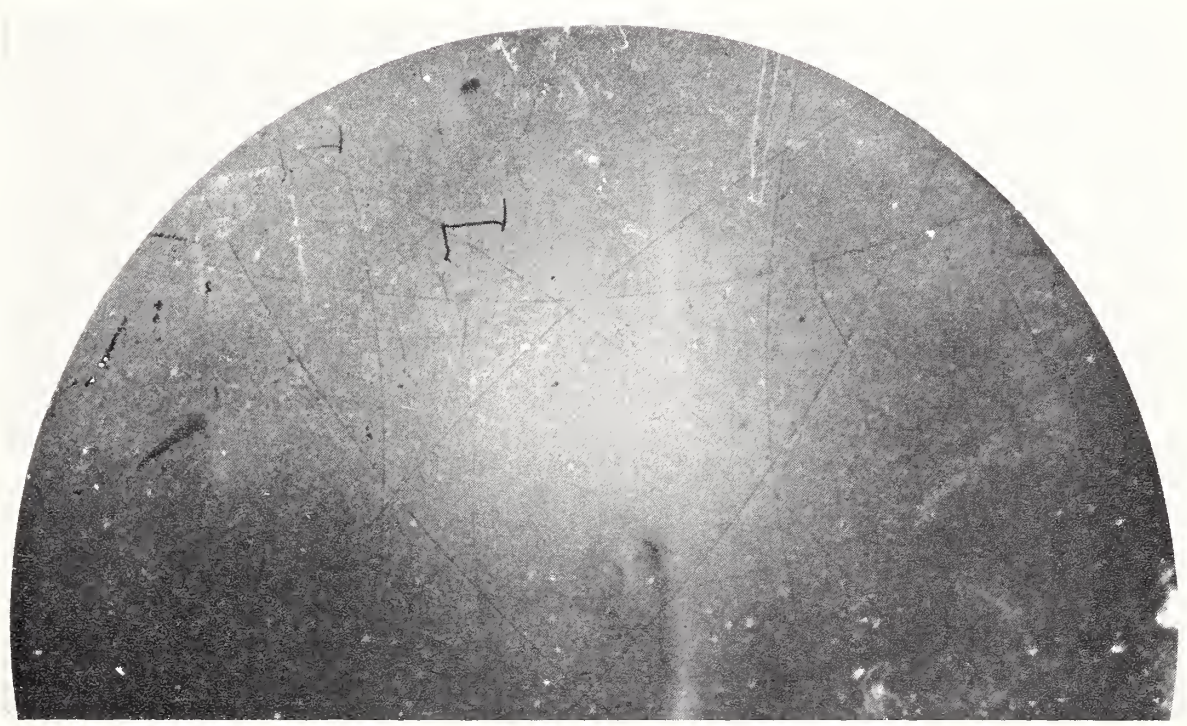


FIGURE 35.7.

36. Ions in Cathode Ray Tubes

By C. H. Bachman [1]

The existence of both positive and negative ions in cathode-ray tubes has been known for some time; both kinds contribute to shortening of tube life. The negative ions cause screen burning (decreased or altered fluorescence), and the positive ions not only burn the screen but bombard the cathode with a resultant loss in emission.

By magnetic separation of the focused-ion components in standard cathode-ray tubes the identities of the negative ions were established some fifteen years ago.[2] The components may be seen in fluorescence and recorded as burned spots. So far some thirty negative ions have been indicated in studies of this kind.[3 4 5] (This total is made up from the reports of several investigators). Some of these studies were made with demountable systems; others with sealed-off tubes. Some studies were made under the best possible vacuum conditions; in others contamination was purposely introduced.

Negative-ion currents of 10^{-10} amp to 10^{-13} amp have been reported. No identification of the positive-ion components has been reported despite the fact that evidence indicated that there are many more positive ions than negative ones.

One of the features of positive-ion destruction of the cathode is that the ions apparently arrive at the cathode as a well focused beam. Utilizing this feature we constructed a "hollow" cathode electron gun as shown in figure 36.1. The "electron" side of this gun is a typical electrostatically focused cathode-ray tube structure of the type with which positive-ion beam effects have been observed at the cathode.

In this structure the beam arriving at the cathode passes on through and is examined by suitable means on the "positive ion" side. The beam may be accelerated, focused, and deflected for analyzing, or bucking potentials may be used to study energies.

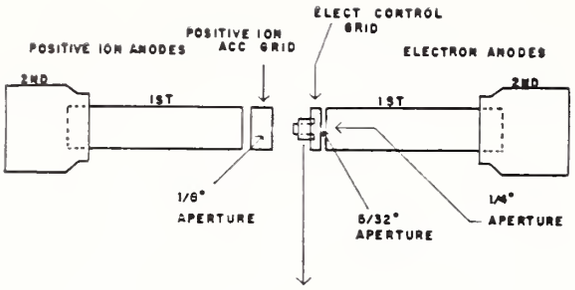
A description of this system and preliminary results have been published.[6] From work with this setup we learned that the positive ions are apparently formed on or close to the surface of the first accelerating electrode passed by the electron beam. They arrive at the cathode with energies corresponding closely to the potential of the electrode. No ions could be detected coming from the fluorescent screen or any electrode other than the first. We also learned that these positive ions could be directed on to a phosphor on glass without the charging difficulties that one might anticipate although there was instability in some cases. Positive ions seemed about as efficient as negative ions in burning the phosphor.

As a result of this information we turned to standard-production-type cathode-ray tubes with guns of the type shown in figure 36.2.

---

[1] Syracuse University, Syracuse, New York.
[2] C. H. Bachman and C. W. Carnahan, Proc. Inst. Radio Engrs. **26,** 529 (1938).
[3] L. F. Broadway and A. F. Pearce, Proc. Phys. Soc. London **51,** 335 (1939).
[4] H. Schaefer and W. Walcher, Z. Phyik. **121,** 679 (1943).
[5] R. H. Sloane and C. W. Watt, Proc. Phys. Soc. London **61,** 217 (1948).
[6] C. H. Bachman, H. Eubank, and G. Hall, J. Appl. Phys. **22,** 1208 (1951).

257899—54——21



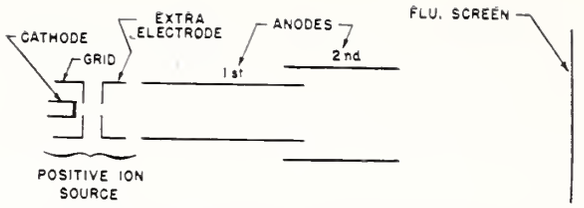


FIGURE 36.2. *Electrostatic gun for positive and negative ion studies.*

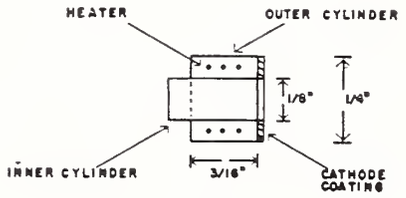


FIGURE 36.1.

This is an elctrostatically focused gun containing an extra electrode between control grid and the first anode. This electrode normally operates at a few hundred volts. It is seen that the cathode, control grid and extra electrode serve as a positive ion source. By running the first and second anode potentials at suitable negative voltages, positive ions are attracted to the screen instead of to the cathode. By focusing these and applying a transverse magnetic field, positive-ion spectra can be observed and recorded as radial rows of spots just as has been done with negative ions. In fact, by shifting the polarity of the anode voltages the negative-ion spectra can be placed beside the positive-ion spectra of the same tube.

Such studies have been made on many tubes of various kinds. The most persistent ions, positive and negative, are listed in table 36.1.

TABLE 36.1.

M/e number	Positive ions		Negative ions
	Identity	Estimated Intensity	Identity
1	H^+	9	H^-
2	H_2^+	10	----------
4	O^{+++}	4	----------
6	C^{++}	7	----------
7	N^{++}	4	----------
8	O^{++}	6	----------
12	C^+	8	----------
14	N^+	6	----------
16	O^+	8	O^-
26	--------	--------	$(CN)^-$ or $(C^2H^2)^-$
28	CO^+	10	----------
35—	--------	--------	Cl^-

There was an inconsistent appearance of a few other negative ions between mass numbers 40 and 100. The strongest negative ion is mass 1, having an estimated intensity of about 6 on a scale of 1 to 10 with 10 as the most intense. The rather violent type of ionization depicted for the positive ions leads one to suspect that these basic atoms were originally part of a more complex system.

It is a pleasure to acknowledge the help of the General Electric Company, which has partially supported this investigation at Syracuse University.

Discussion

Mr. M. Krawitz, Sylvania Electrical Products, Inc., Bayside, L. I., N. Y.: In conventional cathode rays we use ion traps in order not to burn the screen, and I have noticed time and again that on the second grid, where you have the ion trap, you have a well-worn spot due to ions coming back. That means they must have been formed somewhere beyond that. That is, beyond the second grid.

Dr. Bachman: There are many instances of positive-ion burns on the back of apertures. This we encountered many years ago. With the particular setup we are using, utilizing bucking-potential methods, we tried to find evidences in our demountable tube of anything coming from beyond that first aperture. We were unable to do it in this particular case. The fact that you find them, and we have already found them in other instances, is a problem we are still concerned with. You also get burns on the control grid surface. I think I know what that is, but I don't have it ready for discussion yet.

Question: How many ions and positive ions of oxygen are there?

Dr. Bachman: The only thing that we have is this—the majority of the ions are hydrogen. That is 10 on our estimated scale. The other estimates are given. We are making measurements now with d-c amplifiers.

Dr. V. E. Cosslett, University of Cambridge, Cambridge, England: Your table gave only the estimate of the positive ions.

Dr. Bachman: I stated that hydrogen in the case of the negative was the strongest, and it would be about 6 on that same scale. The others would be way below that.

Mr. C. J. Calbick, Bell Telephone Laboratories, Murray Hill, N. J.: Well, is that to say that the negative ion acts in the same way on the phosphor as the positive ion?

Dr. Bachman: Yes. We have noticed one other thing: A given phosphor will burn two different colors, depending sometimes upon the type of ion. In other words, the burn that you might get from oxygen would give a different type of change as evidenced in fluorescent color, from one you would burn with carbon. What is back of this, I don't know. As to the relative intensities and currents, the workers in England, I believe, found negative ions to be to the order of 10^{-10} and 10^{-13} amperes. We have checked that roughly, and we are getting positive-ion currents that are of the order of 1,000 times greater.

Mr. Krawitz: Do you get these peculiar positive-ion burns on the cathode-ray tubes in the form of crosses?

Dr. Bachman: If you will look back at a paper I published in 1938, I showed pictures of some of rectangles, crosses, and other shapes. The crosses, I am pretty sure, are due to positive ions originating in this region and being drawn toward the screen, which may be at a lower potential than the surrounding anode. I think there are two kinds of crosses involved. In the case I referred to years ago, we were scanning with a sine-wave type of raster and I think my solution is right as given then.

Mr. Krawitz: You get the same thing in television rasters.

Dr. Bachman: Yes, but this is a little different thing. You don't get a true cone projection in television rasters. You get peculiar shapes, including crossed arcs. This I can't explain yet.

Mr. Calbick: Does the use of metalized screens or metalized phosphors affect the ion spotting very much?

Dr. Bachman: We have made some tubes in which half of the screen area will be aluminized. We can put negative or positive ions on one side or the other. In general, we find that the ion burning is decreased. In other words, it takes a longer time to get ion burns through aluminum, but at least up through mass 14 or 16 you can still get them. The thickness of aluminum films that are used are not sufficient to completely stop them. It merely decreases the effect.

Mr. M. Knoll, Princeton University, Princeton, N. J.: You mentioned that you expected positive ions to charge up the screen to an amount you did not observe.

Dr. Bachman: We observed it, but it was not so bad but that we could carry on. Electrons, in this case, have no place to go if they are knocked out. This does not rule out the possibility that the positive ions might be striking these surfaces and knocking out positives. If what you mean is positive-ion secondary emission, I think it could very well be a factor.

37. High-Voltage Electron Guns[1]

By M. E. Haine [2] and P. A. Einstein [2]

Throughout any electron optical system free from absorption or scattering, the product of the brightness or current density per unit solid angle and the square of the refractive index is invariant. Thus, the brightness produced by an electron gun is given by

$$\beta_i = \frac{\rho_0 \phi_0}{\pi \phi_e},$$

where ρ_0 is the cathode emission density, ϕ_0 is the accelerating voltage and $e\phi_e = KT$, the mean emission energy of electron (K=Boltzmann's constant and T the absolute temperature). This value is independent of the focusing system and its aberrations. It follows that the current density at any point is given by

$$\rho = \rho_0 \frac{e\phi_0}{KT} \alpha^2, \ (\alpha \text{ small})$$

where α is the semiangle of the cone of electrons at the point under consideration. α is, of course, dependent entirely on the focusing system and its geometric aberrations.

In the electron microscope the illuminating beam angle is determined by imaging requirements and hence the current density of the object is left dependent only on the accelerating voltage, cathode current density, and temperature. The beam angle must be adjusted to the correct value through suitable variation of the focusing system.

The current density on the final screen on the electron microscope is given by

$$\rho_f = \rho_i / M^2 = \frac{\rho_0 \phi_0}{M^2 KT} \alpha^2$$

For satisfactory vision of the fluorescent screen this current density must equal at least 5×10^{-11} amp/cm^2.

Figure 37.1 shows the required cathode current density plotted against illuminating semiangle for various magnifications of an instrument operating at 50 kv. Running the cathode at a temperature of 2,800° K should provide an emission of 10 amp/cm^2; this is about the maximum temperature that will give a usable life (10 hr). For an objective lens of the spherical aberration constant of 0.1 the optimum semiaperture angle is 2×10^{-3}. With critical illumination, focusing by direct vision should therefore be possible up to 300,000x magnification with the cathode current density quoted. With an

[1] The subject matter of this paper will appear in extended form in the British Journal of Applied Physics, by whose permission figures 37.2 to 37.10 are reproduced.

[2] The Associated Electrical Industries Research Laboratory, Aldermaston, Berkshire, England.

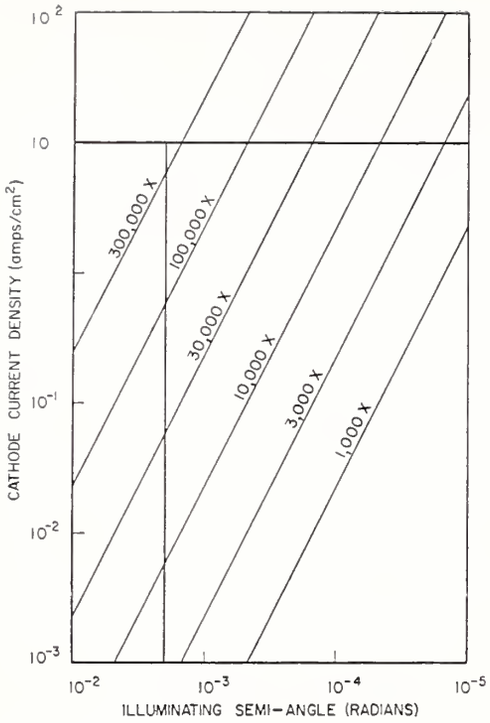


FIGURE 37.1. *Cathode current density to give 5 x 10^{-11} amp/cm² on final screen.*

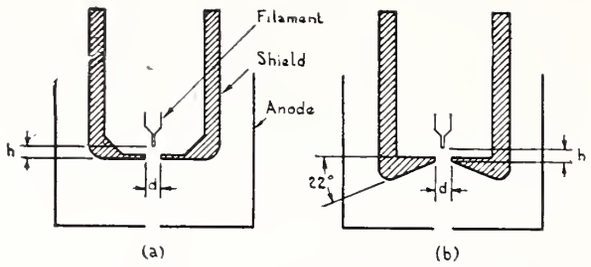


FIGURE 37.2. *Electron gun geometries investigated.*

illumination angle of one-tenth critical as used in practice, this is reduced to 30,000x. Thus we see that a perfect gun should give adequate illumination for normal use of the instrument. The question arises as to how nearly perfect is the type of gun in common use. In a recent investigation we have been able to show that under the correct operating conditions this type of gun does, in fact, give the theoretically predicted brightness. One may well ask how this can fail to be the case as the brightness is invariant. It may be worth considering this briefly before describing some of our results in more detail. The most obvious possible cause of brightness falling short of the theoretical value is the limitation of emission by space charge. It is well known that this occurs if the electric field in front of the cathode is insufficient to draw off all the available electrons. That this does not, as has often been supposed, apply in the electron microscope gun except at extreme cathode temperatures will be explained later. The effect that does reduce the apparent brightness under certain conditions results from chromatic aberration, or better, chromatic-geometric aberration.

The experiments we have performed were set up to measure the variation of brightness, beam angle, and current density of a range of electron gun geometries in order to relate these with theory and, if possible, to obtain improved performance by change of geometry or operating conditions.

The types of electrode structure investigated are illustrated in figure 37.2. They comprise a hairpin filament, circular grid, or shield and anode with central apertures. Two shapes of shields were investigated, one being flat and the other a reentrant cone. The overall apparatus is shown schematically in figure 37.3. The general arrangement is similar to that used by several other authors. It comprises the experimental electron gun, which is designed to allow ready interchange of the electrodes, and centering and adjusting the height of the filament while the gun is operating. The high voltage is variable up to 100 kv and a variable bias power unit is incorporated.

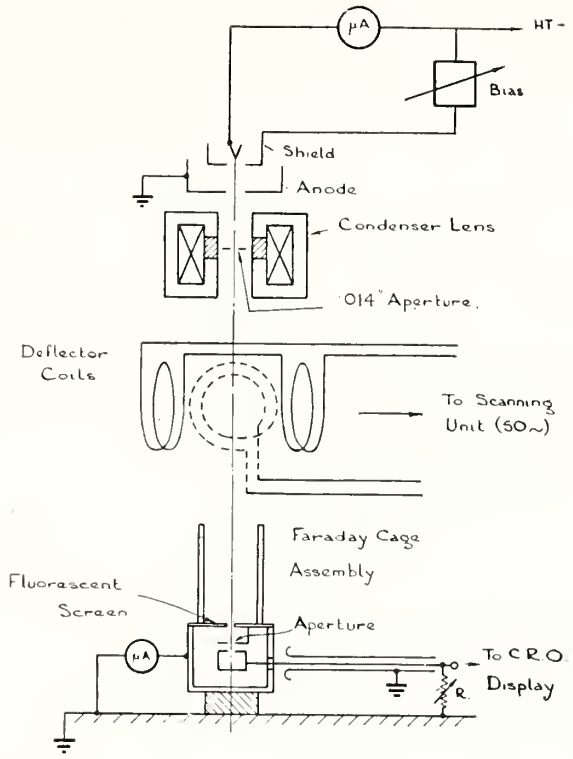


FIGURE 37.3. *Experimental apparatus.*

Following the electron gun is a magnetic focusing lens, which can image the beam on to a small aperture in a Faraday cage situated at the bottom of the apparatus. In the lens is placed an aperture that can be centered or removed from outside the apparatus. The focused or unfocused beam can be swept across the aperture in the Faraday cage by means of two perpendicular pairs of deflector coils fed with 50 cycles alternating current from a Selsyn. The current collected in the Faraday cage is passed through a load-resistance and the voltage developed amplified and displayed on a cathode ray oscillograph against a time base synchronous with the current in the deflector coils. Thus, the current density distribution across the focus spot or across the unfocused beam can be recorded. If the beam-limiting aperture is in position, and the beam focused, the oscillogram gives the brightness and size of the focused spot. If the aperture is out and the beam unfocused, the distribution of current density with beam angle is recorded. In addition, means are provided for measuring the total beam count, the bias voltage and the high voltage.

If the results obtained on this apparatus are to be related to the theoretical predictions, then it is of the greatest importance that the cathode temperature shall be known to a reasonable degree of accuracy. In our experiments, the cathode temperature was measured by a modified disappearing-filament optical pyrometer. The accuracy of the latter measurements was checked with the aid of a demountable cylindrical diode having two guard rings. The saturated current density obtained from samples of wire as used in the electron gun filaments were measured as a function of temperature. The good agreement found with figures published by Jones and Langmuir is shown in figure 37.4. This gave us reasonable confidence in the accuracy of our measurements. In figure 37.5 is shown a series of results plotted graphically. These curves show the brightness in

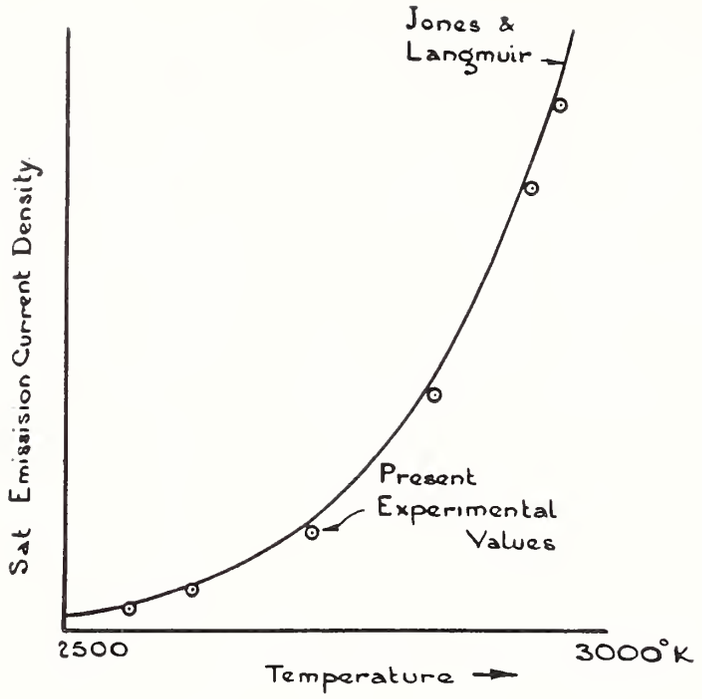


FIGURE 37.4. *Comparison between intrinsic emission from tungsten wire as measured by the authors and by Jones and Langmuir.*

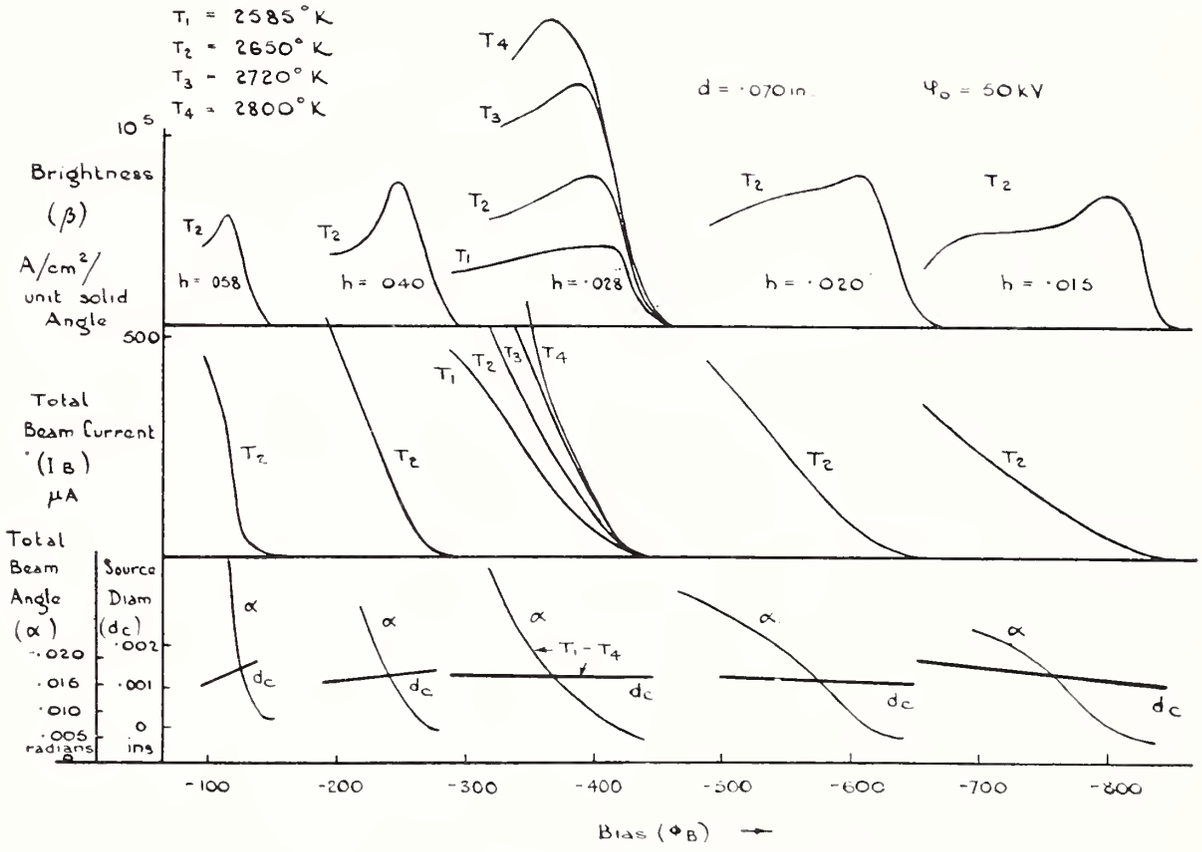


FIGURE 37.5. *Electron gun characteristics.*
Flat shield.

amperes per square centimeter per unit solid angle, the total beam current in microamps, the beam angle, measured as a half width, and a source diameter, again measured as a half width. All these are plotted against the bias voltage for a constant high voltage of 50 kv. The series of curves running across the bias scale are for different filament heights for a constant aperture diameter in the grid. It will be seen that for all these geometries a similar type of characteristic is obtained. The brightness rises with increasing bias to a maximum and then falls rapidly down to zero, the total current falling over the

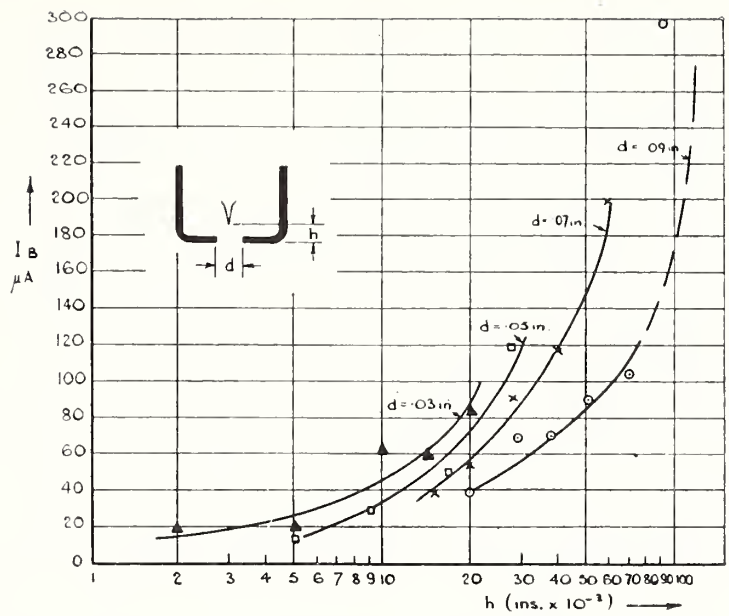


FIGURE 37.6. *Filament height and shield hole size versus current at maximum* $\beta(=80{,}000$ *amp/cm²/steradian) (constant temperature).*

whole range of bias; the beam angle also falls as bias is increased while the source size remains substantially constant. At one height a series of characteristics are shown for different cathode temperatures ranging from 2,500 to 2,800° absolute. It will be seen that the brightness goes up with temperature as would be expected and that up to a temperature approaching 2,800° the value of bias for the maximum brightness is unchanged. Thereafter the bias for maximum brightness is decreased. The total current increases also with temperature but, on the other hand, the beam angle and the source size remain independent of temperature.

It will be seen that variation of height has little or no effect on the maximum brightness. Indeed, this was found to be true also for variation in shield aperture diameter and shield slope for which similar series of curves were drawn. Such changes have no effect either on the focused spot size, which remains remarkably constant whatever parameter is varied.

Thus, as far as the maximum brightness is concerned, electron gun geometry is of little importance. It is, however, of some importance if one is interested in obtaining this maximum brightness with as small a total current as possible. In figure 37.6 is plotted the total current required in the gun to give the maximum brightness with different geometries. It will be seen that for a minimum total current it is desirable to use as small a height as possible; that is to say, the filament should be as far forward in its shield as possible and the diameter of the aperture in the shield should be as small as possible. The desirability of reducing the total current to an absolute minimum should not be over stressed. Except in very high-voltage electron guns it is of comparatively little significance whether the total current is 10 microamps or 100 microamps. It is, in fact, very desirable that some current should be thrown away in the condenser aperture. This results from the fact that the current density distribution across the unfocused beam is approximately Gaussian, and if all the beam is included in the condenser aperture as would be obtained if the gun is designed to work with very high current efficiency, then a lower

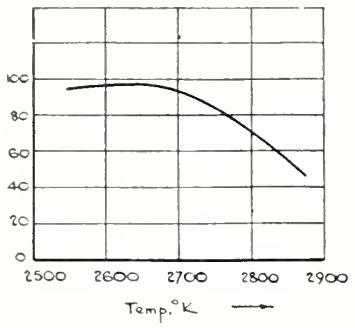


FIGURE 37.7. *Plot of brightness efficiency versus filament temperature at maximum β points.*

Bias not constant. Flat shield, 50 kv, d=.050 in., h=.009 in.

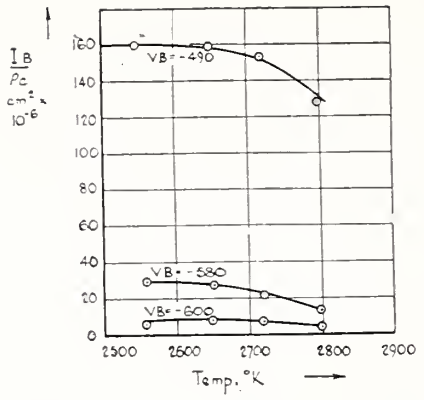


FIGURE 37.8. *Plot of beam-current-cathode-current density ratio versus filament temperature at different values of bias.*

Flat shield, 50 kv, d=.050 in., h=.009 in.

brightness is obtained at critical focus, because the outer part of the aperture is filled with electrons of comparatively low density.

In figure 37.7 the maximum value of the brightness divided by the theoretical brightness is plotted against temperature and it will be seen that for temperatures up to 2,700° the maximum brightness obtained is substantially 100 percent of the theoretical value, thereafter up to 2,800°, which is the maximum practical operating temperature of the cathode, there is a small drop in this ratio. Thus, provided the gun is run under the optimum conditions, it is possible to obtain the theoretical brightness from it; and no improvement can be obtained except by improving upon the intrinsic emission of the cathode.

Part played by space charge. In the past, considerable emphasis has been placed on the part played by space charge in this type of electron gun. From the results described so far there is no particular reason to suppose space charge is of any significance below a cathode temperature of about 2,800° K., before which the brightness is equal to the theoretical, and at which the filament life is limited to about 10 hr. by evaporation. To confirm this, some auxiliary experiments were carried out. Space charge can be expected to have one or both of two effects. It can limit emission and/or cause beam spreading, which may show up as an increase in beam angle or possibly an increase in the focused-spot size. Now any of these effects can be detected relatively easily. If a gun is operated with all parameters except the cathode temperatures fixed, then, in the first place, the variation of total current with cathode temperature should follow the known law for a temperature limited cathode if space charge effects are not present, and should vary less if space charge effects are present. In the limit the current will be almost entirely independent of temperature. Any beam spreading effect due to space charge will be observed by the change in beam angle or source size as the temperature is changed.

In figure 37.8 is plotted the ratio of the total current to the saturated cathode current density against the temperature. It is seen that the ratio is substantially constant until a temperature approaching 2,800° K. is reached. The next figure shows a series of oscillograms showing

the distribution of current density across the unfocused beam for different values of temperature. Again, the beam angle is constant up to temperatures approaching 2,800°. Clearly space charge is not significant at temperatures normally used. Not only is this so for the optimum operating bias but for the whole range of bias over which the gun could be operated, even very close to cutoff. It is true that space limitation of emission must inevitably take place before the gun is completely cut off, but it was not possible to detect it by the methods described. Thus, the cut-off part of the characteristic is not determined by the field strength in front of the cathode. This may seem somewhat surprising, particularly when one considers the well-known phenomena which takes place with most electron microscope guns when as the cathode temperature is raised a point is reached where the total current ceases to rise just as if emission had reached a space-charge limited condition. The effect is in fact due to the feedback action of the cathode bias resistor. This is illustrated in figure 37.9. Curve A shows the total gun current plotted against bias for different temperatures. The straight lines have slopes corresponding to various values of the bias resistor. The operating point for a particular value of resistor and temperature is the intersection of the corresponding straight line and the total current curve. The way in which the total current varies with temperature for a fixed bias resistor is illustrated at B. It is seen that the saturation effect results not from space-charge limitation but from the form of the beam current-bias curves.

The optimum value of bias resistor for a given geometry and temperature can be determined from the curves by drawing a vertical line from the maximum of the brightness curve (C) to intersect the corresponding total current curve (A). The slope of the line from the zero of coordinates to the intersection corresponds to the required

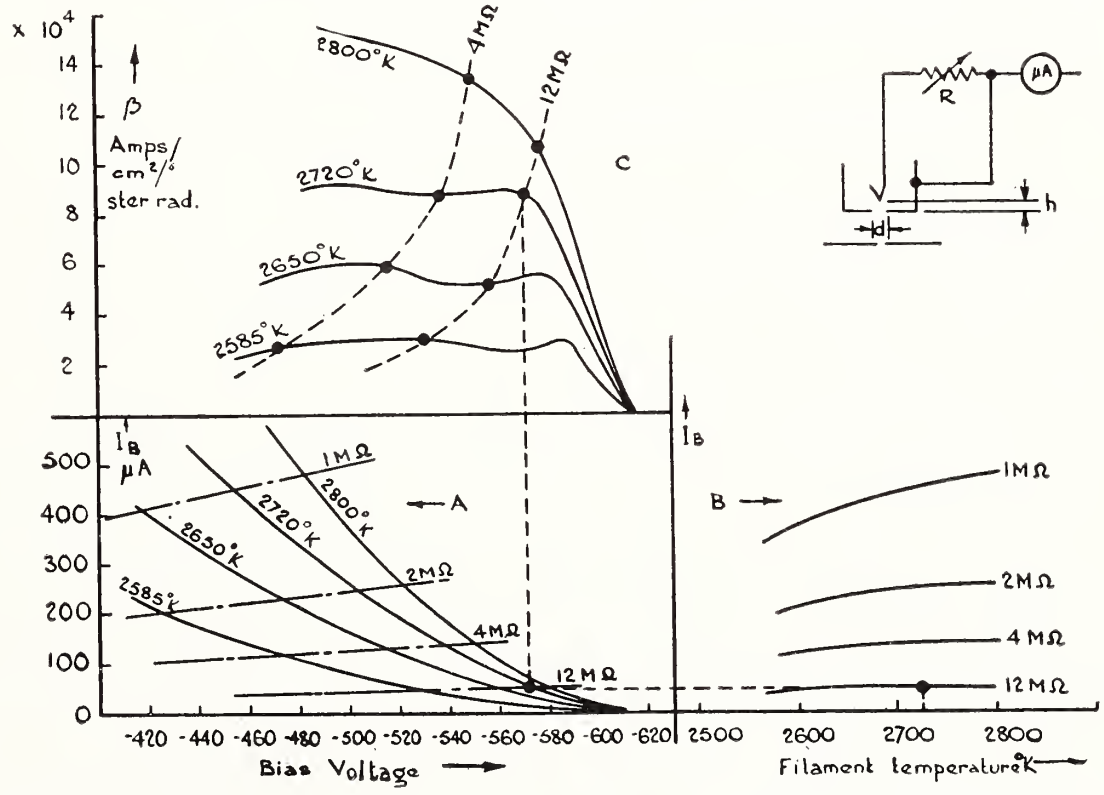


FIGURE 37.9. *Diagram illustrating operations of automatic bias with bias resistance and filament temperature varied.*

Fixed geometry, d=.050 in, h=.009 in., kv=50.

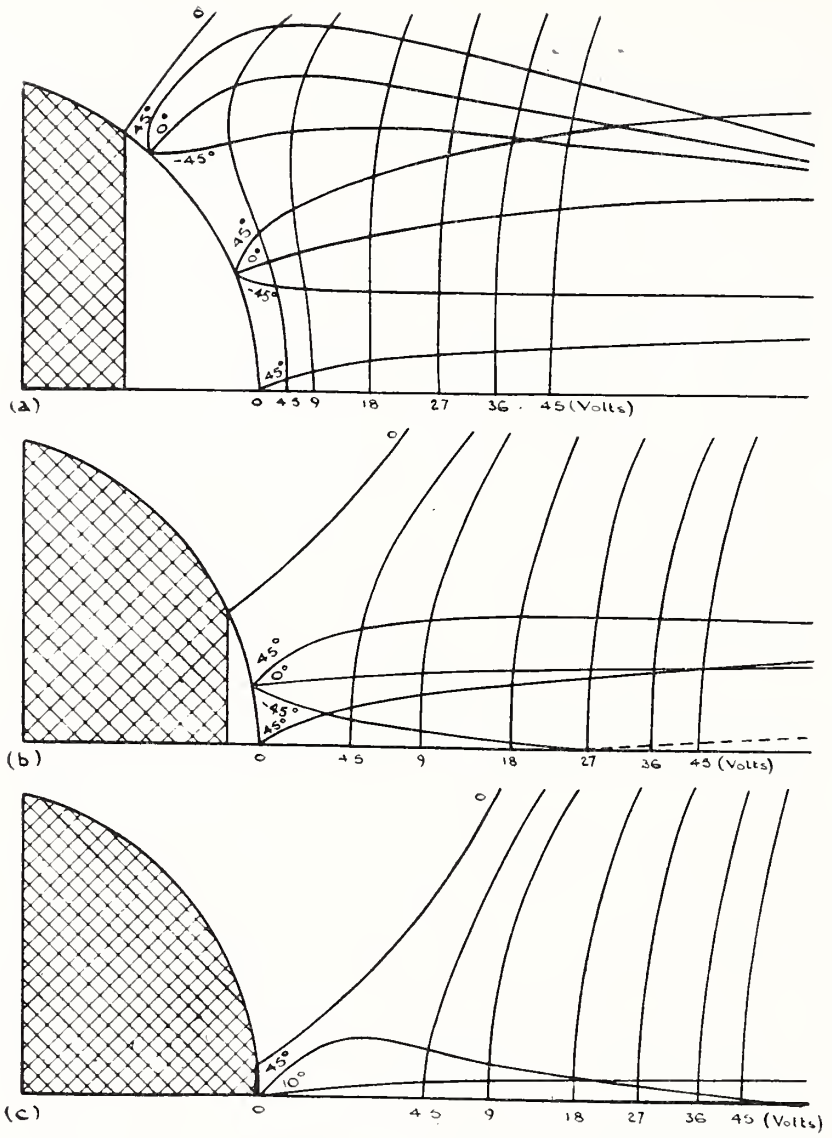


FIGURE 37.10. *Potential field and electron trajectory plots near cathode tip.*

Initial velocity 0.3 volt. a, At low bias (beam current, 500 μa); b, at optimum bias (beam current, 100 μa); c, near cutoff (beam current, 10μa).

value of the bias resistor. In practice the bias for a given gun is so sensitive to changes in filament height that it is not normally possible to work from the curves in this way, since the height cannot simply be set with adequate accuracy; it is therefore desirable to arrange for the bias resistor to be variable.

