1

BASIC CONCEPTS

1.1 Charged particles, force and field

1.1.1 *The force between charged particles*

At a certain level of sophistication much of physics can be conceived in terms of elementary particles, such as the electron, proton and neutron, between which there are forces of interaction. These forces are classified as gravitational, electromagnetic and nuclear; and broad areas of physics can be distinguished by the type of force with which they are chiefly concerned. In this sense the immensely diverse phenomena of electromagnetism, ranging from the orientation of a compass needle to the reception of radiation from the remotest regions of the universe, are basically manifestations of the electromagnetic force.

The electromagnetic force of interaction that exists between certain elementary particles is regarded as a force between electric charges; electric charge is thus intrinsic to these particles. There is some analogy with the gravitational force between masses, and in particular the magnitude of the force between two charged particles is proportional to the product of the charges. However the characters of the two types of force are in general quite distinct. One obvious difference is in the reversibility of the direction of the electromagnetic force, and this is accounted for by attributing alternative signs to charge. The electron and proton have charges equal in magnitude but opposite in sign, and by convention the electron charge is negative: the neutron is uncharged. It may also be noted that the repulsive electromagnetic force between two protons at rest exceeds the attractive gravitational force between them by a factor of the order of 10^{36}; and the proton mass is 1836 times the electron mass.

In laboratory scale phenomena the electron, proton and neutron are usually the only fundamental particles that need be considered, because together they constitute overwhelmingly the greater part of charge and matter. Moreover, they are effectively immutable, so that charge, of each sign, and matter are conserved. Thus, for example, the amount of positive charge in some specified region of space is determined by the number of protons the region contains; and any increase in the amount, synonymous with an increase in the number of protons, can only arise from a corresponding influx of protons across the boundary of the region. The net charge in a region is, of course, represented by the excess of protons over electrons; and an increase in net charge can arise from an efflux of electrons as well as from an influx of protons.

1.1.2 *The electromagnetic field*

Although in principle the law of force between elementary charged particles can be taken as the ultimate physical basis of electromagnetic theory, it goes without saying that development of the theory to bring it into close contact with applications is required. This is the more important because the general force law is extremely complicated; the familiar Coulomb law, that the force between two particles is inversely proportional to the square of the distance between them and acts along the line joining them, only applies to the special case when the particles are at rest.

A major step in forwarding the development of the theory is the introduction of the concept of the electromagnetic field. Any specific charged particle interacts with other charged particles, and the force on it due to these other particles is the vector sum of the individual forces of interaction. It is convenient to ascribe this force to the fact that the 'test' particle, as it may be called, is in an electromagnetic *field* that arises from the presence of the other charged particles. This field at position $\mathbf{r}$ at time t is defined in terms of the force per unit charge on a test particle, located at $\mathbf{r}$ at time t, in the following way.

Consider, first, a test particle of charge e held at rest; then the electromagnetic force on it is written

$$e\mathbf{E}, \tag{1.1}$$

and this defines the electric field vector $\mathbf{E}$. Suppose, next, that the particle is moved with uniform velocity $\mathbf{v}$. Then it is a matter of experience that the electromagnetic force can be written

$$e(\mathbf{E}+\mathbf{v}\wedge\mathbf{B}), \tag{1.2}$$

where $\mathbf{B}$ is independent of $\mathbf{v}$; and the expression (1.2), known as the *Lorentz force*, defines the magnetic field vector $\mathbf{B}$, often called the magnetic induction.

There need be no hesitation in claiming that a charged particle at rest or in uniform rectilinear motion exerts no electromagnetic force on itself. The quantities $\mathbf{E}$ and $\mathbf{B}$, in general functions of position and time, can therefore be granted an existence independent of the test charge. Together they represent the electromagnetic field due to all the charges except the test charge.

If the velocity $\mathbf{v}$ of the test particle is not constant the expression (1.2) still represents the force on it to an adequate degree of approximation for nearly all purposes. More precisely, though, a contribution should be included from the field of the accelerated charged particle reacting on the particle itself; and this is no simple matter to investigate, since it involves the question of the 'structure' of the particle. However the contribution is usually very small, and the cases for which it is significant are outside the scope of this book, so there is no occasion to refer to it again.

