# PART I. STATIC ELECTRON OPTICS

CHAPTER 1

## THE VARIATIONAL EQUATION

### 1.1. Introduction

The first part of this monograph deals with 'static' electron optics, i.e. the study of the optical properties of beams of electrons, or other charged particles, which are in *steady* motion so that, although the electrons making up a beam are in motion along their individual trajectories, the appearance of the beam as a whole does not vary in time.

In constructing a theory of static electron optics, one must take into account some or all of the following factors:

(*a*) the corpuscular properties of electrons,
(*b*) the wave properties of electrons,
(*c*) long-range interaction, and
(*d*) radiation reaction and short-range interaction.

Although the final examination of the performance of an electron-optical instrument will probably entail consideration of two or more of these factors, the initial investigation and basic design are frequently carried out on the assumption that (*a*) is predominant. The major part of the design of an electron microscope would entail this approximation, although it would be impossible to calculate the resolving power without consideration of (*b*). Similarly, the design of a cathode-ray tube can be based largely on the same approximation, but a more exact estimate of the performance for high-beam currents would necessitate the introduction of (*c*). It will be seen that (*d*) may always be neglected. Indeed, it is only if the first factor is dominant that one may expect to obtain image formation; diffraction effects, high-current densities and, of course, appreciable variation of the field during the time of transit of electrons, all tend to mar an image.

The theory of static electron optics which takes account of (*a*) and, if necessary, (*c*), but necessarily neglects (*b*) and (*d*), is known as *geometrical electron optics*, since, as will be seen in Chapter 2, much of the theory may be conveniently expressed in terms of

geometrical concepts. Since the fields are static, no interest attaches to the motion of electrons with time but only to their spatial trajectories so that our study may be further restricted by the definition:

*Geometrical electron optics is the study of the spatial trajectories and of the associated image formation of electrons moving in static electromagnetic fields, radiation reaction and short-range forces being neglected and dynamical laws taken in their classical form.*

It is natural to regard the spatial trajectories as *electron rays*, and it will be shown in this chapter that, under the conditions laid down, electron rays satisfy a variational equation formally identical with Fermat's principle of light optics.‡

There are four observations concerning our definition which should be made at this point. The first is to explain that radiation reaction becomes important only in particle accelerators such as the betatron and electron-synchrotron where electrons of very high energies follow curved paths, and that short-range interactions, by which we mean forces arising from statistical fluctuation of the electromagnetic field due to 'granulation' of the beam,§ are never significant in electron-optical instruments, since the beams are not sufficiently dense.

The second point concerns the long-range interaction (*c*), by which we mean the Coulomb and Lorentz forces exerted by the beam, regarded now as a uniform fluid, upon its constituent elements. These are represented by the space-charge and space-current distributions, but if their effect is important the technical difficulty arises that the electron trajectories are determined by the electromagnetic field while the field is determined partly by the space charge which in turn depends on the trajectories. Hence although general rules, such as those to be established in Chapter 2, are implicitly valid even for high-density beams, explicit calculation of the effect of space charge will be considered only briefly in §4.5.

The third remark is that, although time itself is of no interest, it is essential to introduce time as independent variable if electrons

‡ The term 'light optics' is unfortunately both objectional and indispensable.

§ For an estimate of the forces due to 'granulation' see, for instance, *Microwave Electronics*, ed. G. B. Collins (McGraw-Hill, New York, 1948), pp. 221, 222. For a deeper analysis of the separation of Coulomb and Lorentz forces into 'short-range' and 'long-range' components, see D. Bohm and D. Pines, *Phys. Rev.* **82** (1951), 625–34; **85** (1952), 338–53.

are ever stationary along their paths, as they are in electron mirrors. It is not proposed to investigate electron mirrors‡ in this monograph, but, if it were, this investigation would fit more naturally into the second part, which deals with dynamic electron optics.

The fourth remark is that we have, by our definition, excluded from consideration certain calculations which are essential to the investigation of any electron-optical instrument, namely, the initial calculation of the electromagnetic field. However, the only general and practicable method for numerically calculating the fields of electron lenses appears to be the relaxation method,§ which it would be inappropriate to reproduce in these pages. Moreover, it is seldom that the fields are computed by numerical methods; they are more easily determined by means of an analogue computer such as the electrolytic tank‖ or the more accurate resistance network.‡ In some cases it is even possible to measure the field directly.

