Chapter 1
Introduction

Although all electromagnetic phenomena can be studied in empty space, an important part of any introductory course on electricity and magnetism is a proper understanding of the nature of matter. We shall therefore discuss dielectric behaviour in the chapter on electrostatics, conduction in metal wires in that on magnetostatics, and magnetism in matter (whether para-, dia- or ferromagnetism) in the chapter on magnetism. In this first chapter the nature of matter is summarised.

All matter is composed of elementary particles, some charged positively (protons), some charged negatively (electrons) and some without charge (neutrons). The forces between these particles are of three different sorts, – gravitational, electrical and nuclear – which differ enormously in their strength and range.

The *gravitational force* was made famous by Newton in his studies of the planets and expressed by him in 1665 in the inverse square law of force between two masses m_1 and m_2

$$F_G = \frac{G m_1 m_2}{r^2} \qquad [1.1]$$

where r is the distance between m_1 and m_2 and G is the gravitational constant. The *electrical force* will be familiar as the law Coulomb found in 1785 for the force between electrical charges. This is another inverse square law of force. If r is now the distance between the charges q_1 and q_2, and K is an electrical constant

$$F_E = \frac{K q_1 q_2}{r^2} \qquad [1.2]$$

The third type of force between the elementary particles that constitute matter is a comparatively recent discovery. In 1932 Chadwick found that the nuclei of atoms and molecules contained not only protons but new particles – neutrons – and so there had to be a third type of force, the *nuclear force*, that held these particles together in the nucleus.

This nuclear force, composed of both weak and strong interactions, is exceedingly short range, falling off as $r^{-2} \exp(-r/r_0)$, where r_0 is about 10^{-15} m. In contrast the gravitational and electrical forces are comparatively long range (Fig. 1.1). It is

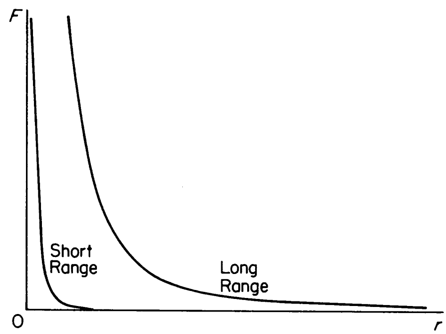

Fig. 1.1 Short-range and long-range forces.

obvious from the motion of the planets round the sun that gravitational forces are long range. It is not so obvious that electrical forces are similarly long range because electrical charges are usually screened by other charges of opposite sign at comparatively short range, so that the overall effect at long range is negligible.

Although both the gravitational and electrical force obey an inverse square law, their size differs enormously. For the proton-electron pair which comprises the hydrogen atom, the electrical force F_E is about 10^{39} or one thousand million million million million million million times as strong as the gravitational force F_G, as can be easily shown from equations [1.1] and [1.2] and a knowledge of the constants. So we can nearly always neglect gravitational effects in the presence of electrical forces. The exception is experiments like Millikan's oil drop, where the enormous mass of the earth acts on the oil drop with a force comparable to the electrical one exerted on the tiny charge of the oil drop as it moves between the charged plates.

The gravitational and electrical forces also differ in one other important respect: the gravitational force between particles of ordinary matter is always attractive, whereas the electrical force is repulsive (positive) between like charges and attractive (negative) between unlike charges. The net result is that large masses have large gravitational attractions for one another, but normally have negligible electrical forces between them.

The paradox is that although all matter is held together by electrical forces, of the interatomic or intermolecular or chemical-bond types, which are immensely strong forces, large objects are electrically neutral to a very high degree. The electrical balance between the number of protons and electrons is extraordinarily precise in all ordinary objects. To see how exact this balance is, Feynman has calculated that the repulsive force between two people standing at arm's length from each other who each had 1% more electrons than protons in their bodies would be enormous – enough in fact to repel a weight equal to that of the entire earth! So matter is electrically neutral because it has a perfect charge balance and this gives solids great stiffness and strength.

The study of electrical forces, electromagnetism, begins with Coulomb's law, equation [1.2]. All matter is held together by the electromagnetic interactions between atoms, between molecules and between cells, although the forces holding molecules and cells together are more complicated than the simple Coulomb interaction. The studies of condensed state physics, of chemistry and of biology are thus all dependent on an understanding of electromagnetism. This text develops the subject from Coulomb's law to Maxwell's equations, which summarise all the properties of the electromagnetic fields, in free space and matter. But if you ask why does the strong electrical attraction between a proton and an electron result in such comparatively large atoms rather than form a small electron-proton pair, you will not find the answer in Maxwell's equations alone. The study of electrical forces between particles at atomic or subatomic distances requires a new physics, quantum mechanics, which is the subject of the first book in the series.