Another point concerning particle structure is that in some sense the charge of the electron, say, should be regarded as 'spinning', so that the charge would be in circulatory motion even when the electron as a whole is at rest. Again, however, this feature, though fundamentally important, has a quite negligible effect in most circumstances. It is therefore convenient to maintain the concept of a charged particle free of spin, and this is understood throughout the book unless there is an explicit statement to the contrary.

The second term of (1.2) indicates that the force associated with the magnetic field is proportional to the speed and at right angles to the motion of the particle. These distinctive features are peculiar to electromagnetism, and play a major part in many characteristic phenomena. The predictions of (1.2) have been tested countless times in experiments on charged particle dynamics, some simple but important examples of which are indicated in problems 1.1 to 1.9.

That the force on a charged particle is the vector sum of the forces of interaction with other charged particles implies, of course, that the electromagnetic field $\mathbf{E}$, $\mathbf{B}$ is the vector superposition of the fields of individual charges. Consideration of the field of a single charged particle now follows. This is in effect a discussion of the law of force between a pair of particles based on a knowledge of the dependence of the force on the motion of one of the pair as expressed in (1.2).

1.1.3 *The field of a charged particle*

The inverse square law of force between charged particles at rest was established by direct methods (Coulomb 1785), and contemporaneously by indirect methods (Cavendish) which were subsequently developed to great accuracy; from an experiment in 1936 it was concluded that the index in a power law could not differ from -2 by more than one part in 10^9. In terms of the field concept the law can be stated as follows: the field of a particle of charge e at rest in an unbounded vacuum is purely electric, and the electric vector $\mathbf{E}$ is proportional to

$$e\mathbf{r}/r^3, \tag{1.3}$$

where $\mathbf{r}$ is the position vector of the field point from the particle.

There is no reason to doubt that, in the context of classical physics, this statement is in effect exact. The experiments just referred to involve familiar 'laboratory scale' measurements, but results from atomic and nuclear physics indicate that the law continues to hold for microscopic values of r, perhaps even down to 10^{-15} m.

For a charged particle in *motion* no such exact statement of comparable simplicity can be made about the field, which is both electric and magnetic. However in certain circumstances, the background to which will be clarified shortly, it can be said that, for a particle of charge e with velocity $\mathbf{v}$, it is approximately true that the additional field due to its motion is magnetic, and the magnetic field $\mathbf{B}$ is proportional to

$$e\mathbf{v}\wedge\mathbf{r}/r^3. \tag{1.4}$$

This vector product containing the velocity of the particle producing the field may be compared with that in (1.2) containing the velocity of the particle on which the field acts.

The result (1.4) cannot claim as direct an experimental basis as Coulomb's law, chiefly because, as will be seen, the magnetic force between charged particles is likely to be negligible compared with the electric force. It is introduced here mainly to give an immediate insight into the fundamental connection between electric and magnetic fields. The historical development of the theory of the magnetic field of moving charges followed experiments by Ampère, Biot and Savart (1820); these are taken up in due course in Chapter 2, where the reasoning that links them to the present statement will be found.

1.1.4 *Units*

The constant of proportionality by which the expressions (1.3) and (1.4) must be multiplied to give, respectively, the electric and magnetic field of a charged particle depend on the units adopted. In one conventional notation, used in this book, the formulae are written

$$\mathbf{E} = \frac{e}{4\pi\epsilon_0}\frac{\mathbf{r}}{r^3}, \tag{1.5}$$

$$\mathbf{B} = \frac{\mu_0 e}{4\pi}\frac{\mathbf{v}\wedge\mathbf{r}}{r^3}. \tag{1.6}$$

There is no incentive to debate whether or not to include the factors $1/(4\pi)$ explicitly; if they are omitted here, 4π factors appear elsewhere, and it is hardly more than a matter of taste where one prefers to see them. If they are retained in (1.5) and (1.6) the units are said to be *rationalized*.