## 1.2. Electron-optical units

Since we shall be concerned with only one type of charged particle,‡‡ it will be convenient to introduce units of field potentials so chosen that no physical constants appear explicitly in our formulation of the variational equation or, consequently, in the subsequent theoretical considerations.

However, let us first notice that since the trajectory of an electron depends, among other things, upon its energy on entering the field, the optical properties of a field should always be referred to a given monokinetic or—in optical terminology—*monochromatic* beam, i.e. a beam whose electrons all have the same energy on crossing an arbitrary equipotential surface of the electric field. Any departure from this condition in an electron-optical instrument will result in image defects which are classified as *chromatic aberration*. Since we are neglecting the short-range interaction of electrons, it is possible to consider electrons of various energies independently

‡ The theory of electron mirrors was first developed by A. Recknagel (*Z. Phys.* **104** (1937), 381–94), but a modern treatment is to be found in ref. (1).

§ R. V. Southwell, *Relaxation Methods in Theoretical Physics* (Oxford University Press, 1946).

‖ See refs. (1) and (3).

‡ G. Liebmann, *Brit. J. Appl. Phys.* **1** (1950), 93–103.

‡‡ Except in § 5.5, where we deal with mass spectrographs.

so that chromatic aberration may be calculated within the framework of geometrical electron optics.

Let us now consider a monochromatic beam of electrons moving in an electric field whose scalar potential, measured in e.s.u., is $\phi^*$. We may take account of the initial energy of the electrons, as well as of the energy which they acquire in the field, by adjusting $\phi^*$ so that *electrons are at rest at zero potential.* If $-e$ is the charge of an electron, measured in e.s.u., and if $m_0$ and $m$ are the rest mass and relativistic mass, respectively,

$$e\phi^* = (m - m_0)\, c^2, \tag{1.2.1}$$

where $c$ is the velocity of light; $m_0$, $m$, $c$ and $v$, the speed of an electron, as measured in c.g.s. units.

If there were no electric field but only a magnetic field of strength $H^*$, measured in e.m.u., normal to the direction of motion of an electron of the beam, the radius of curvature $R$ of the electron trajectory could be found by equating the centrifugal force $mv^2/R$ to the Lorentz force $evH^*/c$. We see that $H^*R = p^*$, where $p^*$, defined by

$$p^* = mcv/e, \tag{1.2.2}$$

is a measure of the *scalar kinetic momentum.*

We shall see in the next section that the variational equation determining electron rays may be expressed in terms of the momentum $p^*$ and the magnetic vector potential $\mathbf{A}^*$ only. The constants $e$, $c$ and $m_0$ will therefore be eliminated from our calculations if they are eliminated from the relations between $p^*$ and $\phi^*$ and between $p^*$ and $H^*$. On using the well-known relation

$$m = m_0/\sqrt{(1 - (v/c)^2)}, \tag{1.2.3}$$

we find that the first of our relations becomes

$$p = \sqrt{(2\phi + \phi^2)}, \tag{1.2.4}$$

and the second retains the form $p = HR$ if we write

$$\phi = (e/m_0c^2)\,\phi^*, \quad p = (e/m_0c^2)\,p^*, \quad H = (e/m_0c^2)\,H^*. \tag{1.2.5}$$

Hence we should measure electric potential in units of $e/m_0c^2$ e.s.u. and magnetic field strength in units of $e/m_0c^2$ e.m.u.; thus:

*Unit of electric potential* = 511,200 volts,
*Unit of magnetic field strength* = 1,704 gauss.

Provided that related units are derived appropriately from these units, the usual relations between the electric and magnetic field

vectors and the electric scalar and magnetic vector potentials remain valid:

$$\mathbf{E} = -\operatorname{grad}\phi, \quad \mathbf{H} = \operatorname{curl}\mathbf{A}. \tag{1.2.6}$$

Electron momentum, expressed as a magnetic quantity, will be measured in the same units as the magnetic potential (i.e. in units of 1704 gauss cm.).

The above units will be used in all subsequent calculations, but it should be noted that *one may at any time return to e.s.u. and e.m.u. by means of the formulae* (1.2.5).

Since our calculations have been based upon relativistic mechanics, the relation (1.2.4) is *relativistically correct*. However, if the beam energy is small compared with our unit of 511,200 volts, (1.2.4) may be approximated by

$$p = \sqrt{(2\phi)}, \tag{1.2.4a}$$

which is its *non-relativistic* form. Calculations will generally be relativistically correct.