In conclusion, it is now possible to explain in a somewhat qualitative way the mechanism of beam formation in the gun. Figure 37.10 shows the field conditions near the cathode for three values of bias and a typical geometry. The field plots were obtained on the Liebmann resistor network and show how the zero equipotential strikes the anode when bias is applied and restricts the emission area more and more as the bias is increased, thus accounting for the cut-off characteristics of the gun. The cause of the aberration effect is also clear; at large bias near cut-off (C) the electrons of high radial velocity move into a strongly convergent field. At optimum bias (A) the field is mostly weakly convergent, only a small proportion near the periphery being strongly convergent. At low bias (B) the field near the axis is divergent causing small aberrations and a corresponding small reduction in brightness. The field near the periphery is again

strongly convergent and leads to the well-known hollow beam effect. At low bias the distribution across the beam is approximately Gaussian. As the bias increases the hollow beam is formed, eventually breaking up to give a complicated structure due to electrons emitted from all parts of the filament, including the back, being permitted to enter the beam by the receding zero equipotential.

Discussion

Dr. J. Hillier, RCA Laboratories Division, Princeton, N. J.: You said that aberrations did not affect brightness, and yet you proceeded to show an effect where it was aberrations that were affecting brightness.

Mr. Haine: Brightness is invariable and cannot be affected by geometric aberrations alone. It is the effect of the combined chromatic-geometric aberration in filtering out a proportion of the high-emission velocity electrons that results in the apparent fall in brightness.

Dr. Hillier: You are, in other words, saying that the large-angle electrons are always ones of different emission velocity.

Mr. Haine: The effect is to cut out all those electrons with high *radial* emission velocity. These, of course, will have a high total emission velocity, but only part of the high emission velocity electrons are lost, those emitted in a direction close to the exis are not lost. The effect on the Maxwellian velocity distribution is to reduce the height of the high velocity tail of the curve.

Dr. Hillier: What is the brightness given by a biased gun operating near the saturation current value; is this near the maximum or on either side of it?

Mr. Haine: The brightness at current saturation depends on the value of bias resistor used. A suitable choice will give maximum brightness. The correct choice of bias resistor can be obtained from the curves of figure 37.5.

Mr. M. Krawitz, Sylvania Electric Products, Inc., Bayside, L. I., N. Y.: How did you determine the position of the minimum beam diameter near the cathode?

Mr. Haine: The position of the source of minimum diameter is not necessarily the position of the minimum beam diameter. The position was obtained from the calibration of the focal length of the imaging lens.

Mr. J. A. Reisner, RCA Victor Division: At what pressure did you carry out these experiments with the gun?

Mr. Haine: The gas pressure was varied from 10^{-6} up to about 10^{-3} mm of Hg; no effect was observed due to space charge.

Mr. Reisner: At what pressure did you notice a difference in the beam that was coming out of your gun? If you indefinitely increase the pressure, you ultimately reach a point where you do have a good focusing effect. Where did you notice this in the gun? We have noticed this effect very definitely at about 10^{-3} mm of Hg. It is a serious effect in microscopy where you have a leaky system giving a pressure above that value because you can make a self-bias gun; but it seems to be rather continuous below that pressure at 10^{-3}, or perhaps two or three times lower pressure. It is noticeable, I believe, from the experiment we have carried out, so you must have been well below that value.

Mr. Haine: I can only repeat that careful experiments were carried out to check the effect of pressure and up to 10^{-3} mm of Hg; no effect was observed. We should not attempt to operate a microscope at a pressure even as high as this. We looked particularly for some effect, because we had observed a very big effect in the cylindrical diode used to calibrate the emission-temperature relation of the cathode material.

38. Electron-Lens Raster Systems[1]

By M. Knoll[2]

1. Principle

The conventional electron-optical imaging system is used for devices that magnify, demagnify, convert, or amplify electron images. In order to avoid electron-optical errors, the diameter of the electrostatic or magnetic electron lenses that produce the imaging field must usually be several times greater than the beam diameter. Thus, for reasons of convenience, electron image systems have almost always been built hitherto with rather small beam diameters.

In cases where large images are desired, a different principle may be used. Instead of *one* large beam and *one* large lens, a two-dimensional array of minute, identical, electrostatic lenses side by side is employed, each lens focusing a small element of the object. Because of this, the sizes of the object and image are equal. Also, the electron paths do not cross each other at a large angle as in conventional imaging systems, but form a great number of nearly parallel elementary beams, in which the electron paths cross only at a very small angle. An electron optical system (fig. 38.1) using such an array may, for example, consist of three electrodes: One grid for collecting secondaries and providing uniform field strength for all apertures (100 to 1,000 v positive versus cathode), one lens-raster grid (near cathode potential, to minimize secondary emission fog in the picture), and one conducting luminescent screen (several thousand volts positive versus cathode). In order to maintain the tiny beams narrow as they emerge from the individual lenses, it is essential that the field gradient E_{21} between lens raster and collector grid be much smaller than the field gradient—E_{23} between lens raster and luminescent screen ($E_{23}/E_{21}=10$ to 100). Practically no secondary electrons can reach the luminescent screen from the lens-raster grid or from the collector grid, due to their potential.

One use for such as system is for post-deflection acceleration. In contrast to known post-deflection accelerator systems with only *one* grid, however, where secondary emission and the irregular action of grid openings have been considered a great hindrance,[3] these double-grid systems provide a relatively high resolution. Analogous to the lenticular rasters in color photography, such devices may be called electron-lens raster systems.

In addition to amplifying larger images, such systems have a further advantage: By covering the lens-raster screen on one side with an insulating layer, image *storage* or image *integration* becomes possible, because a charge pattern deposited on the surface of such a layer may control locally a high-energy electron viewing beam.

[1] Research carried out under contract with the Evans Signal Corps Engineering Laboratories, Belmar, N. J.

[2] RCA Laboratories Division, Princeton, N. J.

[3] L. S. Allard, An ideal post deflection accelerator C. R. T., Electronic Engineer **22**, 461 (Nov. 1950).

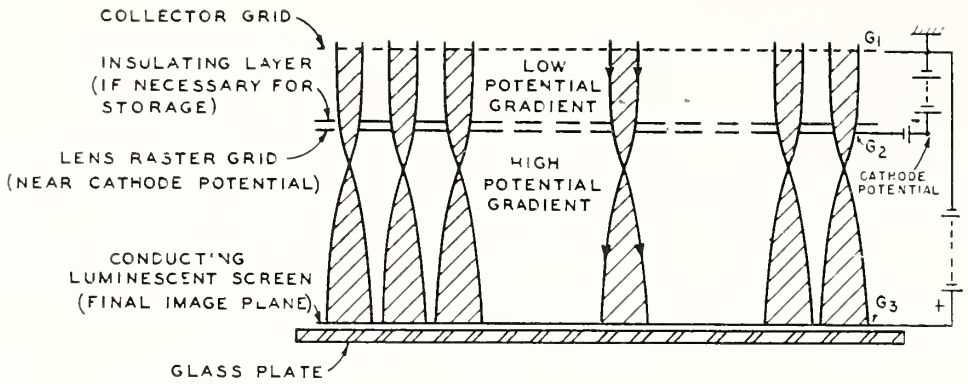
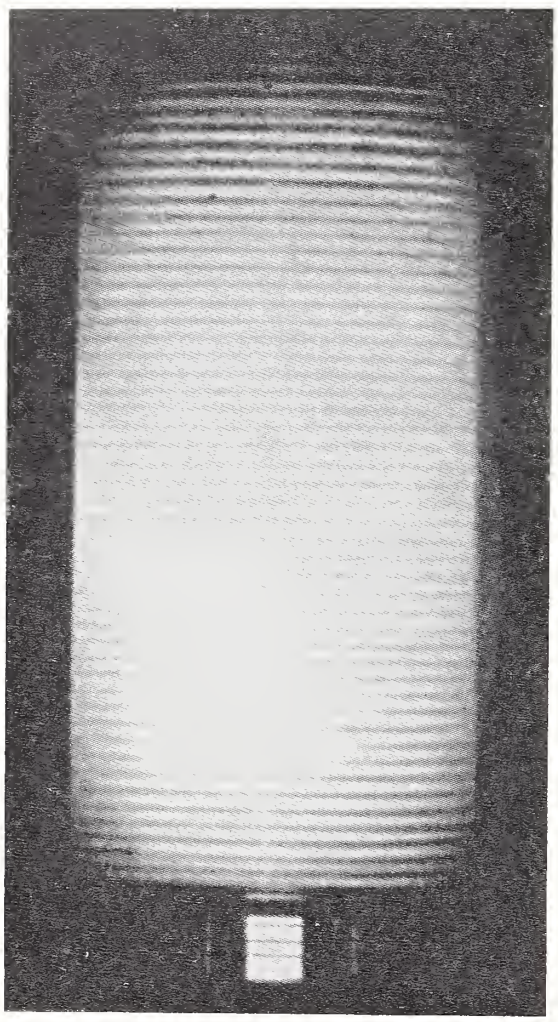


FIGURE 38.1. *Principle of electron lens raster system.*
(Intensifying, integrating, or storing of images.)



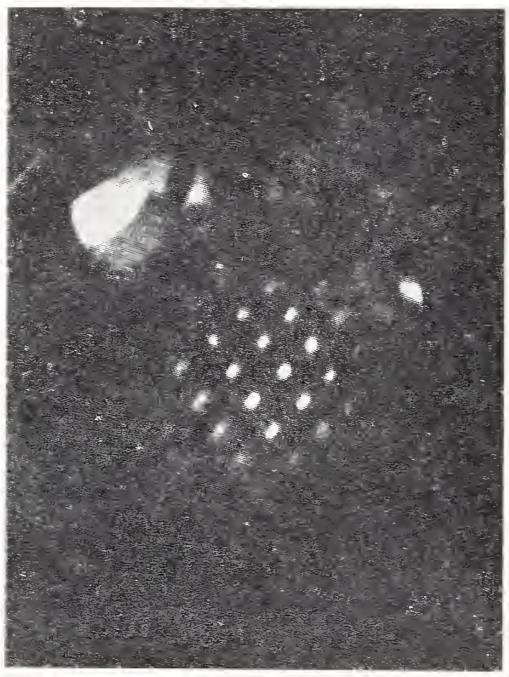


FIGURE 38.2b. *Electron optical dot raster.*
Produced by focusing with mesh grid (made visible by electron microscope).

FIGURE 38.2a. *Electron optical line raster.*
Produced by focusing with parallel grid wires of a triode (made visible by using anode with low heat conduction).

2. Earlier Experiments

The general electron-optical conditions under which line and dot rasters are formed by wire and mesh grids have been investigated for triode and tetrode systems in early experiments. It was shown that an array of apertures in a stretched parallel wire or woven grid, with electrons entering it from one side, and provided with a plane parallel positive final anode on the other side, produces a raster of focused lines or dots in the space between grid and anode. As an example, figure 38.2A shows a line raster, made visible by electron bombardment of a very thin molybdenum anode,[4] and figure 38.2B a dot raster, produced by a woven grid with rectangular meshes, and made visible by a succeeding electron lens.[5]

[4] M. Knoll, Z. Tech. Physik. **15,** 584 (1934).
[5] M. Knoll and G. Lubszynski, Physik. Z., **34,** 671 (1933).

In addition, it has been shown by graphical construction of the electron paths in triodes that adjacent single beams can be distinctly separated from each other within a certain range of grid potentials V_G near cathode potential (fig. 38.3). This picture also shows that near $E_{21}=0$ the diameter of the elementary beams has a maximum at the anode and that for low positive as well as low negative values of E_{21} the beams become narrower. This can be understood for *positive* grid potentials on account of the converging lens action. For *negative* grid potentials, however, the beams cross over just beyond the grid. One would therefore expect them to become wider at the anode. The beams actually become narrower at the anode, however, with increasing negative grid potentials, the reason being that the action of the negative field around each hole cuts off the outside rays. The positive grid voltage range $(-E_{21})$[6] proved to be important for the development of beam power tubes and tetrodes with low positive grid current. In the case of electron lens rasters, as shown below, the negative grid voltage range $(+E_{21})$ is more advantageous, especially for obtaining suitable current modulation and for proper storage and prevention of secondary emission.

3. Conditions for Optimum Beam Focus (Elementary Beams Separated)

We must first define the electrical and mechanical conditions for focusing. The field for circular and slit apertures was calculated and plotted by Glaser and Henneberg[7] for field strength ratios $\frac{E_{23}}{E_{21}}\approx 3$ to

[6] Negative, since by definition, $E=-\frac{dV}{ds}$.

[7] A. Glaser and W. Henneberg, Die Potentialverteilung in Schlitzblende und Lochblende, Z. tech. Physik **16**, 222 (1935), fig. 7. See also A. Recknagel, Hochfrequenztechnik und Elektroakustik **51**, 66 (1938), fig. 7.

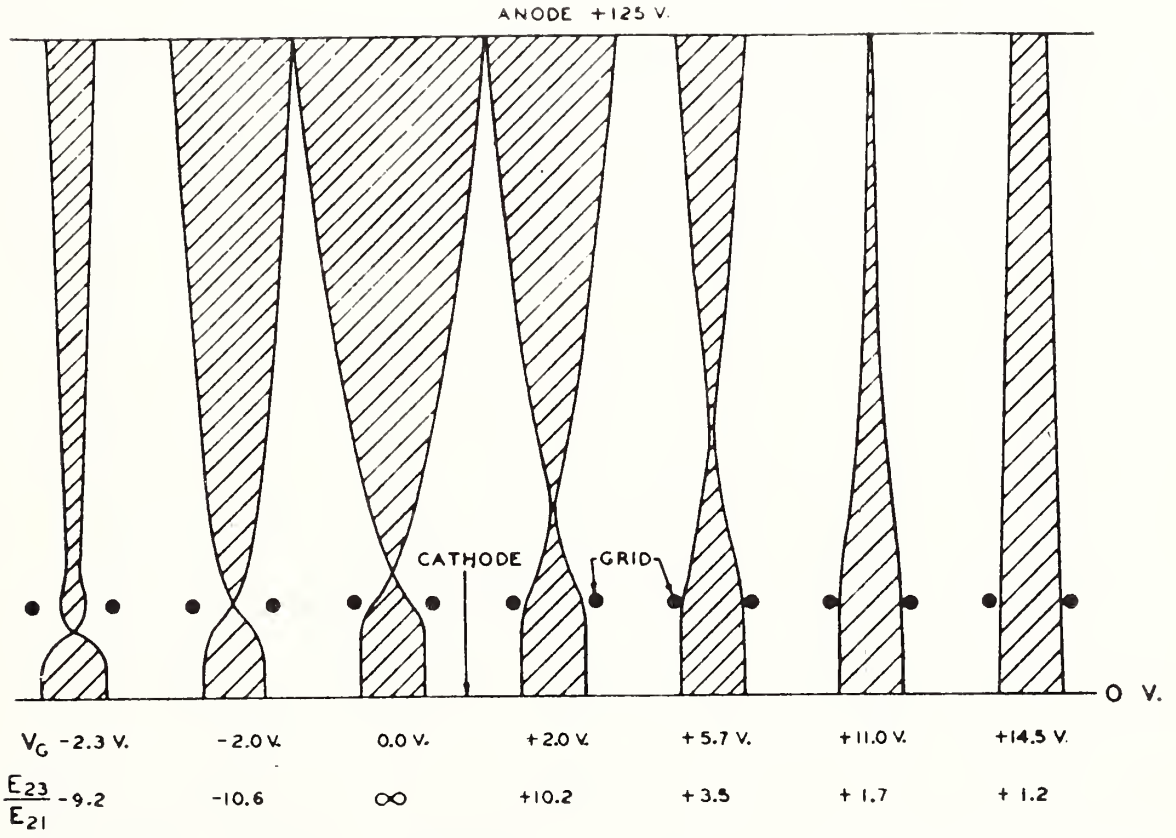


FIGURE 38.3. *Diameter of elementary beams in a triode as a function of potential gradients.*

150. Along the optical axis z of circular apertures it is given by the equation [8]

$$\phi\left(\frac{z}{R}\right)=V_2-R\left(\frac{E_{21}+E_{23}}{2}\right)\frac{z}{R}+R\left(\frac{E_{21}-E_{23}}{\pi}\right)\left(\frac{z}{R}\tan^{-1}\left(\frac{z}{R}\right)+1\right), \quad (1)$$

where $\phi\left(\frac{z}{R}\right)$ is the potential along the axis at a distance of $\left(\frac{z}{R}\right)$ from the plane of the aperture of radius R, and V_2 the potential of the aperture.

The minimum (saddle) point of the potential along the axis can be found by differentiating this equation

$$\frac{d\phi}{d\left(\frac{z}{R}\right)}=-R\left(\frac{E_{21}+E_{23}}{2}\right)+R\left(\frac{E_{21}-E_{23}}{\pi}\right)\left[\frac{\frac{z}{R}}{1+\left(\frac{z}{R}\right)^2}+\tan^{-1}\left(\frac{z}{R}\right)\right]=0. \quad (2)$$

From (2) follows for the saddle point potential: (for $E_{23}/E_{21}=15$, $V_2=0$ volt, $z/R=1.22$), $\phi=22.5$ v.

The corresponding electron paths are given by numerical integration of the equation for paraxial rays [9]

$$\rho''=-\frac{3}{16}\left(\frac{\phi'}{\phi}\right)^2\cdot\rho, \text{ where } \rho=r\cdot\phi^{\frac{1}{4}}. \quad (3)$$

Such electron paths (calculated and drawn by H. Borkan) for a lens raster of 40 holes/cm, $-E_{23}/E_{21}=15$, $V_2=0$ and distances of the luminescent screen of 1.5 or 3 mm are shown in figure 38.4.

As can be seen, the distance of the focal plane is only about one hole diameter distant from the lens raster screen. It should also be noted that the principal plane for the elementary lenses, because of the relatively high field strength ratio, lies nearly 2.3 hole radii away from

[8] V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier and A. W. Vance, Electron Optics and the Electron Microscope (New York 1945), p. 385, eq (11.95).

[9] V. K. Zworykin et al., Electron Optics and the Electron Microscope, p. 402, eq (12, 11) (New York, New York, 1945).

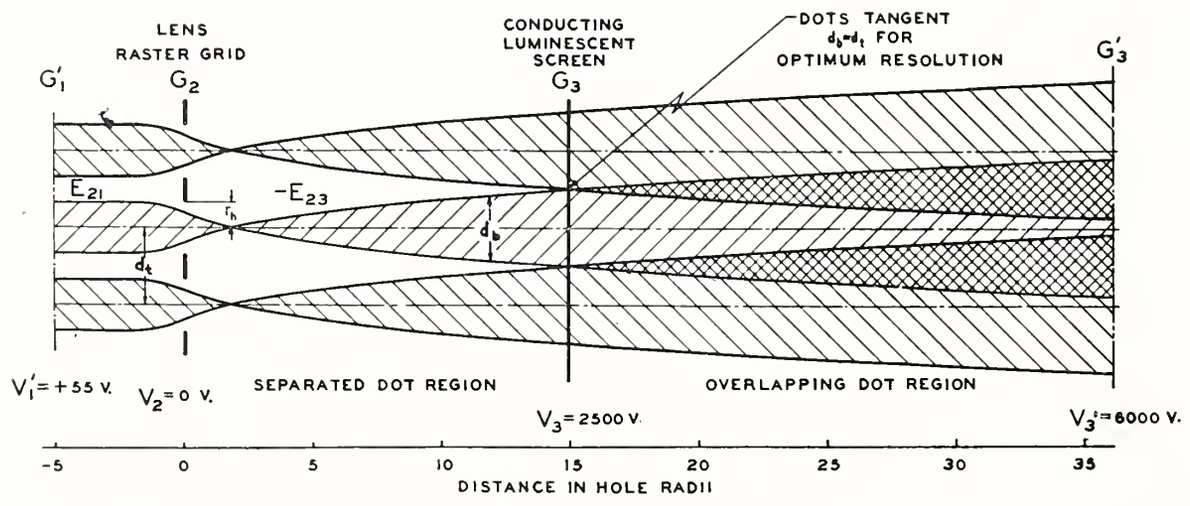


FIGURE 38.4. *Beam paths in electron lens raster systems* $\left(\frac{-E_{23}}{E_{21}}=15,\ V_2=0\ v\right)$.

the lens raster plane on the cathode side of the system, as follows from graphical construction of the tangent to the calculated electron paths.

The overlapping effect of the elementary beams, which is to be avoided,* can be seen very clearly from this picture. For the given field strength ratio, dots become tangent at a plane 15 hole radii away from the lens raster grid, and in a plane 30 hole radii away the overlapping effect is such that the resolution suffers seriously. We define the "tangent spot" condition as "optimum beam focus," considering the fact that still shorter distances of the luminescent screen, which would not deteriorate the resolution, are usually undesirable due to voltage breakdown caused by the increased gradient. In the case of figure 38.4, therefore, the luminescent screen is placed 15 hole radii distant, which for a 40 mesh/cm grid requires a 1.5 mm distance between lens raster and luminescent screen.

For brighter pictures with correspondingly higher final electron energies, greater distances of the luminescent screen may be required because of breakdown. As shown in figure 38.3 for the case of the triode, less divergent beams to prevent overlapping may then be obtained by using a more negative (especially for storage) or a more positive (especially for intensifiers) lens-grid voltage.

4. Focusing Characteristics for Tangent Dot Condition

Knowing the general shape of the elementary electron beams, and applying the electron-optical laws of similarity, it is possible to predict the optimum potentials with respect to resolution and brightness, for the design of electron lens raster systems for various purposes. According to these known laws, for constant electrode potentials of an electrostatic system the paths of the corresponding electron beams grow and shrink as electrode dimensions are geometrically increased and decreased, as if they were a part of the system. On the other hand, the shape of these paths remains the same for constant electrode dimensions if the electrode potentials with respect to cathode (in our case V_1, V_2 and V_3) are increased or decreased proportionally. If, therefore, dot tangency has been observed at the luminescent screen of a system with particular dimensions and particular electrode potentials, V_1, V_2, and V_3, this condition should also hold, for example, for other anode potentials V_3, where

$$V_1 : V_2 : V_3 = K_1 : K_2 : 1 \tag{4}$$

and therefore

$$V_3/V_1 = K_{31} \tag{5}$$

and

$$V_3/V_2 = K_{32}, \tag{6}$$

where K_1, K_2, K_{31} and K_{32} are constants, and $K_{31} = \frac{1}{K_1}$, and $K_{32} = \frac{1}{K_2}$.

For a given electron lens-raster system, we can observe the tangent dot condition (optimum focus) for $V_2 < 0$ with a microscope at the luminescent screen for two particular voltage ratios (5) and (6).

*It may be noted, however, that overlapping is undesirable only if a resolution comparable to the hole distance is desired. Otherwise, overlapping may be admitted to any degree which does not reduce the resolution of the picture to be amplified, which is given for a post-deflection accelerator oscilloscope, for example, by the focus size of the scanning beam. This holds especially in cases where fine, mesh grids can be manufactured with a higher uniformity than coarser ones.

Knowing thus the constants K_{31} and K_{32} we may find other values of V_1, V_2 and V_3, which also satisfy (5) and (6), and therefore correspond to the same electron paths that define the condition of focusing.

Using V_3/V_1 or E_{23}/E_{21} as a parameter,[10] and V_3 and V_2 as coordinates, V_3 plotted as a function of V_2 for focusing conditions should therefore follow a straight line for each field strength ratio E_{32}/E_{21}, all such lines pointing to the origin: $V_3=0$; $V_2=0$.

This is done in the diagram figure 38.5, indicating the electrode potential ratios (observed microscopically) for obtaining spot tangency at the luminescent screen in a particular system. As to be expected, one finds straight lines with different slopes depending on the E_{23}/E_{21} ratio, which cross each other, however, at $V_3=0$, $V_2=2.5$ v, instead at $V_3=0$, $V_2=0$ as electron optical laws of similarity would require (dotted line). As can be seen, the actual characteristics are shifted, parallel to themselves, to more positive V_2 values, as if the lens grid would have a surface layer with an additional negative potential of $V_2=-2.5$ v on its surface.

Probably [11] this effect is produced by the contact potential, V_{ct}, between lens-raster grid and viewing-beam cathode, which is equal to the difference of their work functions. For this case, (5) and (6) may be written, following a suggestion of E. G. Ramberg, in the form

$$V_3/V_1=K_{31} \tag{5}$$

$$V_3/(V_2-V_{ct})=K_{21}. \tag{7}$$

Besides the shift due to contact potential, we may conclude from the focusing characteristic three significant facts:

[10] To retain free choice of V_1, which can be accomplished by changing the distance between collector and lens grid, the use of E_{23}/E_{21} as a parameter is often preferable.

[11] More accurately: $(V_3-V_{ct})/(V_1-V_{ct})=K_{31}$. Usually, V_3 and V_1 are>100 v, and here V_{ct} becomes negligible.

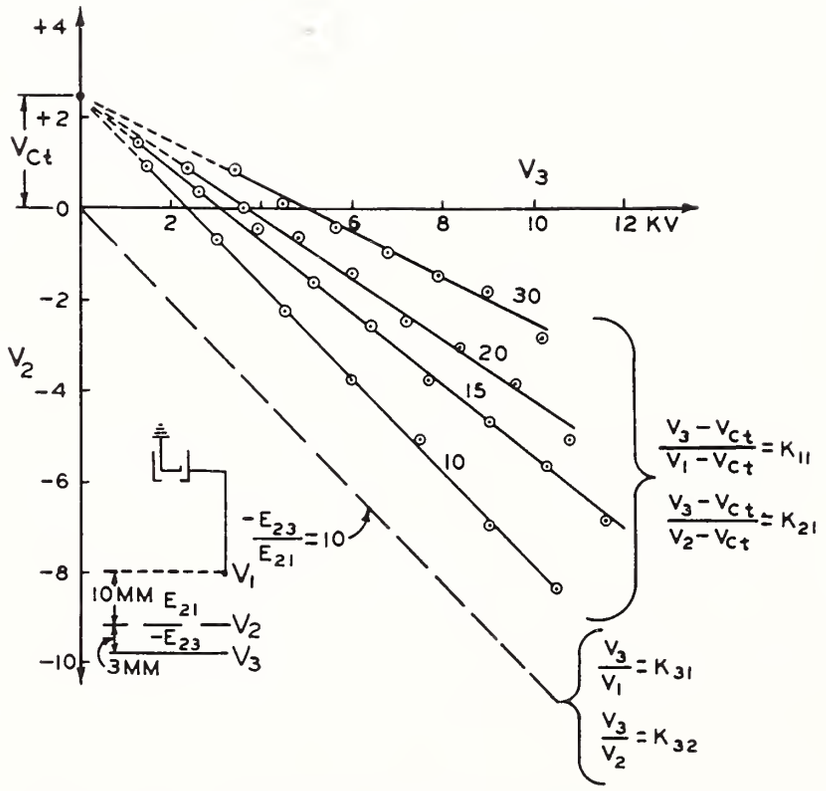


FIGURE 38.5. *Focusing limit characteristics (dots tangent) for electron lens raster system.*

40 holes per linear centimeter. V_{ct}=contact potential.

(a) Separation of elementary beams can be obtained for a 40 dots/cm system at final electron energies which are sufficient to produce bright electron images (1 to 12 kv).

(b) Focusing can be accomplished with only slight changes of the negative raster-grid potential.

(c) The field strength ratio E_{23}/E_{21} and V_3 being chosen, the curves determine the value of V_2 for focusing. A choice of suitable collector-grid voltage V_1, and therefore of the energy of the beam entering the lens-raster system, can still be made by using a suitable value of collector-grid distance a.

So far, focusing conditions have been discussed only for *negative* lens-grid potentials, which are essential for storage and current modulation purposes. If the lens-raster grid is used for intensifying only, as indicated in section 2, tangent spots can be obtained also at *positive* V_2 potentials, which are high enough to satisfy the focusing conditions but low enough to avoid detrimental secondary emission from the lens-raster grid reaching the luminescent screen. In the system shown in figure 38.5, the action of the collector grid makes possible the use of positive lens-raster potentials as high as 40 to 400 v for focusing ($V_1=1{,}000$, $V_3=7{,}000$ v) without significant deterioration of contrast.

5. Resolution and Applications

Due to the short length of the elementary beams, the resolving power of an electron lens-raster system is usually not limited by space charge, but rather by three other conditions:

(a) The hole spacing of the raster grid.

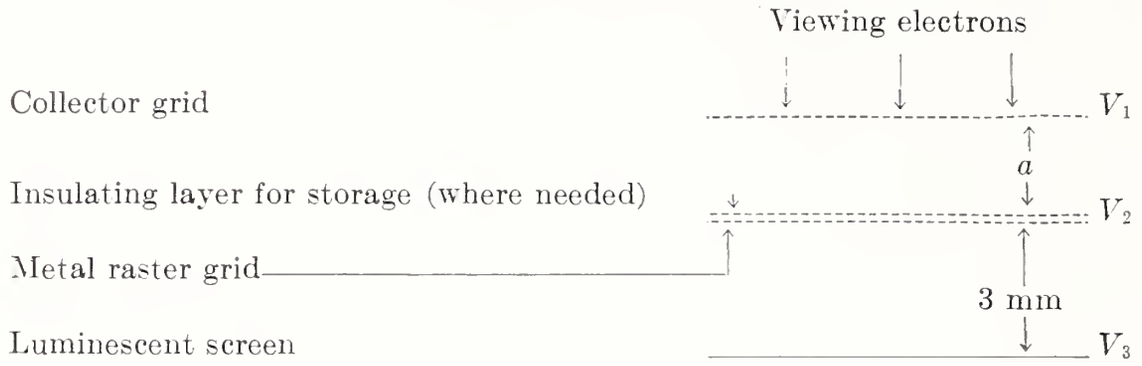
(b) The focusing (tangent spot) condition, which implies that the elementary beams should remain separated and not overlap. This means that, in cases where fine grids are used, there should be a small distance between lens raster and luminescent screen, which may be limited by electrical breakdown.

(c) Suitable values of lens-raster grid potentials V_2, which permit an adequate voltage range along the useful portion of the viewing current characteristic.

Tentative considerations have shown that the resolution of electron lens-raster systems is limited by the hole spacing (a), rather than by (b) or (c). At present, the mechanical limit for larger pictures (>10 cm total diameter) of etched or electroformed lens-raster grids seems to be of the order of magnitude of 200 holes/linear cm, and similar values hold for stretched parallel wires. As may be seen from table 38.1, a fraction of this value is already sufficient for images according to present television standards. In this table, in agreement with experience, the actual resolution is assumed to be approximately 1.6 times less favorable than the hole spacing.

The table shows that, for example, *post-deflection acceleration oscilloscopes* and *kinescopes* may be built with $V_1=1{,}000$ v, and $V_3=10{,}000$ v, which combine high deflection sensitivity and high brightness. For *viewing storage tubes*, in order to obtain a sufficiently small spot size with the writing gun, V_1 has to be chosen usually above 3,000 v. An example of such a tube is given in figure 38.6. Besides viewing of a single picture during several minutes, it offers also the following advantages for television or radar standard scanning, because each picture element may be viewed during a large fraction of the frame

TABLE 38.1. *Calculated dimensions of electron lens raster systems for typica cathode ray tubes*



V_3	V_2	V_1	a	Image size	Lines resolved	Application
Volts	*Volts*	*Volts*	*mm*	*cm*	*(25/cm)*	
5,000	0	1,000	20	15×15	375	Post-deflection accelerator oscilloscopes and kinescopes.
10,000	−2.5	1,000	10	20×20	500	
8,000	−6	3,600	15	20×20	500	Storage oscilloscopes and storage viewing tubes.
(b) Raster with 80 holes/cm						
12,000	−3.5	400	3	32×40	*(50/cm)* 1,600	Large-size image tube, for intensifying, integration, or storage.

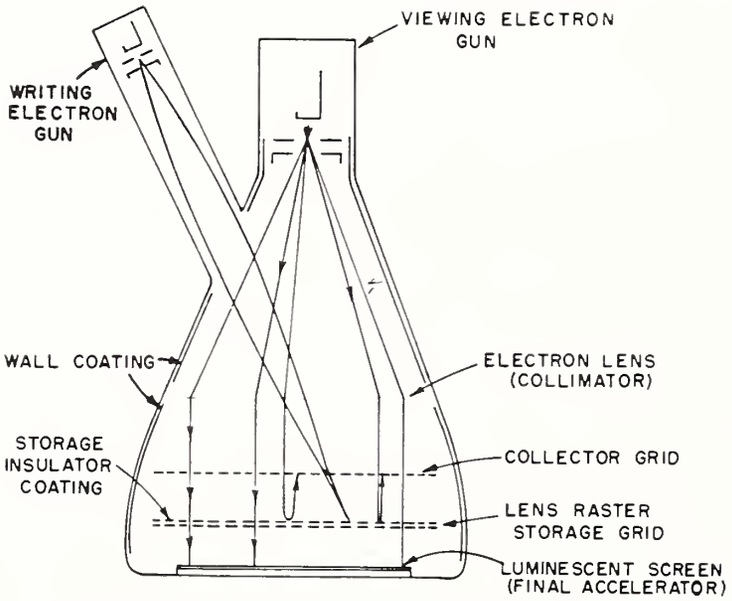


FIGURE 38.6. *Direct view storage tube with electron lens raster system.*

time: Increase in brightness by several orders of magnitude, lower anode voltage for the same brightness, and a smaller signal bandwidth due to reduction in frame repetition frequency without flicker effect. A detailed description of such a tube is given in the following paper.** With only one gun and without the storage layer, the same kind of tube may be used as a post-deflection oscilloscope.

**See also: M. Knoll and B. Kazan, Storage tubes and their basic principles, p. 79, figure 22.

For *image tubes*, due to the desired low field strength near the photocathode, V_1 values below 500 v are desirable. It may be interesting to determine the hole spacing for a photoelectric image tube with lens-raster grid for which the same resolution may be expected, as compared to the known image tube of Holst, deBoer, Teves and Veenemans,[12] which consists simply of a luminescent screen parallel to a photocathode at a distance d. For such a tube, the diameter of the "circle of confusion" and thus the resolution, is given by

$$2r_0 = 4d\sqrt{\frac{U_0}{U_A}} \tag{8}$$

where U_0 is the initial and U_A the final electron energy in volts. For a practical tube ($d \approx 0.3$ cm, $U_0 \approx 1$ v, $U_A \approx 10{,}000$ v) $2r_0$ is approximately 0.15 mm, which corresponds to a resolution already attained with electron lens-raster grids. Therefore, their use for large-area image tubes seems promising. For image *storage* tubes, the capacity of the storage elements has to be sufficiently small to be charged in a sufficiently short time by the writing beam, whereas the writing-beam cathode, irradiated by a local light source, may serve as viewing-beam cathode during the viewing time.

Electron lens-raster systems may also be useful as a final stage for electron microscopes of the emission and transmission type, and for electron-diffraction *cameras*. Using scattering data given by Von Borries[13] for example, for a transmission-type microscope, good contrast should be expected at voltages of 10 kv and lower, with a specimen thickness of the order of 200 A, if sufficient brightness is provided by an intensifying, integrating, or storing lens-raster system. Because the resolution of these systems is limited, a higher magnification than usual may be necessary for the preceding electron-lens system in many cases.

I wish to express my appreciation to E. G. Ramberg and P. Rudnick for helpful discussions and to H. Borkan and H. O. Hook for assistance in measurements and calculations.

[12] C. Holst, J. H. deBoer, M. C. Teves, and C. F. Veenemans: Physica **1,** 297 (1934).
[13] B. von Borries, Z. Naturforsch. **4a,** 51 (1949), fig. 7.

39. Direct View Storage Tube [1]

By M. Knoll [2] and P. Rudnick [2]

Introduction

Although a few proposals for direct view storage tubes are to be found in the literature,[3] [4] [5] [6] [7] the continuous viewing of bright pictures for several seconds or minutes by storing transient signals such as a few television or single radar frames, facsimile pictures or oscillograms has not heretofore been accomplished. For many applications, especially in airborne equipment, such a technique is of considerable importance.

Besides viewing a single picture during several minutes (practically without decay), a viewing tube capable of such storage would also offer advantages for television-type picture reception, as compared with the conventional kinescope, if to the conventional scanning (now used for writing) is added a locally modulated flooding beam. It would have greater brightness for a given anode voltage or lower anode voltage for the same brightness; it would also have less flicker, for a given signal bandwidth, or allow a smaller bandwidth without flicker effect due to reduction in frame repetition frequency.

As compared to a combination of a signal converter storage tube with a kinescope at the receiving end of a picture transmitting system, it would require lower anode voltage, smaller deflection circuits, and no amplifier.

Description of Tube

An electron-lens raster system, as described in the preceding paper, was considered to be very suitable for the purpose indicated above because it combines the possibilities of grid control storage,[8] increased brightness, and half-tones, with the possibilities of simultaneous writing and viewing. The construction of a developmental viewing storage tube with such a system may be seen from figures 39.1 and 39.2.

The bulb (figure 39.1) has the shape of an ordinary kinescope but with two necks, one for the viewing (flooding) gun, which is normal to the luminescent screen, and another for the writing gun, which is slightly inclined to the axis of the tube. The target (figure 39.2) consists of a fine-mesh collector grid (near ground potential), a metallic storage grid covered with an insulating layer (near viewing-beam

[1] Research carried out under contract with the Signal Corps Engineering Laboratories, U. S. Army.
[2] RCA Laboratories Division, Princeton, N. J.
[3] T. F. Adams, The Krawinkel image storing cathode ray tube, Fiat Final Rep. 1021, P. B.–78, 273 (April 1947).
[4] A. V. Haeff, A memory tube, Electronics **20,** 80 (Sept. 1947).
[5] F. Schroeter, Image storage in television reception, Optik **1,** 406 (1946); Image storage problems, Bull. Schweiz. El. Vercin **40,** 564 (1949).
[6] M. Knoll and J. Randmer, Control grid reading type storage tubes, Archiv f. el. Uebertragung **4,** 238 (1950).
[7] R. C. Hergenrother and B. C. Gardner, The reading storage tube, Inst. Radio Engrs. **38,** 740 (1950).
[8] Thus, the electrical charge delivered at the storage elements in writing is not removed during reading or viewing as in earlier storage tubes that use capacity discharge reading or viewing. For a discussion of various methods of writing and reading see B. Kazan and M. Knoll, Fundamental processes in charge controlled storage tubes, RCA Review **12,** 702 (Dec. 1951).