On the other hand, what values and dimensions are to be ascribed to ϵ_0 and μ_0 does offer scope for argument, and has been widely debated.

In discussing this point it should first be noted that, since the ratio of the electric to the magnetic force on a particle is a dimensionless number, an implication of the combination of (1.2), (1.5) and (1.6) is that $(\epsilon_0\mu_0)^{-\frac{1}{2}}$ is a velocity (to be pedantic, a speed). There is, then, a fundamental velocity

$$c = (\epsilon_0\mu_0)^{-\frac{1}{2}} \tag{1.7}$$

built into electromagnetic theory. Development of the theory in fact shows that c is the velocity of propagation, in vacuum, of electromagnetic effects; briefly, it is the speed of light. The specification of one of ϵ_0, μ_0 therefore fixes the other through the experimental determination of c.

Of the available systems of units only two are currently in wide use; the mks (rationalized), and the gaussian (unrationalized). The basic distinction between them concerns the unit and dimensions of charge.

In the unrationalized gaussian system, which omits the $1/(4\pi)$ factors in (1.5) and (1.6), the unit of charge is fixed by specifying $\epsilon_0 = 1$; and since ϵ_0

is also assumed dimensionless, the force law saddles charge with dimensions (mass)$^{\frac{1}{2}}$ (length)$^{\frac{3}{2}}$ (time)$^{-1}$. With $\epsilon_0 = 1$ the implication of (1.7) is $\mu_0 = 1/c^2$. However it happens that in the gaussian system the second term of Lorentz force (1.2) is conventionally written $e\mathbf{v} \wedge \mathbf{B}/c$, giving $\mathbf{B}$ the same dimensions as $\mathbf{E}$; hence (1.5) and (1.6) appear as $\mathbf{E} = e\mathbf{r}/r^3$, $\mathbf{B} = e\mathbf{v} \wedge \mathbf{r}/(cr^3)$.

On the other hand, in the mks system charge is assigned dimensions independent of those of mass, length and time, so that the force law ascribes dimensions to both ϵ_0 and μ_0, those of μ_0 being (charge)$^{-2}$ mass length. Moreover the numerical value of μ_0 is determined by the choice of the amp for the unit of current, as described in § 1.3.2.

The rationalized mks system is adopted in this book, and it is worth emphasizing the simplicity with which any relation can be converted into the unrationalized gaussian system if required. From what has just been said, the basis of the conversion is merely the replacement of ϵ_0 by $1/(4\pi)$, of μ_0 by $4\pi/c^2$, and of $\mathbf{B}$ by $\mathbf{B}/c$. A few other consequential changes must be included, but only to cater explicitly for quantities not yet defined; the complete list is given in § A.3 of the Appendix.

The mechanical units in the gaussian system are the centimetre, gram and second (cgs). The unit of charge is therefore g$^{\frac{1}{2}}$ cm$^{\frac{3}{2}}$ sec^{-1}, called the electrostatic unit (esu). The main reason for introducing the mks system was to bring the units for electromagnetic quantities into line with those commonly used in applied calculations, the so-called practical units; specifically, the amp for current, the coulomb (amp sec) for charge, the volt for potential difference, and so on. It was noticed that this aim would be largely achieved if the mechanical units were chosen to be the metre, kilogram and second; hence the appellation mks.

The charge of an electron (or proton, or positron) is the smallest quantity of charge known to be recognizable as an individual element. The electrostatic unit of charge is a very much greater quantity, and the practical unit, the coulomb, much greater again, reflecting the fact that commonplace 'laboratory scale' electromagnetic phenomena involve enormous numbers of electrons and protons. In such a situation it is, of course, out of the question to consider keeping track of each particle; some 'averaging' or 'smoothing out' process must be part of the description, and this requires the concept of charge density, which will now be introduced.