Let us now consider how we should take into account the existence of a steady space charge of density $\rho^*$, measured in e.s.u., and a steady space current of density $\mathbf{j}^*$, measured in e.m.u. If we write

$$\rho = (4\pi e/m_0c^2)\rho^*, \quad \mathbf{j} = (4\pi e/m_0c^2)\mathbf{j}^*, \tag{1.2.7}$$

the inhomogeneous field equations take the simple forms

$$\nabla^2\phi = -\rho \tag{1.2.8}$$

and

$$\operatorname{curl}\operatorname{curl}\mathbf{A} = \mathbf{j}. \tag{1.2.9}$$

It is therefore proposed that we adopt the following units:

*Unit of charge* $= 4{\cdot}524 \times 10^{-8}$ *coulombs*,
*Unit of current* $= 1356$ *amperes*.

The unit of charge density may be expressed alternatively as

*Unit of charge density* $= 2{\cdot}834 \times 10^{11}$ *electronic charges/c.c.*

It may be established that if a beam of electrons has, at any point, energy $\phi$, momentum $p$, space-charge density $\rho$ and space-current density $\mathbf{j}$, then

$$\mathbf{j} = \left(\frac{p}{1+\phi}\right)\rho\mathbf{l}, \tag{1.2.10}$$

where $\mathbf{l}$ is the unit vector in the direction of motion.

The most important application of electron optics is, of course, to systems involving beams of electrons, but it may also be applied, with only minor modifications, to problems concerning beams of protons or ions. Since these particles carry a charge of the opposite

sign to that of the electron, it is necessary to replace $\phi$ and $\mathbf{A}$ by $-\phi$ and $-\mathbf{A}$. The difference in specific charge—the ratio $e/m_0$—is reflected as the difference in the appropriate units; for instance, the units appropriate to proton beams are as follows: *unit of electric potential* $= 9{\cdot}391 \times 10^8$ *volts*; *unit of magnetic field strength* $= 3{\cdot}130 \times 10^6$ *gauss*; *unit of space charge* $= 8{\cdot}310 \times 10^{-5}$ *coulomb/c.c. or* $5{\cdot}187 \times 10^{14}$ *electronic charges/c.c.*; and *unit of current* $= 2{\cdot}491 \times 10^6$ *amperes*. An alternative is to make all measurements in e.s.u. and e.m.u. and to regard $\phi$, $p$, $\mathbf{H}$, $\rho$ and $\mathbf{j}$ as abbreviations for the expressions given in (1.2.5) and (1.2.7); this would be necessary in the study of mass spectrographs whose purpose is the determination of the specific charges of ions.

It is perhaps unnecessary to state that if one is considering the effect of the space charge of one beam of particles upon the optical properties of a second beam, the space charge and current of the first beam should be measured in units appropriate to the second, i.e. the 'focused', beam. However, it is worth noticing that the relation (1.2.10) would remain valid provided that one measures the energy and momentum of the first, i.e. the 'space charge', beam in units appropriate to that beam.

## 1.3. Derivation of the variational equation‡

We shall now proceed to derive from the principle of least action§ the variational equation which will be taken as the basis of geometrical electron optics.

If the velocity vector $\mathbf{v}$ has Cartesian components $v_r$, where $r = 1, 2, 3$, the principle of least action, as applied to a single electron, may be written as

$$\delta \int_A^B \mathbf{v} . \frac{\partial L}{\partial \mathbf{v}} \mathrm{d}t = 0, \tag{1.3.1}$$

where $L$ is the Lagrangian function, $t$ is time and the notation $\partial/\partial\mathbf{v}$ is adopted for the vector operator with components $\partial/\partial v_r$. The variational operator refers, in this case, to all variations of the trajectory which leave the terminal points $A$ and $B$ and the function $\mathbf{v}.(\partial L/\partial\mathbf{v}) - L$ invariant. It should be noted that (1.3.1) *and all*

‡ Since it would seem unreasonable to grace every variational formulation of the ray equations and of the equations of motion with the title of 'principle', such a formulation will normally be referred to as a *variational equation*.

§ See ref. (12), p. 207, eq. (65.9).