Chapter 2
Electrostatics

Electric charge has been known since the Greeks first rubbed amber and noticed that it then attracted small objects. Little further progress was made until the eighteenth century when du Fay showed that there were two sorts of charge. One sort followed the rubbing of an amber rod with wool, the other a glass rod with silk. It was Benjamin Franklin who arbitrarily named the latter a positive charge and the original amber one a negative charge. He also showed that the total charge in a rubbing experiment was constant.

2.1 Coulomb's law

In 1785 Coulomb succeeded in discovering the fundamental law of electrostatics. A brilliant experimenter, he was able to invent and build a highly sensitive torsion balance with which he could measure precisely the relative force of repulsion between two light, insulating, pith balls when charged similarly and placed at different distances apart. He showed that this electrostatic force:

1. acts along the line joining the particles;
2. is proportional to the size of each charge; and
3. decreases inversely as the square of the distance apart.

It is therefore a long-range force (Fig. 1.1) and is given by the vector equation for the force $\mathbf{F}_1$ on charge q_1 due to charge q_2:

$$\mathbf{F}_1 = \frac{K q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} \qquad [2.1]$$

where $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$, $r_{12}^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$

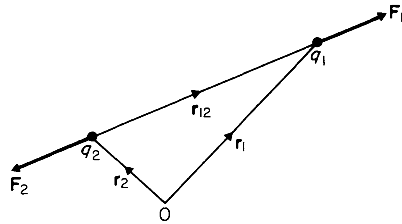

Fig. 2.1 Electrostatic forces between electric charges.

and $\hat{\mathbf{r}}_{12}$ is a unit vector drawn *to 1 from 2* (Fig. 2.1) given by $\mathbf{r}_{12}/r_{12}$. Fig. 2.1 shows that Newton's laws must apply and the force $\mathbf{F}_2$ on charge q_2 due to q_1 is $\mathbf{F}_2 = -\mathbf{F}_1$. When both charges have the same sign, the force acts positively, that is the charges are repelled, while between a negative and a positive charge the force acts negatively and the charges are attracted.

For historical reasons the constant of proportionality K in equation [2.1] is not one, but is defined as:

$$K = \frac{1}{4\pi\epsilon_0} = 10^{-7}c^2 \qquad [2.2]$$

where ϵ_0 is the *electric constant* (permittivity of free space) and c is the velocity of light. The constant has to be determined from experiment. A recent value of $c = 2.997925 \times 10^8 \mathrm{m\,s^{-1}}$ is accurate to better than 1 in 10^6, but for use in problems can be taken as $3.0 \times 10^8 \mathrm{m\,s^{-1}}$. On the same basis $K = 9 \times 10^9 \mathrm{N\,m^2\,C^{-2}}$, using the SI unit coulomb (C) for electric charge.

It is important to note that we have written in equation [2.1] Coulomb's law for charges in a *vacuum*; we have not mentioned the effects of a dielectric or other medium.

Principle of superposition

The only other basic law in electrostatics is the principle of superposition of electric forces. The principle states that if more than one force acts on a charge, then all the forces on that charge can be added vectorially into a single force. Thus for the total force on a charge q_1 due to charges q_2 at $\mathbf{r}_{12}$, q_3 at $\mathbf{r}_{13}$, etc., we have:

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} + \ldots$$

$$\mathbf{F}_1 = K\left\{\frac{q_1 q_2}{r_{12}^2}\,\hat{\mathbf{r}}_{12} + \frac{q_1 q_3}{r_{13}^2}\,\hat{\mathbf{r}}_{13} + \ldots\right\}$$

$$\mathbf{F}_1 = \frac{1}{4\pi\epsilon_0}\sum_j \frac{q_1 q_j}{r_{1j}^2}\,\hat{\mathbf{r}}_{1j}. \qquad [2.3]$$

That the electric force between two small particles can be enormous is readily seen by estimating the force produced in Rutherford's scattering experiment when an alpha particle ($^4_2$He nucleus) makes a direct approach to a gold ($^{197}_{79}$Au) nucleus. The distance of closest approach is 2×10^{-14} m and so the maximum electrostatic repulsion is, from equations [2.1] and [2.2]:

$$F = \frac{2e \times 79e}{4\pi\epsilon_0(2 \times 10^{-14})^2} = \frac{9 \times 10^9 \times 2 \times 79e^2}{4 \times 10^{-28}}\ \text{N}.$$

Since the charge on the proton, $e = 1.6 \times 10^{-19}$ C, the force on a single nucleus is about 100 newtons and a very strong force.