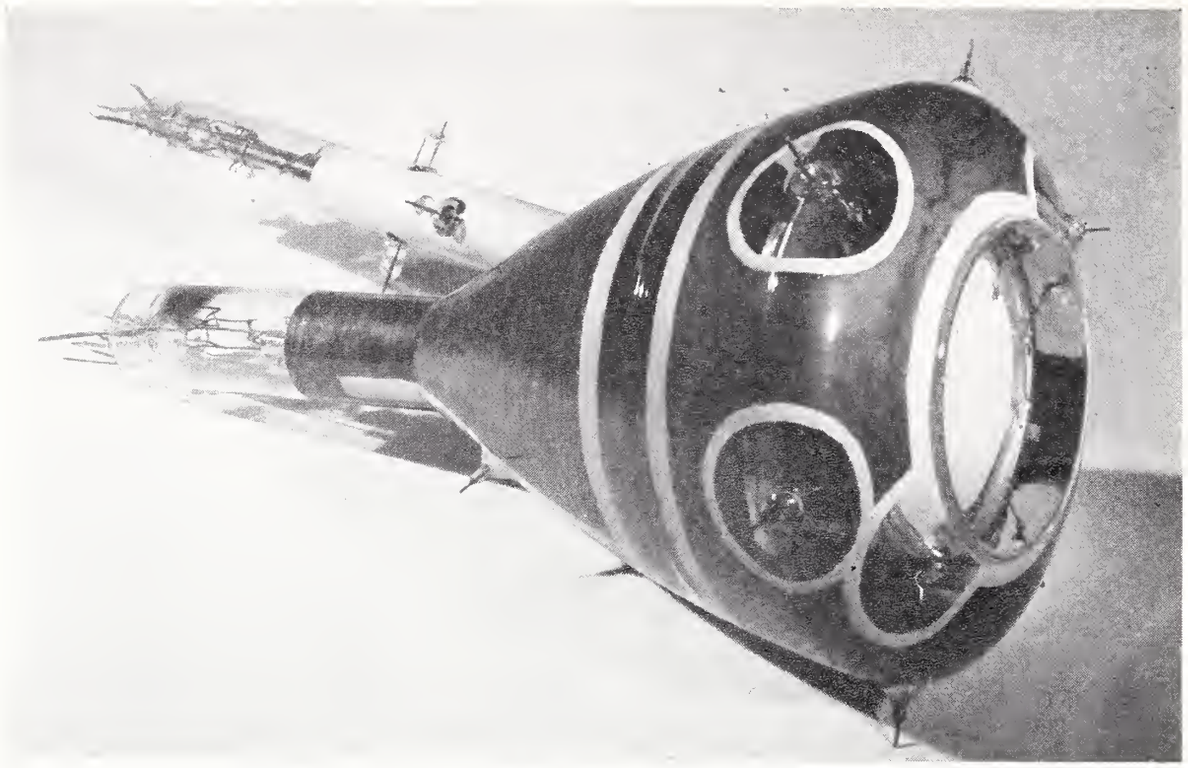


FIGURE 39.1. *Developmental direct-view storage tube with electron lens raster system.*

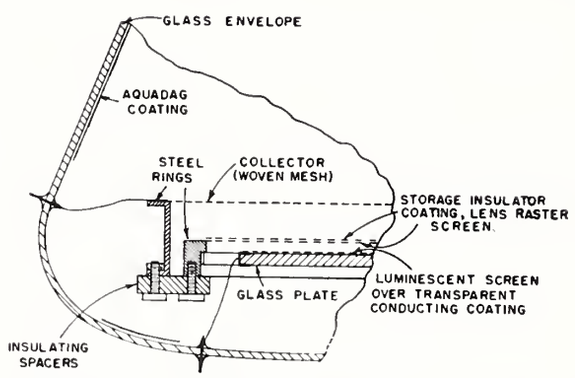


FIGURE 39.2. *Storing image amplifier system using electron lens raster.*

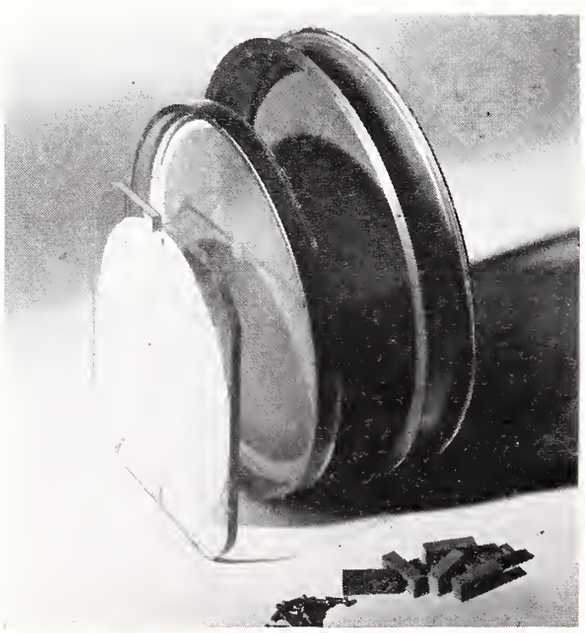


FIGURE 39.3. *Viewing storage system.*

Left to right: Luminescent screen, storage grid, collector grid.

cathode potential) and a conducting luminescent screen (7 to 10 kv positive), all three of which are stretched parallel, at a short distance from each other. Figure 39.3 shows separately the two grids and the luminescent screen (left).

Operation

During operation, the viewing beam is collimated and directed towards the storage grid, maintaining the storage-layer surface at viewing-beam cathode potential (approx. −1 kv). Due to the high field strength on the other side of the storage grid, usually a large amount of viewing current will pass through this grid, and will then be focused at the luminescent screen by each hole of the mesh (electron-lens raster principle). If, then, during writing, a negative charge pattern is established at the storage surface by the writing beam (cathode potential approx. −5 kv),[9] this negative pattern will modulate the viewing current so that a bright negative (black on white) image of it appears on the luminescent screen. By a similar process, a positive (white on black) image may also be achieved. Erasing is accomplished by temporary connection of the storage grid with the collector grid, or by the writing beam using a cathode potential slightly lower than the viewing cathode.

Preliminary Results

In contrast to most previous storage tubes, the stored charges are here not discharged during reading (viewing), so that maximum viewing durations of several minutes have been obtained. With the present tube, writing and erasing of television type pictures can be accomplished within 1 sec; further reduction of these values seems probable. Although not yet satisfactory, due to nonuniformities of the storage surface, half-tone presentation is possible. The resolution of the picture is approximately equal to the storage grid hole spacing. At present, a 100-mesh-per-inch screen with 4-in. diameter is used, the resolution being about 250 lines. Figure 39.4 shows an enlarged highlight portion of an image visible in bright daylight. Due to the low current density and short length of the elementary beams, as compared with the scanning beam in a conventional kinescope, there is no spreading of the image dots by space charge.

The storage layer consists of silica or magnesium fluoride, evaporated onto a carefully etched metal grid support. With a storage layer thickness of approximately 1 μ and a writing-beam current of the order of 10 μamp, this will give writing, viewing, and erasure durations as indicated above and maximum writing speeds of the order of 10^5 spot diameters per second.

Figures 39.5 and 6 show black on white and white on black oscillograms (60 cycles) after 0.5 to 1 min. storage. The retention time of the charge pattern (without viewing) was found to be several days. Figure 39.7 shows stored television pictures, which have been written in a fraction of a second.

We wish to express our appreciation to D. W. Epstein, F. H. Nicoll, L. Pensak, and E. G. Ramberg for many interesting and stimulating discussions, and to E. Apgar, H. Borkan, L. Freedman, H. O. Hook, and M. Topke for assistance in measurements.

[9] For example, using a silica storage layer with a second crossover at 3.5 kv on the secondary emission curve.

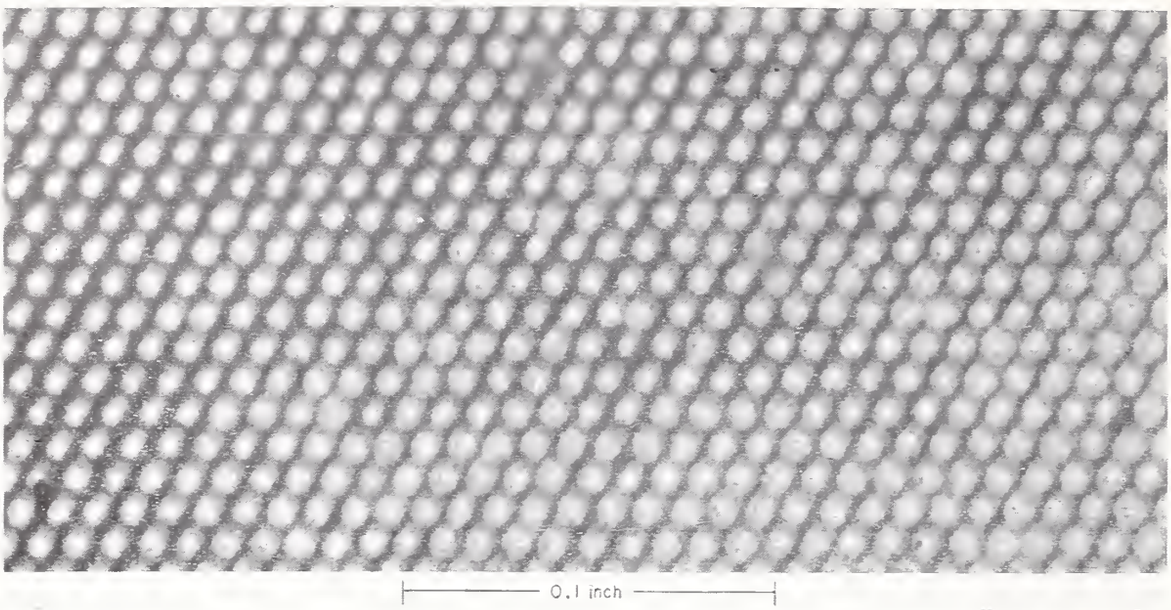


FIGURE 39.4. *Enlarged highlight of amplified image showing full resolution at brightness of 200 foot-Lamberts.*

FIGURE 39.5. *Single transient, black on white, 60-cycle wave, after 30-second viewing duration.*

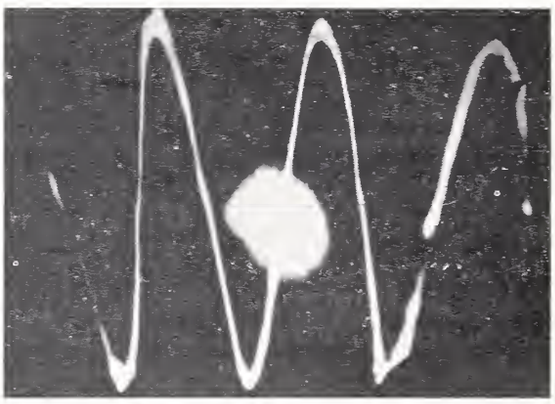
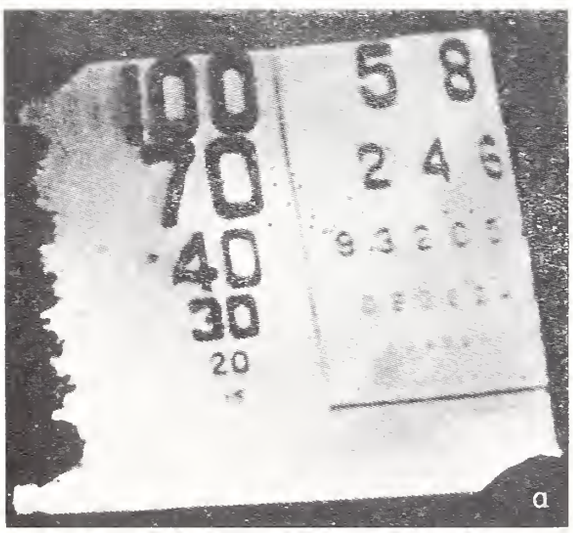


FIGURE 39.6. *Single transient, white on black, 60-cycle wave, after 1-minute viewing duration.*



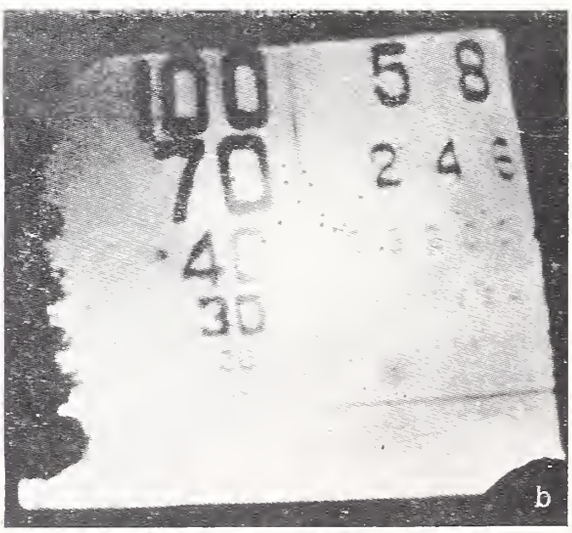


FIGURE 39.7. *Stored television pictures (flying spot scanner).*

White on black writing with 3 to 10 frames. a, Immediately after writing; b, 2 minutes after writing

Discussion

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: How is the storage grid alined, and what is the writing speed of the storage system?

Dr. Knoll: There is no mechanical alinement necessary. The insulating material which forms the storage layer is evaporated directly onto one side of the metallic storage grid support. As to writing speed, we are able to obtain a picture (to be stored for several minutes) by writing with a few standard television frames (1/30 sec. each) at present. Higher writing speed is possible with a thicker storage layer and therefore less capacity between the storage surface and the metal support.

Dr. M. E. Haine, Associated Electrical Industries, Aldermaston, England: In the last device you showed, do you still have to aline the holes of the first grid and the storage grid in the main illuminating beam?

Dr. Knoll: This is not necessary. The first grid merely collects the secondary electrons and does not have to be alined with the second (storage) grid.

Dr. Haine: So the main illuminating electron beam is falling on your insulating surface?

Dr. Knoll: Yes; that is correct.

Dr. R. G. E. Hutter, Sylvania Electric Products, Inc., Bayside, L. I., N. Y.: Do not the viewing electrons disturb the charge pattern deposited on the storage layer?

Dr. Knoll: No; because they are reflected by the charge pattern and cannot land.

Dr. Hutter: How thick could the storage layer be made without danger of preventing the flow of the viewing-beam electrons?

Dr. Knoll: Possibly up to an order of magnitude of tenths of millimeters, or 100 times thicker than at present.

Dr. Hutter: The beams would not charge up the insulator on the inside of each hole?

Dr. Knoll: No; if you determine the surface potential of your insulator by proper use of the secondary emission curve.

Dr. D. Gabor, Imperial College, London, England: You mentioned the post-acceleration tube. I suppose everybody will know Pierce's paper [10] in which he shows the possibilities of post-deflection acceleration.

Dr. Knoll: Are you also referring to L. S. Allard's paper, An Ideal Post Deflection Accelerator Tube? [11]

Dr. Gabor: What they recommend is not valid here because if you break up the system, using instead of one lens system a number of small ones, then the theorem does not apply and you cannot consider the usual sort of refractive index. So it is clear in principle. Practically, what do you gain in post acceleration?

Dr. Knoll: In post-acceleration with conventional cathode ray tubes, there is one large accelerating lens near the luminescent screen. This tends to improve the deflection sensitivity on one hand, because for the same brightness, as in the simple cathode ray tube, the beam can be deflected at low energy before entering this lens. On the other hand, the large acceleration lens tends to decrease deflection sensitivity

[10] J. R. Pierce, After-acceleration and deflection, Proc. Inst. Radio Engrs. 29, 28 (1941).
[11] Electronic Eng. **32**, 461 (Nov. 1950).

to a certain degree by its demagnifying action. With an electron lens raster system, this negative effect does not occur. This is expressed by the fact that the deflection sensitivity is constant with respect to final anode voltage in our case, whereas in the case of the large post-deflection accelerating lens the sensitivity decreases as a function of final anode voltage.

Dr. Gabor: Once you break the beam up into small beams, the theorem simply is not valid, but that is what you get practically.

Dr. Knoll: There is no special theory necessary because the length of the accelerating field is very small (3 mm in the tube discussed here), and therefore the simple rules for the deflection of a low-energy beam apply.

Dr. Gabor: What sort of amplifications do you get?

Dr. Knoll: There is no theoretical limit. It depends how bright you want your tube to be. We obtain the same sensitivity as described by Allard. He made an experiment with an accelerating system using *two* parallel positive planar mesh grids, one of which was in contact with the luminescent screen. In this combination, he found high sensitivity, but unsatisfactory resolution and contrast due to secondary emission. In our case, we avoid this by addition of a third electrode near cathode potential between two positive electrodes (collector grid and conducting luminescent screen).

Dr. Gabor: There was a German paper, also on this subject, 25 years ago. There are at least three old ideas worked up there. I am glad it has come to such a success.

Dr. J. Hillier, RCA Laboratories Division, Princeton, N. J.: I wonder if I might ask a question in which I will imply an answer to Dr. Gabor's practical question. The use of the high voltage, I presume, is to get higher efficiency and higher luminosity from the luminescent screen. At the same time, this system permits use of lower deflecting powers.

Dr. Knoll: Yes.

Dr. Gabor: I should still like to have a numerical answer.

Dr. Knoll: For reasons of electron optics, one usually does not deflect a beam of less than 1000 volts energy. For our tube, 20 or 40 kilovolts on the final anode is possible, so that a voltabe ratio up to 1 to 40 is possible. If one deflects at less than 1000 volts, this ratio may be increased. Of course, a ratio of 40 is not necessary in many cases; often 10 is enough.

Dr. V. E. Cosslett, University of Cambridge, Cambridge, England: It occurs to me that the corresponding optical system is used in X-ray technique, similar to a fly's eye. Isn't it possible to call this a fly's-eye electron lens system? The eye of a fly is built the same way.

Dr. Knoll: Yes. You are still free to use any suitable name.

Dr. Gabor: There is an even older expression for it, lenticules.[12]

[12] This word, stemming from the Latin "lenticulum" (little lens) should, indeed, be a suitable name for electron lens raster systems. However, the adjective form "lenticular" lost this particular meaning in the English language and is now used indiscriminately for lenses of any size.

40. Point Projection Microscopes

By Erwin W. Müller [1]

A simple device to get pictures of a metal surface with a high resolving power and a large magnification is the point projection microscope. Until now it was only used as an electron microscope with field electron emission.[2] But now it is possible too to operate with positive points and to get a sufficient current density of positive ions, thus achieving an improved resolving power.[3]

The field electron microscope tube contains a well-smoothed metal point opposite to a screen (fig. 40.1). With a voltage of some 3,000 to 10,000 v applied to the anode ring an electrical field strength of the order of 50 million v/cm is achieved on the tip. The field electrons are preferentially emitted perpendicular to the hemispherical surface of the tip and fly directly to the screen, giving a picture of the emitting surface due to the different current densities of the various crystal faces.

With a point radius of about 10^{-5} cm the magnification is in the range of 500,000 times. Opposite to every emitting spot on the surface appears a scattering disc on the screen corresponding to an objectside diameter of about 20 A, thus determining the resolving power. A special advantage of every emission microscope is the extreme sensitivity regarding very thin absorbed layers that influence the work function.

The clean surface of the point metal itself with the different crystal directions as well as the absorption of many substances on this surface may be observed in a wide range of temperature. The tip may be cooled by immersing the whole tube in liquid air or hydrogen, or it may be heated to more than 1,400° K during the observation.

The picture of a clean tungsten tip contains only the different crystal faces, but when it is covered with a thin layer of barium as in figure 40.2, there appears a large number of single dots. Every point of the granulation is assumed to be the blurred picture of a single protruding barium atom. Similar pictures may be obtained with a number of other large atoms, e. g., strontium, zirconium, sodium, or uranium.

The question, if these blurred grains with an apparent diameter of some 20 A are really pictures of single atoms or of small clusters of 10 or 100 atoms, may be decided in several ways. One is to measure the activation energy of surface migration on a defined crystal face and compare with the theoretically calculated value.[4]

Another method is the discussion of the observed contrast. Clusters of barium atoms, for instance, forming a small area with a degree of

---

[1] Kaiser-Wilhelm Institut, Berlin-Dahlem, Germany. Now at Pennsylvania State College, State College, Pa.

[2] E. W. Müller, Z. tech. Phys. **17,** 412 (1936); Z. Phys. **106,** 541 (1937); **108,** 668 (1938).

[3] E. W. Müller, Z. Phys. **131,** 136 (1951).

[4] E. W. Müller, Z. Phys. **126,** 642 (1949).

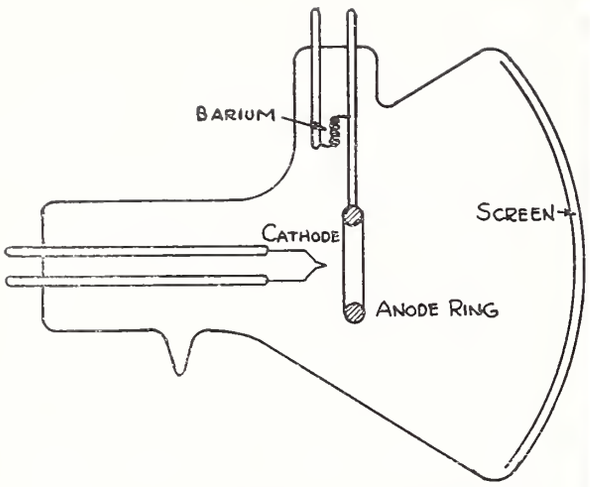


FIGURE 40.1. *Field electron microscope.*

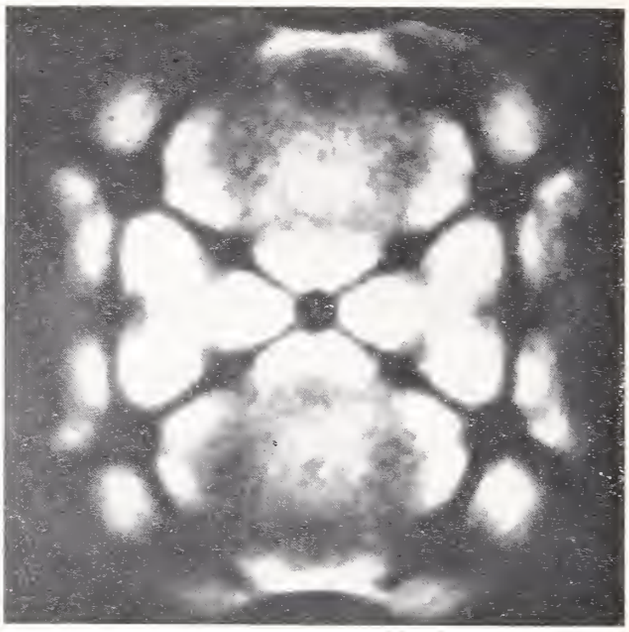


FIGURE 40.2. *Barium adsorbed on tungsten.*

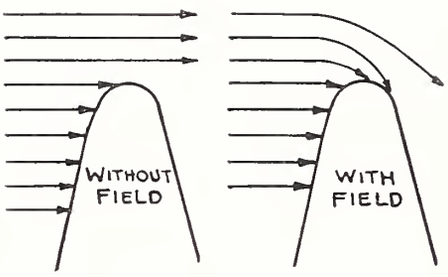


FIGURE 40.3. *Evaporation of atoms onto the tip.*

covering about monoatomic, should have a low work function. These areas should have an about 10,000 times enhanced current density compared to the surrounding clean tungsten surface with a high work function. The observed contrast is only in the range of 10 to 1 or 30 to 1, corresponding to a calculation for the contrast of a single atom.[5]

Another method, used by M. Drechsler and E. W. Müller,[6] is to evaporate a well known number of barium atoms and to count the appearing number of blurred dots. About 60 to 80% of the expected amount is to be found as single dots, showing that only the few bright ones can be clusters of two or more atoms. During this work Drechsler and Müller [6] developed a new method of determining the polarization factor α of single particles at high field strengths. When barium atoms are evaporated on the tip from one side, their pictures naturally appear only on the one side towards the source of evaporation. But when an electrical field is applied during the evaporation, the induced dipoles will experience an attractive force in the inhomogenious field, and they are drawn to the back of the tip also (fig. 40.3). By measure ment of the voltage necessary to get atoms to the back side and by some calculations of orbits and consideration of the Maxwellian velocity distribution, the polarization factor of barium was found to be $\alpha = 6.0 \times 10^{-23}$ cm^3 in good agreement with other methods.

Adsorbed layers with a large degree of covering are not resolved into single particles, as the distances between the molecules are too small. Some examples of epithaxy may be discussed by observing adsorbed layers of Cu_2O on the 310 face of the underlying tungsten

[5] E. W. Müller, Z. Naturforsch, **5a**, 473 (1950).
[6] M. Drechsler und E. W. Müller, Z. Phys. **132**, 195 (1952).

tip (fig. 40.4) or of carbon layers on the 334 faces of the tungsten crystal (fig. 40.5).

Large molecules, of course, are easier to recognize than single atoms, especially when they are adsorbed in protruding positions. It seems to be due to a special field distortion effect, that even the shape of some flat molecules like phthalocyanine may be recognized within a resolving power of 5 A (fig. 40.6).

As the pictures show atomic and molecular dimensions, there can be seen several kinds of movement on the surface, caused by the surface migration or by cathode sputtering by impinging positive ions. The technique of observation has been improved to make possible taking movie pictures from the screen. The 16 mm film shows the adsorption of barium evaporated onto the tip. The expansion of a monoatomic layer over the whole hemisphere by surface migration at higher temperatures and the different speed of surface

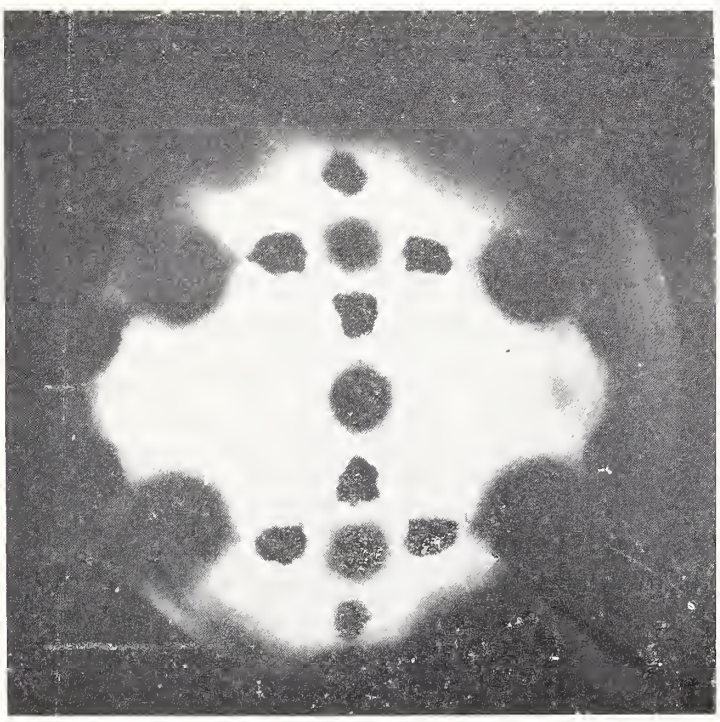


FIGURE 40.4. *Cu_2O adsorbed on the 310 faces of a tungsten tip.*

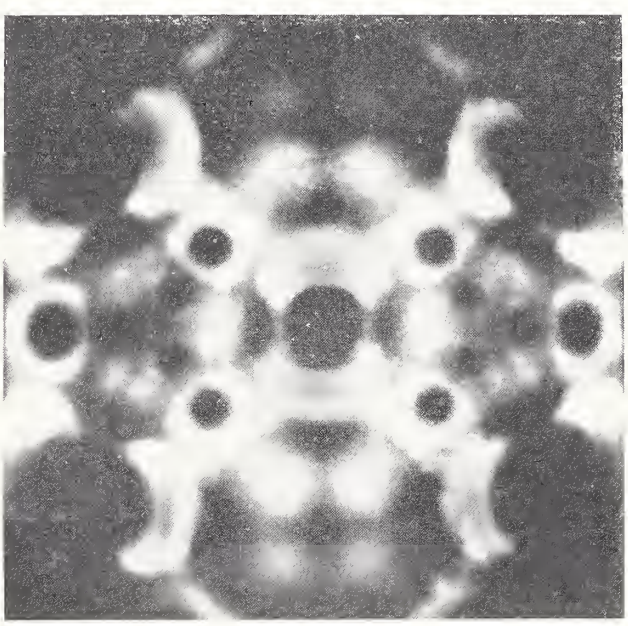


FIGURE 40.5. *Carbon adsorbed on tungsten.*

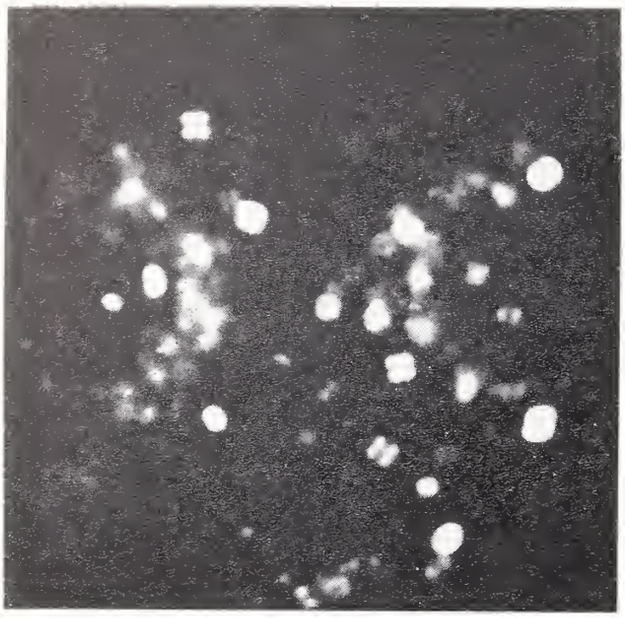


FIGURE 40.6. *Phthalocyanine molecules.*

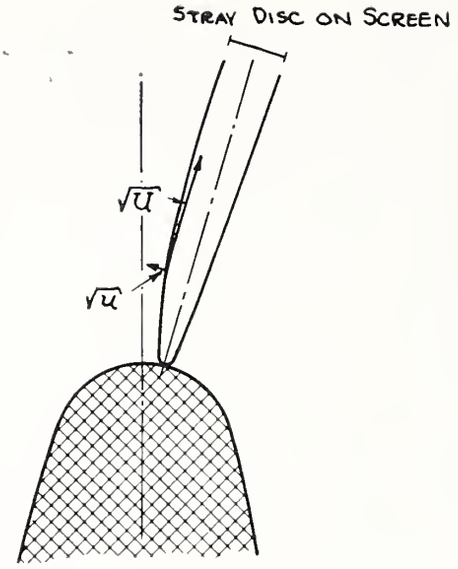


FIGURE 40.7. *Electron beam from a point of atomic dimension.*

migration on various crystal faces can be observed. Then the microscope tube with a clean tungsten tip is filled with oxygen (10^{-6} Torr). Within seconds the surface is covered with a thick adsorption layer, while the current density decreases rapidly by increase of the work function. Several states of the surface are shown after being heated to higher temperatures in steps of 100 deg up to 2,200° K, where a clean surface is obtained again. Below 900° the pictures are always taken with a cold tip, which has been heated previously to the indicated temperature. A thin layer of Al_2O_3 condensed at one side of the tip spreads over the whole hemisphere at temperatures of about 1,000° K. Suddenly monomolecular layers with protruding oxygen atoms and a high work function are built up around the 100 faces of the underlying crystal. The delay of the growth due to the nuclei formation at the beginning of a new lattice row can be observed. Thin layers of SrO seem to be adsorbed preferentially on the steps of the tungsten lattice, at least at temperatures between 1,200 and 1,400° K, thus marking these steps during their movements and dissolution by the reduction of the 110 lattice layers, which occurs when the point radius increases by the heating. At last the adsorption of phthalocyanine molecules is shown with a number of fourfold scattering discs, rotating on the surface and being shot off by impinging ions. When the tip is heated, only a part of the molecules evaporates, the other part becomes cracked and remains adsorbed on special crystal faces in an orientation which is typcial for adsorbed pure carbon.

The pictures on the screen of these insensitive objects are bright enough to allow a direct demonstration by projecting the screen on the wall of the auditorium at a diameter of 2 m. Thus the surface migration of barium layers and the granulation due to single atoms are demonstrated.

The resolving power of this field electron microscope amounts to only 15 to 20 A except in the very special cases of adsorbed flat molecules. The scattering disc on the screen may be described by the tangential velocity u of the emitted electrons (fig. 40.7), which remains unchanged during the acceleration due to the anode voltage

(U). The object-side stray scattering has a diameter of

$$d = 2 \cdot r \sqrt{\frac{u}{U}}.$$

Thus the resolving power cannot be improved essentially because the velocity u is determined by the inner velocity of the metal electrons and the shape of the potential barrier before the surface, and U cannot be increased very much in order to prevent a too large current density, which would destroy the fine point.

Considering the resolving power as a wave diffraction problem the value of about 20 A is again reasonable. An exact calculation is difficult due to the fact that the wavelength near the surface itself cannot be expressed in the case of the tunnel effect. Within a distance of 12 to 15 A from the surface the wavelength drops below 10 A due to the acceleration, thus giving a possible explanation of the observed resolving power.

An improved resolving power is to be expected by the use of positive ions instead of electrons. Unfortunately, there exists no field ion emission comparable to the field electron emission. Ten years ago Müller [7] showed how to tear off adsorbed barium atoms from a tungsten point, using positive field strengths of 80 to 120 million v/cm. Assuming a desorption of positive ions one would achieve a current pulse of 10^{-14} Coulombs by tearing off a monoatomic layer. In order to get a visible picture on the screen a one-million-times enhanced emission is required. This is possible by repeating the adsorption and the desorption act very rapidly.

A simple method to do this is the following: The tube is filled with hydrogen, the pressure being about 5×10^{-3} mm of Hg. With a negative point the well-known field electron picture may be observed practically undisturbed by the presence of the hydrogen, which seems to give no cathode sputtering at all when quite pure. Then the voltage supply is reversed and the positive voltage at the tip is increased about 4 to 6 times. Now the adsorbed hydrogen atoms become desorbed by the field forces and fly to the screen on nearly the same orbits as the electrons before, giving there a picture of the places at which their ionization and desorption occurred. The free places on the surface are rapidly filled up by the impinging neutral gas molecules, these being desorbed again as ions and so on. After the conception of the idea it was somewhat questionable if the intensity would really be sufficient. It is not possible to increase the gas pressure over the value mentioned before in order to prevent a gas discharge breakdown. With this pressure an ion current of only 5×10^{-10} amp is expected by the calculation of the gas-kinetic frequency of molecular collision to the surface. Thus the picture should be very faint, when sufficient magnification is obtained.

But when this experiment was actually performed, it was encouraging to observe an about 20-times enhanced emission. The reason is, of course, that the molecules in the neighborhood of the tip are attracted by the polarization of dipoles in the very strong inhomogenous field. A 20-times enhanced capture of molecules is estimated to be reasonable. Thus the picture becomes bright enough to be observed visually.

---

[7] E. W. Müller, Naturwiss. **29**, 533 (1941).

The ion picture is very sharp and has good contrasts. The diameter of the scattering discs is observed to be about 3 Å. This is in accordance to the tangential velocity corresponding to the temperature of the point, that is only 0.04 v at room temperature, using again the equation $d = 2 \cdot r \, (\sqrt{u/U})$.

After cooling the whole tube to the temperature of liquid air one expects only a scattering disc of 1.6 Å. The De Broglie wavelength of the ion seems not to be important. It amounts to 1.4 Å for a proton of 0.04 volts and 0.2 Å for a proton of 2 v, which are reached in a distance of 3 Å above the surface.

Already at room temperature interesting lattice steps with indentations and kinks, having the size of atomic diameters, become visible. The well smoothed surface of the tip has still many dislocations, which may be observed visually. As the pictures are very faint, an exposure time in the order of 1 min is required for the photographic record. As the screen had no conductive coating there were some movements during the exposure and the photographic picture is not as sharp as the visible picture on the screen itself. Fig. 40.8 shows an electron picture of a tungsten point contaminated with a

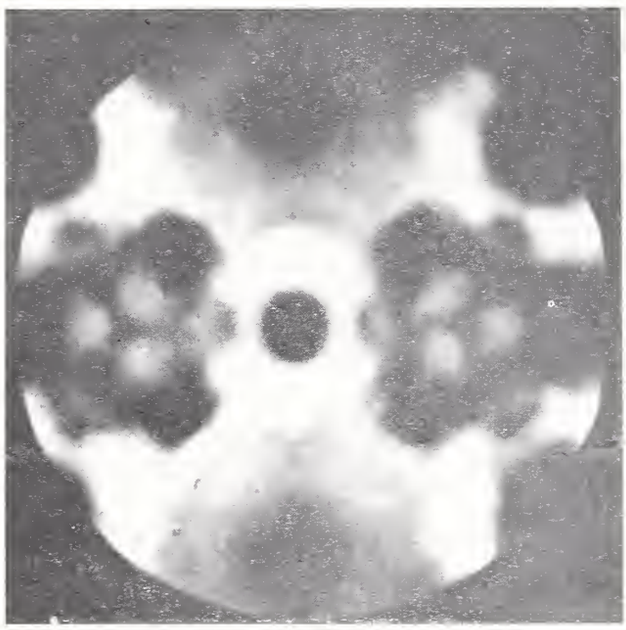


FIGURE 40.8. *Electron image of a tungsten point with adsorbed carbon and hydrogen.*

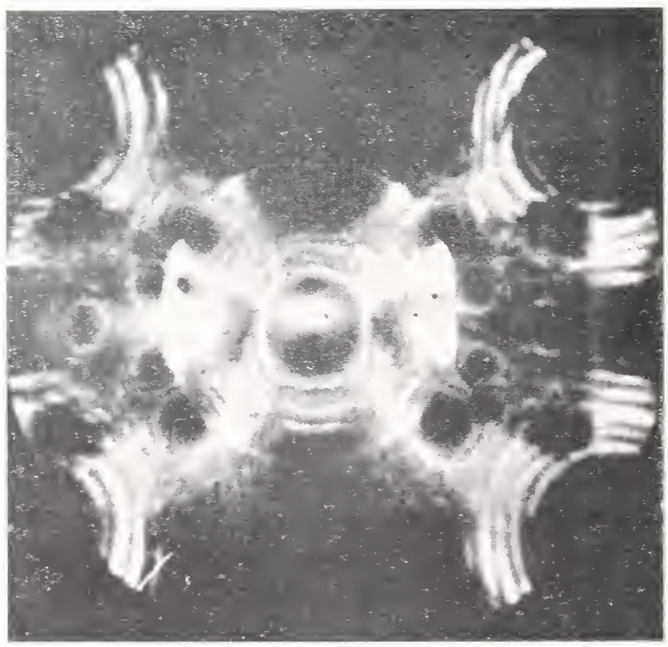


FIGURE 40.9. *Ion image of the same point as on figure 40.8.*

very small amount of carbon. The radius is about 940 A. Fig. 40.9 shows the same tip, recorded one minute later with positive hydrogen ions. The improvement in the resolving power compared with the electron picture is obvious. The bright lattice steps have a distance of 30 A from each other, but the resolving power is higher. Points with a distance of 8 A are separated.

Further improvements in photographic record may be obtained by working with conductive screens or with direct photographs by introducing the plate into the tube. This is not quite simple in a tube that must be baked out. Maybe the technique of image converters can be applied for the amplification of the picture.

The ion picture on the screen becomes visible for the dark adapted eye at field strengths of 200 million v/cm. This must be the field strength for the beginning of field desorption. The picture grows brighter until 300 million v/cm are reached. Then suddenly the whole picture is quite blurred. Obviously the ionization occurs no longer at a place that is determined by the surface structure, but in the free space before the impinging molecules have reached the surface itself. The mechanical forces on the metal surface $F^2/8\pi$ are, of course, enormous at 300 million v/cm, about 400 kg/mm^2. It is supposed there will be a possibility, too, to work with other ions, e. g., helium or lithium, having a larger polarization factor α and thus requiring a smaller field strength. In this way it is to be assumed that this first method of getting microscopic pictures with a resolving power of atomic dimensions may become applicable for the solution of some surface problems.

Discussion

Dr. M. Knoll, Princeton University, Princeton, N. J.: I would like to know whether Dr. Müller has any suggestions on the fringes in the ion picture.

Dr. Müller: These are the steps of the lattice. This area consists of a number of atoms in the surface of the 110 face. We must have a step here and again another here or we cannot get the lattice. These steps now become visible, and they are not smooth but they have kinks and prongs.

Dr. Knoll: Are there any lattice defects?

Dr. Müller: There are certainly dislocations. Some are very small. They are the interruptions in the fringes.

Dr. V. E. Cosslett, University of Cambridge, Cambridge, England: Have you, in fact, examined the points as far as you can in the electron microscope? What warrant have you in any case for assuming that the surface is a smooth curve?

Dr. Müller: In an ordinary electron microscope you cannot see very much of the shape. You see always a quite exact spherical, smoothed point, never any sharp corners.

Dr. Cosslett: Can you be sure of that? One would imagine, would one not, from the thermal etching which occurs on some metals, that after the treatment you give the surface you would in fact have quite pronounced ridges?

Dr. Müller: No, then the intensity of these fringes must be greater. Of course, we have not only single layers but steps of several layers. But these steps may not be higher than about three atom layers.

Dr. Cosslett: No, but it is quite remarkable because in metals which have been investigated for thermal etch, the steps are of the order of 100 A.