*subsequent equations involving the symbol $\delta$ are exact only to the first order in the increments due to $\delta$.*

If, for the purposes of this section, the unit of time is so chosen that the velocity of light is unity, the Lagrangian function for an electron‡ is

$$L = 1 - \sqrt{(1 - v^2)} + \phi - \mathbf{v}.\mathbf{A}. \tag{1.3.2}$$

The invariance under the variational operation of the function $\mathbf{v}.(\partial L/\partial \mathbf{v}) - L$ now leads to the invariance of the equation

$$1 + \phi = 1/\sqrt{(1 - v^2)}. \tag{1.3.3}$$

If $s$ measures arc length along the trajectory, $\mathrm{d}s = v\,\mathrm{d}t$, so that the time integral of (1.3.1) may be replaced by a line integral. If we now eliminate $v$ by means of (1.3.3), we obtain as the variational equation defining the rays of a given monochromatic electron beam in a static electromagnetic field

$$\delta \int_A^B \{\sqrt{(2\phi + \phi^2)} - \mathbf{l}.\mathbf{A}\}\,\mathrm{d}s = 0, \tag{1.3.4}$$

where $\mathbf{l}$ is a unit vector along the tangent to the trajectory and $\delta$ now refers to all variations of the path which leave the terminal points invariant.

Equation (1.3.4) will be adopted as the basis of our theory of geometrical electron optics.

It is interesting (though quite irrelevant) to consider the range of values of $\phi$ for which (1.3.4) is physically significant. It is clearly necessary, for practical electron optics, that $\phi \geqslant 0$. For small negative values of $\phi$, the radical becomes imaginary so that, in classical mechanics, $\phi = 0$ represents a lower impassable boundary.

However, let us suppose that by some process outside the scope of classical mechanics $\phi$ were depressed below $-2$. Let us write

$$\phi = \phi^* - 2, \tag{1.3.5}$$

and assume that $\phi^*$ is negative. When $\phi$ is positive the radical takes a positive value; when $\phi$ is negative let us give the radical a negative value. We then find that (1.3.4) may be written as

$$\delta \int_A^B \{\sqrt{(-2\phi^* + \phi^{*2})} + \mathbf{l}.\mathbf{A}\}\,\mathrm{d}s = 0. \tag{1.3.6}$$

If we now refer to § 2 and find the variational equation which applies to 'ions' whose specific charge is the same as that of

‡ See ref. (12), p. 349, eq. (99.6).

electrons, we obtain exactly the form (1.3.6). This shows that *when the energy of the electron is reduced by more than twice its rest energy, it behaves as a particle with the same mass but positive charge.* This is in agreement with Dirac's positron theory.‡

Let us now apply the equation (1.3.4) to the simple problem of slow electrons moving in a strong magnetic field. In this case we can neglect the first term of (1.3.4) so that it becomes

$$\delta \int_A^B \mathbf{A}.\mathrm{d}\mathbf{s} = 0. \tag{1.3.7}$$

Now the integral (1.3.7), taken around a closed circuit, gives the magnetic flux enclosed by the circuit, as is readily seen from (1.2.6) and application of Stokes's theorem. The equation (1.3.7) therefore states that the closed circuit formed by a ray and an arbitrary slight displacement of the ray embraces no magnetic flux. This is possible only if there is no component of field strength normal to the ray, so that slow electrons in a strong magnetic field must follow the lines of field strength.

If we take into account small but finite energy of the electrons, their paths will differ slightly from the lines of field strength. In any small neighbourhood their motion must resemble the motion of electrons in a uniform magnetic field. The latter, as we shall see later,§ is a helix so that *slow electrons in a strong magnetic field move in helices about the lines of field strength.*

The variational equation is expressed in (1.3.4) explicitly in terms of the potentials, but it will be convenient for many general discussions to write it in the shorter form

$$\delta \int_A^B \{p - \mathbf{l}.\mathbf{A}\}\,\mathrm{d}s = 0, \tag{1.3.8}$$

where $p$, the scalar momentum of the beam, is given by (1.2.4). We may observe, incidentally, that for a purely magnetic field, for which $p$ is a constant, it is not necessary to measure $p$ and $\mathbf{A}$ in electron-optical units although they must be measured in the same units; if the field is purely electric, the units may be left arbitrary only if the treatment is non-relativistic. We may also note that the *direction* of the ray enters only by way of the unit vector $\mathbf{l}$ and that

‡ P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press, 3rd ed. 1947), p. 272.

§ See pp. 19, 20.

reversal of $\mathbf{l}$ is equivalent to a change in sign of $\mathbf{A}$. It follows that *electron rays are reversible only if the field is purely electric*; if the field is partly or wholly magnetic, rays can be reversed only if the sense of the magnetic field is reversed.