Electric field

The electric forces due to a distribution of electric charges, and particularly those due to a uniform distribution of charge, are best described in terms of an electric field vector, **E**, defined as the electric force per unit charge at a point. It can be visualised as the total force on a positive test charge, q at r_1, which is then allowed to become vanishingly small so as not to disturb the electric field. Using equation [2.3] we have:

$$\mathbf{E}(1) = \lim_{q\to 0} \frac{\mathbf{F}_1}{q} = \frac{1}{4\pi\epsilon_0}\sum_j \frac{q_j}{r_{1j}^2}\,\hat{\mathbf{r}}_{1j}. \qquad [2.4]$$

This vector equation is a shorthand version of three much longer equations, which are nevertheless needed when a particular case has to be worked out.

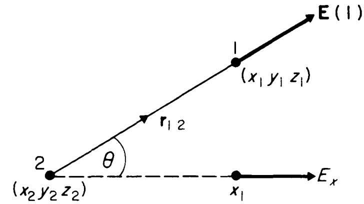

Fig. 2.2 The x-component E_x of electric field vector **E**.

For each coordinate plane there is a component of **E**(1) such as the x-component $E_x = E \cos\theta$ shown in Fig. 2.2. These components are therefore, for a charge q_2 at $(x_2 y_2 z_2)$:

$$E_x(x_1 y_1 z_1) = \frac{q_2}{4\pi\epsilon_0} \frac{(x_1 - x_2)}{\{(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2\}^{3/2}},$$

$$E_y(x_1 y_1 z_1) = \frac{q_2}{4\pi\epsilon_0} \frac{(y_1 - y_2)}{\{(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2\}^{3/2}},$$

$$E_z(x_1 y_1 z_1) = \frac{q_2}{4\pi\epsilon_0} \frac{(z_1 - z_2)}{\{(x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2\}^{3/2}}.$$

Just writing out equation [2.4] in this way for Cartesian coordinates shows how useful vector equations are in saving time and space in print.

In a similar way we can write a vector equation for a charge distribution, using the notation shown in Fig. 2.3, where

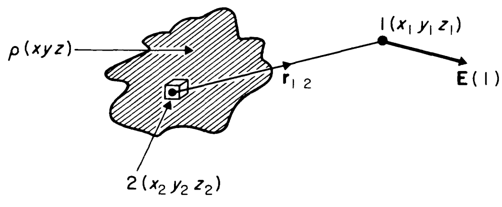

Fig. 2.3 The electric field **E**(1) at point 1 due to a distribution of charge.

$\rho(x, y, z)$ is the charge density, which produces a charge $\rho d\tau$ in a small volume $d\tau$. From equation [2.4] the electric field at point 1 is now:

$$\mathbf{E}(1) = \frac{1}{4\pi\epsilon_0} \int_{\substack{\text{all}\\ \text{space}}} \frac{\rho(2)\,\mathrm{d}\tau_2}{r_{12}^2} \hat{\mathbf{r}}_{12} \qquad [2.5]$$

where r_2 is the variable and the integral $\int \mathrm{d}\tau$ stands for $\iiint \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ in Cartesian coordinates and similar triple integrals for other co-ordinate systems. To apply this equation one must again evaluate each component, for example:

$$E_y(x_1 y_1 z_1) = \frac{1}{4\pi\epsilon_0} \int_{\substack{\text{all}\\ \text{space}}} \frac{(y_1 - y_2)\,\rho(x_2 y_2 z_2)\,\mathrm{d}x_2\,\mathrm{d}y_2\,\mathrm{d}z_2}{\{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\}^{3/2}}.$$

The equations [2.4] and [2.5] show, in principle, how all electrostatic fields can be obtained. Until the charges move there is no more to electricity: it is just Coulomb's law and the principle of superposition. In practice there are some clever tricks to avoid such horrible calculations that are only fit for computers. Remember too that, whatever happens, electric charge is always conserved *in toto*, since it depends ultimately on the stability of the electrons and protons in the universe. (Recent theories of elementary particles and of cosmology imply that the proton is not absolutely stable, but has a half-life $\sim 10^{31}$ years.)