Dr. Müller: Such surfaces were not perfectly clean. If you have absorption layers, you get larger steps. These layers are growing out. But if you have a clean surface, expecially of tungsten, you get smoothed surfaces.

Dr. W. T. Scott, Smith College, Northampton, Mass.: Do these spots that were shown on the sides of the point mean there will be essentially a plain layer for the given face approximately parallel to the side of the sphere so that further down you would have faces on the side?

Dr. Müller: Well, you have here the lattice and here the atoms lying in a direction where (011) is the axis. Then again you have the same faces. $01\bar{1}$ and $0\bar{1}1$ even with an angular distance of 90° from the axis.

Dr. J. B. Johnson, Bell Telephone Laboratories: In the electron emission picture, where strontium oxide covered the surface, you had waves travel in and out on the surface. What was the nature of those?

Dr. Müller: I believe it was the following: If you are annealing a point in vacuum, its radius always increases. This is only possible by dissolving these single layers of the (110) face. The corners are marked by the oxide as bright absorption stripes. The layer disappears by surface migration to this side and the strontium oxide remains at these steps. When the steps are going away, we see concentric rings with decreasing diameters.

Dr. J. Hillier, RCA Laboratories, Princeton, N. J.: I am afraid I am horribly confused on the matter of resolving power here. In the first place, you said the radius of curvature of this point is 1,000 A, is that correct? If that is so, then between the (110) direction, which is this direction, and the (100) direction, which looks like this, there is 45-degree angle. That would imply 750 A on the surface, and yet in that range you had only maybe 10 structures in the ion pictures with the rings.

Dr. Müller: There was a distance of about 30 A between the individual rings in those pictures, but the resolving power is better. In the picture the diameter of the scattering disk is about 8 A and along the rings we see such fine details.

Dr. Hillier: I get that impression from the picture, so it is almost an order of magnitude discrepancy.

Dr. Dyke, Linfield College, McMinnville, Oreg.: Was the point, which you showed in ion microscope picture, flashed at high temperature and therefore as smooth and as spherical as possible?

Dr. Müller: There was a not quite clean surface. Therefore we have bigger lattice steps.

Dr. Dyke: One always has the question of whether it is adsorbed impurities or ridges and high fields. Was this a built-up type?

Dr. Müller: No, there were some contaminations. I believe there was some carbon as the directions (334) were somewhat preferred.

Dr. Dyke: What was the highest temperature at which this point glowed prior to use?

Dr. Muller: Oh, between 1,400 and 1,600 degrees. Then it was some time between taking the hydrogen picture and the electron

picture, before I reversed the voltage. The electron picture itself is not a picture of a clean tungsten surface; this was not the intention in this case. It is rather tungsten with carbon and hydrogen.

Dr. Dyke: Were you able to get any quantitative data on the surface migration?

Dr. Müller: It would be possible, of course, to measure the velocity of surface migration, but I have no quantitative data with me. I have measured this, particularly of tungsten atoms in the (110) plane of tungsten, a long time ago. You cannot measure the activation energy, or the movement of a single barium atom because you see it only when it is adsorbed in a protruding position. You see only the resting atoms and not the movement itself, because the thermal movement would be too fast.

Dr. Dyke: I would like to report some of the work we have done on a closely related subject.

We have been interested in large field current densities and in part have used pulse techniques. The field emission process offers relatively direct access to the copius conduction electrons that are normally contained in the metal by its surface potential barrier. This barrier is thinned and reduced in height by the presence of a large surface electric field according to the theory. The current density, J, is of the order of 10^{10} amp/cm^2 as predicted by the theory prior to the point at which the classical barrier is lowered to the Fermi level beyond which the emission becomes primarily a transmission over the barrier instead of transmission through the barrier.

Most of the voltage drop in such a system exists in the immediate vicinity of the hemispherical field emitter. If one has a centimeter electrode spacing, in nine-tenths of this there is a drop of 2,000 v and the rest of the potential difference occurs near the emitter tip so that the electrons receive most of their energy near the emitter and the electron trajectories are more or less radial. When the field is of the order of 3×10^7 v/cm, Prof. Müller has shown that the distribution in energy among the field emitter electrons, is around 1 v. Theoretically when the field is three times this value, as it was in our pulse work, the distribution of energy increases somewhat. It is approximately 3 to 5 v.

In order to examine the theory one has to measure or determine simultaneously the variables current density, electric field, and work function. One may use the Müller projection tube to examine the emission process for clean tungsten to check the theory. If one has a phosphor which is metal-backed, the electrons penetrate the metal anode and excite the phosphor. Light is given off, and this is monitored photographically during a microsecond pulse. One then records the total current, voltage, and learns the distribution of current density over the surface from the photographic negative, which may be calibrated and analyzed with a densitometer. If one plots relative values of current density versus polar angle relative to the (110) crystal direction of the hemispherical emitter, radius R, one observes that in the (110) direction there is a minima and there are maxima in the (116), (310), and (111) directions. If the emitter apex, the region of highest field, is asymmetrically located with respect to crystal directions, and if we now take an azimuth that includes the (110), the (310) directions, so that the azimuth then includes the (100) direction, there are large indicated differences in current density for clean tungsten.

These may be compared with those predicted by M. H. Nichols values for the work functions of clean tungsten with good agreement, using the field emission theory.

From these data the percent of current versus percent of area is obtained. Thirty percent emission originates from 6 percent area. One gets an experimental curve that looks something like this, which may be of use for those who might consider this emitter as an electron source. As the most profusely emitting directions are in a region that surrounds (100) directions rather symetrically (it appears as a doughnut and you have seen this in the pictures), the possibility exists that if this direction is made coincident with the emitter apex, desirable distributions in current density for practical uses may be obtained.

Currents are stable at densities greater than 10^7 amp/cm^2 for microsecond pulses (single microsecond pulses). There is little data available as to higher repetition rates.

There is considerable experimental evidence that average current densities larger than 10^7 amp/cm^2 may be drawn.

Those are the high points of our progress to date, which has been largely building a laboratory and getting an experimental method worked out. We are continuing along these lines. Thanks very much for this opportunity to report this.

Dr. M. E. Haine, Associated Electrical Industries, Aldermaston, England: When talking about emission of 10^7 amp/cm^2, is it necessary to modify in any way the conduction processes going on just inside the metal where the current density is of the same order?

Dr. Dyke: We have examined the heating that one would get on a simple classical picture by using known tungsten constants and the observed geometry and find that an emitter of this size should support a direct current of density of 10^7 amp/cm^2, the implications being that the pulse current could be larger. The voltage drop in the emitter is small under these conditions; the temperature rise is also small. However, both become appreciable at a current density of about 10^8 amp/cm^2 for the tungsten emitter at pulse lengths greater than 1μsec.

41. On the Breakdown Between Metal Electrodes in a Dynamic Vacuum

By F. Bertein [1] and P. Grivet [1]

In this paper, Professor Grivet and I shall give some observations about electrodes in electrostatic microscope lenses and their ability to support high voltages. The question is considered from the point of view of the electrical circuit and the thermal properties of the metal electrodes.

In figure 41.1, c is the cathode at high voltage V; a is the grounded anode. V is connected to c through a resistor R, which is of the order of 50 megohms. It is necessary to take account of C, the interelectrode capacitance or, more accurately, the capacitance to ground by the electrode c and the conductor connecting it to R.

It is well known that the applicable high voltage V is limited because of the appearance of a discharge current i taking place between a and c as soon as V exceeds some value, for instance, 10 kv/mm.

This current is caused by the electronic field emission at the cathode c. Actually, i is the result of this electronic emission and also of an ionic current running in opposite direction; the ions are ejected from the anode a (and its impurities) by the electron beam.

Such a complication due to the anode influence is well pointed out by the following feature: i is dependent upon the surface field on c and also on the gap length ac, that is to say, the applied voltage V. An increase in voltage V will, indeed, give an increase of energy in the electron collisions on the anode and accordingly in the ionic current and i, possibly causing sudden breakdown.

Indeed, the capacity C is storing the energy $W = \frac{1}{2} CV^2$, and this energy can be wasted within the gap ac by sudden discharge; the current i arcs during a short time, owing to the volatilization of impurities or metal.

A localized spark is observed, able to eject the cathode and anode impurities that caused it. This type of spark allows the lens to reach a higher voltage V. It is this "seasoning" process that allows up to 100 kv/mm with electrodes of small area.

Let us see first the possibility for the d-c current i causing a sudden voltage breakdown. This will depend on the volatilization of impurities and therefore the electrode heating due to the incident current in vacuum, the value of which has i as upper limit. If this current is striking an area of radius r_0, it is found that the electrode has a temperature increase on this area in degrees centigrade,

$$\theta \leq 0.76 \frac{Vi}{r_0 K},$$

[1] École Normale Supérieure, Paris, France.

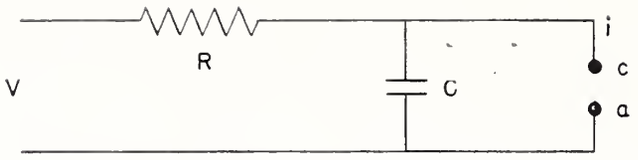


FIGURE 41.1.

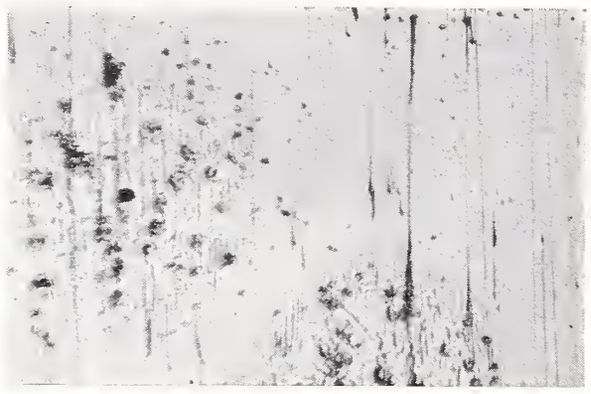


FIGURE 41.2.

K denoting the thermal conductivity of the metal. The units KV, μA, and μ are used for V, i, r_0.

Let us consider, for instance, stainless steel ($K \approx 0.3$); supposing $r_0 \approx 1\mu$, a current i=a few μA (under 50 kv) gives θ reaching a few hundred degrees.

Under the same conditions, the heating would be one-thirtieth as large in a copper electrode ($K=1$). A metal with poor thermal conductivity sparks more easily than a good conductor.

Let us consider now the effect of a sudden breakdown at anode a. This electrode, indeed, will take almost the whole energy W, inasmuch as the discharge is carrying chiefly electrons because they have greater speed than the positive ions.

It thus happens that the anode a is melted around the area of electron impact. The melted areas are left visible after cooling, for during this melting ripples appear on the metal surface (fig. 41.2). These ripples arise from the polishing irregularities due to electrostatic pressure. One of us established a corresponding theory.

On the two plates, the radii r_1 of melted areas is about 300 μ, with $V=50$ kv, $C=50$ $\mu\mu$f.

Let us try to compute theoretically the value of r_1. Suppose the sudden discharge strikes a on a point O, the melted area will be nearly hemispherical, its center being O. An upper limit can be obtained for r_1 (microns) by writing the whole energy W; this gives the heating and melting of this hemisphere at the metal-melting temperature

$$r_1 = 38\left(\frac{C}{\delta Q}\right)^{1/3} V^{2/3},$$

where δ=metal density, Q=total melting heat, r_1 changes very little as C, δ, Q vary. The same order of magnitude is found with the different metals under the same voltage V. Typical values are 150 or 200 μ if $V=50$ kv. The illustration (fig. 41.3) gives a larger value; this shows the anode area struck by breakdown cannot be

FIGURE 41.3.

considered as a point, but its radius is of the order of magnitude at least 100 μ. That indicates possibly the area struck and eventually cleaned by breakdown.

We shall see now how the anode temperature varies after such a spark, supposing there is no melted metal.

In the preceding hypothesis of a discharge striking a on a point O, the variation of temperature in any point M of the anode surface would be:

$$\theta = A \frac{CV^2}{(Kt)^{3/2}} e^{-r^2/5Kt},$$

θ=temperature increase in degrees centigrade, $r=OM$, t=time, and K=thermal conductivity. A is a coefficient, almost the same for the different metals, within our approximation.

From this equation, W being given, the thermal evolution changes according to the metal only in its speed, that is, varying as K^{-1}. The phenomenon is, for instance, 30 times more rapid in copper than in stainless steel. However, the area struck by the beam is not a point O, although the influence of K is the same.

Taking the numerical values as before and considering a very narrow area around the center, its temperature exceeds 500° C during 10^{-4} sec in the case of copper, 30 times more in the case of stainless steel. One may think the breakdown will be more effective in the latter case, because of longer volatilization of impurities.

Thus, as far as the bulky properties of the anode are concerned the "seasoning" process is favored by increasing capacity C and having a metal with poor thermal conductivity, as we have seen. This last property gives two results:

1. It tends to make the breakdown easier by increasing temperature under the action of field emission currents.

2. It makes the heating of the anode of longer duration (as long as the metal surfaces are not damaged by the melting).

If the vacuum is poor, residual gases continuously contaminate the electrodes and sudden breakdowns take place according to statistical rules. It is quite clear, however, under these circumstances that the breakdowns have little effect.

42. Some Studies of the Contamination Induced by Electron Bombardment in Kinematic Vacuum Systems

By S. G. Ellis [1]

Introduction

"Contamination" is the term used in this paper to describe the solid material, usually insulating, that is formed on a surface in a demountable vacuum system when the surface is under electron bombardment. Contamination presents the chief technical obstacle to the further development of precision electron-optical instruments. In the electron microscope the contaminating layers on diaphragms in the system eventually become thick enough that the charges produced upon them by the electron beam cause serious potential disturbances in the column with a consequent reduction in performance of the instrument. This is particularly true of the objective diaphragm wherein asymmetric charge distributions destroy the symmetry of the objective lens and hence, make it very difficult to use small diaphragms successfully. The growth of contamination on the specimen leads in time to a loss of contrast and in some cases can also be seen as an apparent growth in size of the specimen.

In the electron diffraction camera, the growth of contamination leads chiefly to a loss of contrast. This loss of contrast is more particularly noticeable in probe work when one attempts to obtain diffraction patterns from small single microcrystals.

In the electron microanalyzer [1],[2] contamination at present makes it impossible to obtain the energy losses of electrons passing through small regions of thin specimens, for in the time required to record the electron-energy distribution on a photographic plate, the contamination will have grown until carbon (in the contamination) forms a major percentage of the solid through which the electrons are being driven.

The phenomena of contamination was discovered prior to its recognition in the electron microscope. It has, for example, been observed in demountable tubes for the generation of soft X-rays [2]. Contamination was also a source of difficulty in precision determinations of e/m for electrons and in this connection the phenomena has been given some attention by A. E. Shaw [3, 4]. A very good summary of the information available in 1934 was given by R. L. Stewart [5]. This paper is still one of the best on the subject and since that date very little fundamental information has been added to what was already known. Hillier [6] has summarized the observations made with the electron microscope and with the probe type of electron optical instrument, for example, the microanalyzer. These observa-

[1] RCA Laboratories Division, Princeton, N. J.

[2] Figures in brackets indicate the literature references on p. 368.

tions emphasized the complexity of the phenomenon and at first sight suggested certain inconsistencies in its behaviour. For example, it was found that in the electron microscope a collodion film could be bombarded without the production of contamination, providing that the illuminating beam did not strike the supporting copper mesh. On the other hand, in the microanalyzer contamination was invariably formed on bombarding a collodion film whether or not the mesh was being simultaneously illuminated. It was also impossible to suggest any relation between the rate at which the contamination formed and the pressure within the vacuum system, and the current density of bombardment of the substrate.

It was therefore decided to make further studies of the contamination phenomena and in particular to concentrate attention upon these apparent contradictions. Before describing the experiments it will be well to say something of the conditions under which they were done and to follow this by a description of the contamination.

The contamination has been studied by three methods:

(a) As a layer of measurable thickness on filmless mounts of carbon or magnesium oxide in the electron microscope. In some experiments the particles were mounted on collodion films, and the thickness of the added layer determined at different times.

(b) Small amounts of contamination have been produced by removing the condenser diaphragm from an electron microscope and bombarding films of the order ½ cm diameter with currents of the order 100 microamps at 50 kv.

(c) It has been studied in the electron microanalyzer as small mounds of material built up on collodion films, or as projections built up on the copper mesh.

There is a marked difference in the conditions of bombardment in the electron microscope and in the electron microanalyzer. The differences are illustrated in figure 42.1 and may be described in the following terms:

(a) With peak illumination in the electron microscope, the beam strikes an area 20 or 30 μ in diameter. The total current reaching the specimen is of the order 1 microamp and the current density at the specimen of the order 1 amp/cm^2.

(b) When the condenser current is increased to spread the illumination at the specimen in the electron microscope, figure 42.1b, the current density drops to the order 10^{-2} to 10^{-4} amp/cm^2, and the area of illumination can be made large enough to insure that the beam somewhere strikes at least one mesh wire.

(c) In the microanalyzer with the probe formed at the film, figure 42.1c, the total current reaching the film is of the order 10^{-9} to 10^{-10} amp. The diameter of the illuminated area is of the order a few hundred A and the current density may be between 1 and 10 amp per square cm.

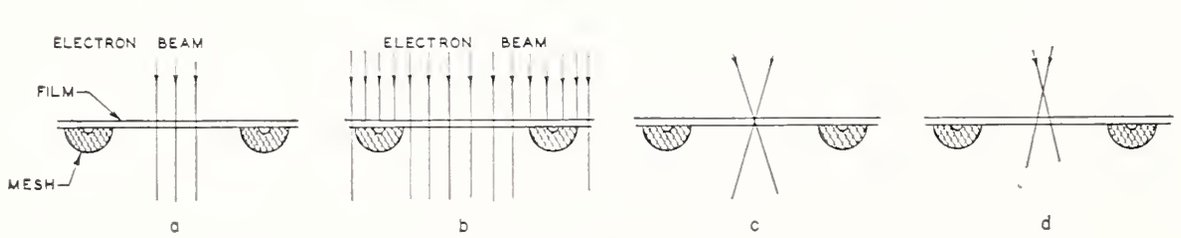


FIGURE 42.1.

(d) If the probe lens is weakened or strengthened in the microanalyzer, figure 42.1d, so that the area of illumination increases, the diameter of the illuminated area may become several microns and the current density of illumination may fall to 10^{-2} amp/cm^2 or less.

The most important difference between the conditions of bombardment in the electron microscope and in the electron microanalyzer is that in the former instrument the temperature of the film may rise to the order 1,000°C when the illumination is peaked, whereas in the latter the temperature will seldom rise above 100°C and will probably be much less under normal conditions of operation. When the illumination is spread in the electron microscope (condition of fig. 42.1b) the temperature of the film may fall to less than 100 deg centigrade in excess of the temperature of the walls of the instrument. I am indebted to Dr. E. G. Ramberg, who calculated these temperatures.

Another factor which may be of importance in understanding bombardment induced contamination is the potential of the bombarded area. So long as the illuminating beam does not strike the mesh, or any other grounded conductor in the vicinity of the film, the film is maintained at a positive potential with respect to the rest of the instrument. For films a few hundred angstroms thick, the primary electrons are absorbed by the film to a negligible extent, and many of the secondary electrons escape from it leaving the film with a net positive charge. If now the specimen is moved so that the mesh enters the bombarding beam, the potential of the film drops to a lower value more nearly equal to that of the walls of the instrument. The sign but not the magnitude of the potentials reached has been deduced from the nature of the deflections produced in the illuminating beam.

It is concluded that when the film and mesh are simultaneously bombarded, the film is kept near ground potential because of bombardment by low-velocity secondaries from the mesh. We shall return to this point when considering the mechanism, whereby contamination is produced on the film.

Other phenomena relevant to the present discussion are the chemical and physical changes that take place in a collodion film under electron bombardment. These were first studied by König [7], who claimed that the film became chemically inert and thermally very stable because it had lost hydrogen, oxygen, and nitrogen to become graphite. The matter has also been investigated by Hillier, et al. [8], who conclude that the film does not necessarily lose the lighter elements, but becomes chemically stable and thermally resistent under electron bombardment. This question, like that of contamination, is complicated by the fact that under electron bombardment there is usually some rise of temperature of the film. It is known from previous microanalyzer studies, that upon bombardment material can be lost from the collodion film, and it had been considered that this was loosely bound water. In this connection the work of Cowley [9] on the electron diffraction by fatty-acid layers on metal surfaces is of interest. Cowley concluded that as a result of electron bombardment, changes were produced in fatty-acid layers and that, irrespective of the substrate and the fatty acid that had been bombarded, there resulted another material that he found to be thermally resistant. He suggests that the material is akin to polythene being a long chain polymer with strong lateral binding between small "crystallite" regions.

Perhaps the most direct evidence for the nature of the chemical and

physical changes that take place under electron bombardment can be obtained by placing films of organic materials on glass or metal substrates and bombarding them in the electron microscope with the condenser diaphragm removed. Collodion films on glass became white after bombardment. Those on copper were darker after extended bombardment, but in no case was the resistence of the film less than 10^9 ohms between two points a millimeter apart. The conductivity of much thinner films of evaporated graphite was readily observable, and it was concluded that though chemical changes had been produced by bombardment they were not extensive graphitization.

While the process is undoubtedly complicated, the conclusions from this work may be summarized in the following terms: When an organic material is bombarded in the electron microscope, we may imagine the valence bonds to be broken repeatedly by the incident electrons. Moreover, one will expect considerable ionization to take place. With an organic material, it will be unlikely that all the four valence bonds of one carbon atom will remain simultaneously broken for a sufficient time that the carbon atom can migrate very far or escape from the molecule in which it is combined. At the same time that this general ionization and dissociation is going on, we may expect combination and neutralization to take place as well. The electron beam also heats the system. Under this electronic excitation (the electronic energies of the incident electrons being far higher than the activation energies required for chemical combination) we may expect the system of atoms to rapidly approach the condition of thermodynamic equilibrium corresponding to the temperature of the system. By chemical standards this will be a highly stable and hence inert system, and when the bombardment is stopped the system remains in this condition. It is only in the case where the temperatures produced by bombardment are very high that the thermodynamically most stable state of this system is that of graphite, and where, as is commonly the case, the temperatures reached are only a few hundred degrees centigrade, the resultant product is an inert hydrocarbon.

Experimental Work on Bombardment Induced Contamination

A balanced evaluation of the factors outlined in the introduction had not been attained when the experimental work presently to be described was started. At that time the hypothesis due to Stewart, which had been supported by Watson [10] was considered to be substantially correct and was used as a working hypothesis in designing the experiments. As the work progressed, it became apparent that the temperature of the substrate was playing a hitherto unrecognized part in phenomena. It was also suspected that the material which was polymerized to form the contamination was arriving in the bombarded region, not wholly from the vapour state but in part, if not in major part, by surface migration. Attempts were then made to test these two hypotheses. As a chronologic presentation of the experimental evidence would be confusing, it will not be attempted. Hindsight permits a more coherent account to be given.

If the beam is concentrated in the electron microscope so that it does not hit the mesh, then after an initial period, no contamination will form. If the beam is spread so that the mesh is struck, contamina-

tion will form chiefly on the side of the film that faces the mesh, and this is true whether the film is above of below the mesh. This suggests three possibilities:

(a) that the material from which the contamination forms is driven from the mesh as a result of heating caused by the electron bombardment,

(b) that low voltage secondaries from the mesh cause the contamination, and

(c) that the material from which the contamination if formed migrates on the side of the mesh away from the beam.

In the microanalyzer, contamination forms on both sides of the film whether the mesh is being struck or not. It was considered possible that the higher temperature to which the film is heated when the beam is concentrated in the electron microscope is the cause for the lower rate of contamination as compared with the microanalyzer.

When a diaphragm with a small hole (of the order 1 micron cross-section diameter) was placed above the specimen film in the electron microscope so as to reduce the cross section of the beam striking it, it was found that contamination was caused at an appreciable rate even though the beam did not hit the supporting mesh. Moreover, by shadowing this comtamination, it was shown that some of it had formed on the side away from the supporting mesh.

Some attempts were made to study the rate of contamination (rate of increase of thickness of the contaminating layer with time) as a function of specimen temperature. Magnesium-oxide smoke was mounted on a tungsten wire that could be electrically heated when placed at the specimen level in the electron microscope. There are two sources of difficulty in such an experiment. The thermal expansion of the specimen causes image drifts that make accurate quantitative measurements well-nigh impossible. Moreover, there is often some doubt as to whether the specimen is not at a lower temperature than the supporting wire due to radiation losses from the specimen. Using carbon for the specimen, it was found that the contamination that formed at lower temperatures could be removed by heating to 900° C. The contamination could be reformed on allowing the specimen to cool, but there was some indication that the rate of contamination was reduced when the specimen was heated.

As the electron-microscope results had suggested that the mesh was a source of contamination, studies were made using vacuum-fired platinum mesh, copper mesh, stainless steel mesh and fired stainless steel mesh. The platinum mesh that had been vacuum fired to a temperature in excess of 1,000° C. still showed contamination. No treatment has been found that will reduce the rate of contamination when copper mesh is used. Neither washing the mesh with amyl acetate nor coating it with Formvar has reduced the contamination rate, nor, indeed, has the contamination rate shown any tendency to decrease over long periods of use in the microscope and over long periods of bombardment. The stainless steel mesh has, it appears, a somewhat lower rate of contamination, though no significant improvement was found on firing it before use.

Magnesium oxide mounted on copper mesh in a spare specimen cartridge was heated to 500° C. for 1½ hours in an air furnace. With the specimen cartridge carrying tool, the cartridge was transferred while warm to the electron microscope. It was found that at first

the contamination rate was much reduced, but after some time it returned to the normal value. While this experiment was complicated by condensation of zinc from the brass of the specimen holder on to the mesh, it nevertheless suggested that the contamination rate was reduced when the specimen and specimen cartridge were warm and that it returned to its normal rate as these cooled, or some short time afterwards. It has been noticed by Turner [11], that the contamination rate in the electron microscope increases to a maximum with increase in beam intensity and thereafter declines. Turner also reports that prior heating of the screens was ineffective in reducing the rate of contamination. These experiments suggest that the rate of contamination can be considerably reduced if the temperature of the substrate is raised, and moreover they suggest that temperatures of a few hundred degrees centigrade, may be effective in reducing the rate of contamination. Attempts to heat specimens in the electron microscope and in the electron microanalyzer have so far proved unsuccessful for technical reasons connected with thermal expansion of the stage mechanisms and possibly because the parts under electron bombardment were considerably cooler than the heating elements, owing to radiation.

These results suggest that heating the substrate will reduce contamination, presumably because the reacting molecules either are not present in the area under bombardment, or are in proximity for too short a time to be polymerized.

The experiments described above give little or no information on how the reacting material arrives at the bombarded region. In this connection it is instructive to consider the conditions in the vacuum system. If the pressure is p in millimeters of mercury, the absolute temperature is T, and the molecular weight of the residual gas is M, then the mass G of residual gas incident per second on unit area of the substrate is given by

$$G = 5.83 \times 10^{-2} p \sqrt{\frac{M}{T}}$$

If we take $M = 390$ (the molecular wt of Octoil), $T = 298°$ K (or 25° C), $p = 10^{-5}$ mm of mercury, $G = 67 \times 10^{-8}$ gm/cm^2/sec. If then all this material were deposited as contamination, and the density of the contamination was 1 gm/cm^3 the contamination rate would be 67 A/sec.

Under probe conditions a contamination rate of 60 A/sec has been observed in the microanalyzer, and there is no indication that a higher rate would not be observed if the current density (approx. 4 amp/cm^2) could be increased. These figures may be stated in different units: the contamination is being built up at approximately 30 atoms/A^2/sec and the current density is approximately 3,000 electrons per A^2/sec.

While it seems probable that the vapour pressure of the diffusion pump fluids used in these experiments was much less than 10^{-5} mm of mercury, the total pressure in the system remained in doubt since ionization gauges were employed.

In the microanalyzer a liquid air trap between the diffusion pump and the instrument had a flat, polished, metal face. This face was cleaned and the instrument was run for 3 hr. The condensed film in the trap was estimated, from the interference fringes, to have a thickness of approximately 2 microns. This is a rate of deposition of 2 A/sec.

This suggests then that the rate of contamination in the probe instrument is too high to be accounted for by the polymerization of diffusion pump oil arriving from the vapour state alone.

It was therefore supposed that the material that polymerizes to form the contamination arrives in the bombarded region by surface migration over the substrate. A simple calculation shows that to explain the contamination under probe conditions in the micro-analyzer, the migration rate of molecules would have to be of the order 1 μ/sec. The migration hypothesis would also explain the result reported by Hillier [6], that surrounding the specimen with a liquid air trap did not reduce the rate of contamination, for in this experiment the mesh and a rod supporting the mesh were not at liquid air temperature. Had the mesh alone been cooled to liquid air temperature, it is not certain that one should expect a reduction in the contamination rate since it would take several minutes under probe conditions to polymerize a monomolecular film from the 1,500 or so square microns of the film in one mesh opening. Even had the film been cooled it is not certain that an appreciable reduction in the rate of contamination might be secured. The reason for this is that it requires very much less energy of activation to cause migration than to cause evaporation of a molecule from a surface and therefore even at liquid air temperatures the thermal energies of the substrate may be sufficient to excite migration.

While in a 3000–A diameter probe the rate of arrival of material across the perimeter by migration may be considerably in excess of the amount that arrives in the bombarded area directly from the vapour state, it must not be assumed that the material in the vapour state does not play an important part in contamination, for it may well be that the supply of material causing the contamination is built up outside the bombarded region largely from the vapour state. In this connection the observation of Hillier [6], that the rate of contamination within a small, fairly well isolated section of the vacuum system decreased with time of bombardment, suggest a vapour source of supply.

Where the bombarded area is larger as in the electron microscope the direct vapour contribution to the bombarded area may exceed the migration contribution since the former increases as the area of the region while the latter increases only as the perimeter.

Another observation tending to favour the migration hypothesis is that the rate of contamination under probe conditions appears to be very little dependent upon the pressure or state of vacuum of the system, at least within the range of pressures used in these studies. In most of the work the pressure ranged between 10^{-4} and 10^{-6} mm of mercury equivalent air pressure, and studies were made both with and without liquid nitrogen traps on the system. In some further studies the specimen was enclosed in a small chamber with apertures just big enough to permit the entrance and exit of the electron beam. Air, water vapour, and helium were admitted to this chamber in turn. In some cases the local pressure became high enough to give observable electron scattering, but in no case was there a marked change in the rate of contamination.

Finally it should be noticed that Dunoyer [12] has obtained independent evidence of migration phenomena on the walls and surfaces in a vacuum system.

We have considered the nature of the contamination process and the means whereby the material which is polymerized to form the contamination arrives at the region under electron bombardment. We shall now consider the origin of this material. The evidence which has been obtained is largely of negative character, in that while it suggests a number of sources which by themselves are not sufficient to account for the contamination, it does not provide quantitative information which will permit one to pick out a major source of the contaminating material.

The work reported above has been done on kinematic demountable vacuum systems employing rubber gaskets and pumped by oil diffusion pumps backed by mechanical pumps. In the microanalyzer it is possible to use a liquid nitrogen trap between the diffusion pump and the system. It has also been run with one additional liquid nitrogen trap in the system and with a liquid nitrogen trap between the fore pump and the diffusion pump. In no case was a significant reduction in the rate of contamination observed. It has been suggested that the oil diffusion pumps are a major source of contamination. There are two reasons for believing that this is not the case. First of all in the work reported by Stewart [5], contamination was produced in a system pumped by a mercury diffusion pump, and secondly there is a more recent study by Blears [13] in which the atmosphere of a vacuum system has been studied with the mass spectrometer. Blears finds a mass distribution which is continuous to 300 and states that the atmosphere was essentially the same whether using mercury or oil diffusion pumps since it was due in a large part to the oil vapour from the backing pump.

It has been possible to make some interesting observations of surface contamination in a demountable vacuum system pumped by a mechanical pump and an oil diffusion pump by a rather simple though qualitative technique. This depends on the fact that the coefficient of friction between glass surfaces that have been cleaned by flaming them is very high [14].

Since the chattering of a flamed glass rod upon a flamed glass microscope slide is a sensitive indication of the degree of cleanliness of the microscope slide, it was decided to see if this phenomena could be used to test for the contamination of such slides in a vacuum system. A flamed slide was placed in a demountable vacuum system (one used for metal evaporation and shadow casting) which was then pumped to a pressure of 10 microns of mercury with the backing pump alone. On removal from the system it was found that the cleanliness of the surface had been markedly reduced and the flamed rod no longer chattered upon its surface. A similar glass slide placed in the vacuum system, but not pumped down, and also a control glass slide left on the laboratory bench for the same length of time still showed a marked chattering when tested with a clean glass rod. This experiment shows that originally clean surfaces will be contaminated during the prepumping with the backing pump in such a vacuum system. This contamination is not to be confused with that which is produced by electron bombardment.

In a second experiment, it was found that if the previously flamed glass slide was subjected to ion bombardment in the vacuum system while being pumped to a pressure of 10 microns (by means of a Tesla coil) this contamination was very much reduced, though there was still a

suggestion of some contamination. I am indebted to Dr. James Hillier for repeating these experiments and confirming the necessarily qualitative estimation of the changes observed.

If now a flamed slide is introduced into the vacuum system, and an ion discharge is directed upon it during the prepumping operation, and if then the system is pumped to approximately 10^{-4} mm of mercury and left under these conditions for 20 minutes, the diffusion pump continuing to operate and no liquid air trap being used on the system, it is found that the slide has contaminated to a greater extent than can be explained by its exposure to the fore vacuum alone. This experiment indicates therefore that the contamination continued even when the operating vacuum had been attained.

In further experiments it was observed that this contamination takes place even when the slide has been flamed all over and is suspended from a tungsten wire maintained at red heat. This observation suggests that the layer deposited upon the glass arrives from the vapour state since otherwise it would have had to migrate over the red hot tungsten wire to arrive at the glass. It should also be noticed that in this experiment the glass, because it was supported from a red hot tungsten wire, became slightly warmer than the walls of the vacuum system in which it was suspended. Nevertheless it contaminated. Two other isolated observations are relevant to the present discussion. As the vacuum system mentioned above, was pumped down it was noticed that the colour of the discharge obtained with the Tesla coil changed from the purple colour indicative of air, to a gray colour, indicative of organic vapours, as the ultimate vacuum of the fore pump was achieved in the pump-down period. Secondly it has been found that if one smells the interior of a vacuum system immediately after air has been let into it, specific odors may be detected. For example, in an experimental electron microscope employing rather soft rubber gaskets, a tarry smell similar to that of the gaskets could be detected in the system immediately after it had been let down to atmospheric pressure. The contamination rate on this microscope appeared to be higher than that on other instruments in which the smell of the interior of the system was less noticeable.

Conclusions

We may now summarize the foregoing experimental results and discussion in the following terms: The contamination induced by electron bombardment is an organic material which results from the polymerization of simpler organic molecules whose vapor pressure is of the same order as, or is less than the pressure achieved in the system. The source of these materials is in part the fore pump and diffusion pump and in part gaskets, grease, fingerprints and dirt introduced into the vacuum system, either at its manufacture, assembly or during its use.

These materials arrive at the bombarded region partly by migration along the inner walls of the vacuum system and partly by the vapor phase. We can imagine this migrating film on the walls of the vacuum system to be replenished from the vapor phase, and at hot points in the system we can imagine the organic molecules to vapourize from this migrating film and condense elsewhere.

Under electron bombardment the molecules in this film are poly-

merized by processes closely akin to those which occur during the bombardment of a collodion film. The end product is that which is most stable at the temperature at which it is formed. If, because the electron bombardment is sufficiently intense and extends over a sufficient area, the substrate becomes very hot, the population of this migrating film is reduced by evaporation, or by the reduction in density which would be expected on heating a two dimensional gas with a result that the rate of polymerization is greatly reduced. It is not clear from the results at present available, whether the activation energy for this polymerization reaction comes mainly from the primary electrons, or whether the reaction is activated, mainly by secondary electrons. It requires between ten and a hundred primary electrons to fix one atom of the contamination, when the contamination is being formed at room temperature or thereabouts. It seems probable that the contamination rate can only be considerably reduced by heating to a temperature in the range 300° C to 500° C or perhaps even higher.

Attempts to improve the vacuum, or to trap vapors have not resulted in a significant decrease in the contamination rate. It seems unlikely that the migration could be very much reduced by cooling the substrate film. A partial reduction in contamination has been obtained in the microanalyzer by running an auxiliary spray gun at 1000 volts and 20 to 40 milliamps for five minutes [15]. The contamination rate was reduced to a third or a seventh of its previous value and this improvement was maintained for five or ten minutes after the auxiliary spray gun was turned off.

The most certain way of reducing contamination therefore seems to be to heat the parts which are to be kept clean. If the specimen contamination is to be reduced, the specimen must be mounted on a conducting substrate such as a graphite or aluminium film. If the specimen is of organic material, it should be fixed by electron bombardment before heating and then raised in temperature by several hundred degrees centigrade before being subjected to the intense bombardment of the illuminating beam. The technical difficulties involved in this procedure have been mentioned above and do not make it a very attractive solution of the contamination problem. Similarly the reduction of the contamination of the objective diaphragm in the electron microscope by heating this diaphragm would add most undesirable complexities to the design and operation of the objective lens.

References

[1] J. Hillier, J. Appl. Phys. **15,** 663 (1944).
[2] E. L. Webster, W. W. Hansen, and F. B. Duvenek, Rev. Sci. Instr. **3,** 729 (1932).
[3] A. E. Shaw, Phys. Rev. **44,** 1009 (1933).
[4] A. E. Shaw, Phys. Rev. **54,** 193 (1938).
[5] R. L. Stewart, Phys. Rev. **45,** 488 (1934).
[6] J. Hillier, J. Appl. Phys. **19,** 226 (1948).
[7] H. König, Naturwissenschaften **34,** 261 (1948).
[8] J. Hillier, S. Mudd, A. G. Smith, and E. H. Beutner, J. Bacteriol. **60,** 641 (1950).
[9] J. M. Cowley, Trans. Faraday Soc. **302,** 60 (1948).
[10] J. H. L. Watson, J. Appl. Phys. **18,** 153 (1947).
[11] T. J. Turner, M. A. thesis, Clemson Agricultural College, Clemson, S. C., (1949).

[12] L. Dunoyer, Le Vide **27,** 793 (1950).
[13] J. Blears, Vacuum Physics, Suppl. No. 1 to J. Sci. Instr., p. 36 (1951).
[14] J. Strong, Procedures in Experimental Physics, p. 166 (Prentice-Hall, Inc., New York, N. Y., 1945).
[15] N. Davidson and J. Hillier, J. Appl. Phys. **18,** 499 (1947).

Discussion

Mr. M. E. Haine, Associated Electrical Industries Research Laboratory, Aldermaston, Berkshire, England: Contamination in the electron microscope becomes of considerable importance when very high resolutions are sought, as in the diffraction microscope. One of my collaborators, A. E. Ennos, has been investigating the mechanism of contamination and has been able to show with reasonable certainty that the deposit comes from condensation of hydrocarbon vapor and not surface migration. He has accomplished this by irradiating a plate suspended on thin wires. The supposition is that a stable monomolecular layer of the hydrocarbon is formed, a thicker layer being unstable. The electron beam decomposes this layer to form carbon on which a further layer can then condense. The monolayer is only stable below about 100° C; above this temperature the fraction of a complete layer that is stable falls rapidly. Correspondingly, the rate of contamination also decreases. Thus contamination can be greatly reduced by holding the bombarded surface at 200° C.

Attempts to reduce contamination show that a very clean system is necessary. The contamination comes not only from the oil of the condensation pump, but the presence of any considerable area of metal can also lead to contamination even when careful cleaning methods are applied. Ultimately it is quite possible that the carbon may come from the metal carbonyls.

Dr. P. A. Clavier, Sylvania Electric Research Laboratory, Bayside, L. I., N. Y.: Brass and copper are both highly absorbent for carbon monoxide. Nickel-plated brass might also be good.