Let us now write (1.3.8) in the form

$$\delta \int_A^B n \, \mathrm{d}s = 0, \tag{1.3.9}$$

where the function $n(\mathbf{x}, \mathbf{l})$ is defined by

$$n = p - \mathbf{l}.\mathbf{A}. \tag{1.3.10}$$

It is now obvious that electron rays possess optical properties, for (1.3.9) is formally identical with Fermat's principle of light optics. We shall, by analogy, call the quantity $n$ defined by (1.3.10) the *refractive index* of the field. It is to be remembered that, just as the refractive index of glass depends on the colour of the light beam, so the electron-optical refractive index depends implicitly upon the energy with which the electron beam enters the field.

It is well known that the electric scalar potential and the magnetic vector potential are to some extent arbitrary. We have made use of the indeterminacy of the former to combine the initial energy of the beam with the potential of the field, but the magnetic potential $\mathbf{A}$ is still arbitrary in that we may add to it the gradient of an arbitrary scalar distribution $\chi(\mathbf{x})$, say. Let us investigate briefly the consequences of this indeterminacy in $\mathbf{A}$.

If we change $\mathbf{A}$ to $\mathbf{A} + \operatorname{grad} \chi$, (1.3.8) becomes

$$\delta \int_A^B \{p - \mathbf{l}.\mathbf{A}\} \, \mathrm{d}s + \delta\chi_a - \delta\chi_b = 0, \tag{1.3.11}$$

where the suffix $a$ or $b$ will generally denote that a function is evaluated at $A$ or $B$, respectively. Since $\delta$ refers to variations for which $\delta\mathbf{x}_a = \delta\mathbf{x}_b = 0$, the integrated terms of (1.3.11) vanish. Hence *the variational equations formed from equivalent potential distributions are themselves equivalent.*

We see from (1.3.10) that, at any point, the refractive index is *isotropic* or *anisotropic*—i.e. it does not depend upon or depends upon the direction vector—according as the magnetic potential vanishes or does not vanish, respectively, at that point. It is obvious that, by replacing a distribution $\mathbf{A}$ by a distribution $\mathbf{A} + \operatorname{grad} \chi$, where $\chi$ is suitably chosen, *the refractive index may be made isotropic*

*at any finite number of points* so that the notion of isotropy *at a point* is without physical significance.

Let us now consider the refractive index over a surface (or finite number of surfaces). If we resolve the vector potential at each point of this surface into components $\mathbf{A}_n$ and $\mathbf{A}_t$ which are normal and tangential, respectively, to the surface at that point, there will clearly be no difficulty in eliminating $\mathbf{A}_n$ by adding to $\mathbf{A}$ the gradient of some potential distribution. However, it will not be possible simultaneously to eliminate $\mathbf{A}_t$ unless it is possible so to arrange the distribution of $\chi$ in the surface that $\mathbf{A}_t + \mathrm{grad}_t \chi$ vanishes, where $\mathrm{grad}_t \chi$ is the tangential component of $\mathrm{grad}\, \chi$. By considering integrals along arbitrary closed curves lying in the surface and applying Stokes's theorem, we find that the necessary and sufficient condition for this to be possible is that the normal component of $\mathrm{curl}\, \mathbf{A}$ vanishes. Hence *the refractive index may be made isotropic at all points of a given surface if and only if the normal component of the magnetic field strength is zero at all points of the surface.*

The necessary and sufficient condition that, given a vector potential distribution $\mathbf{A}$, we may find a scalar distribution $\chi$ such that $\mathbf{A} + \mathrm{grad}\, \chi = 0$ throughout a given volume is that $\mathrm{curl}\, \mathbf{A}$ vanishes throughout the volume. It follows that *the refractive index may be made isotropic over a given volume if and only if the magnetic field strength vanishes throughout the volume.* If we say that a *field* is isotropic at a point if it is possible to make the refractive index isotropic throughout a small volume containing the point and anisotropic otherwise, we see that *an isotropic field is one which is purely electric whereas a field which is partly or wholly magnetic is anisotropic.*

## 1.4. The electrostatic lens

Now that the variational equation has been established, it will be instructive to make a preliminary investigation of one or two typical electron-optical systems in order to determine how the equation may be applied to certain simple but important calculations, and in order to decide what further calculations we shall ultimately wish to make. In this section we shall establish the paraxial imaging properties of an electric field of rotational sym-