2.2 Gauss's law

Gauss's law is about electric flux. The idea of the flux of a vector field arises from the flow of a fluid. We need a measure of the field lines (lines of force) coming out of a surface. We know that if we tilt the surface it has a maximum 'flow' when it is normal to the field lines and minimum (zero) when it is parallel to them (Fig. 2.4). If we describe the size of the surface by $\mathrm{d}S$ and its orientation by its unit normal vector $\hat{\mathbf{n}}$, then the flux of the vector $\mathbf{E}$ from the element $\mathrm{d}S$ is defined as $\mathbf{E}.\hat{\mathbf{n}}\mathrm{d}S$. This is commonly abbreviated to $\mathbf{E}.\mathbf{dS}$, where $\mathbf{dS} = \hat{\mathbf{n}}\mathrm{d}S$ is a vector along the outward normal for outgoing flux. For a point charge q at the origin using equation [2.4] we have therefore:

$$\mathbf{E}.\mathbf{dS} = \frac{q}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}}.\mathbf{dS}. \qquad [2.6]$$

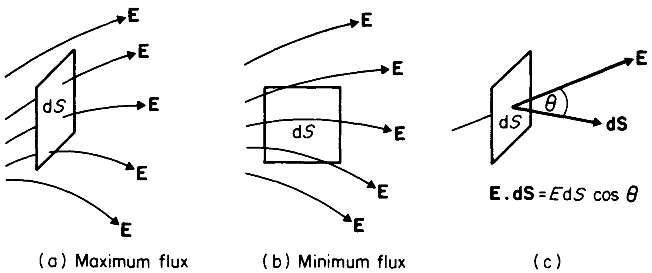

Fig. 2.4 Electric flux through various surfaces: (a) maximum for dS normal to **E**; (b) minimum for dS in plane of **E**; (c) flux of **E** is scalar product **E.dS**.

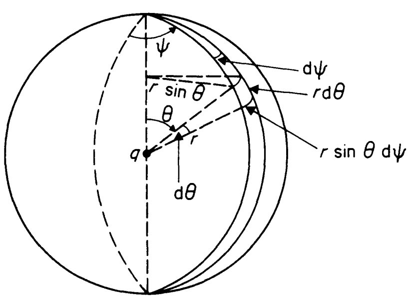

Fig. 2.5 Elementary surface area on a sphere around a point charge q.

The simplest way to evaluate this vector equation is to draw a sphere of radius r round the point charge and use spherical polar coordinates (r, θ, ψ), as in Fig. 2.5. The elementary area $\mathrm{d}S = r\sin\theta\,\mathrm{d}\psi . r\mathrm{d}\theta$ and so the flux due to q through dS is:

$$\frac{q}{4\pi\epsilon_0 r^2}\,\hat{\mathbf{r}}.\mathbf{dS} = \frac{q}{4\pi\epsilon_0}\sin\theta\,\mathrm{d}\psi\,\mathrm{d}\theta$$

since $\hat{\mathbf{r}}$ and **dS** are parallel. Thus for an inverse square law of force, the flux from a point charge is independent of the distance r of the sphere from the point charge. The total flux through the sphere is:

$$\int_S \mathbf{E}.\mathbf{dS} = \int_S \frac{q}{4\pi\epsilon_0}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\psi$$

where the integral is taken over the surface S of the sphere. This is easily evaluated:

$$\int_S \mathbf{E}.\mathbf{dS} = \frac{q}{4\pi\epsilon_0}\int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\psi = \frac{q}{\epsilon_0}. \quad [2.7]$$

Of course a sphere is a particularly symmetrical surface to have chosen to find the total electric flux. Does it have to be so symmetrical to get such a simple answer? To find out we must evaluate equation [2.6] for a surface of arbitrary shape (Fig. 2.7) and this is best done by using the concept of a solid angle dΩ (Fig. 2.6). For any surface area dS whose normal $\mathbf{dS}$ makes an

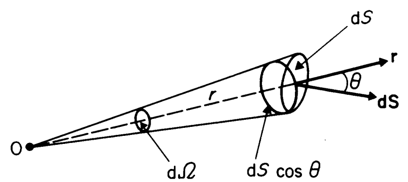

Fig. 2.6 Solid angle dΩ for cone of base dS.