Mr. Haine: Copper, brass and aluminum are being tried.

Dr. J. Hillier, RCA Laboratories Division, Princeton, N. J.: In Mr. Ennos' experiment, what was the area of the plate compared with the area irradiated by the electron beam?

Mr. Haine: The fraction of the total area irradiated was between 25 and 50 percent. The amount of contamination produced was far in excess of what would have been produced by a monomolecular layer of oil over the entire surface of the plate.

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: What current densities were used in the experiments which have been described?

Dr. Ellis: In the electron microscope the current density was of the order of 1 amp/cm^2, it may have been up to 5, the beam diameter was about 10 microns. In the microanalyser case, the beam diameter was 1,000 to 2,000 A and the current density of the order 1 to 10^{-2} amp/cm^2.

Dr. Marton: Was the contamination effect related to the current density?

Dr. Ellis: Not alone. Temperature is also of importance. If the object is thin, and hence its absorption low so that its temperature is also low, the contamination rate will be high. As the con-

tamination increases, the energy absorption and the temperature will increase. Eventually an equilibrium condition may be reached.

Dr. V. E. Cosslett, Cavendish Laboratory, Cambridge University, Cambridge, England: It is significant that in the X-ray microscope that I have described, where the current density is 10 amp/cm^2 for a 1 micron probe, contamination is only observed in a ring around the spot after a long exposure. This suggests that the central part of the bombarded region is too hot to contaminate.

Dr. Hillier: There is one other observation agreeing with this. In examining a very heavy carbon black deposit under very high intensities no contamination could be seen over long periods. At first I thought the carbon black was absorbing the oil which was migrating in, but I finally decided that the specimen thickness and the intensity were so great that the resulting high temperature prevented the migration.

Mr. Haine: For a long time the temperature effect confused our experiments. The effect is quite critical. One may find that no contamination appears in an experiment until the beam intensity is greatly reduced.

43. Controlling of Electron Beam by Means of a Rotating Electric Field and Its Application

By Kenichi Owaki [1]

Introduction

Since A. M. Skellett [1, 2][2] announced a method of controlling an electron beam in a tube in which the radiated electron beam is controlled by a rotating field formed when voltage is supplied to several grids surrounding the cathode or by the use of rotating electric and magnetic fields, improved types of such tubes have been developed in the Electrical Communication Laboratory [3] and Kobe Kogyo Corporation [4, 5], Japan. The purpose of these tubes is to find application to the commutation of channels in multiplex communication by time division.

The author, with coworkers,[3] proposed to utilize the tubes for frequency multipliers, and encouraging results were obtained in the comparatively low frequency region. In order to extend the principle to the region of ultrahigh frequencies (particularly the microwave), however, it is necessary greatly to improve the construction and the operating conditions of the tubes.

In this report, the new types are described mainly from the standpoint of electron optics. The tubes for the low frequency possess the same problems as those found in tubes for microwave frequencies. About the former, detailed explanation is given.

Tubes for Low Frequencies Developed by Kobe Kogyo Corporation

Construction and Operating Principles

Figure 43.1 shows the arrangement of the electrodes for a six-channel commutator tube SP–350 developed by the Kobe Kogyo Corporation. Six focusing grids G_{f1}, G_{f2}, . . ., G_{f6}, which swell out toward the center with the shape of an arc, are arranged surrounding the cathode. Around them is placed a screening electrode with shielding fins, which have six rectangular windows on the same radii as that of the focusing grids. Furthermore, outside the screening electrodes, six anodes A_1, A_2, . . ., A_6, each of which is surrounded by a respective control grid G_{c1}, G_{c2}, . . ., G_{c6}, are set up in each section.

A six-phase sinusoidal voltage is applied to the focusing grids G_f, thus forming a rotating electric field between the cathode and grids,

[1] Kobe Kogyo Corporation, Akashi City, Japan.
[2] Figures in brackets indicate the literature references on p. 387.
[3] K. Takashima, T. Takagi, T. Misugi, and T. Nakamura at Kobe Kogyo Corporation, and S. Okajima at Nagoya Institute of Technology.

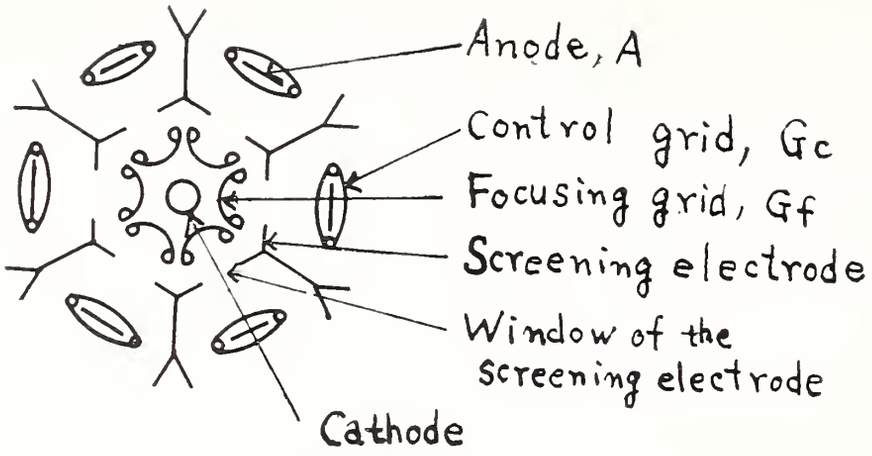


FIGURE 43.1. *Arrangement of the electrodes for 6-channel commutator tube.*

which causes the electron beam to rotate. Therefore, current flows into anodes A_1, A_2, . . ., A_6 in turn, and a current whose waveform is as shown in figure 43.9 can be obtained in a circuit in which the six anodes are connected in parallel. It is evident from figure 43.9 that a frequency component of six times (6f) that of the focusing grids f constitutes most of the anode current; hence, frequency multiplication can be performed very efficiently. Moreover, in this multiplication, frequency components other than $6n$f (n is a positive integer) are very small. In other words, spurious frequency is found to be very small when compared with that obtained in other methods of frequency multiplication.

Effective Angle of Cathode

The previous paragraph described how the commutator tube can be used in frequency multiplication. In such a case, it is desirable that, during the time that the current flows into one anode, it does not flow into the others because the d-c component of the current must be as small as possible in the output anode current; that is, it is important to build up such an electric field that the electron current may flow in only one direction at a time. Since it is difficult to obtain such a potential distribution or electric field by calculation, the results obtained by experiments using an electrolytic bath are described here.

It is supposed that the voltage of sinusoidal distribution, as shown in figure 43.2, is applied to each of the focusing grids, and that the d-c voltages are applied respectively to the screening electrode, control grids, and anodes as is shown in figure 43.3. For such a case, the equipotential lines and the trajectories of the electrons are shown in the same figure; it is clear that the width d of the electron beam passing through the window of the screening electrode is about five-eighths of the width b of the window. When the phase of the voltage is applied to the focusing grids, the electron beam rotates successively from the first section to the adjoining one, and when the phase leads by 30°, the voltage distribution on each electrode, the equipotential lines, and the electron trajectories become as is shown in figure 43.4. In this case, the electron current almost flows into the screening electrode.

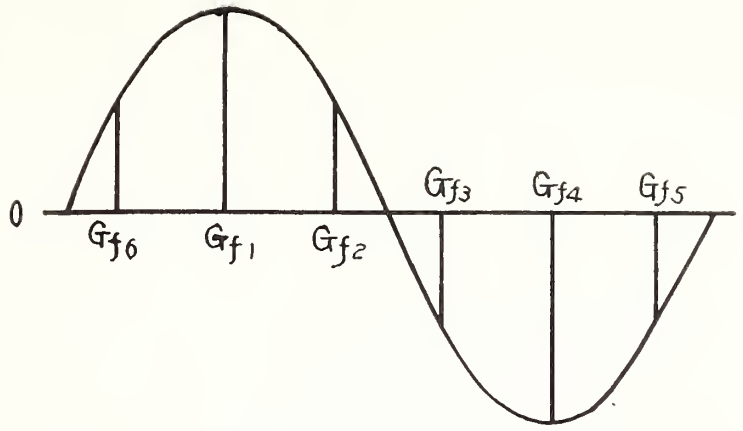
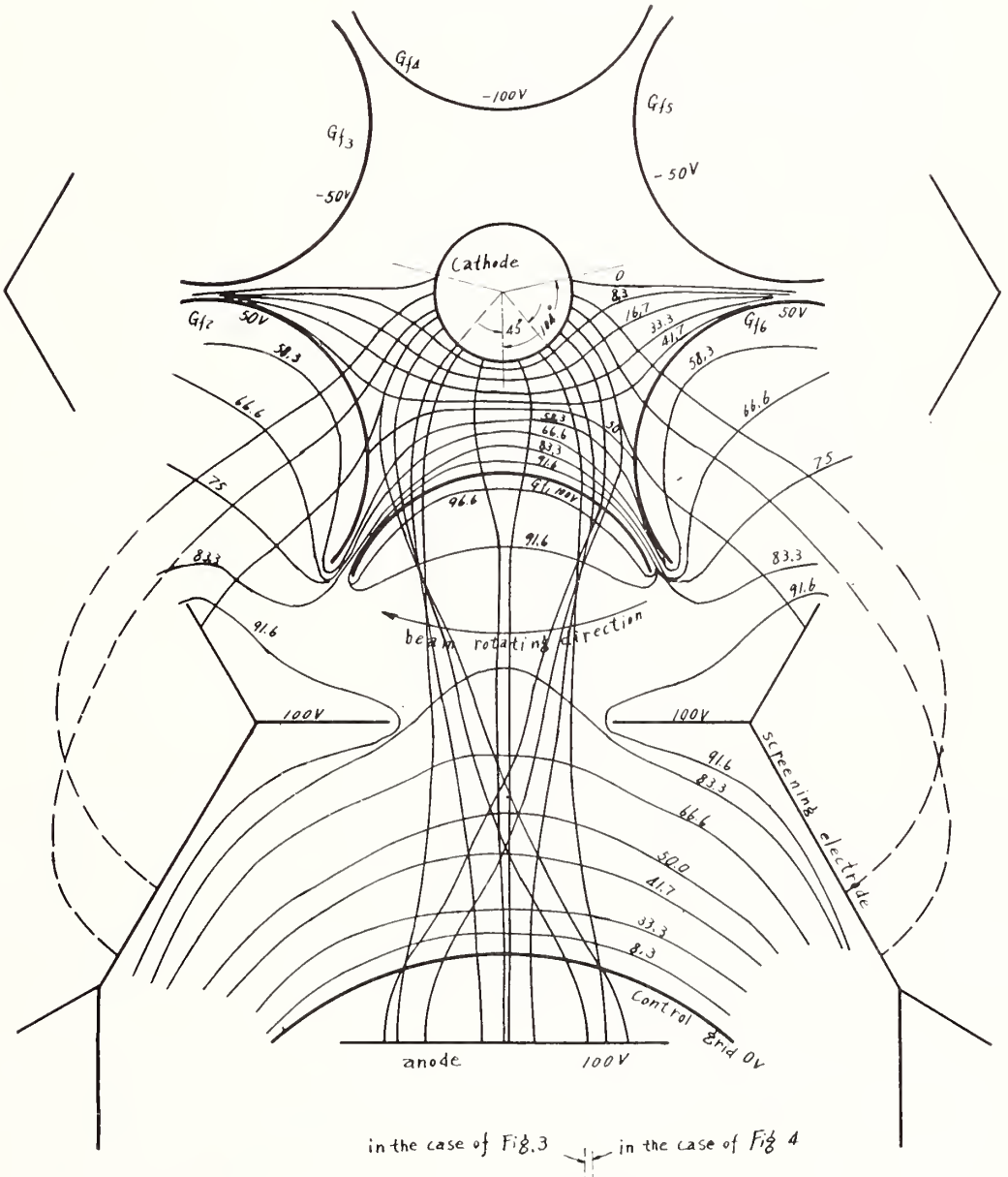


FIGURE 43.2. *A sinusoidal voltage waveform applied to the focusing grids.*



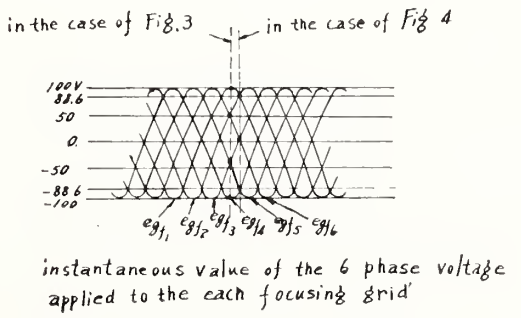


FIGURE 43.3. *Equipotential lines and electron trajectories in the tube.*

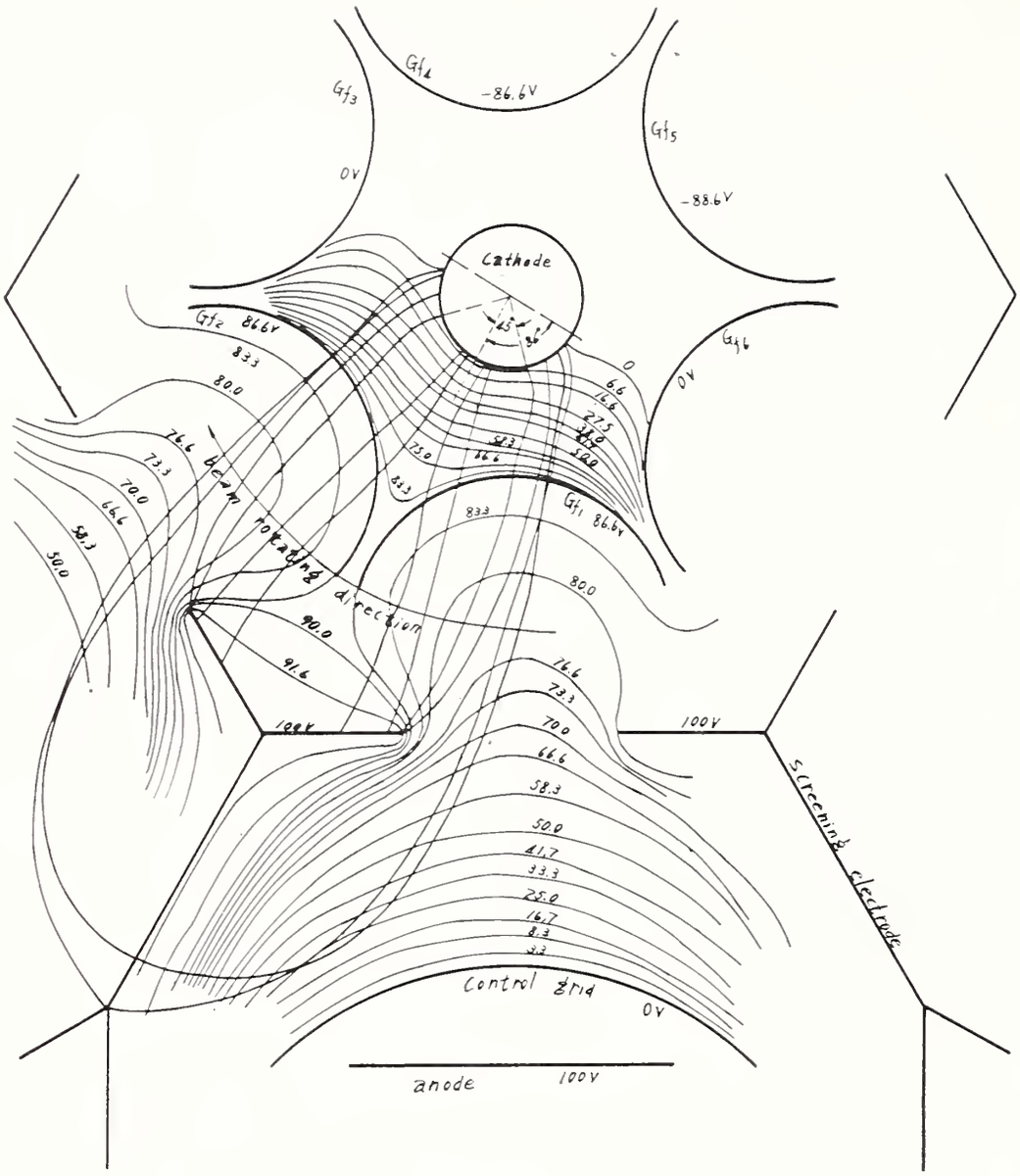


FIGURE 43.4. *Equipotential lines and electron trajectories in the tube.*

In the above consideration, the d-c voltages of the cathode and the control grids are zero. One example in which the bias voltage is applied to the cathode is taken into consideration in the next stage. The equipotential lines and the electron trajectories, when the cathode is positively biased by a d-c voltage of half of the peak value of the sinusoidal voltage applied to the focusing grids, is shown in figure 43.5 and figure 43.6 (The voltages written in these figures are the potential differences between the cathode and other electrodes). In this case d is equal to one-tenth of b.

The above-mentioned is described concerning the cases when the electron beam is at the center of one section and at the boundary of the adjoining two sections. The electron emitting angle of the cathode can be obtained from both cases. The electrons are emitted from the cathode surface at which the potential gradient is positive if the inititial velocities of the electrons are neglected; hence, when the cathode is zero-biased, θ is about $\pm 104°$ from figure 43.3, and θ is about $\pm 86°$ from figure 43.4, meaning, in both cases that the electron emitting angle θ is about $\pm 95°$. In other words, 53 percent of the

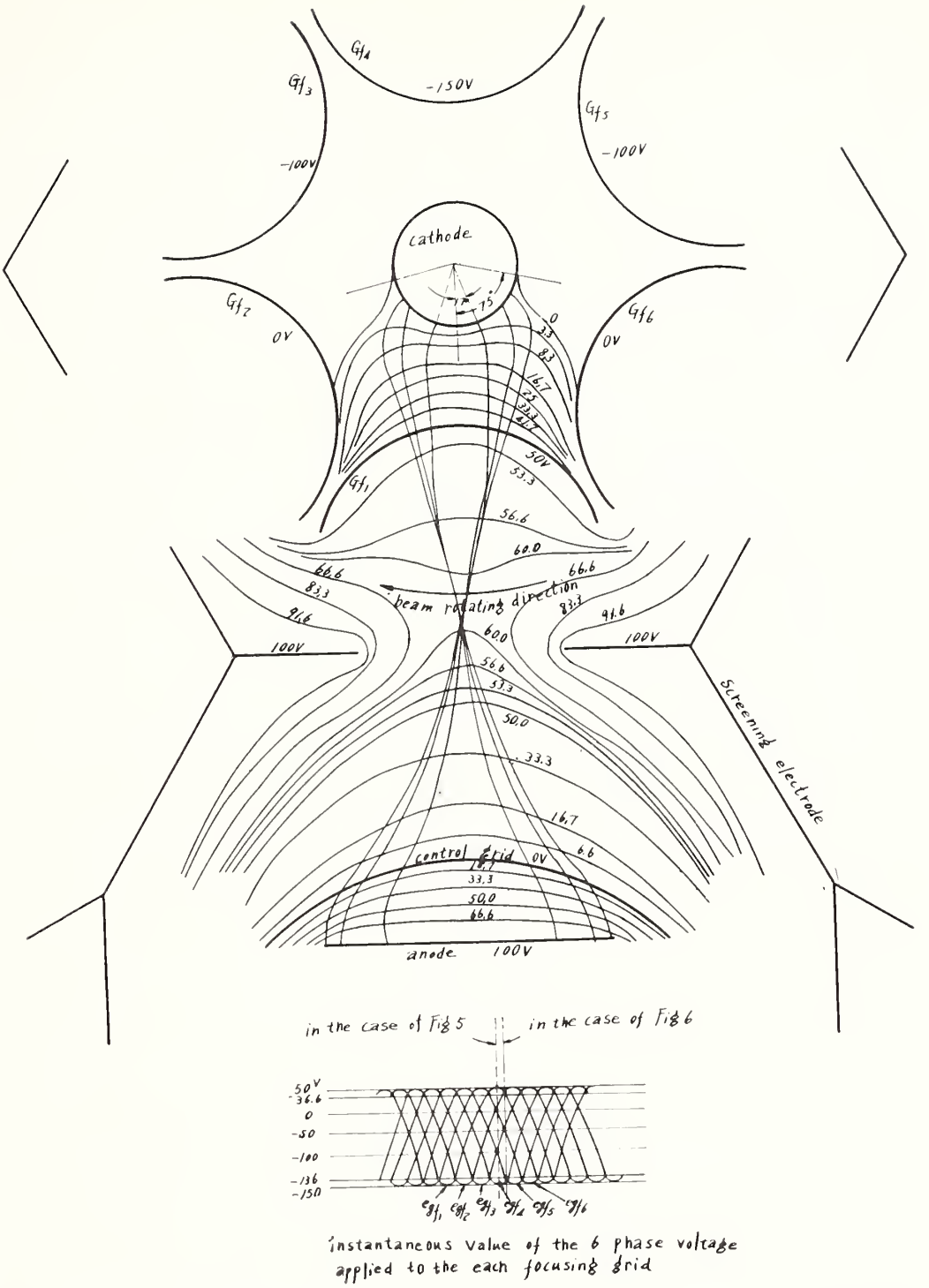


FIGURE 43.5. *Equipotential lines and electron trajectories in the tube.*

whole surface area of the cathode emits useful electrons. When the cathode is biased, θ is about $\pm 75°$ in figure 43.5 and figure 43.6; that is, the electron emitting surface of the cathode is about 42 percent of the whole area.

It can be seen from these results that the higher the bias voltage of the cathode, the better the focusing of the electrons, but the less the anode current, because of the reduction in the useful area of the cathode. On the other hand, the ratio of b and d influences the waveform of the output current as described later, therefore the operating conditions must be decided by taking into account the desirable value of the anode current and the waveform of the output current, etc.

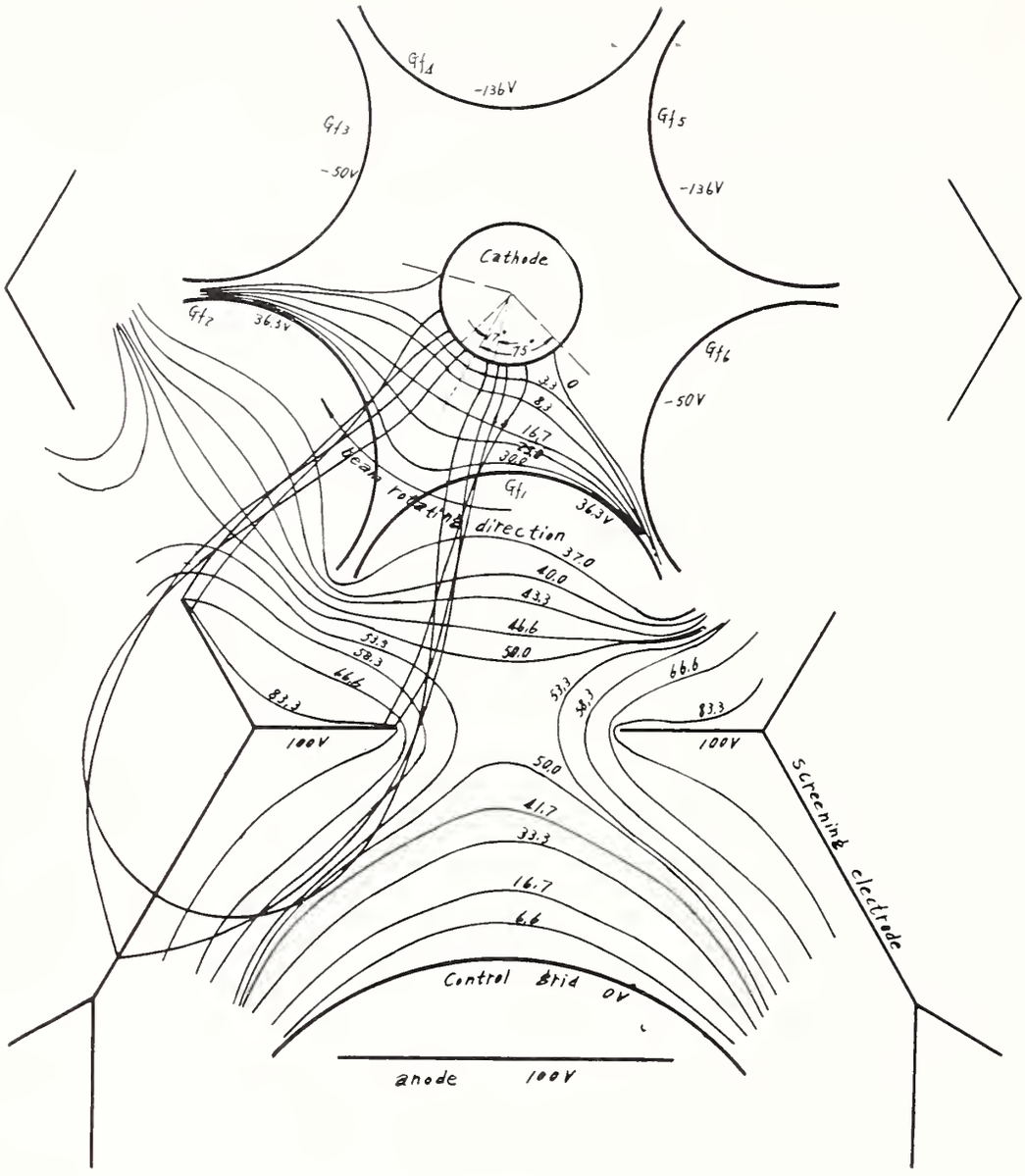


FIGURE 43.6. *Equipotential lines and electron trajectories in the tube.*

Relation Between Width of Electron Beam and Shape of Window of Screening Electrode

Before the investigation of the relation between the width of the electron beam and the shape of the window of the screening electrode, the mechanism by which the output current waveform is formed is considered.

The following factors are taken into account as the elements determining the output waveform: (a) Effect caused by rotating the electron beam, (b) effect of electron emission control caused by change in voltage on the focusing grids, (c) effect caused by change of the width of the electron beam, and (d) effect of the distribution of the electron density in the beam current.

In the ideal case, when the voltage of sinusoidal distribution is applied on each element of the infinitely split cylinder concentrically arranged around the cathode and the phase of the voltage rotates with time, the electron beam of constant width and electron density is considered to rotate around the cathode with constant angular velocity. In such a case, there exists only the effect of (a), and effects of (b), (c), and (d) do not exist.

In the present case, the focusing grids are separated into six distinct elements having curvature; moreover, they are placed near the cathode so that the effect of the grids on the cathode surface is not uniform; hence the effects from (b) to (d) must be considered. To simplify the consideration, only the effect of (a), which is greatest, is to be taken into account: Namely, such a condition is considered that the rectangle electron beam of constant width d and of constant electron density J rotates with constant velocity v. Then, $v=rw$, where w is the angular velocity of the sinusoidal voltage applied to the focusing grid, and r is the radius, where the screening electrodes are considered as sections of a cylinder. In this case, the waveform of the output current is influenced by the relation between the shape of the window and the width of the electron beam.

When the shape of the window is a rectangle or a right triangle, the relation between the shape of the window and the width of the electron beam can be obtained by a simple calculation, the calculated results are shown in tables 43.1 and 43.2. Figure 43.7 and figure 43.8 indicate the calculated relations between the width of the electron beam and the waveform of the output current, and examples of the experimental results are shown in figures 43.9 and 43.10.

From the results described above, it is concluded that various current waveforms can be obtained from various combinations of the width of the electron beam and the shape of the window; that is, the

TABLE 43.1. *Calculated results of the output current when the shape of the window is a rectangle*

Time	Output current (I)
(a) $b>d$	
$0\leqq t\leqq\frac{d}{v}$	$J\cdot a\cdot v\cdot t$
$\frac{d}{v}\leqq t\leqq\frac{b}{v}$	$J\cdot a\cdot d$
$\frac{b}{v}\leqq t\leqq\frac{b+d}{v}$	$J\cdot a[b+d-v\cdot t]$
(b) $b=d$	
$0\leqq t\leqq\frac{b}{v}$	$J\cdot a\cdot v\cdot t$
$\frac{b}{v}\leqq t\leqq\frac{2b}{v}$	$J\cdot a[2b-v\cdot t]$
(c) $b<d$	
$0\leqq t\leqq\frac{b}{v}$	$J\cdot a\cdot v\cdot t$
$\frac{b}{v}\leqq t\leqq\frac{d}{v}$	$J\cdot a\cdot b$
$\frac{d}{v}\leqq t\leqq\frac{b+d}{v}$	$J\cdot a[b+d-v\cdot t]$

Note: The coordinate is taken as is shown in figure 43.7a. J, the electron density of the electron beam that is supposed to be constant; I, current by the electron beam passing through the window.

TABLE 43.2. *Calculated results of the output current when the shape of the window is a right triangle*

Time	Output current (I)
(a) $b>d$	
$0 \leqq t \leqq \frac{d}{v}$	$\frac{1}{2} c \cdot J(v \cdot t)^2$
$\frac{d}{v} \leqq t \leqq \frac{b}{v}$	$\frac{1}{2} c \cdot J[2dv \cdot t - d^2]$
$\frac{b}{v} \leqq t \leqq \frac{b+d}{v}$	$\frac{1}{2} c \cdot J[b^2 - (v \cdot t - d)^2]$
(b) $b=d$	
$0 \leqq t \leqq \frac{b}{v}$	$\frac{1}{2} c \cdot J(v \cdot t)^2$
$\frac{b}{v} \leqq t \leqq 2\frac{2b}{v}$	$\frac{1}{2} c \cdot J[b^2 - (v \cdot t - b)^2]$
(c) $b<d$	
$0 \leqq t \leqq \frac{b}{v}$	$\frac{1}{2} c \cdot J(v \cdot t)^2$
$\frac{b}{v} \leqq t \leqq \frac{d}{v}$	$\frac{1}{2} c \cdot J \cdot b^2$
$\frac{d}{v} \leqq t \leqq \frac{b+d}{v}$	$\frac{1}{2} c \cdot J[b^2 - (v \cdot t - d)^2]$

Note: The coordinate is taken as is shown in figure 43.8a and $c=a$. J, the electron density of the electron beam that is supposed to be constant; I, current by the electron beam passing through the window.

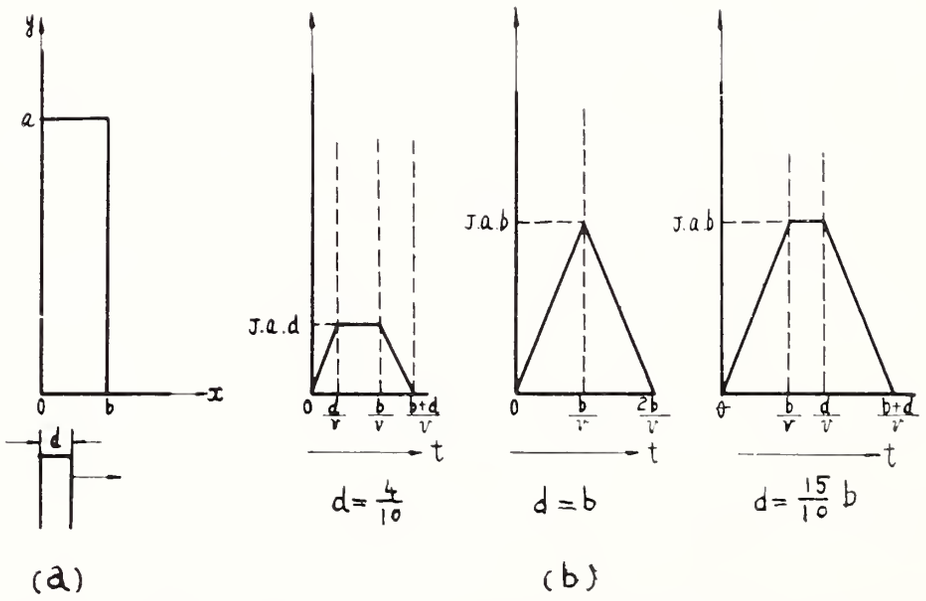


FIGURE 43.7. *Relation between the shape of the window and the output-current waveform when the shape of the window is rectangular.*

(a), the shape of the window and (b), the output-current waveform for various values of d/b. (Calculated results.)

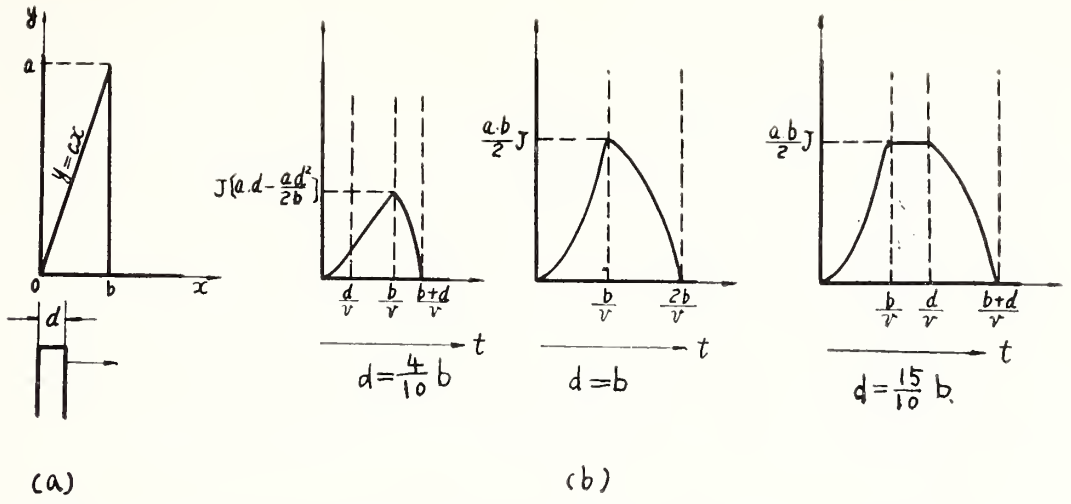


FIGURE 43.8. *Relation between the shape of the window and the output-current waveform when the shape of the window is that of a right triangle*

(a), the shape of the window and (b), the output-current waveform for various values of d/b. (Calculated results.)

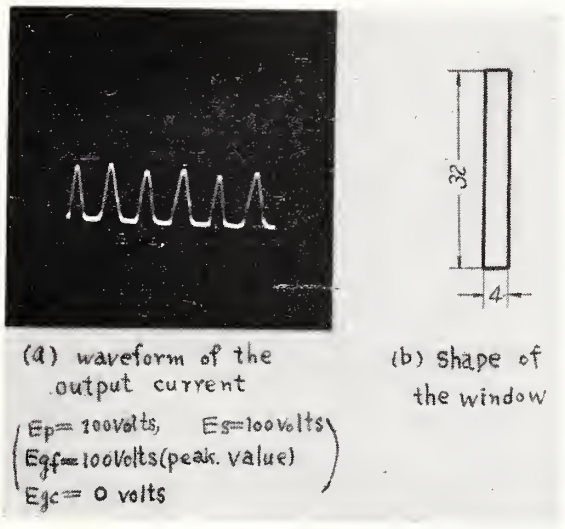


FIGURE 43.9. *The output-current waveform obtained by an experiment.*

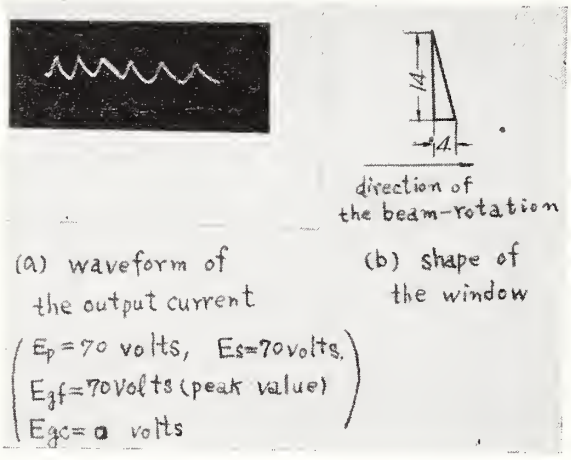


FIGURE 43.10. *The output-current waveform obtained by an experiment.*

fact that the output waveform is changed by the width of the electron beam even if the shape of the window is held constant means that the various voltages applied to the focusing grids, the screening electrodes, and the anodes cause the various shapes of the output current waveform. Therefore, the voltages applied to the electrodes and the shape of the window must be chosen appropriately in order that a desirable current waveform in practical use can be obtained.

New Type of Tube for Microwave Region

Controlling an electron beam by means of a rotating electric field in the comparatively low-frequency region is described in the previous section. In the microwave region, however, it is impossible to utilize a phase shifter or a goniometer to obtain a polyphase sinusoidal voltage. To overcome this difficulty, the author proposed the new method in which a traveling wave is utilized. As is shown in figure 43.11, the cathode is surrounded by n grids, which are made by folding one wavelength of rod or bar repeatedly; the traveling wave propagates along them. The phase difference between the adjoining grids is $360°/n$, and thus an n phase sinusoidal electric field is formed

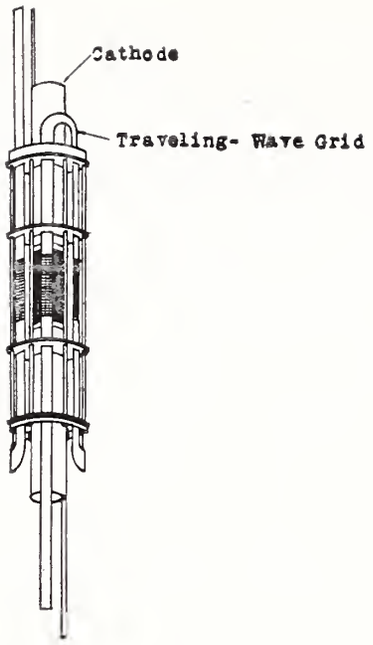


FIGURE 43.11. *Assembly of cathode and traveling-wave grid.*

around the cathode. Thus electron beam emitted from the cathode rotates with the same period as that of the traveling wave. As one example of the applications of a rotating electron beam controlled by means of the traveling wave, the examination of possibilities of a frequency multiplier tube for microwave frequencies was commenced. Summary of the results obtained are described mainly from the standpoint of electron optics.

Construction and Operating Principles

Construction of the tube is shown in figure 43.12. The traveling wave propagates along the traveling-wave grid, which operates as a control grid. The inner wall of the output cavity serves as the screening electrode.

A radio-frequency input signal is applied between the cathode and the traveling-wave grid, and by matching the circuit, this signal propagates along the grid as the traveling wave. Thus an electric field of n-phase is produced around the cathode. In figure 43.12 ($n=7$), the electron beam emitted from the cathode rotates with the same period as that of the input voltage when a suitable negative bias voltage is applied to the grid. The rotating electron beam excites the output cavity through the window of the cavity wall, and an output current whose frequency is n times that of the input signal can be obtained. (The output cavity is excited as a coaxial resonator and a part of the window is made to be the voltage loop.)

Since a large electron transit time between the cathode and the traveling-wave grid is unfavorable, a virtual cathode is proposed to be utilized in the tube.