1.2 Charge and current density

1.2.1 *Charge density*

The concept of charge density is entirely analogous to that of mass density. The charge density ρ at any point is the charge per unit volume at that point; that is to say, if $\rho'\,\delta\tau$ is the charge in an element of volume $\delta\tau$ that includes a fixed point P, then ρ at P is the limit of ρ' as $\delta\tau \to 0$ through the element shrinking to the point P.

What emerges if this mathematical definition is applied to the 'actual' physical situation? The charge density is zero everywhere except in the regions occupied by charged particles. Since the linear dimensions of the electron and proton, when described in classical terms, are estimated to be of the order of 10^{-15} m, these regions are tiny and comparatively isolated; moreover, their locations are continually changing.

To adopt the actual charge density is, of course, tantamount to accepting the impossible task of keeping track of each particle. A reasonable alternative procedure is to work in terms of a macroscopic density ρ, from which spatial variations on the scale of interparticle distances are removed by taking an average of the actual density over regions of space that are small on the macroscopic view, but whose linear dimensions are large compared with the particle spacing. This means that $\rho\,\delta\tau$ continues to give, with negligible error, the charge in a small volume element $\delta\tau$, provided the element contains a large number of charged particles.

It should be mentioned that there is a difficulty in calculating the average charge density that does not arise when considering mass density. The reason for the difference lies in the existence of charge of either sign. Commonly, of course, positive charge (protons) and negative charge (electrons) are closely knit; and if they are in the form of atoms whose net charge is zero it may well be that volume elements containing a large number of charged particles will simply contain a large number of neutral atoms. It would then appear that the average charge density is zero. However this is not necessarily the case, basically because the individual amounts of positive and negative charge involved are so enormous (see § 1.3.3) that their separation by distances small even on the scale of an atom gives a macroscopic effect. An adequate method of averaging has therefore to be quite sophisticated; it needs explicit consideration in Chapter 6, but not before.

It is because the difference between the fields associated with the actual and the macroscopic charge densities would not ordinarily be detectable that comparatively simple deductions from electromagnetic theory can claim to given an adequate account of electromagnetic phenomena. For example, the mathematical theory of electrostatics can safely proceed on the basis of a charge density assumed constant, even though the actual charge density must vary because of the thermal agitation of electrons and protons.

The advantage of a description of electromagnetic theory in terms of charge density rather than charged particles is thus clear. Fundamentally both descriptions carry the same information, but the former enables problems to be treated macroscopically by permitting the charge density to be idealized in a manner consistent with the other idealizations inherent in any macroscopic treatment. The field calculated from a macroscopic charge density is, of course, correspondingly an 'average' field; it is what would be found by a laboratory scale measurement.

1.2.2 *Current density*

The special cases in which, on average, the charge may be considered at rest are of comparatively limited interest. In general, charge is in motion, and it is the associated current that becomes the physical quantity of prime significance. This is the reason why the practical units are determined by a definition of the amp (see § 1.3.2) rather than the coulomb.

For the 'actual' physical situation, described in terms of elementary charged particles, each element of charge has a specific velocity, and the current density is defined as the product of charge density and velocity; in symbols $\mathbf{J} = \rho\mathbf{v}$.

Current on the atomic scale has no measurable effect unless there is some macroscopic ordering of the motion; in electrostatic phenomena, for example, no measurable magnetic field is produced, despite the thermal motion of individual electrons and protons. The macroscopic current density is given by averaging the actual microscopic current density, just as for charge.

For particles of a single species the contribution to the average current density is the product of the average charge density and the average velocity, the latter being simply the vector sum per particle of the individual velocities of the particles in a volume element that is macroscopically small but contains many particles. If, then, both electrons and protons, say, are involved, contributing macroscopic densities ρ_e and ρ_p, and having average velocities $\mathbf{v}_e$ and $\mathbf{v}_p$, respectively, the corresponding net current density is

$$\mathbf{J} = \rho_e\mathbf{v}_e + \rho_p\mathbf{v}_p. \tag{1.8}$$

It should perhaps be emphasized that in general $\mathbf{J}$ is not, of course, the product of $\rho = \rho_e + \rho_p$ with some average velocity. In particular it may well be the case that $\rho = 0$ but $\mathbf{J} \neq 0$; for example, current might be constituted solely from the mean motion of electrons ($\mathbf{v}_e \neq 0$, $\mathbf{v}_p = 0$) with the electron charge density annulled by protons ($\rho_p = -\rho_e$).