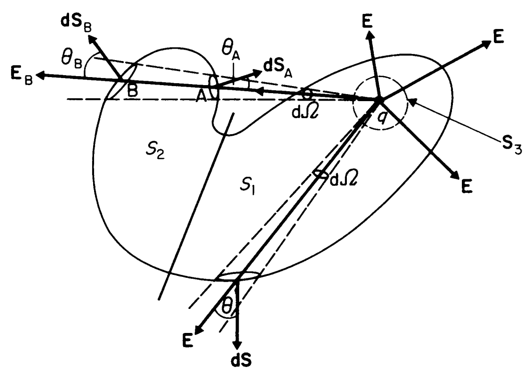

Fig. 2.7 Arbitrary surface around a point charge q.

angle θ with the radius vector $\mathbf{r}$ from an arbitrary point O, the solid angle dΩ is dΩ = dS cos θ/r^2. We can therefore write equation [2.6] as

$$\mathbf{E}.\mathrm{d}\mathbf{S} = q\,\mathrm{d}\Omega/4\pi\epsilon_0$$

and use this expression for the electric flux through $\mathrm{d}S$ to evaluate the total flux through the arbitrary surface S.

We first divide S into two parts: S_1 that does enclose q; and S_2 that doesn't. Then for S_1 we must compute the $\int_{S_1} \mathrm{d}\Omega$. Since q is a point charge, any surface S_1 that completely encloses q will subtend the same solid angle at q as the sphere S_3. Therefore:

$$\text{The flux through } S_1 = \frac{q}{4\pi\epsilon_0}\int_{S_3} \mathrm{d}\Omega = \frac{q}{\epsilon_0}.$$

For the surface S_2 that does not enclose q, any flux cone must cut the surface twice, once on entry (e.g. at A) and once on exit (e.g. at B). The total flux flowing *out* of the surface bounded by A, B and the cone in between is therefore $\mathbf{E}_\mathrm{B}.\mathrm{d}\mathbf{S}_\mathrm{B} - \mathbf{E}_\mathrm{A}.\mathrm{d}\mathbf{S}_\mathrm{A}$. But

$$\frac{E_\mathrm{B}}{E_\mathrm{A}} = \frac{r_\mathrm{A}^2}{r_\mathrm{B}^2}$$

by the inverse square law for electric fields and

$$\frac{\mathrm{d}S_\mathrm{B}\cos\theta_\mathrm{B}}{\mathrm{d}S_\mathrm{A}\cos\theta_\mathrm{A}} = \frac{r_\mathrm{B}^2\,\mathrm{d}\Omega}{r_\mathrm{A}^2\,\mathrm{d}\Omega}$$

by definition of solid angles. The flux $\mathbf{E}_\mathrm{A}.\mathrm{d}\mathbf{S}_\mathrm{A}$ that flows into this region is thus exactly the same as the flux $\mathbf{E}_\mathrm{B}.\mathrm{d}\mathbf{S}_\mathrm{B}$ that flows out of the region and the net flux for the surface S_2:

$$\int_{S_2} \mathbf{E}.\mathrm{d}\mathbf{S} = 0.$$

It follows, then, that the total flux for *any* surface surrounding q is

$$\int_S \mathbf{E}.\mathrm{d}\mathbf{S} = \frac{q}{\epsilon_0} \qquad [2.8]$$

exactly as obtained from the sphere in equation [2.7].

By the principle of superposition the flux due to two charges q_1 and q_2 is just:

$$\int_S \mathbf{E}_1.\mathrm{d}\mathbf{S} + \int_S \mathbf{E}_2.\mathrm{d}\mathbf{S} = \frac{1}{\epsilon_0}(q_1 + q_2)$$

and it follows that the flux due to any charge distribution is:

$$\int_S \mathbf{E}.d\mathbf{S} = \frac{1}{\epsilon_0}\int_V \rho d\tau. \qquad [2.9]$$

where V is the volume enclosed by S.

This is *Gauss's law:* the total flux out of any closed surface is equal to the total charge enclosed by it divided by the electric constant, ϵ_0.

Applications of Gauss's Law

Gauss's law is particularly useful for finding the electric field due to a symmetrical distribution of charge. In each case the Gaussian surface is chosen to suit the symmetry of the problem, as will be seen from three examples.

1. **E** *for a sphere of charge*

Suppose we have a sphere of uniform charge density ρ_0, then

$$Q = \int_\tau \rho d\tau = \frac{4}{3}\pi a^3 \rho_0,$$

as illustrated in Fig. 2.8, is its total charge.

We first draw an imaginary Gaussian surface S of radius R through the point P, where we wish to find **E**. Since the charge is uniformly distributed throughout Q, by symmetry **E** is everywhere radial from Q.