As is shown in figure 43.13, an accelerating grid is placed around the cathode on which a positive d-c voltage is impressed, and outside of it is arranged the traveling-wave grid. When a suitable d-c voltage of negative value is applied to the traveling-wave grid, a virtual cathode is formed before it. Such construction of the tube is supposed to be favorable in reducing the effect of the transit time of electrons [6] and in making the perveance of the tube high. The radio-frequency input voltage is applied between the accelerating grid and

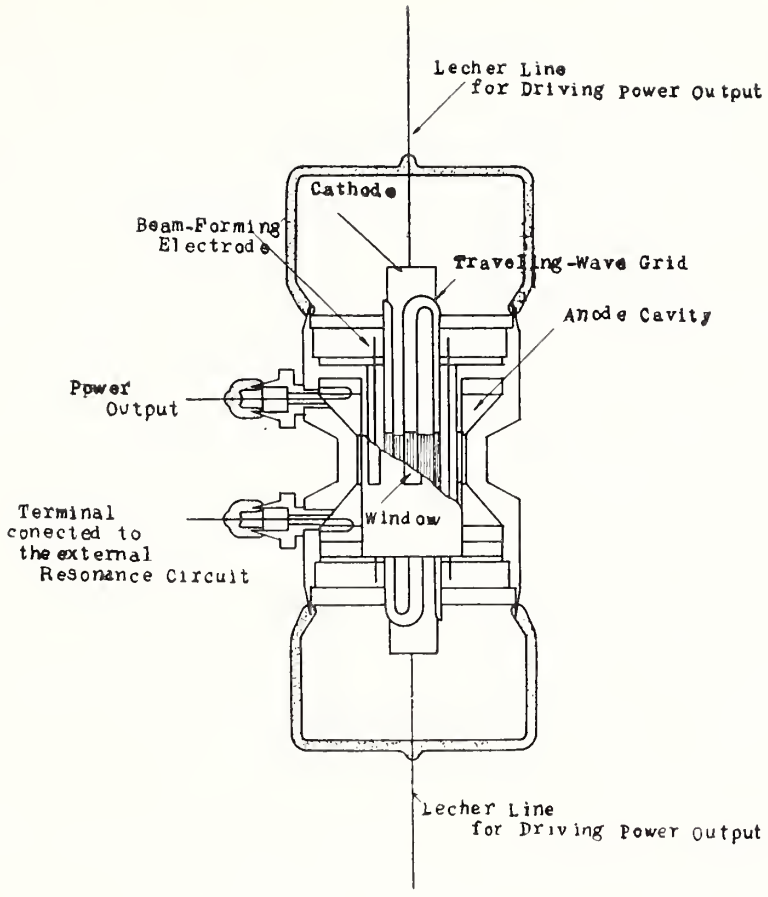


FIGURE 43.12. *Test tube No. 1.*

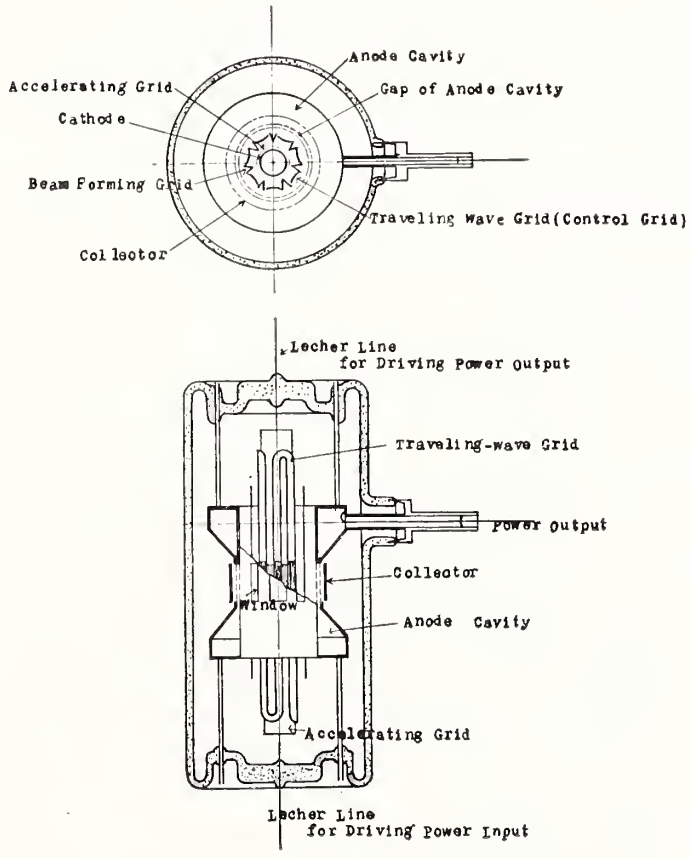


FIGURE 43.13. *Test tube No. 2.*

the traveling-wave grid, and the negative resistance of the electrons is supposed to be able to be utilized by adjusting the transit time of electrons suitably between the accelerating grid and the virtual cathode.

In the latter sentence, the term "control grid" means that part of the traveling-wave grid that is opposed to the window of the inner wall of the output cavity.

Investigation From the Standpoint of Electron Optics, and Static Characteristics of the Tube

In order to obtain a high power output and a high efficiency in the tube, an electron beam with as large a component of the radio-frequency current as possible must be allowed to flow into the window of the output cavity; that is, the difference between electron currents that flow into the window when the voltage of one of the control grids is maximum, and that when the adjoining two grids are at the same voltage, is desirable to be as large as possible. From this point of view, an electron-optical investigation of the tube has been performed. As the mathematical analysis of the tube is very difficult, the potential distribution in the tube was obtained by using an electrolytic bath at first.

Figures 43.14 and 43.15 show examples of the potential distribution when the voltage on one of the control grids is maximum. In figure 43.14, the peak value of the input radio-frequency voltage is 30 v; the d-c voltage of the screening electrode is 300 v with respect to the cathode. With respect to the cathode, the accelerating grid is 30 v; the beam-forming electrode, −30 v; and the control grid −27 V. In figure 43.15 the peak value of the input radio-frequency voltage is 20 v and the d-c voltage of the control grid is 18 v, the d-c voltages applied to other electrodes being the same as that in figure 43.14.

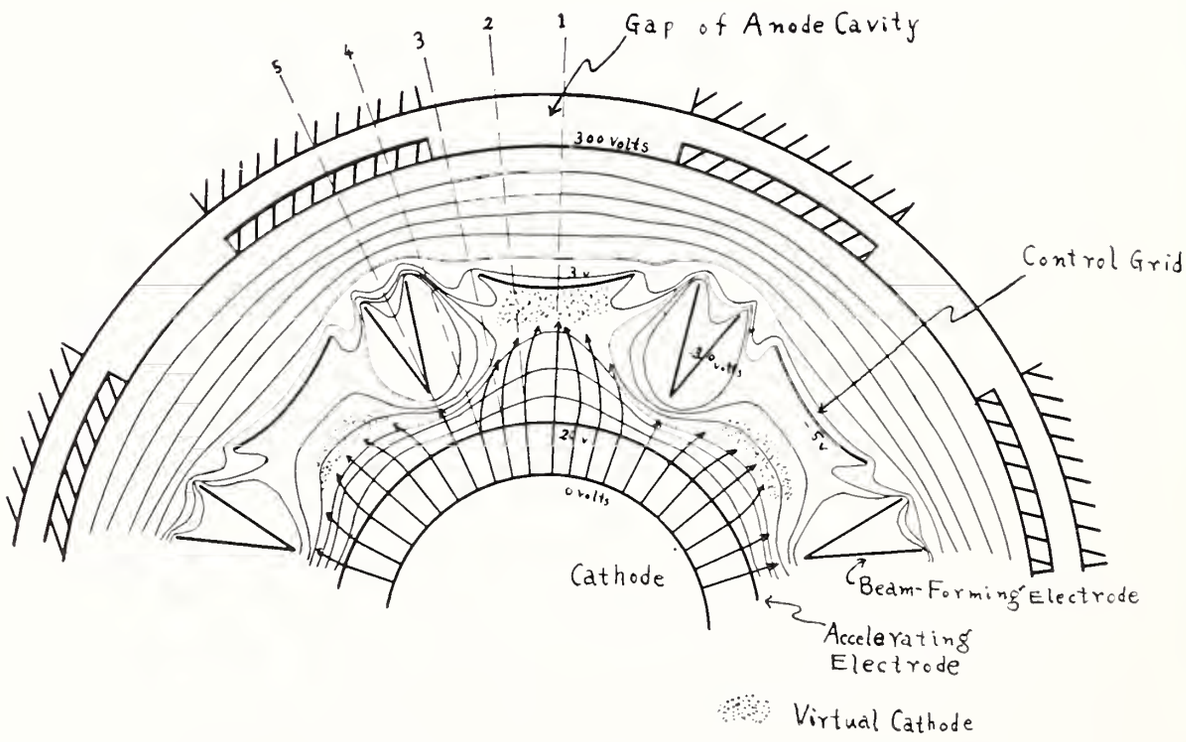


FIGURE 43.14. *Equipotential lines and electron trajectories.*
Driving voltage is 30 v (peak).

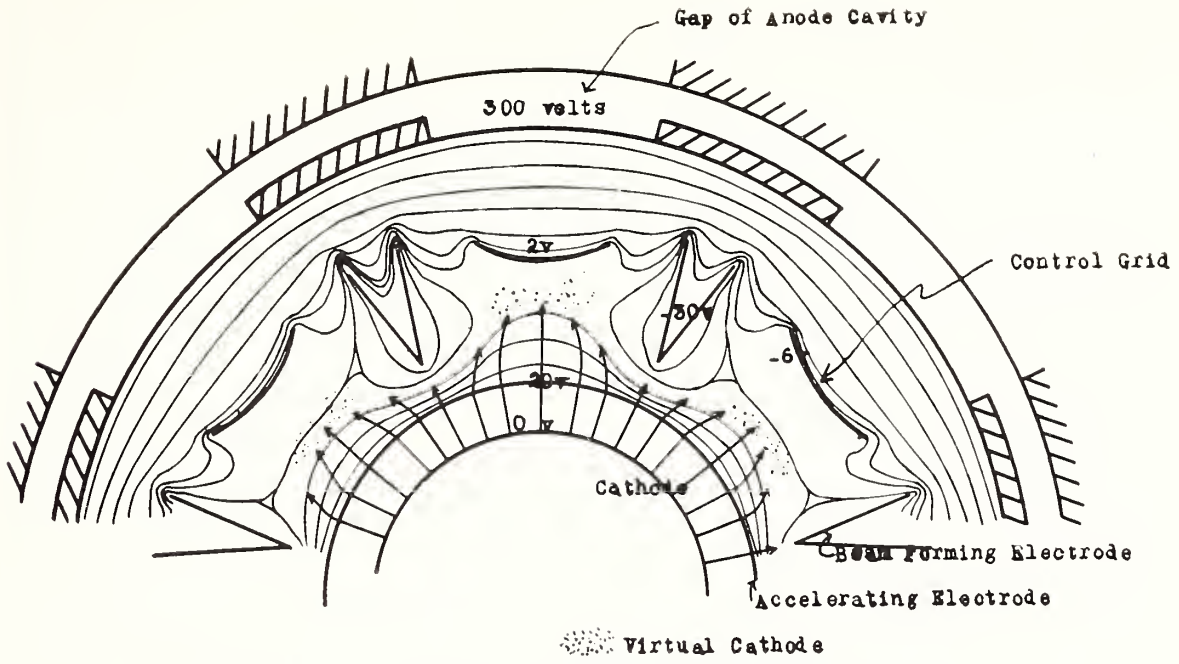


FIGURE 43.15. *Equipotential lines and electron trajectories.*
Driving voltage is 20 v (peak).

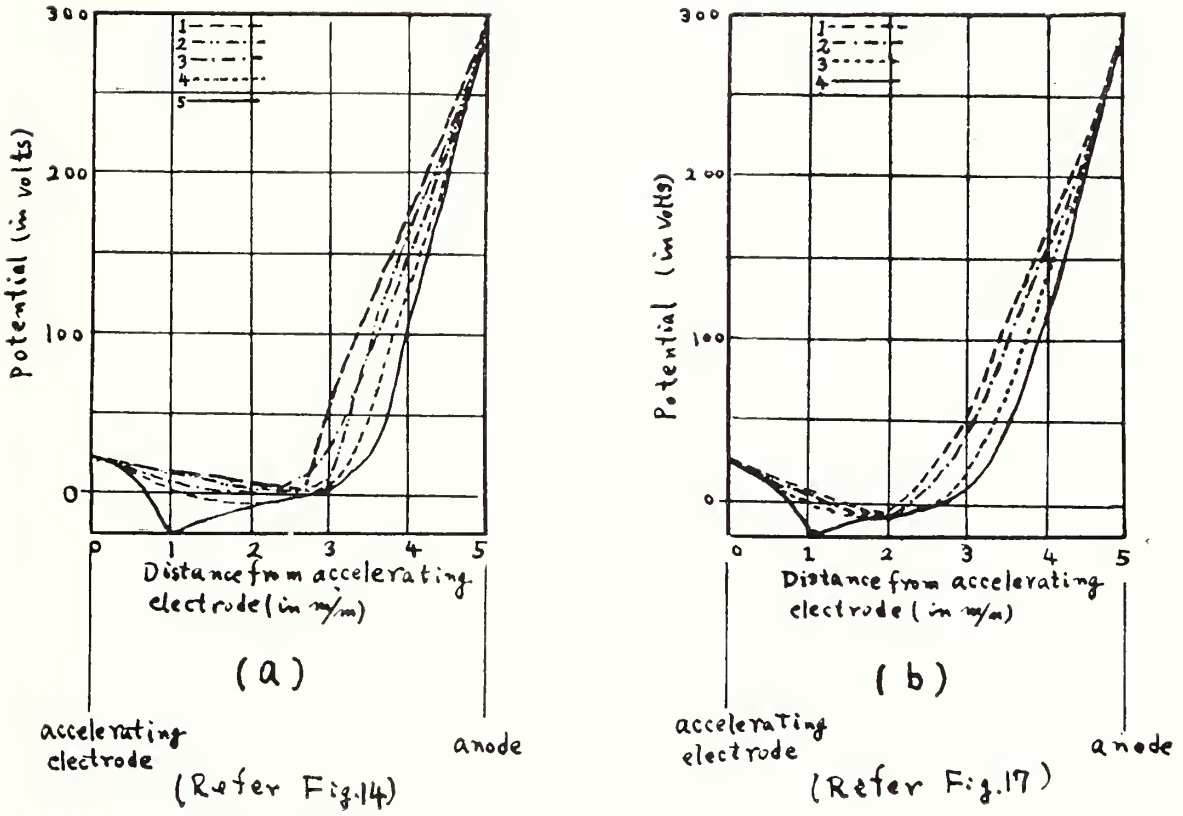


FIGURE 43.16. *Potential distribution on the several radii.*

Figure 43.16a shows the potential distribution along the radii 1, 2, 3, 4, and 5, in figure 43.14.

It is supposed from these figures that the electron emitted from the cathode are accelerated by the accelerating grid, and then the virtual cathode is formed before the control grid by the negative voltages of the beam-forming electrodes and the control grids The electrons from the virtual cathode are controlled by the applied radio-frequency voltage to the control grid, and they effectively flow into the window of the inner wall of the cavity by means of the electron lens formed between the control grid and the inner wall of the cavity.

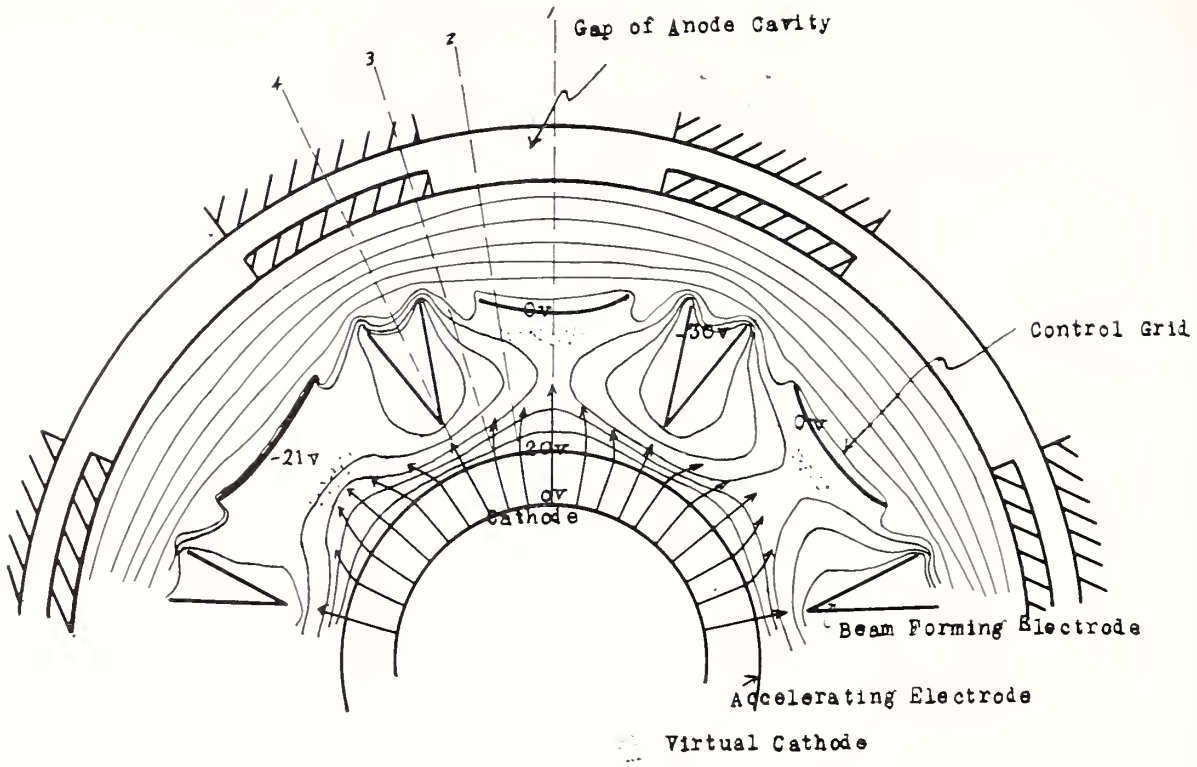


FIGURE 43.17. *Equipotential lines and electron trajectories.*
Driving voltage is 30 v (peak).

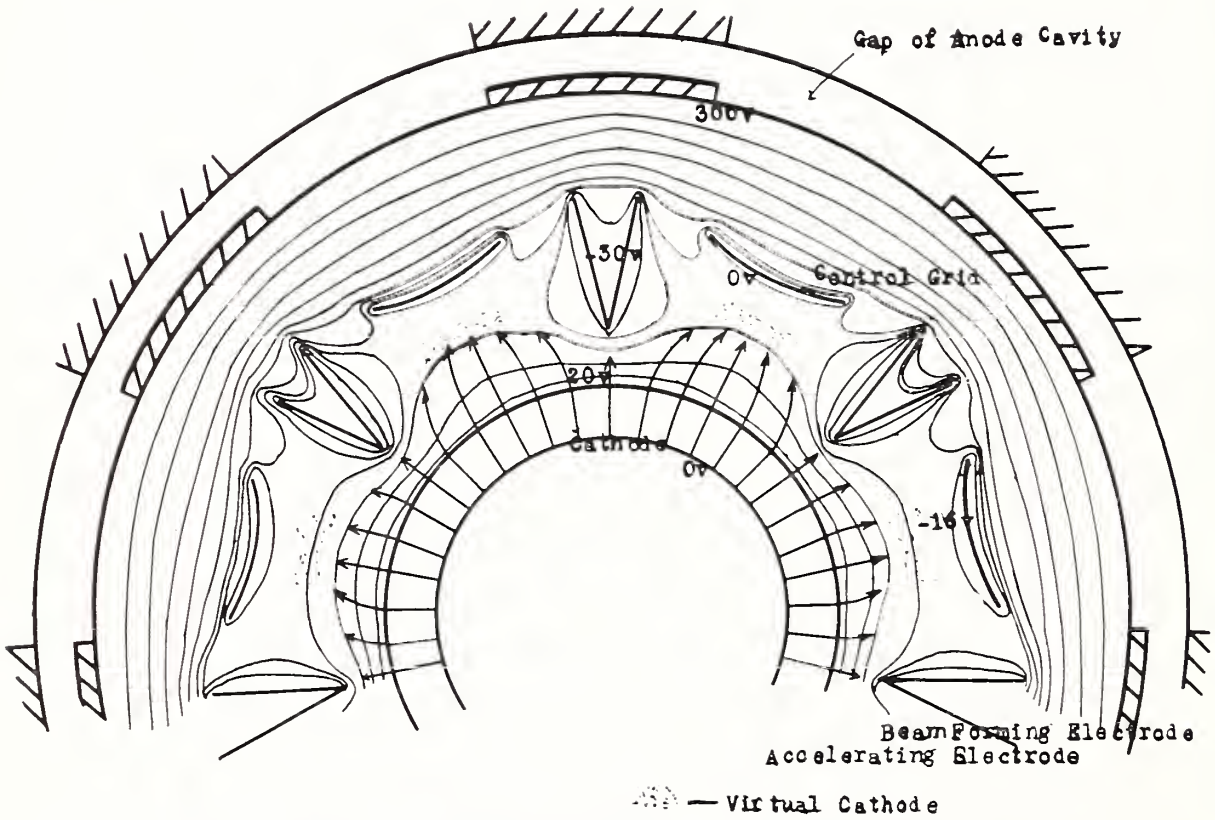


FIGURE 43.18. *Equipotential lines and electron trajectories.*
Driving voltage is 20 v (peak).

Figures 43.17 and 43.18 show the potential distributions when the radio-frequency voltages of the two adjoining control grids are equal, namely, when the phase leads about 25.7° from the previous state in figures 43.14 or 43.15. Input radio-frequency voltages in figures 43.17 and 43.18 are 30 and 20 v, respectively, and figure 43.16b shows the potential distribution along the radii 1, 2, 3, and 4 in figure 43.17.

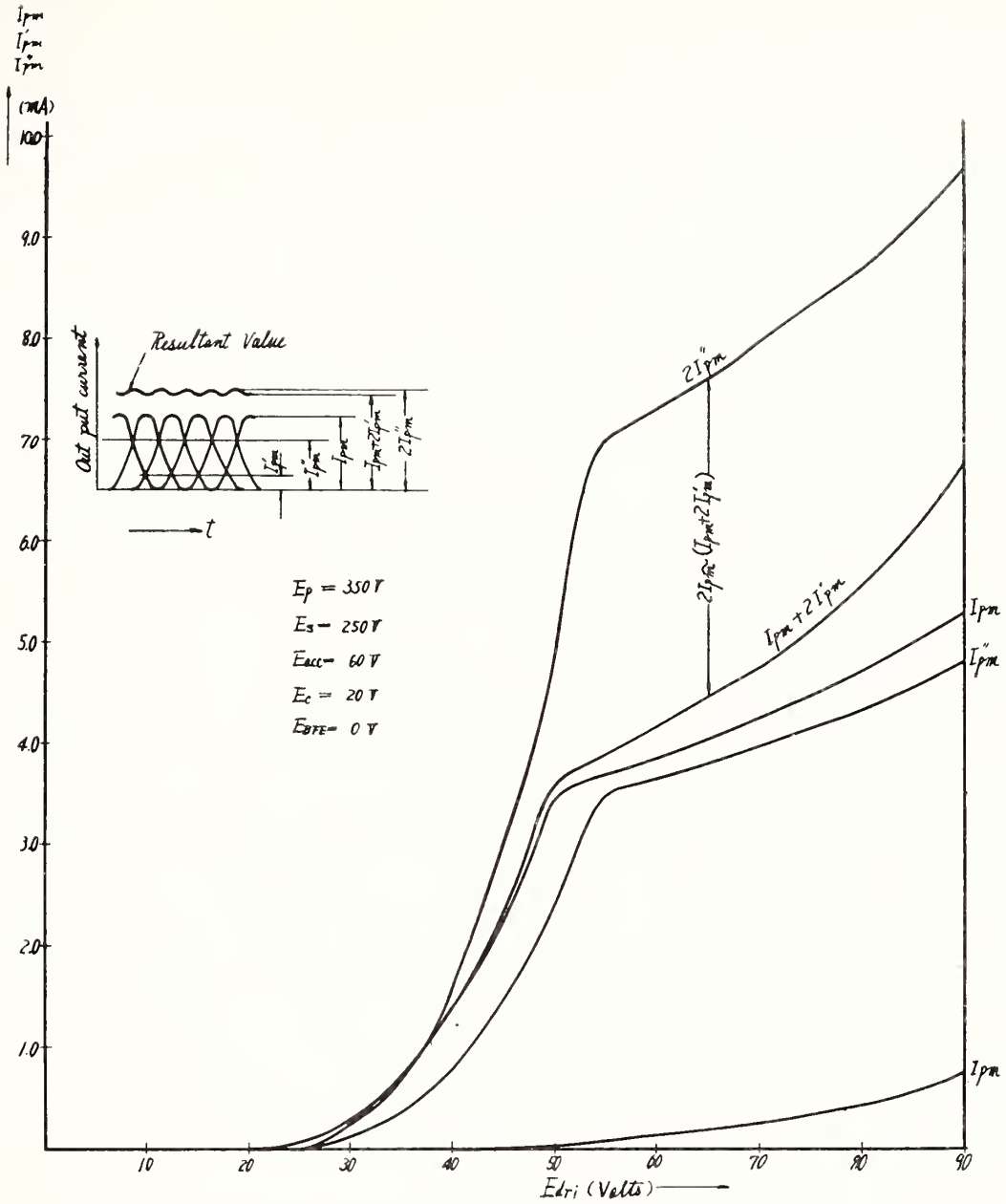


FIGURE 43.19. $I_{pm}.I_{pm}'.I_{pm}''$—*Edri curve.*

From the above results it becomes clear that the negative equipotential lines swell out before the control grid and that by such an electric field the electrons are made to enter the window. Furthermore, the beam forming electrodes serve to prevent the effect of the d-c voltage of the output cavity on the virtual cathode and to reduce the d-c component of the electron current.

Based on results obtained in the experiments using the electrolytic bath, test tubes were made to examine the static characteristics of the tube. In these test tubes, seven electrodes separated from each other are used as the control grids instead of ones described above; d-c voltages, whose magnitudes are different by as much as 360°/7 of one period of the sinusoidal voltage in phase with each other, are applied to them. On the other hand, seven anodes that are connected in parallel are placed behind each window of the inner wall of the output cavity, and both the outer and side walls of the cavity are omitted.

I_{pm}, I'_{pm}, and I''_{pm} are defined as follows:

I_{pm}, the anode current flowing through one window when the maxi-

mum voltage is applied to the control grid on the same radius as that of the window;

I'_{pm}, the anode current flowing through one window when the maximum voltage is applied to the control grid on the adjoining radius of that of the window; and

I''_{pm}, the anode current flowing through one window when the voltage of the grid on the same radius as that of the window is the same as the voltage of the adjoining control grid.

Then, the larger the difference between $(I_{pm}+2\ I'_{pm})$ and $2\ I''_{pm}$ is, the more the radio-frequency component of the anode current is. Figure 43.19 indicates the relation between the input radio-frequency voltage and I_{pm}, I'_{pm}, I''_{pm}, $I_{pm}+2\ I'_{pm}$, and $2\ I''_{pm}$ under the following conditions:

d-c anode voltage	350 volts
d-c voltage on the inner wall of the cavity	250 volts
d-c voltage on the beam forming electrode	0 volts
d-c voltage on the accelerating grid	60 volts
d-c cathode bias voltage	20 volts

In the previous investigation of the results obtained using the electrolytic bath, it was supposed that the anode current begins to flow through the window after the anode current through the adjoining window reduces to zero, i. e., $I'_{pm}=0$, and $I''_{pm}=0$. These results can be concluded only when the input radio-frequency voltage is small, and as the increase of the input voltage $2\ I''_{pm}$ becomes larger than $(I_{pm}+2\ I'_{pm})$; therefore the maximum radio-frequency component of the anode current can be obtained when the voltages of the adjoining control grids are equal. Such a result is shown in figure 43.19; in such a condition the efficiency falls off because the d-c components of the anode current are many. Therefore, to increase the efficiency of the tube, it is necessary that the maximum anode current can be obtained when the control grid voltage is maximum, concerning this problem further investigations and experiments are needed.

Conclusion

In the previous sections, controlling an electron beam by means of a rotating electric field in both low and microwave regions of frequency was described.

In the frequency multiplier tube for low frequencies, the characteristics were much improved by introducing electron optical systems into the electrode structure.

On the other hand, increasing the input frequency makes it difficult to generate the sinusoidal polyphase voltage, so adequate measures must be considered. Moreover, the distortion of the sinusoidal polyphase voltage gives different a shape to each output waveform; spurious frequencies of the order of multiples of the input frequency have appeared. Besides, in controlling an electron beam using the traveling wave, it is necessary that the traveling wave propagate along the traveling-wave grid. The experimental and theoretical investigations on this problem are now being continued.

Moreover, it is supposed that the electron optical examination on this tube must be performed not only from the standpoint of static consideration but also from that of the dynamic state in which the

existence of the virtual cathode and the transit time effect of the electrons must be considered. In the experimentally manufactured frequency multiplier tubes for the microwave region, favorable characteristics have not yet been obtained.

References

[1] A. M. Skellett, Electrostatically focused radial-beam tube, Proc. Inst. Radio Engrs. **36,** 1354–1357 (1948).
[2] A. M. Skellett, The magnetically focused radial beam vacuum tube, J. Appl. Phys. (1944).
[3] Electrical Communication Laboratory, unannounced.
[4] K. Takashima and T. Nakamura, Radial-beam type commutator tube, Lecture-meeting of Engineering in Kansai, October 1949.
[5] K. Owaki, K. Takashima, T. Takagi, and T. Nakamura, Radial-beam type commutator tube, unannounced.
[6] K. Miyaji, Virtual cathode control tube, J. Inst. Elec. Comm. Eng. of Japan **34,** 316–325 (1951).

44. Experiments With Adjustable Magnetostatic Electron Lenses

By E. Ruska [1]

Purpose and Basic Result of Experiments

Among the fields in axially symmetric and space-charge-free electron lenses, magnetic lens fields distinguish themselves from electric lens fields by their ability to produce particularly short focal lengths as well as by their inherent low degree of aperture defect, chromatic aberration, and astigmatism. This superiority is increased by several advantages of operation; for instance, by the more convenient adjustability, by the lower expenditure required for the reliability of operation and, last but not least, by the advantage of their better suitability for use with faster electrons. The essential drawback of the electromagnetic electron lenses generally used up to now consists in the considerable expenditure to secure a stable current source, whether it be in the form of a battery or an electronically regulated power supply. Permanent-magnet-electron lenses were originally used for cathode ray oscillographs [1] [2] requiring relatively weak magnetic fields. However, an attempt was soon made to produce strong fields required for lenses with a short focal length as, for example, Siemens and Halske did in 1940 when designing a magnetostatic transmission electron microscope [2] that, with 76-kv electrons, yielded a direct magnification of 5,200 times. Figure 44.1 is a cross section through this lens arrangement that consists of an externally disposed, axially magnetized cylinder casing with connected pole pieces [3], and this lens is shown to have a strong, external stray field. The electron microscope makes very high demands on the exactly circular shape of the lens fields and the mutual alinement of the lenses. With lens arrangements having an external stray field, interferences of the ideal lens field may occur if iron components are brought into the stray field and moved within it. Hence lenses of this type are not very suitable for use in the electron microscope. Therefore it was tried to transfer the axially magnetized cylinder into the interior of the lens arrangement [4] (fig. 44.2). In this assembly a second extensive field is produced behind the lens field at the pole-piece gap within the cylindrical magnet. On the one hand, this second field increases the minimum attainable focal length to an undesirable extent, and, further, its axial symmetry is not sufficient due to the insufficient homogenity and the difficult machining of the magnetic material. Some proposals enable both disadvantages to be avoided and lead to stray-field-free double-lens systems. In this case, a radially magnetized annular magnet [5] or an axially magnetized cylindrical magnet or two such magnets connected in opposition [6] are housed in a soft-iron casing with two annular gaps, each of which

[1] Wernerwerk für Messtechnik der Siemens u. Halske A. G., Berlin-Siemensstadt, Germany.
[2] Figures in brackets indicate the literature references on p. 406.

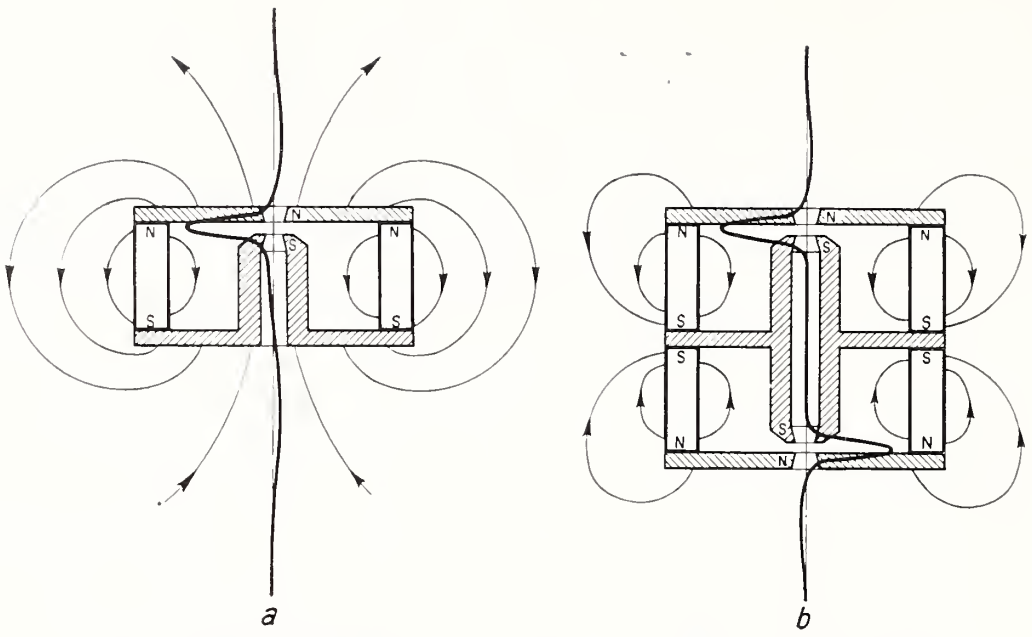


FIGURE 44.1. *Permanent magnetic electron lenses with external stray field.*

a, Single; (b) two opposite axial magnets.

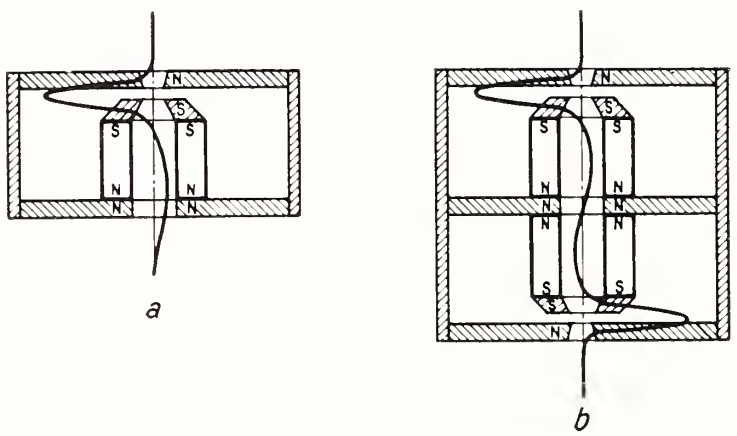


FIGURE 44.2. *Permanent magnetic electron lenses with internal stray field.*

a, Single; (b) two opposite axial magnets.

produces a lens field. Electron microscopes with axially magnetized lens systems have recently been carried into effect by v. Borries [7] and by Reisner and Dornfeld [8]. In the latter arrangement, the magnetostatic lens fields of about 1,000 ampere-turns are not variable. Focusing is achieved by variation of the velocity of the electrons. With the first-mentioned electron microscope, lens fields of 2,000 and 3,200 ampere-turns were obtained, and the shapes of the fields in the objective lens and projector lens are varied to change the magnification.

For electron microscopes, but also for other electron-optical devices, the adjustability of magnetostatic fields seems to be essential for operational reasons. Field adjustment within narrower limits is required for focusing with constant accelerating potential of the electrons and a given distance between lens and specimen; within a wide range, adjustment of lens fields is necessary if a higher variability of the accelerating potential of the electrons is desirable, while the focal length is maintained, or if the magnification of the electron optical system must be varied within a wide range with constant accelerating potential of the electrons.

We therefore decided to investigate differently designed adjustable magnetostatic lens arrangements with reference to their minimum obtainable focal lengths (maximum attainable field strength in the lens center equal to $y_{0\ (max)}$), and to their adjustability to longer focal lengths (minimum attainable field strength in the lens center equal to $y_{0\ (min)}$). Different constructional designs should, at the same time, enable us to acquire experience in connection with the most useful methods of magnetization and the adjusting devices. The investigations were carried out in collaboration with Mr. M. Eisfeldt, Miss Dipl. Phys. K. Müller, and Dr.-Ing. O. Wolff and led to the result, that the demands made on an extensive and continuous adjustability of the fields and focal lengths, respectively, may be met satisfactorily by different arrangements. Therefore, the prospects for the development of high-performance electron microscopes with adjustable magnetostatic lenses seemed to be favorable.

Design and Efficiency of Investigated Adjustable Magnetostatic Lens Arrangements

The investigated lens systems can all be reduced to two basic forms, the radially magnetized annular disk [9] and the axially magnetized tubular cylinder [1]. It is evident that both basic forms, as well as hollow conical forms, can be more or less approximated by means of bar magnets. Since, for all permanent magnetic systems, the integral of the field strength along a closed path must be equal to zero, the distribution of the field strength on the axis of symmetry z of both magnet types is given qualitatively as shown in figures 44.3a, 4a, and 5a. The relatively wide field ranges of the symmetrical arrangements in figures 44.3a and 5a, which are opposite to one another with respect to the field direction, are equal in magnitude. In order to obtain lens fields of strong refractive power, the fields must be confined to a small region of space, so that strong maximum field strengths H_0 are produced on the axis. This is achieved by the internal and external stray magnetic flux being passed through a soft-iron casing of suitable form [5] (figs. 44.3b, 4b, and 5b). The two oppositely directed

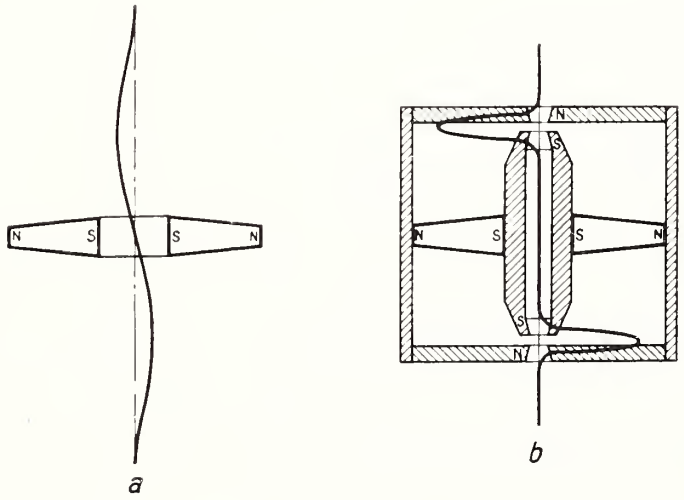


FIGURE 44.3. *Pair of electron lenses, produced by means of a radial magnet.*

a, Without; b, with soft-iron casing.

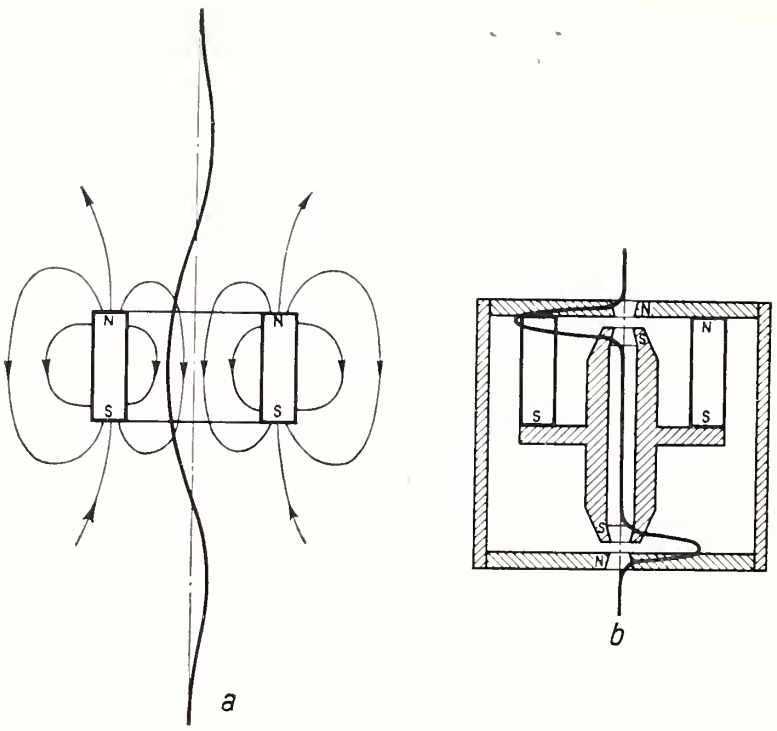


FIGURE 44.4. *Electron lens and pair of electron lenses, respectively, produced by means of an axial magnet.*

a, Without; b, with soft-iron casing.