It is evident that, at any point P, $\mathbf{J} \,.\, \mathrm{d}\mathbf{S}$ is the rate at which charge crosses an infinitesimal plane surface element represented by the vector $\mathrm{d}\mathbf{S}$. This statement refers equally to the microscopic and macroscopic situations, the net rate of flow of charge being understood in the latter case.

Once the relation of the mathematical formalism to the physics has been grasped there is for the most part no need in the development of the theory to distinguish explicitly between actual and average densities. Their role in the theory will be based on the assertions that $\rho\,\mathrm{d}\tau$ gives the charge in an infinitesimal element of volume $\mathrm{d}\tau$, and $\mathbf{J} \,.\, \mathrm{d}\mathbf{S}$ gives the rate at which charge crosses an infinitesimal element of area $\mathrm{d}\mathbf{S}$. On this basis mathematical expression is now given to the fact, to which attention was drawn in § 1.1.1, that charge is conserved.

1.2.3 *The charge conservation relation*

Conservation of charge is synonymous with the fact that the rate of increase of charge within an arbitrary closed surface is equal to the net rate at which charge flows across the surface. That is,

$$\frac{d}{dt}\int_V \rho \, d\tau = -\int_S \mathbf{J}.d\mathbf{S}, \qquad (1.9)$$

where the volume V is bounded by the closed surface S, and $d\mathbf{S}$ is along the outward normal to the surface.

An important alternative form of the relation is obtained in the following way. Transform the right hand side to a volume integral by the divergence theorem (A. 18),

$$\int_S \mathbf{J}.d\mathbf{S} = \int_V \operatorname{div}\mathbf{J}\, d\tau,$$

and combine it with the left hand side after the time derivative has been taken underneath the integral sign. Thus

$$\int_V \left(\frac{\partial \rho}{\partial t} + \operatorname{div}\mathbf{J}\right) d\tau = 0,$$

where the partial time derivative of course signifies the rate of increase of ρ at a fixed point. Since the integral vanishes when taken over any region it follows that

$$\operatorname{div}\mathbf{J} + \dot{\rho} = 0, \qquad (1.10)$$

where the dot stands for $\partial/\partial t$; for if the left hand side of (1.10) were not everywhere zero there would be a region throughout which it was of one sign, and its integral over this region would not vanish.

Equation (1.10) is the differential relation between ρ and $\mathbf{J}$ equivalent to the integral relation (1.9), and gives a concise mathematical expression of the fact that charge is conserved.

Of course the wider statement that electrons and protons separately are conserved contains (1.10) by implication. In the notation of (1.8), conservation of electrons means

$$\operatorname{div}(\rho_e \mathbf{v}_e) + \dot{\rho}_e = 0; \qquad (1.11)$$

the corresponding equation for protons has suffix p replacing e, and the addition of the two equations gives (1.10).

1.3 Practical units and magnitudes

1.3.1 *The field of a steady rectilinear current*

Current flows readily in metals, therefore said to be good *conductors*, because the atomic structure is such that some of the electrons, though confined to the metal, are not bound to parent atoms and can move comparatively freely through the structure. Under the influence of a electric field these so-called *conduction* electrons acquire an average velocity in the direction of the field, and thereby constitute a macroscopic current density.

Many man-made electrical devices depend on the flow of current in wires. Often the value of the cross-sectional area of a current carrying wire is immaterial, and it is permissible to adopt the idealization of a *line* current. The concept of a line current implies an infinite current density, and its magnitude must of course be specified simply by the total current that flows along the line at each point on it.