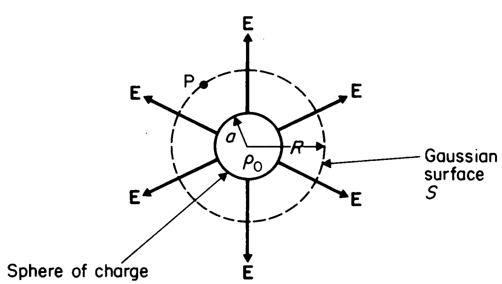

Fig. 2.8 Electric field of a sphere of charge.

Applying Gauss's law we obtain:

$$\int_S \mathbf{E}.d\mathbf{S} = \frac{Q}{\epsilon_0}.$$

Hence $\quad E.4\pi R^2 = Q/\epsilon_0$

or $\quad E = Q/4\pi\epsilon_0 R^2.$

This is exactly the same as the field due to a point charge Q at the centre of the sphere of charge, a result that is quite hard to prove without Gauss's law.

2. **E** *for a line charge*

To find the electric field at a point distance r from an infinite line charge λ C m^{-1}, we note the cylindrical symmetry of the problem and draw a Gaussian cylindrical surface S (Fig. 2.9)

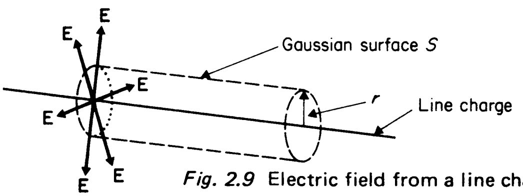

Fig. 2.9 Electric field from a line charge.

of radius r and of length 1 m. The electric vectors will be everywhere radial by symmetry and the same at all points along the line. Applying Gauss's law:

$$\int_S \mathbf{E}.d\mathbf{S} = \int_{\substack{\text{cylindrical}\\ \text{surface}}} \mathbf{E}.d\mathbf{S} + \int_{\substack{\text{end}\\ \text{faces}}} \mathbf{E}.d\mathbf{S} = \frac{\lambda}{\epsilon_0}.$$

Hence $\quad 2\pi r E + 0 = \lambda/\epsilon_0$

or $\quad E = \lambda/2\pi\epsilon_0 r.$

3. **E** *for a plane sheet of charge*

To find the electric field near a plane sheet of charge σ C m^{-2}, we first note that **E** must be everywhere normal to the sheet and that the field $\mathbf{E}_1$ on one side must be the same size as the field $\mathbf{E}_2$ on the other side (Fig. 2.10). By symmetry our Gaussian surface S is a rectangular box whose sides parallel to the sheet of area A contain all the flux that the charges are producing. By Gauss's law:

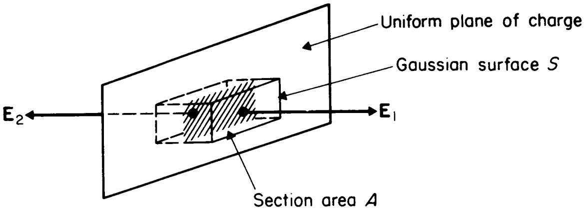

Fig. 2.10 Electric field from a sheet of charge.

$$\int_S \mathbf{E}.\mathbf{dS} = \int_A \mathbf{E}_1.\mathbf{dS} + \int_A \mathbf{E}_2.\mathbf{dS} = \frac{\sigma A}{\epsilon_0}.$$

Here both these integrals refer to outward fluxes, so:

$$EA + EA = \sigma A/\epsilon_0$$

or

$$E = \sigma/2\epsilon_0.$$

2.3 Electric potential

The electrostatic field, like the gravitational field, is a conservative field. The concept of potential energy used in gravitational problems can therefore be applied to electrical problems. In mechanics the work done, dW, on a particle travelling a distance ds along a path ab by an applied force $\mathbf{F}$ (Fig. 2.11) is given by the component of $\mathbf{F}$ acting in the direction s times ds,

$$\mathrm{d}W = F \cos\theta\, \mathrm{d}s.$$

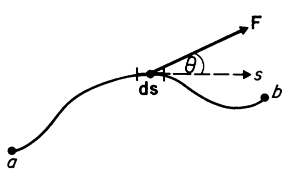

Fig. 2.11 Force **F** acting on a particle as it moves along the path *ab*.