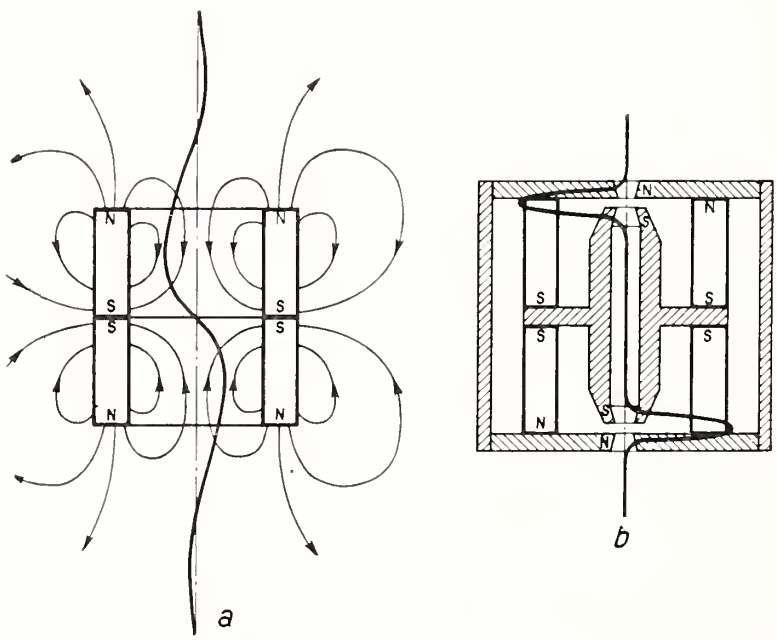


FIGURE 44.5. *Pair of electron lenses, produced by means of two opposite axial magnets.*

a, Without; b, with soft-iron casing.

field regions on the axis are concentrated at the two gaps of the pole-piece units with simultaneously increased maximum field strength, with the result that two lenses adequately separated from one another and of short focal length are produced. The axis of symmetry for each externally stray-field-free permanent magnetic lens system is given by

$$\int_{-\infty}^{+\infty} H(z)dz=0.$$

Thus, with the encased system (fig. 44.4b) as well, which is modeled on the basic form (fig. 44.4a) with only one axially magnetized cylinder, two oppositely directed field regions of equal integral are produced on the axis. Hence, every externally stray-field-free permanent-magnet-lens system is not subjected to image rotation, if only external beam cross sections on both sides of the lens systems are considered. Since the refractive power of a magnetic lens increases with the square field-strength integral on the axis of rotation, we can make the focal lengths of the lenses different in magnitude by giving different dimensions to the pole-piece systems (pole-piece diameter and gap width determine the half-value width of the field distribution) even if the $\int H(z)dz$ has the same value for the two lens fields. Permanent magnet lens systems, as shown in figures 44.3b, 4b, and 5b therefore are sufficient for such electron optical arrangements in which, for a given axial distance, two lenses of fixed focal lengths are required that may be equal in magnitude or different from one another. In systems of this type with unvariable magnetic fields, a synchronous enlargement, or reduction of the focal lengths of the two lenses can be obtained by an increase or decrease of the electron velocity.

In some important electron-optical instruments, however, the lens fields must be adjustable, for instance, if, for any reason whatever, an extensive variation of the electron velocity is not permissible, but the refractive power of the two lenses is to be varied synchronously, or if the variations in the refractive power of two successive lenses are required to be independent of one another. Thus, in the electron microscope for example, the accelerating potential of the electrons has been fixed within certain limits to prevent deterioration of the specimens. The focal lengths of the objective and the projector of the electron microscope need not be adjusted or only slightly, whereas the focal lengths of the intermediate lenses must be adjusted within a wider range to obtain variation in magnification. We therefore tried to develop stray-field-free permanent-magnet double-lens systems in in which at least one lens field should be adjustable within a wider range wih a reaction as slight as possible on the second lens field.

To reduce the field of a permanent-magnet lens, we can use four different methods, singly or together:

(1) Decrease of the internal magnetomotive force of the permanent magnet by reduction of the reluctance of the magnetic circuit. In this case, the magnet can be loaded more or less (a) along its whole length or (b) a variable part of it. The coercive force, however, prevents the magnet from being weakened effectively, so that, by this means alone, it is not possible to achieve an adjustment sufficient to cut off the lens effect completely. Of course, a change of the magnetomotive force affects all lens fields produced by the permanent magnet in the same way and, approximately, to the same extent.

(2) Increase of the half-width of the lens field and decrease of the maximum field strength H_0, with the external magnetomotive force between the pole pieces of the lenses remaining constant, by means of the enlargement of the diameter of the pole pieces or the gap between the pole pieces or both dimensions. The adjustment of the field shape can be achieved either by exchanging different pole-piece systems or by continuously varying the distance between the pole pieces (gap width); it is not possible, however, by these methods of adjustment, to make the refractive power of the lens arrangement entirely equal to zero.

(3) Reduction of the external magnetomotive force between the pole pieces of the lens with the magnetomotive force of the permanent magnet remaining constant by displacing a portion of the total external magnetomotive force from the lens gap to other points of the magnetic circuit (adjustments by dividing the external magnetomotive force). If a magnetic adjusting gap is provided in series with the lens gap, the magnetic adjusting field can be (3a) either predominantly an external stray field or (3b) predominantly a second lens field. In the latter case, the second lens can be kept weak by providing suitable dimensions for the gap and the diameter of the pole pieces if its refractive power is not desired. In some cases, however, the additional lens field is desired and can be given suitable strength by means of adequate pole-piece dimensions (3c). An adjustment, by dividing the magnetomotive force, with external stray field permits of practically suppressing the refractive power of the two component lenses. Hence it is of special importance. An adjustment, by dividing the magnetomotive force, can also be affected without adjusting gap by utilizing the drop of the external magnetomotive force always occurring within the soft-iron components. Stray fields that weaken the lens fields at the gaps and act themselves as weak supplementary lenses can be produced at different points along the axis of the magnetic systems, depending on the relative design of the soft-iron circuit and on the manner in which the distribution of the external magnetomotive forces and the dimensions of the soft-iron circuit are varied by the adjusting process.

With adjusting mechanisms used in practice, several of the previously mentioned effects usually occur at the same time or successively during the adjusting process. Up to now, we have investigated the following arrangements, which are shown, together with their qualitative field shapes, in figures 44.6 through 10:

Radial magnet with adjustment by means of the movement of a shorting plug. The magnetomotive force of the magnet is decreased by an additional loading on the whole magnet length.

Axial magnets with adjustment by means of an external annular disk. The magnetomotive force of the magnet is decreased by an additional load on a variable part of the magnet length.

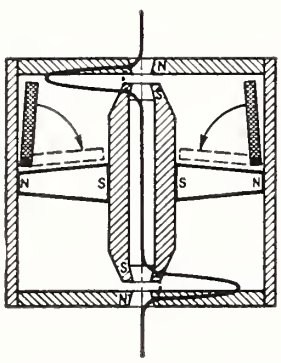


FIGURE 44.6. *Pair of electron lenses with one radial magnet adjusted by means of movement of a shorting plug.*

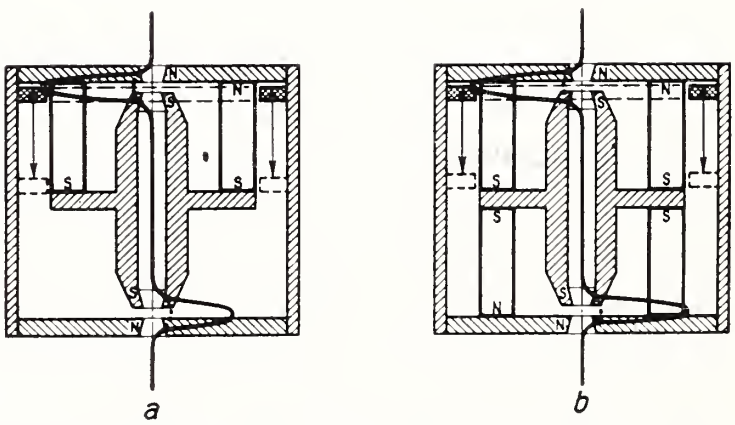


FIGURE 44.7. *Pair of electron lenses with adjustment by means of an external annular disk.*

a, Single; b, two opposite axial magnets.

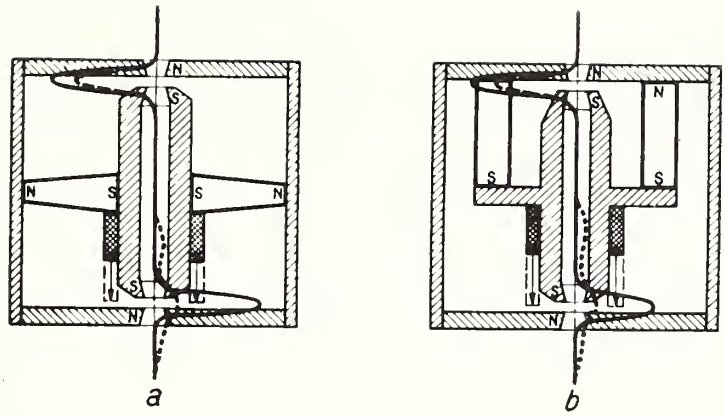


FIGURE 44.8. *Pair of electron lenses with barrel adjustment.*

a, Radial magnet; b, axial magnets.

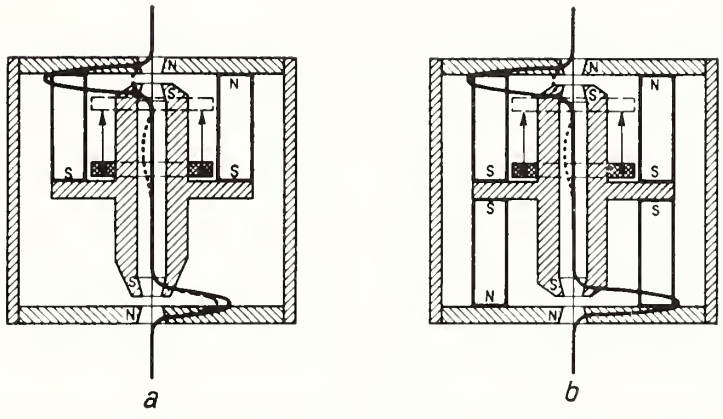


FIGURE 44.9. *Pair of electron lenses with adjustment by means of internal annular disk.*

a, Single; b, two opposite axial magnets.

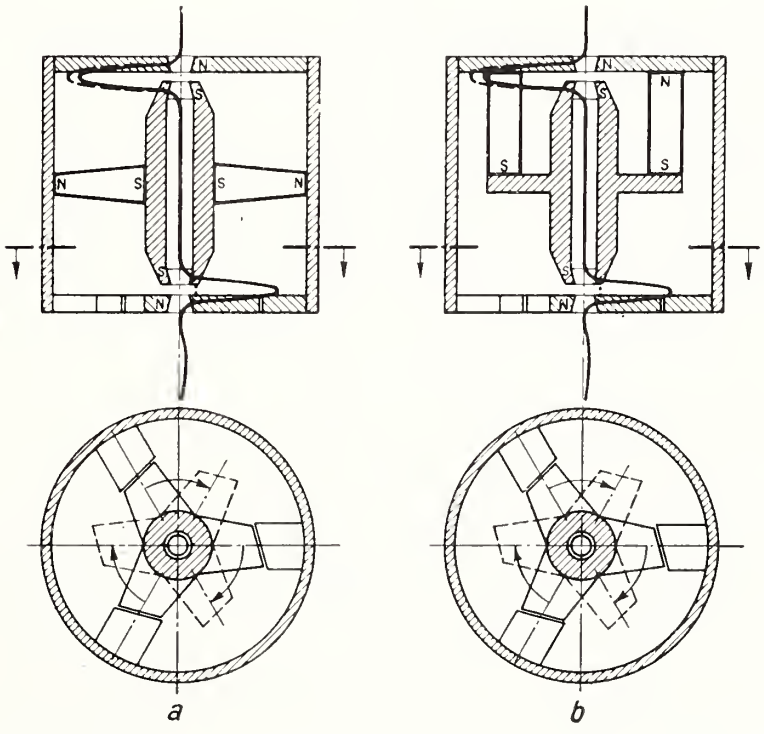


FIGURE 44.10. *Pair of electron lenses with adjustment by means of a wheel-shaped yoke*

a, Radial magnet; b, axial magnets.

Radial or axial magnet with internal adjusting barrel. With a preliminary small decrease in the magnetomotive force of the magnet and in the external magnetomotive force between the lens pole-pieces, the adjusting barrel gives rise only to an unimportant decrease in the maximum field strength. In the last part of the adjusting path, the magnet is weakened more quickly by the more rapidly increasing load, with the result that the magnetomotive force between the pole pieces is decreased in a correspondingly shorter time. This leads to the deviation from the proportionality usually required between the adjusting path and the refractive power of the field. The more the lens gap is bridged by the adjusting barrel, the more extensive the drop of the magnetomotive force within the soft iron will be, which results in the occurrence of stray fields on both sides of the lens gap that further reduce the field strength within the gap.

Axial magnet with adjustment by an internal annular disk. The adjustment is accomplished by partially shortening the magnet. This results in a drop of the magnetomotive force at the gap. The residual magnetomotive force is, with flat field distribution, perceived within the oversaturated internal iron tube.

Radial or axial magnet with adjustment by a wheel-shaped yoke. The adjustment is made by dividing the drop of the magnetomotive force between a lens gap and an adjustable external air gap. According to the position of the wheel-shaped yoke, there occurs in the plane of the yoke a more or less extensive radially directed ring-shaped stray field that is connected in front of the lens and produces also an axially directed stray field outside the soft-iron casing. Due to the large half-value width and the low amount of the axially directed stray field, this, as a lens, has only a weaker refractive power.

The decrease obtainable by the adjusting devices in the investigated arrangements of the magnetomotive force at the lenses and the maximum field strength may be denoted as the adjusting coefficient for the field strength $RH = H_{0\,(\min)}/H_{0\,(\max)}$. The adjusting coefficient for the focal length $Rf = f_{(\min)}/f_{(\max)}$ follows from the adjusting coefficient R in two different ways according as the field shape (half-value width) has been maintained or changed. With the adjusting coefficient R being the same, the focal length of the lens is, for an invariable field shape, adjustable within a wider range than for a widened field shape, since, in the latter case, the field is effective for a longer time, and therefore the refractive power remains greater. The determination of the absolute magnitude of the focal length requires consideration of the dimensions of the pole-piece systems.

Magnetic material and magnetization

For the permanent magnetic lenses we use cast material of the Alnico type having a $(B{\cdot}H)_{\max}$-value of $4{\cdot}10^6$ cgs-units. The magnets were machined to have a preferred direction. For our systems we used 5 to 8 cm long magnets, cylindrical and conical tubes for the axial systems, and molded bodies for the radial systems. In the first experiments, the magnets were magnetized before the lens systems were assembled, which, of course, did not fully utilize the magnet. Therefore we soon changed over to the better method of magnetizing the assembled lens system by means of a powerful electromagnet (30,000 ampere-turns). Axial systems having only one cylindrical

magnet are relatively easy to magnetize. Systems using two oppositely disposed axial magnets were divided, during construction, in such a way, that the resulting two system halves could easily be magnetized. Severer difficulties arise with radial magnet systems, if they are to be magnetized when already assembled. We overcame this inconvenience by designing a star-shaped system of bar magnets. At first the bar magnets were magnetized alone, and later with a shorting plug, which was removed only after the assembly of the closed star system had been accomplished. Magnetization by means of a ring coil (8,000 ampere-turns) disposed coaxially to and beside the magnet star was also tried, but was not sufficiently effective.

Measuring Methods

The focal length of electron optical lenses can be determined, on the one hand, from the measurement of the field distribution along the axis and, on the other hand, from the magnification obtained through an exposure. For the following reasons, we have used both methods. On the one hand, we usually could, for constructional reasons, only determine the two-stage magnification of a pair of magnetostatic lenses, which alone was not sufficient to specify the exact focal length of each single lens, as required in our case, and, on the other hand, we would not merely rely on the measurements of the different lens fields, as our measuring device allowed only the field distribution near the lens gap to be determined, and not the weak residual fields between the two lenses, which also exert an influence on the attainable magnification. Moreover, in order to evaluate the image quality, we attached great importance to an exposure allowing determination of the magnification, since, from the observation of astigmatism, we hoped to be able to draw conclusions concerning the origin of not exactly circular lens fields for whose occurrence either the constructional design or the machining process may be responsible.

Field Measurement by Means of Dropping and Swinging Coils

The field strengths on the lens axis were determined by a well-known method involving the ballistic measurement of the potential induced by the field within small search coils [10]. With electromagnetic lenses, the induction in the search coil is effected by changing the pole of the current that reverses the field direction, the search coil thereby maintaining its position in the field, and the magnitude of the induction produced being twice that which would correspond to the field. For the measurement of permanent-magnet fields, however, the search coil must quickly be brought from a field-free point into the field to be measured or inversely, so that only an induction corresponding to the simple magnitude of the field is produced. We therefore designed a measuring device including a search coil ($n=300$, $\phi=1$ mm) which can drop by about 10 mm, and is kept exactly in the axis of the magnetostatic lens system to be measured. Later we also used a search coil swinging in the axial direction [11] whereby the alternating voltage produced within the coil was amplified (fig. 44.11). In this case, the field gradient at the point of the search coil $H'=dH(z)/dz$ rather than the field itself is measured. The displacement of the swinging center of the search coil along the lens axis enables very

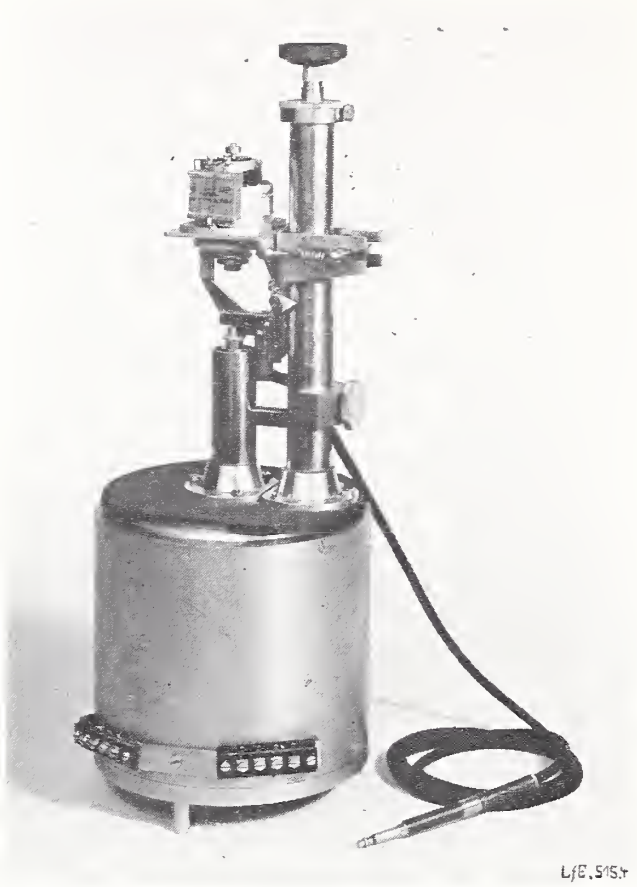


FIGURE 44.11. *Swinging coil with model lens for the measurement of magnetic lens field.*

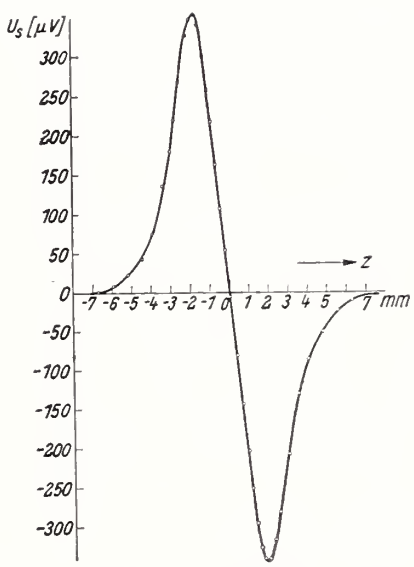


FIGURE 44.12. *Measurement of the axial gradient of the magnetostatic lens field by means of a swinging coil.*

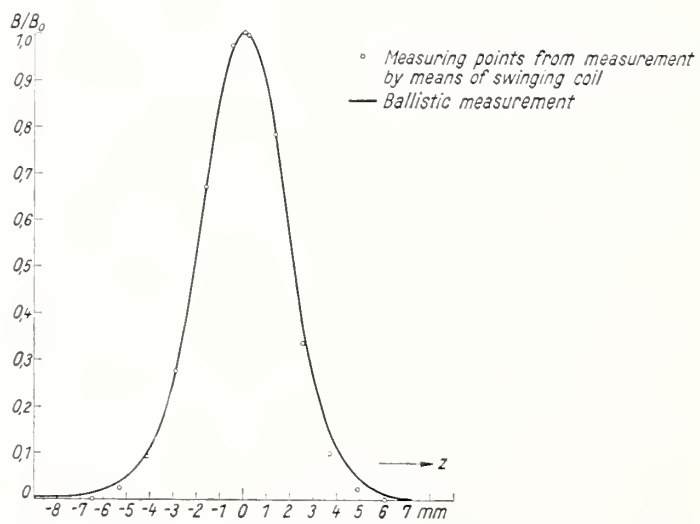


FIGURE 44.13. *Shape of the magnetostatic lens field according to the measurement by means of dropping and swinging coils.*

exact characteristics of the field distribution to be obtained also by this method, as shown in figures 44.12 and 13. From the measured field distribution the half-width value a^* of the field distribution and the maximum field strength H_0^* can be deduced. As is well-known, the measured field distribution largely agrees with the bell-shaped magnetic field more thoroughly dealt with by Glaser [12], where the

focal length with

$$k^2=\frac{e}{8mU}(H_0a)^2$$

has the value

$$\frac{f}{a}=\frac{1}{\sin \pi/\sqrt{k^2+1)}}.$$

We could frequently do without the measurement of the whole field and only determine the maximum field H_0^* alone, since, from extensive previous measurements on electromagnetically excited lens pole-pieces of different diameters and gap widths, we knew already the relation between the maximum field strength and the focal length obtained from magnification measurements.

Determination of Magnification in Electron Optical Bench

To test the values of the focal lengths that have been calculated from the ballistic measurements on H_0 while considering the pole-piece dimensions determined from the half-width values, we have measured the electron optical magnifications obtainable in the magnetostatic lens systems. These experiments were carried out in an electron optical bench which we designed in the course of the last years [13] (fig. 44.14). The development of this arrangement should enable us to test electron-optical beam paths and components within a shorter time than that required by the methods generally used up to now. The elements of an electron optical arrangement are built as separate sections of a vacuum column (fig. 44.15). Electron guns, specimen stages, lenses of different types and focal lengths, tubes used to prolong the image observation, as well as arrangements designed for direct photographic exposures of the electron optical image or such obtained through the intermediary of a fluorescent screen, and finally an optical microscope, serving for the observation, measurement and photographic exposure, are acccommodated in a bench bed consisting of two parallel columns.

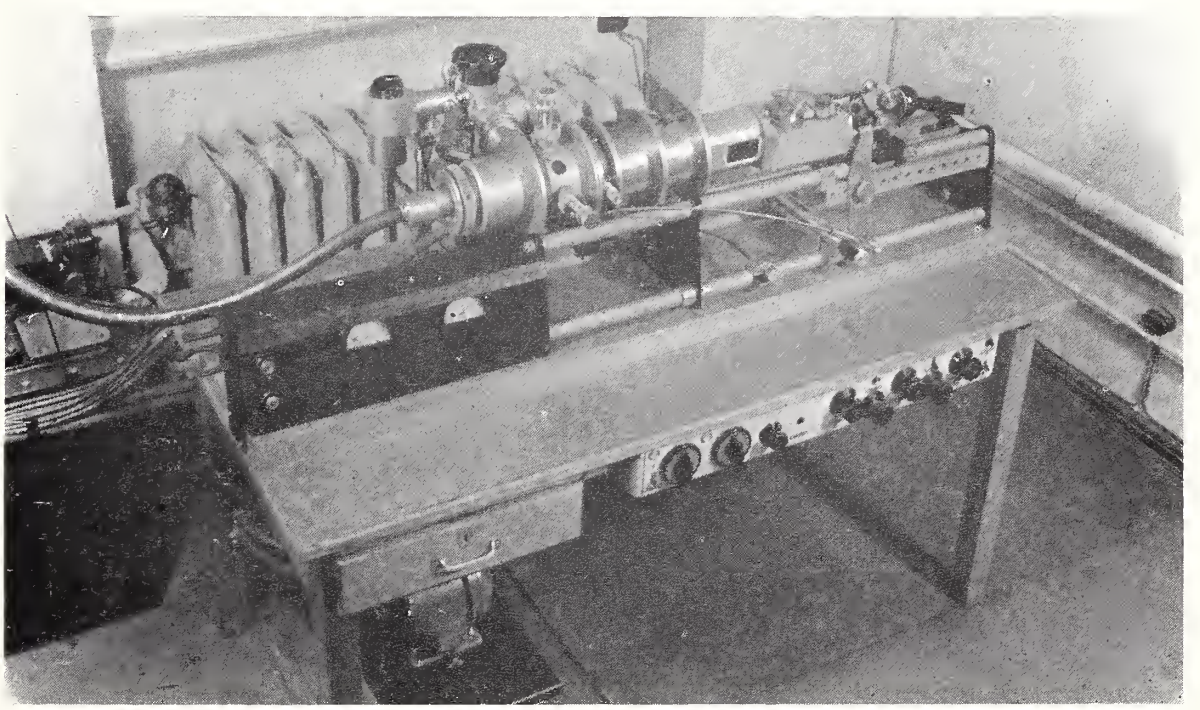


FIGURE 44.14. *View of the electron-optical bench.*

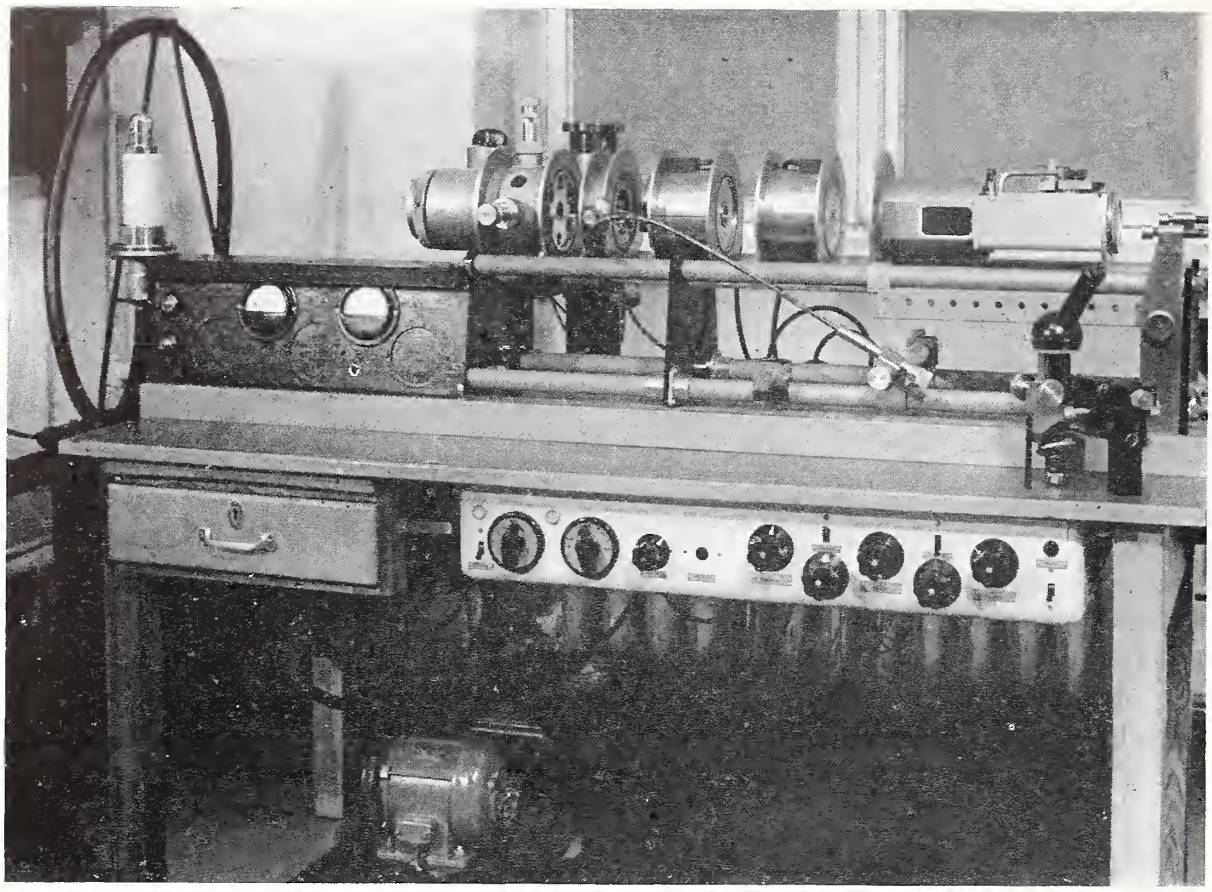


FIGURE 44.15. *Components of the electron-optical bench.*

The column sections are equipped at both ends with annular disks of equal dimensions, in which rubber rings are inserted to provide a vacuum-tight sealing. The sequence of and the mutual distance between the electron optical elements therefore can be varied within a few minutes, which necessarily results in a sufficient coaxial alinement of all the elements of the vacuum column. In operation, the beam can be directed exactly into the bench axis; this is achieved by a universal-stage type displacement of the cathode with respect to the control electrode-anode system. Evacuation is effected by means of a single valve permitting the adjustment to all operating conditions. Pumping down to the operating vacuum requires about two minutes. The high voltage, adjustable in steps and shielded from contact, is applied to the electron gun and the electrostatic lenses by means of an externally grounded cable. The electromagnetic lenses, of which each is equipped with windings for coarse and fine adjustment, are supplied from a battery. Thereby, each lens may be associated, with means for coarse and fine control.

To determine the magnification of the magnetostatic systems, replicas were made from scales scratched in glass, with the grid distance being ½ micron [14], and then photographed in the electron microscope. The specimen stage could be displaced both in the direction of the optical axis and in the universal-stage type manner.

Principles Governing Design of Adjustable Magnetostatic Lens Systems

The comparison of the adjusting coefficients measured for the field strength R_H of the investigated arrangements furnishes simple rules for the design of adjustable magnetostatic lens systems.

Since, for geometrical reasons, the external stray flux of a cylinder is appreciably higher than the internal stray flux, the flux within the magnet is largely dependent on variations of the external magnetic flux. Thus, an effective decrease of the magnetomotive force of the magnet can be obtained only by an increase of the permeability of the external stray flux. (Adjustment by means of the movement of a shorting plug, fig. 44.6, or by means of an external annular disk, fig. 44.7.)

If, during the adjustment of a lens, the magnetomotive force of the magnetic system is to be varied within a range as narrow as possible, for example, to maintain the refractive power of the other lenses unchanged, the flux through the magnet should not be varied appreciably. The external stray circuit, therefore, must not be interfered with, and the internal stray circuit should be modified in such a way that, during adjustment, the interior stray flux also will only vary slightly. (Adjustments by means of dividing the magnetomotive force, fig. 44.16b.)

The least variation possible of the load of the magnet is, at the same time, the fundamental condition for obtaining the highest possible magnetomotive force and lenses of the highest refractive power possible for a magnet system of given dimensions. In such systems, more than 3,000 ampere-turns (that is, about 3,750 gilberts) can still be obtained with a minimum of magnetic material.

The refractive power of a permanent-magnet lens system can be adjusted down to zero, only if the two lenses with oppositely directed field ranges are weakened by the addition of an external adjusting gap in front of them (fig. 44.16a).

Weakening of one lens in a magnetostatic double-lens system without stray fields causes the second lens to be weakened if, at the same time, the flux in the magnet increases, or to be strengthened if the flux decreases. Both lenses are weakened to nearly the same extent in the case of adjustment through the movement of a shorting plug (fig. 44.6) and in the case of adjustment by means of an external disk (fig. 44.7). The second lens not directly adjusted to a lower value is slightly weakened through both barrel adjustment (fig. 44.8) and interior disk adjustment (fig. 44.9), and it is slightly strengthened through yoke adjustment (fig. 44.10). The reaction due to adjustment is very slight if, in arrangements equipped with axial magnets, the lens not directly adjusted has its own magnet. If, therefore, as with the interior disk adjustment, the adjustment is made within a magnet, we must provide a second magnet for the lens not directly adjusted, while leaving it unchanged within the limits possible. With barrel or yoke adjustment, for example, the adjustment can also be made outside the magnet cylinder, in which case, one of the magnets is automatically attached to the lens not adjusted.

The experimental designs allowed some useful experience to be made. The adjustment of electromagnetic lenses has the following advantages:

(1) The adjustment of their refractive power is accompanied with the slight torque inherent in an adjustable electric resistance.

(2) On the reversal of the adjusting movement (for example in the case of focusing) only a slight backlash is noticeable.

(3) The required coarse and fine controls are simple.

(4) Moreover, the current adjustment only involves the variation of the field strength of the lens without affecting the symmetry of the field with respect to the optical axis.

To obtain the same advantages also with the adjustment of magnetostatic lenses, certain conditions must be fulfilled. The adjustment through the movement of flux-conducting iron components requires the less energy the smaller the components to be moved and the more the flux within the adjusting components progresses vertically with respect to the adjusting movement; the easier it is also to avoid backlash in the mechanical arrangement, since there exist no uni-directional forces. The weight of the adjusting element is smaller with adjustment by means of an internal annular disk than with adjustment by means of an external annular disk. The additional provision of fine adjustment would, in general, require more complications for magnetostatic than for electromagnetic lens systems. If, to adjust the half-width value, the pole-pieces of the same lens system are moved oppositely to one another in the axial direction, the lens field is assymmetrically distorted due to the unavoidable radial clearance, so that astigmatic pictures as well as lateral image displacements are liable to occur during adjustment, particularly if the adjusted lens has a great refractive power; that is, it has small bore and small gap.

As is well-known, the pole-piece systems require a very high degree of axial symmetry if the lens is to be free from astigmatism. Since, with the adjustment of magnetostatic lenses, apart from the proper lens fields, stray fields of low refractive power frequently occur on the axis, care must be taken that also at these points there will be no deviations from the axial symmetry in the flux-conducting iron components. If, for constructional reasons, this cannot be avoided, the dissymmetry should have at least two mutually perpendicular axes. The wider the optically utilized beam is at the stray field, the more attention must be paid to the symmetry of the field shape.

Adjustable Magnetostatic Lens Systems for the Electron Microscope

Let us first deal with the conditions of the illuminating system: The intensity of the specimen illumination is usually adjusted through a condenser whose refractive power must be increased from a very low value to a moderately high value which corresponds to the image of the cathode on the specimen. It is advantageous to keep the condenser free from image rotation, so that the shift of the electron beam on the specimen will be negligible, which otherwise, in the case of a

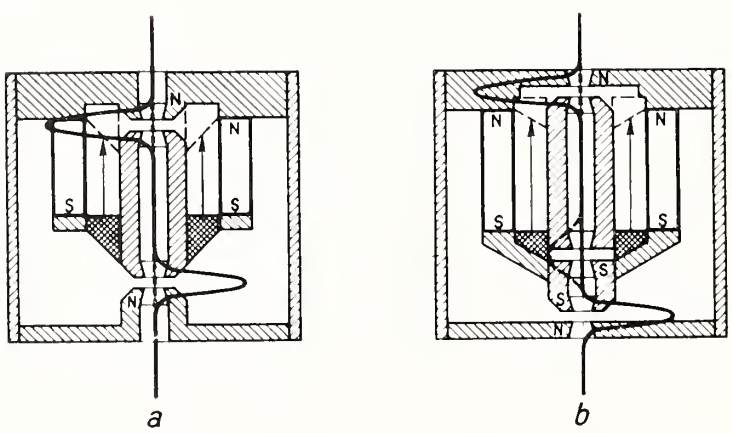


FIGURE 44.16. *Permanent magnetic two-lens systems for adjustment within a wide range.*

a. Both lenses synchronously adjusted; b, only one lens adjusted.

not perfectly coaxial entry of the electron beam into the condenser lens, would be very pronounced. The relatively great focal length of the condenser can easily be obtained with the aid of one of the image-rotation-free systems under consideration that use two single lenses. The refractive power can be extensively reduced in an arrangement as shown for example in fig. 44.16a. When the adjusting barrel is displaced from the center of the system in an outward direction, the value of the magnetomotive force of the internal soft-iron cylinder will more and more approach that of the system housing.

It must be possible to provide all magnifying lenses of the electron microscope, that is, objective, intermediate lenses, and projector, with a high refractive power. For the objective and the projector, the refractive power need not be very variable if the electron velocity remains unchanged. For the intermediate lenses, however, which are used for a continuous reduction of the magnification, the refractive power must be highly variable also in the case of unchanged electron velocity. Since, for each magnifying lens, a short focal length must be attainable, all these lenses, as compared with the condenser whose two gaps act as a lens doublet, can use only single fields, each of which produces a real image. Therefore, as to the two lens fields of a stray-field-free arrangement, the first must be adjustable within a wide range; the second, when used as objective, only within a narrow range; and, as projector, not at all. The two adjustments must be independent of one another as far as possible. (Yoke adjustment, fig. 44.10, or adjustment by means of distribution of the magnetomotive force on two lens fields, fig. 44.16b). Two such double systems connected in series result in a microscopic arrangement which magnifies in four stages and whose intermediate lenses, internally disposed and adjustable within a wider range, permit the magnification to be varied within large limits. These requirements are most satisfactorily met in the system shown in fig. 44.17a in which two external axial magnets connected in opposition excite the objective and the projector directly,

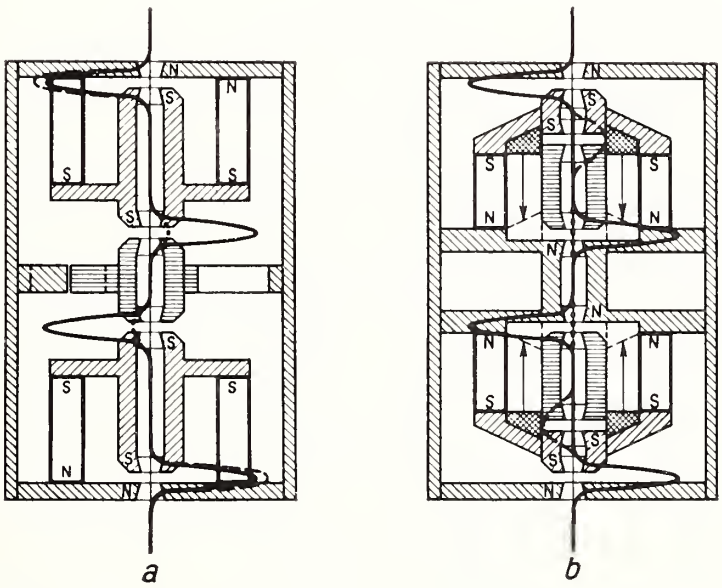


FIGURE 44.17. *Permanent magnetic multilens systems for electron microscopes with continuous magnification adjustment.*

a, Four-lens system and adjustment of both internal lenses by means of a wheel-shaped yoke; b, six-lens system and adjustment of the two pairs of internal lenses by means of dividing the magnetomotive force.

and the refractive power of the two internal lenses is reduced by means of either a yoke adjustment or a barrel adjustment of corresponding efficiency. Both methods of adjustment include the advantage, that, on the reduction of the two internal lenses, the refractive power of the external lenses increases, which is desirable for the objective of the electron microscope, because the sharpness of the final image is then more easily maintained during the adjustment of magnification. As to the projector, the increase in refractive power is not necessarily objectionable. A somewhat different adjustable microscopic arrangement with six single lenses and a four-stage magnification is shown in fig. 44.17b. In this case, the two opposed axial magnets are located internally, and adjustment of the two intermediate lenses is obtained by using two lens gaps for each lens for dividing the external electromotive force.