The experimental convenience and importance of current carried by wires hardly needs stressing. The study of the magnetic fields of such currents, through the measurement of forces of interaction, played a vital part in the development of the theory and practice of electromagnetism. In particular, it happens that the practical electrical units stem from a definition of the unit of current based on the force between line currents. The simple facts necessary for an understanding of the definition are now given.

Consider an infinite straight line along which flows a constant, uniform current I. The line is supposed uncharged, so that the associated field is purely magnetostatic.

The field can be calculated from (1.6) by treating the current as a uniform line distribution of charged particles all moving with the same velocity. If N is the number of particles per unit length of line, v their velocity, and e the charge on each particle, the current is

$$I = Nev. \tag{1.12}$$

Now the application of (1.6) shows that, at any point P, the contribution to $\mathbf{B}$ from a charged particle at Q is at right angles to the plane containing P and the line current, and of magnitude

$$\frac{\mu_0 ev \sin\theta}{4\pi R^2}, \tag{1.13}$$

where R is the distance QP, and θ is the angle between QP and the line current. Since the direction of the contribution is independent of the position of Q on the line it follows at once by superposition, using (1.12), that the magnitude of $\mathbf{B}$ is

$$B = \frac{\mu_0 I}{4\pi}\int_{-\infty}^{\infty} \frac{\sin\theta}{R^2}\,d\xi, \tag{1.14}$$

where ξ is the distance of Q from the foot of the perpendicular from P to the line (see figure 1.1). From figure 1.1 it is evident that $R\,d\theta/d\xi = r/R$; hence

$$B = \frac{\mu_0 I}{4\pi r}\int_0^{\pi} \sin\theta\,d\theta = \frac{\mu_0 I}{2\pi r}. \tag{1.15}$$

The calculation therefore shows that at a point P distance r from a rectilinear line current I (of infinite extent) the magnetic field is inversely proportion to r, and is perpendicular to the plane containing P and the line

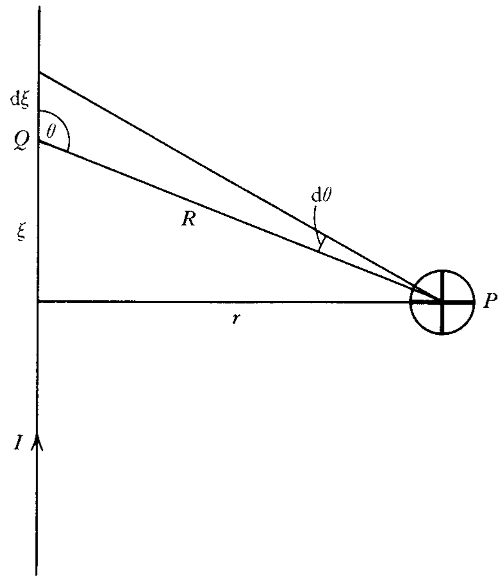

Figure 1.1

current in the sense of a right-handed screw about the current (into the paper, in figure 1.1, as indicated by the arrow tail ⊕ at P).

It should be emphasized that this is an exact result, despite the derivation being based on the approximate equation (1.6). In fact, equation (1.6) is inexact in so far as it implies instantaneity and thus fails to take account of the finite velocity c of electromagnetic effects. But for steady current flow the situation is time independent, and the error is immaterial; it cancels out in the superposition process. The application of (1.6) to steady currents is essentially equivalent to the so-called Biot–Savart law (see §2.2.4, and problem 2.17).

1.3.2 *The units of current and charge*

Suppose two rectilinear currents I and I' are parallel and distance r apart. The magnetic field of one acts on the moving charged particles that constitute the other, and by applying the force formula (1.2) in conjunction with (1.15) it is apparent that the force between the two currents is

$$\mu_0 II'/(2\pi r) \tag{1.16}$$

per unit length, being an attraction if the currents are in the same sense and a repulsion if they are in opposite senses.

The mks unit of current is the practical unit, namely the amp. It is specified in terms of this force of interaction in the following way.

Units based on fundamental formulae are called absolute, and the