The total work done over the path ab is then

$$W = \int_a^b (F\cos\theta)\,\mathrm{d}s = \int_a^b \mathbf{F}.\mathbf{ds}$$

where **ds** is a vector element of the line from a to b. In electrostatics the particle becomes a test charge moving quasi-statically (with zero velocity) from a to b and the applied force must overcome the electric forces acting on the test charge. In Fig. 2.12

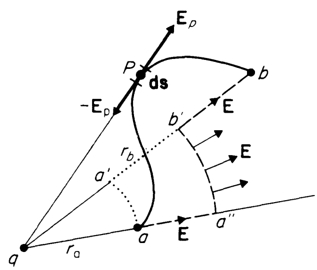

Fig. 2.12 Electric force $\mathbf{E}_p$ and external force $-\mathbf{E}_p$ acting on test charge at P, used in calculating the work done on taking unit charge from a to b.

the electric force on the test charge at P is $\mathbf{E}_p$ and so the external, applied force to move the charge quasi-statically is $-\mathbf{E}_p$. It is this force that is needed to calculate the external work done against the electric field due to the point charge q.

Therefore the work done on unit charge in taking it from a to b is

$$\int_a^b -\mathbf{E}.\mathbf{ds}.$$

Using the definition of **E** in equation [2.4] and noting that $\hat{\mathbf{r}}.\mathbf{ds} = \mathrm{d}r$. this integral becomes:

$$\frac{-q}{4\pi\epsilon_0}\int_a^b \frac{1}{r^2}\,\mathrm{d}r = \frac{-q}{4\pi\epsilon_0}\left(\frac{1}{r_a}-\frac{1}{r_b}\right).$$

Referring to Fig. 2.12 we can see that this amount of work would also be done if the charge was just moved radially from a' to b. Equally it would be the same if it was moved first along the radial path aa'', then along the circular path $a''b'$ and finally along the radial path $b'b$, since **E** is always normal to a circular path about q and therefore no work is done along a circular path.

It thus follows that the $\int_a^b \mathbf{E}.\mathbf{ds}$ is the same along any arbitrary path which can always be considered as the zero sum of normal components along circular paths and the work done by the tangential components along radial paths. If the path is a closed loop C (a to b and back to a) then clearly the integral is zero:

$$\oint_C \mathbf{E}.\mathbf{ds} = 0. \qquad [2.10]$$

This is the *circulation law* for the electrostatic field and is a characteristic of conservative fields that have spherical symmetry and of forces that are radial $f(\mathbf{r})$. It does not have to be an inverse square law force to have zero circulation.

When the path is not closed, the work done depends only on the end points and is independent of the path taken (Fig. 2.12). The work done on unit charge can therefore be represented as the difference between two electric potentials $\phi(b)$ and $\phi(a)$, by analogy with mechanical potential energy:

$$-\int_a^b \mathbf{E}.\mathbf{ds} = \phi(b) - \phi(a). \qquad [2.11]$$

To obtain an absolute value for the electric potential, we must specify its zero. This is taken for convenience to be at an infinite distance from the source q so that

$$-\int_\infty^r \mathbf{E}.\mathbf{ds} = \phi(r) \qquad [2.12]$$

defines the potential $\phi(r)$ at any point of distance r from q. Using equation [2.4] for $\mathbf{E}$, we obtain

$$\phi(r) = \frac{-q}{4\pi\epsilon_0}\left(\frac{1}{r}\right)$$

for a point charge q at the origin.

Electric potentials can be superimposed like electric fields and so for a distribution of charges we obtain similar equations to [2.4] and [2.5], namely:

$$\phi(1) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r_{1i}} \qquad [2.13]$$

and $$\phi(1) = \frac{1}{4\pi\epsilon_0} \int_{\substack{\text{all}\\ \text{space}}} \frac{\rho(2)\,d\tau_2}{r_{12}}. \qquad [2.14]$$

It is important to remember that electric potentials are the work done on unit charges and therefore measured in volts, not joules like potential energy. The volt is defined by: the work done is 1 joule when a charge of 1 coulomb is moved through a potential difference of 1 volt.