The last figures show cross sections through three different magnetostatic magnifying systems for electron microscopes, by the use of which we obtained electron micrographs. The technical design of a two-lens system with a radial magnet is shown in fig. 44.18. The radial magnet is composed of eight spokes. Every spoke consists of two bar magnets that are formed such as to give rise to approximately the same induction in every cross section. Adjustment of the lens field is obtained by means of two flaps of soft iron that more or less bridge the magnet bars. The flaps are adjusted together, which, however, also allows the fine adjustment of every single flap. Figure 44.19 shows a two-lens system whose objective allows barrel adjustment for focusing purposes. A limited magnification adjustment is obtained in the objective stage by means of the axial displacement of the specimen stage. Figure 44.20 shows an arrangement including three lenses and two axial magnets of different length. The magnets are oppositely connected in such a way, that a central lens is formed by their difference. The central lens is set to a lower value by means of a barrel with the reaction, that the objective is weakened (long magnet), and the projector is strengthened (short magnet). Of course, the magnets of different lengths can with advantage also be inverted in position.

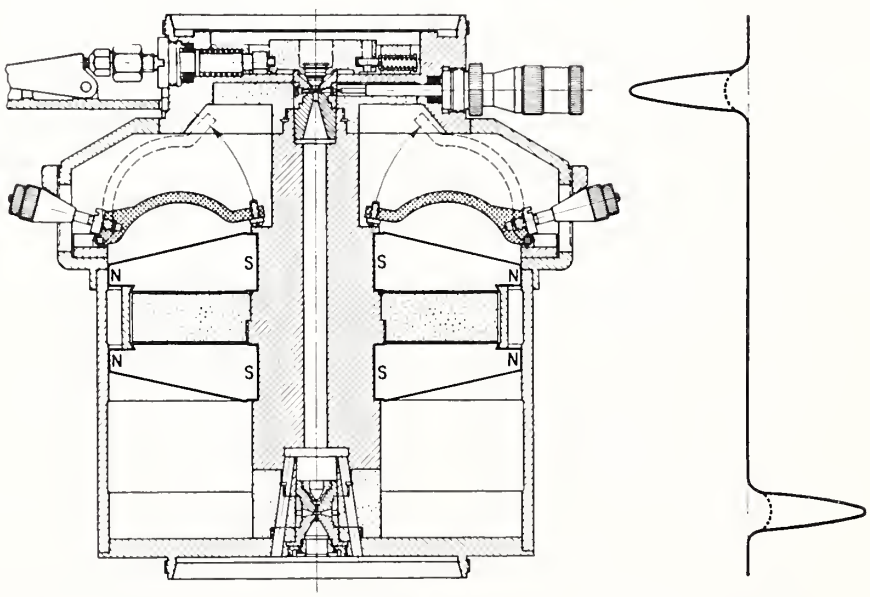


FIGURE 44.18. *Permanent magnetic two-lens system for electron microscopes with fixed magnification allowing adjustment, with the aid of flaps, for focusing purposes when the electron velocity is varied.*

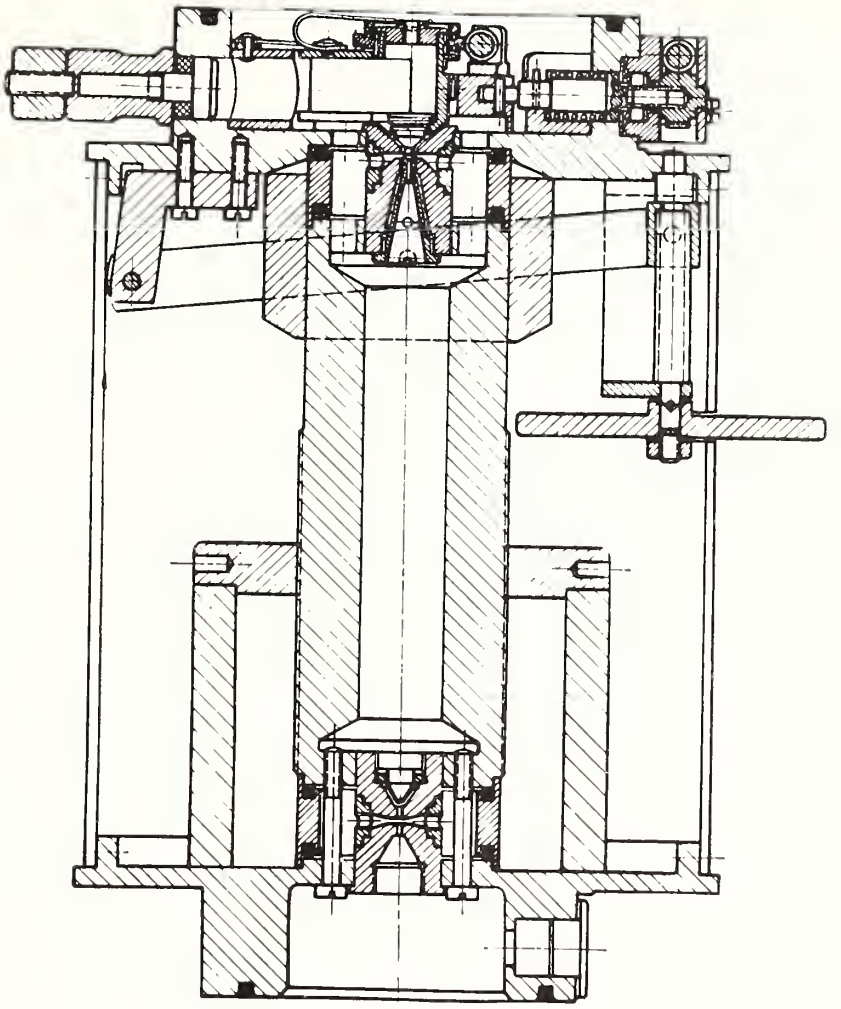


FIGURE 44.19. *Permanent magnetic two-lens system for electron microscopes of variable magnification with barrel adjustment of the objective for focusing purposes, and axially displaceable specimen stage.*

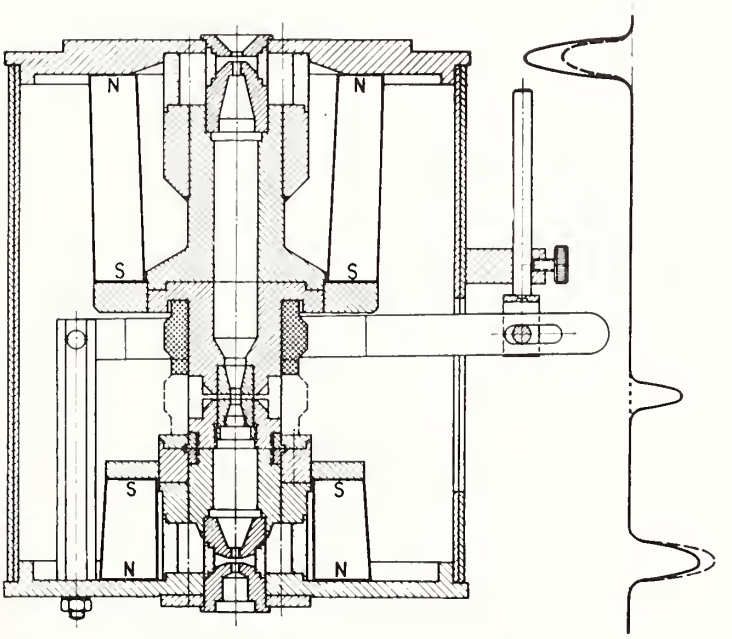


FIGURE 44.20. *Permanent magnetic three-lens system for an electron microscope with variable magnification by means of barrel adjustment of the central lens.*

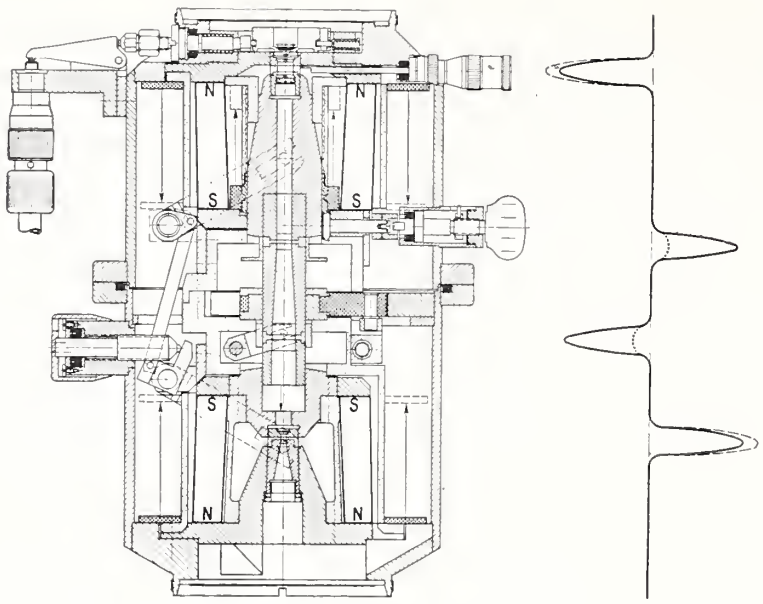


FIGURE 44.21. *Permanent magnetic four-lens system for electron microscopes with variable magnification allowing adjustment by means of an internal disk, an external disk, and a wheel-shaped yoke as well as axial displacement of the pole pieces of the internal lenses.*

Finally, figure 44.21 shows a four-lens system, which we have proved and whose design is in accordance with figure 44.17a; it is seen to use a yoke adjustment for the two central lenses that serves to vary magnification. The adaptation of all lenses to different beam tensions is obtained by means of an external annular-disk adjustment, and the focusing of the objective by means of an internal annular-disk adjustment. Up to now the expenditure on such an optical system seems to be rather high; we hope, however, that an improved constructional design will soon bring us nearer to the goal of a magnetostatic electron microscope of high performance.

References

[1] For example, H. Rumpff, Kathodenstrahloszillographen, Druckschrift 321 of the firm Dr.-Ing. Hans Rumpff, Bonn 1932, 20 p., on p. 14/15, and 16.

[2] B. v. Borries, E. Ruska, J. Krumm, and H. O. Müller, Übermikroskopische Abbildung mittels magnetostatischer Linsen, Naturwissenschaften **28,** 350–351 (1940).

[3] See, for example, W. Schröder, German patent DRP 720927 of February 25, 1937, and B. v. Borries and E. Ruska, German patent application of March 7, 1938.

[4] K. Küpfmüller and B. v. Borries, German patent application of January 15, 1943.

[5] B. v. Borries, German patent application S 153 465 of December 24, 1942.

[6] E. G. Ramberg, U. S. patent 2 369 796 of March 26, 1943; v. Borries, German patent application of October 18, 1944.

[7] B. v. Borries, Die Übermikroskopie (Saenger 1949, p. 104); Ein magnetostatisches Objektiv-Projektiv-System für das Electronenmikroskop, Kolloid–Z. **114,** 164–167 (1949); Ein magnetostatisches Elektronenschattenmikroskop als Elektronenbeugungsapparatur, Kolloid–Z. **118,** 110–112 (1950). Ein magnetostatisches Gebrauchs-Elektronenmikroskop für 60 kV Strahlspannung, Z. wiss. Mikr. **60,** 329–358 (1951).

[8] J. H. Reisner and E. G. Dornfeld, A small electron microspe, J. Appl. Phys. **21,** 1131–1139 (1950); Permanent magnet lenses, J. Appl. Phys. **22,** 561–565 (1951); Permanent magnet electron microscope, Electronics, 86–91 (1951).
[9] W. Henneberg, German patent DRP 689690 of September 2, 1936.
[10] L. Marton, Field measurements and possible correction of aberrations for magnetic electron lenses (abstract), Phys. Rev. **55,** 672 (1939); J. Dosse, Zur Ausmessung des Feldes magnetischer Elektronenlinsen, Z. Phys. **117,** 437–443 (1941).
[11] O. Wolff, Eine Schwingsonde zum Messen der Axialkomponente von Feldern magnetischer Elektronenlinsen, Phys. Verh., No. 4, p. 69 (Third Annual Meeting of the German Society for Electron Microscopy in Hamburg, 1951).
[12] W. Glaser, Strenge Berechnung magnetischer Linsen der Feldform $H=H_0/1+(z/a)^2$, Z. Phys. **117,** 285–315 (1941).
[13] E. Ruska, Über den Aufbau einer elektronenoptischen Bank für Versuche und Demonstrationen, Z. wiss. Mikroskop, **60,** 317–328 (1952).
[14] We are grateful to Mr. H. König, Göttingen, for the production of these specimen micrometers.

Discussion

Mr. M. E. Haine, Associated Electrical Industries Research Laboratory, Aldermaston, Berkshire, England: We should like to include in this discussion the general question of vacuum systems, seals, and methods of leak testing. Would Dr. Hillier say something on the subject of vacuum-leak tracing?

Dr. J. Hillier, RCA Laboratories Division, Princeton, N. J.: We don't have many leaks. When we do, we use a mass-spectrometer leak-tester. This instrument is used by the whole laboratory and is considered a very worth while investment. New pieces of equipment are always tested on it. Starting out with a leak-free system is a big part of the battle.

Mr. S. Newberry, General Electric Company: The differential absorption method of leak tracing is of considerable interest. The method is not as sensitive as the mass spectrometer, but it is much simpler. Two gages are used and connected in a balanced bridge circuit. One of the gages is connected to the system via a trap containing a material that absorbs the tracer gas. For example, carbon dioxide may be used with sodium hydroxide as absorbent. We have found this system very effective. I believe every electron microscope should have a built-in leak detector. It should be possible to achieve this for $100.00.

Mr. Haine: One reason why the differential method is so effective is that fluctuations in background pressure cancel out on the two gages. It is not impossible to fit a leak-tester that is quite effective and costs little more than $100.00. We use such devices on all our experimental equipment. The arrangement used consists of a Pirani gage on the high vacuum side of the diffusion pump. The gage is connected to a simple bridge circuit with an out-of-balance meter calibrated in microns pressure and a jack for insertion of a galvanometer. Also fitted on the high vacuum side of the diffusion pump is a simple but effective flap valve that can be closed to shut off the pump. This arrangement allows us to measure the leak rate of the system in a fraction of a minute by shutting the valve and noting the rate of rise of pressure. Leaks are located with the valve open and by the use of ether as a probe vapor. The leaking of ether into the system causes a deflection of the bridge galvanometer. We do not

fully understand the reason for these deflections that are much larger than can be explained on simple grounds. We think they may be due to the heat of dissociation of the hydrocarbon vapor on the hot Pirani filament. The method is very sensitive and enables us to find leaks down to 0.001 liter-microns/second in systems of not too large a volume.

Mr. R. G. Picard, RCA Victor Division, Murray Hill, N. J.: We have a hydrogen sensitive ionization gage. It does not cost less than $100.00, but it covers a range down very near to that that the mass spectrometer will cover. We have built this into an instrument.

Mr. Haine: I have heard that the palladium diffusion tube in this device tends to get clogged up in time. Is this true?

Mr. Picard: It is certainly possible to choke them with hydrogen. An appropriate circuit arrangement cutting down the palladium temperature when hydrogen is detected largely prevents this.

Dr. L. Marton, National Bureau of Standards, Washington, D. C.: What is the limiting sensitivity of the differential method you described?

Mr. Haine: The differential gage method is capable of detecting leaks down to 10^{-3} or 10^{-4} liter microns per second.

Dr. Marton: Returning to the electron optical bench. Is it possible to introduce objects or apertures and move these around?

Dr. E. Ruska: Yes, such movements can be made by means of translating screws.

Mr. Haine: Do you have electrostatic lenses, and how is the gun insulated?

Dr. Ruska: Yes, electrostatic lenses are incorporated. The gun is insulated by a Plexiglass insulator. The filament is centerable by lateral adjusting screws, so that no other arrangement is necessary to aline the beam on the optical axis.

Mr. Haine: It is probably rather easier to aline the beam by moving the anode plate. The movement of this bends the beam because of the negative lens at the anode. The correction applied this way is independent of voltage.

Dr. Hillier: Do you do this on a bias type gun?

Mr. Haine: Yes.

Dr. J. A. Reisner, RCA Victor Division: Some years ago Dr. Piccard and I published a description of an optical bench in the form of a large open box 24 in. long, 10 in. wide, and 10 in. deep, using a semiautomatic revolving system. We had one advantage in that the use of magnetic lenses required only the use of low voltage connections into the vacuum. We found that in demonstrating optical systems with magnetic lenses we were able to move the lens transversely and axially by means of controls operating through gasket seals.

At that time we did some experiments on a bias type of electron gun using transverse motion of all three of the elements. We obtained results similar to those described.

Dr. Marton: The optical bench that I described in a slide this morning is of the same type Dr. Reisner has described and includes some further refinements.

Dr. E. W. Müller, Pennsylvania State College, State College, Pa.: The tubes in my point projector microscope are pumped with a mercury-in-glass pump and liquid air traps. For adsorption experi-

ments, where a pressure of 10^{-8} mm is sufficient, it is not necessary to seal off as this pressure is obtained on the pumps. At this pressure an adsorbed monolayer is formed in about a second. If a lower pressure is desired, say 10^{-9} or 10^{-10} mm, the tube is sealed off after thorough baking, of course. The vacuum may be further improved by evaporating a tungsten wire. This will give an estimated pressure of 10^{-12} mm. I believe Dr. Gabor has had the same or better experience.

Dr. Newberry: Do you use double liquid air traps?

Dr. Müller: Yes. One trap is baked; the other is not.

Mr. Haine: Have you tried to pump a tube with no traps?

Dr. Müller: No. That would seem quite impossible.

Mr. Haine: I suggest this is not so. I believe the use of liquid air traps is over-rated and even has disadvantages. It is true that without the trap there is a high pressure of mercury in the tube being pumped. However, at the bake out temperature the mercury vapor cannot condense. After sealing off and cooling, the small amount of mercury vapor in the tube will condense, forming a small fraction of a monolayer on the walls of the tube. We have compared two tubes pumped with and without liquid air and find no difference in the ultimate vacuum obtained.

45. Magnetic Lenses of Extremely Short Focal Length

By J. B. LePoole [1] and A. C. van Dorsten [2]

Abstract

A magnetic electron lens having a very high field strength in the gap and a bore appreciably smaller than the gap between the pole pieces, presents interesting possibilities for the construction of objective lenses for electron microscopy of very small focal length. Such a lens, owing to its relatively low chromatic aberration, lessens the requirements on voltage and current stabilization considerably.

The "wings" of the field curve of the lens, which now has become a thick lens of the immersion type, have the effect of acting as an undesired condensor lens in front of the object, at the object side, and as an extra lens in the region beyond the point of intersection of the electron paths with the axis at the image side of the lens. This latter effect reduces the strength of the lens and sets a limit to the minimum focal length obtainable. By reducing the bore in the pole pieces as far as is possible without interfering with the passing of the beam, this effect can be reduced almost to zero. The field along the axis, between the pole pieces, is nearly homogeneous.

A lens is described having a bore/gap ratio of 1:4 and a focal length of under 0.8 mm at 80 kv accelerating voltage.

Reduction of the image side "wing" of the field curve can also be achieved by providing the image side pole piece with a cylindrical insert of magnetic material of high permeability, acting as a magnetic shielding at this part of the lens. Movement of this insert in axial direction presents a continuously variable parameter for experiments.

[1] Technical University, Delft, Holland.
[2] Philips Research Laboratories, Eindhoven, Holland.

46. Correction of Field Measurements in Electron Lenses

By A. Bruaux [1]

Magnetic lenses can be investigated by a method identical to Le Poole's.[2] The experimental setup used here is slightly different, however. The axis of the lens is vertical. The lens itself can be displaced up and down so as to enable us to measure the field at every point along the axis.

The probe consists of a helix of a thin wire suspended at a balance; the axes of both the lens and the probe are brought into coincidence; the sign of the helix changes at a particular location. As can easily be seen, if a given current flows through the wire, the force acting on the probe is proportional to the magnetic flux through the section of the helix. Thus, what is actually recorded is the distribution of flux along the axis of the lens. The problem is then to deduce from that information the distribution of field along the axis.

A mathematical method using matrix calculus has been applied to carry out the correction on account of the finite area of the winding cross section. Thus, given distribution Φ_{r_0} (r_0 is the mean radius of the winding), one has to compute $H(z)$ related to the latter by the expansion

$$\frac{\Phi_{r_0}}{\pi r_0^2}=\sum_{n=0}^{\infty}(-1)^n\frac{1}{n!n+1!}\left(\frac{r_0}{2}\right)^{2n}\cdot\frac{d^{2n}H(z)}{dz^{2n}}. \tag{1}$$

The first step consists in writing down $\Phi_{r_0}(z)$ as a Gram. Charlier series, i. e., an expansion using Hermite functions, which form an orthonormal set.

In this process, two arbitrary constants intervene, which are: (1) the location of the origin, and (2) the parameter a, half the distance between the two inflection points of the fundamental function ψ_0. The origin is so chosen as to cancel the integral $\int\phi\ (z)z\ dz$; a is so chosen as to give the least quadratic error for this function with respect to the given curve. This choice has a simple interpretation: it means that the coefficient of the second-order function ψ_2 vanishes. The calculation of the coefficients of the expansion is performed by numerical integration using the formula

$$C_k=\int_{-\infty}^{\infty}\phi_{r_0}(z)\psi_k(z)dz.$$

[1] Laboratoire de Recherches Physiques, Ateliers de Construction Electriques de Charleroi, Charleroi, Belgium.

[2] M. V. Ments and J. B. Le Poole, Appl. Sci. Research [B] **1**, 3–17 (1947–49).

In practice, the calculation of more than four coefficients is seldom necessary; furthermore, Parseval's theorem gives a convenient means of deciding whether the number of coefficients is sufficient by taking into account the experimental errors.

Thus $\Phi_{r_0}(z)$ can be represented in a Hilbert space by a vector, the components of which are the set of coefficients, C_k. In such a space, derivation of a vector H with respect to z yields another vector, which is the result of the multiplication of a hermitian matrix by the initial vector. Now, an n-fold derivation consists in the multiplication of the nth power of the matrix $\left[\frac{d}{dz}\right]$ by the vector $\overline{H}$. Actually, even powers of $\left[\frac{d}{dz}\right]$ only are needed; it is sufficient, therefore, to calculate the powers of the matrix $\left[\frac{d^2}{dz^2}\right]$. Furthermore, because of the additive property of matrices, the whole second member of eq (1) can be expressed in the form

$$\frac{\Phi_{r_0}}{\pi r_0^2}=[\Lambda]\overline{H}.$$

In this relation Φ_{r_0} (that is, its components) is known, whereas $\overline{H}$ is not; hence, if both members are multiplied from the left by the reciprocal matrix of $[\Lambda]$ the problem is solved.

Direct calculation of $[\Lambda]^{-1}$ is difficult, however. Fortunately, $[\Lambda]$ has the form

$$[\Lambda]=[\lambda_0]+\epsilon[\lambda_1]+\epsilon^2[\lambda_2]+\ \ldots,$$

where $\epsilon=-r_0^2/2$ is usually small. This enables us to calculate $[\Lambda]^{-1}$ with any desired precision by a process somewhat similar to a diagonalization of Heisenberg matrices. Finally, the result has the symbolic form

$$[\Lambda^{-1}]\ \frac{r_0}{\pi r_0^2}=\overline{H}.$$

This equation means that every component of $\overline{H}$ is expressed by a linear combination between the C_k.

From the viewpoint of the electron optics, it would be desirable to develop a similar procedure for solving the paraxial rays equation, in order to take full advantage of the high degree of accuracy. Finally, it seems that other problems could be approached by the same method.

47. New Interference Phenomena in the Electron Microscopic Images of Plate-Like Crystals

By James Hillier [1]

The anomalous intensity distributions that appear in the electron microscope images of small crystal have been often recognized as diffraction phenomena.[2] In the case of thin slightly curved plate-like crystals lying approximately normal to the electron beam the phenomenon appears as in figure 47.1. The dark lines traversing the image of the crystal are extinction contours that coincide with those areas of the crystal in which a set of planes is at the Bragg angle. Since this micrograph was made with no objective diaphragm, the bright reflections corresponding to some of the extinction contours are visible near the crystal. The displacement is due to the fact that the image was taken with the objective focal length somewhat greater than the focused value. More recently, late in 1950, one and possibly two new interference phenomena were observed almost simultaneously in three widely separated laboratories including that of the author.[3] [4]

The more obvious of the observed phenomena is illustrated in figure 47.2. This can be seen to consist of a periodic disturbance of the electron intensity along the extinction contours such that they appear to be traversed by a series of parallel and uniformly spaced fringes. Frequently two superimposed sets of fringes intersecting at 120° are observed. In these cases the spacing of the fringes is not necessarily the same in the two sets. A third, and rather rare example of this phenomenon is indicated in figure 47.2 by the arrows. In this, two sets of superimposed parallel fringes with widely differing spacings are observed within a single extinction contour. This phenomenon is characterized by the high contrast encountered in the fringes, by relatively large spacing of the fringes (usually more than a hundred angstroms) and by the fact that the fringes are confined to the relatively limited area of the crystal defined by the extinction contour.

Mitsuishi, Nagasaki, and Uyeda propose an explanation of this phenomenon that appears quite reasonable. They suggest that the fringes appear when two crystals are superimposed but not lying parallel to each other. It is then possible for the electrons that have been diffracted by the first crystal to be incident on the second crystal at the correct angle to undergo diffraction from a different set of planes. If the arrangement of the system is such that the wavefront representing the doubly diffracted ray leaves the system nearly

---

[1] RCA Laboratories Division, Princeton, N. J.

[2] J. Hillier and R. F. Baker, The observation of crystalline reflections in electron microscope images Phys. Rev. 61, 722–723 (1942).

[3] T. Mitsuishi, H. Nagasaki, and R. Uyeda, A new type of interference fringes observed in electron-micrographs of crystalline substance, Proc. Japan Acad. **27**, 86–87 (1951).

J. L. Farrant and A. L. G. Rees (personal communication).

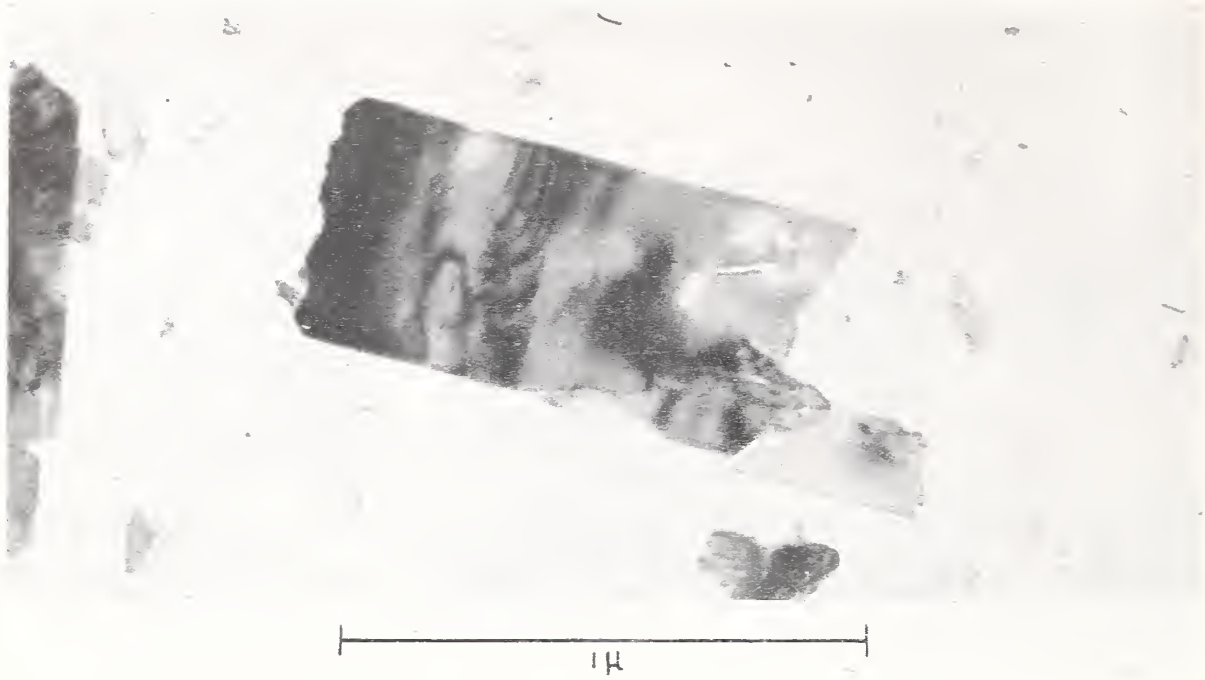


FIGURE 47.1. *Plate-like crystal of iron oxide showing a number of extinction contours and the bright reflections arising in them.*

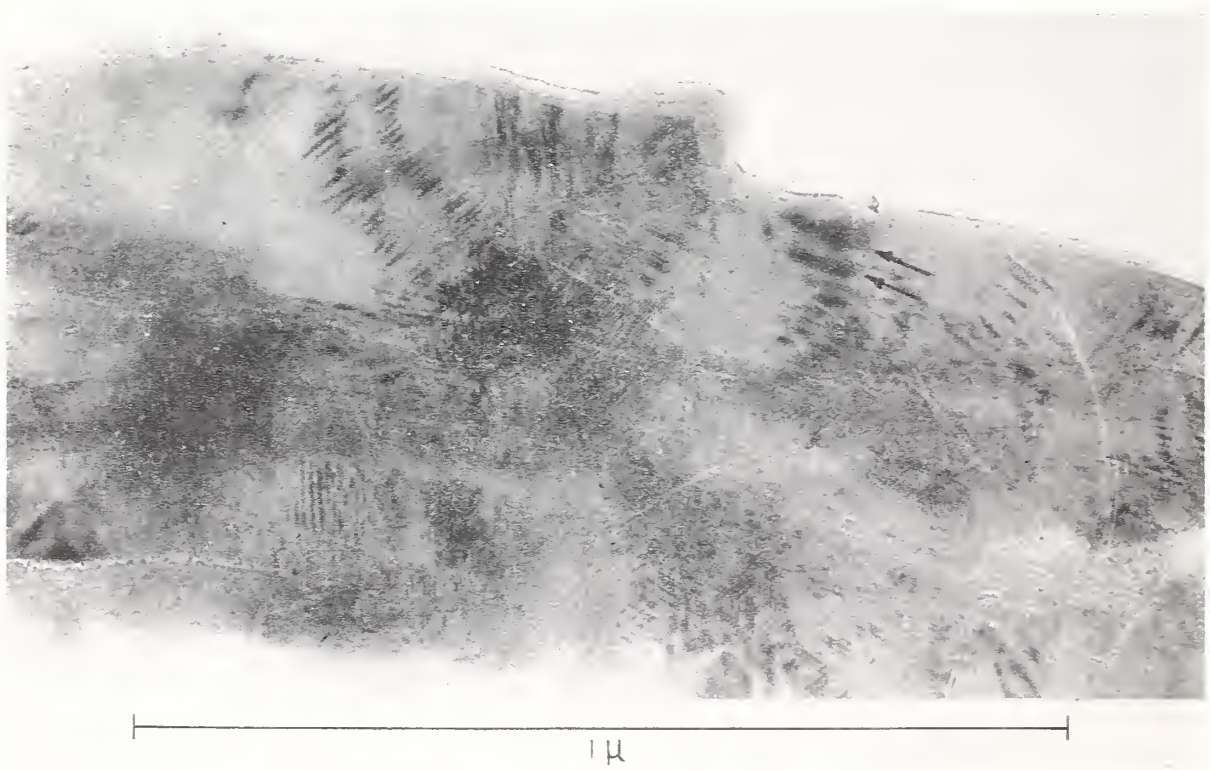


FIGURE 47.2. *Multilayer plate-like crystal showing extinction contours traversed by interference fringes.*

parallel to the wave-front representing those electrons that have passed through both crystals without being diffracted, it is possible for interference to occur. That the observed fringes occur only in areas similar to extinction contours and only when there are two or more superimposed crystals would tend to support this proposal.

The second phenomenon, which is much rarer than the above, is illustrated in figure 47.3. It can be seen to consist of a set of closely spaced fringes extending over very large areas of the crystals in which it appears. Usually, though not invariably, two superimposed sets

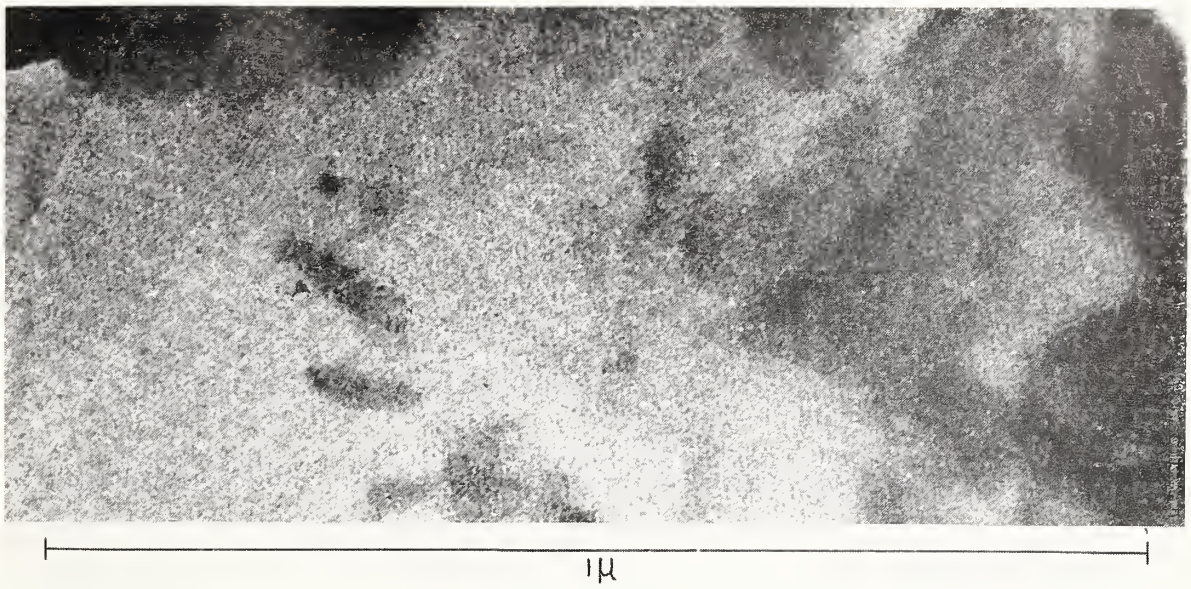


FIGURE 47.3. *Image of multilayer plate-like crystals showing uniform crosshatching over large area.*

Note that crosshatching disappears at boundary of one of the layers. The contrast in the pattern has been greatly enhanced by photographic processing.

of fringes crossing at 120° and with equal spacings are observed. In this laboratory the spacing of the fringes has varied considerably usually being less than 100 A with a minimum value of 40 A. A spacing of 25 A has been reported from another laboratory.[4]

This type of interference has been characterized by the fine spacing, by the uniformity of the fringes over much larger areas than the previous type and by the absence of extinction contours within the area of the crystals showing the effect. As in the previous case it has not been observed except where one or more crystals are superimposed.

It does not seem likely that the explanation proposed above will hold for this case. The main objection that can be raised concerns the physical requirements that must be imposed on the specimen. It can be estimated from the intensity relationships in the images of the crystals in which this phenomenon occurs that the crystals are sheets several microns in extent but of the order of only 100 A in thickness. If a Bragg reflection is to be obtained from a large area, such a crystal must be very flat. The usual mounting of the crystals in, or on, a thin, slightly stretched collodion membrane would favor holding the crystals flat and the frequent absence of extinction contours indicates that this is actually the case. On the other hand, these same conditions would not favor having two superimposed crystals, each nearly planar but with a small angle between them. A very rough calculation based on an estimated "rocking curve" for these crystals indicates that the minimum value for the radius of curvature of both crystals that would permit the above explanation is of the order of 1 mm. This seems unlikely in view of the thin flexible character of the crystals.

A possible explanation of the second type of interference is that it is "Moiré" pattern resulting from superimposing in *close* contact two flat crystals, one of which is slightly rotated relative to the other about a normal to its surface. This would result in a "magnified" image of the hexagonal arrays of atoms in the two crystals, the "magnification" varying inversely as the departure from parallelism of the two arrays. Scattering theory indicates that the contrast

would be very low in this phenomenon. This is in qualitative agreement with experiment.

An attempt was made to distinguish between the two proposed explanations for this second phenomenon by making successive micrographs of a crystal with slightly different angles of incidence of the illumination. The results were unexpected in that they led to the observation that the phenomenon rarely occurred in two successive micrographs even when the angle of incidence was not changed. On the basis of this observation it seems necessary to exclude the possibility that the second phenomenon is the result of a double Bragg reflection since the variation of the angle of incidence was certainly less than 10^{-5} rad while the total half-width of the rocking curve is greater than 10^{-3} rad. Unfortunately, this observation gives even less support to the Moiré pattern theiry. While it could be anticipated that a Moiré pattern produced by overlying crystals with a combined thickness considerably greater than the spacings of the pattern would be somewhat sensitive to changes in angle of incidence, the sensitivity indicated is several orders of magnitude too great. It is possible, of course, that the transitory nature of the second type of phenomenon is not due to instabilities in the angle of incidence of the illumination and that some other explanation should be sought.

Discussion

Mr. J. A. Simpson, National Bureau of Standards, Washington, D. C.: I was wondering if, for a possible explanation of the transitory nature of the patterns, Dr. Hillier had looked into the question of the localization of the fringes. It seems to me that with the angles he is talking about the fringes will be localized to an extreme degree and will fade out on the slightest drift of focus in the instrument.

Dr. Hillier: It does not seem to be a question of localization of fringes and sensitivity to focus. If the large area phenomenon, which is the sensitive case, can be considered as a Moiré pattern, the sensitivity to focus should be the same as for other parts of the specimen. On the other hand, if it is produced by two successive Bragg reflections in very flat crystals, the observed fringes are the result of interference between two or more series of plane waves with a small angle between them. For a fringe spacing of 100 A and 50 kv electrons two series of waves arising in a 1 μ crystal would interfere over a distance of at least 1 mm from the crystal if it were not for the smearing effect of the finite angular aperture of the illuminating beam. Since in our work the angular aperture of the illumination was of the order of 10^{-4} rad, the smearing effect would only become appreciable beyond 20 μ from the crystal. However, this is nearly two orders of magnitude greater than the instability in focus during the exposure time and more than an order of magnitude greater than the change in focus from one exposure to the next.

Mr. Simpson: Have you thought of the dynamical theory instead of the kinematical theory as providing an explanation?

Dr. Hillier: Not as yet, but it should be done. Unfortunately, it has not been possible to form even a rough qualitative idea of the results to which the dynamical treatment might lead.

U. S. GOVERNMENT PRINTING OFFICE: 1954