The calculation of electric fields can often be achieved more simply from electric potentials than from equations [2.4] and [2.5]. To do this we must invert equation [2.11] to obtain a differential equation. For two points distance Δx apart, by the definition of electric potential, the work done on moving charge through Δx is

$$\Delta W = \phi(x + \Delta x, y, z) - \phi(xyz)$$

$$= \frac{\partial \phi}{\partial x} \Delta x.$$

But the work done against the electric field **E** is

$$\Delta W = -\int_x^{x+\Delta x} \mathbf{E}.\mathbf{ds} = -E_x \Delta x.$$

Hence $$E_x = -\frac{\partial \phi}{\partial x}.$$

Similarly, for movements along Δy and Δz we find:

$$E_y = -\frac{\partial \phi}{\partial y} \text{ and } E_z = -\frac{\partial \phi}{\partial z},$$

so that $$\mathbf{E} = -\left(\frac{\partial \phi}{\partial x}\mathbf{i} + \frac{\partial \phi}{\partial y}\mathbf{j} + \frac{\partial \phi}{\partial z}\mathbf{k}\right) \qquad [2.15]$$

where $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ are unit vectors along the Cartesian axes $0x, 0y, 0z$.

In vector calculus the gradient of a scalar field $\Omega(xyz)$ that is continuously differentiable is defined as the vector

$$\text{grad}\,\Omega = \frac{\partial\Omega}{\partial x}\mathbf{i} + \frac{\partial\Omega}{\partial y}\mathbf{j} + \frac{\partial\Omega}{\partial z}\mathbf{k} \qquad [2.16]$$

Comparing equations [2.15] and [2.16] we see that the electrostatic field **E** is just minus the gradient of the electric potential ϕ:

$$\mathbf{E} = -\text{grad}\,\phi \qquad [2.17]$$

It follows that electric fields have the convenient unit of volts per metre, as well as the more fundamental one of newtons per coulomb from equation [2.4].

Conductors

A metal may be considered as a conductor containing many 'free' electrons which can move about inside but not easily escape from the surface. Inside a metal there is perfect charge balance between the positive ions and the negative electrons and, on a macroscopic scale, the net charge density is zero. By Gauss's law the electric field $\mathbf{E}_i$ inside a Gaussian surface that coincides with the surface of a metallic conductor (Fig. 2.13(a)) must therefore also be zero.

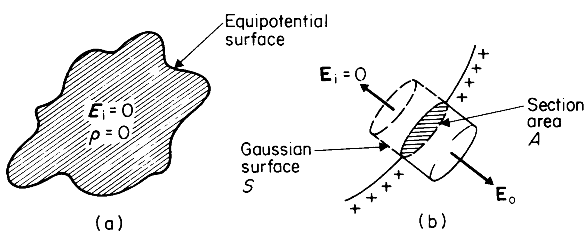

Fig. 2.13 (a) Electric field inside a conductor is zero. (b) Outside a charged conductor it depends on the charge density at the surface σ Cm^{-2}.

By equation [2.17] the gradient of the electric potential at the surface, grad ϕ, must be zero and so the surface of a conductor is an equipotential surface and the interior of the conductor is an equipotential region (ϕ = constant).

When a conductor is charged the excess charges stay on the surface, where they are not completely free. They then produce an external field $\mathbf{E}_0$ just outside the surface (Fig. 2.13(b)), whose value can be obtained from the cylindrical Gaussian surface S of cross-sectional area A and cylindrical axis normal to the surface. Clearly

$$\int_S \mathbf{E}.\mathbf{dS} = E_0 A + 0 = \frac{\sigma A}{\epsilon_0}$$

and $E_0 = \sigma/\epsilon_0$. [2.18]

This is just *twice* the field for a sheet of charge (section 2.2) because the internal electrons have produced zero internal field when the 'sheet of charge' is no longer isolated, but on a conductor. A similar result is seen if we consider two uniformly charged, parallel, conducting plates. In Fig. 2.14 the large

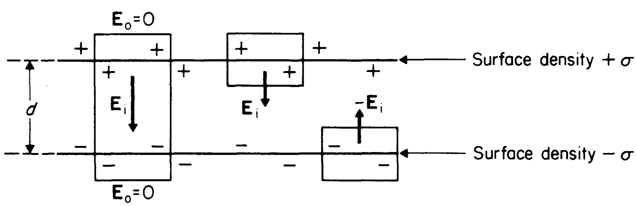

Fig. 2.14 Electric field between two charged plates.

Gaussian surface encloses a total charge of zero and so the external field $\mathbf{E}_0 = 0$. On the other hand the internal field $\mathbf{E}_i$, whether obtained from the positively charged plate, or the negatively charged plate, is just $E_i = \sigma/\epsilon_0$.

An interesting question is: can the inner surface of a hollow conductor be charged? In Fig. 2.15 if there is a surface density of charge σ inside the cavity then there is an electric field in the cavity and for the closed path C through P and Q:

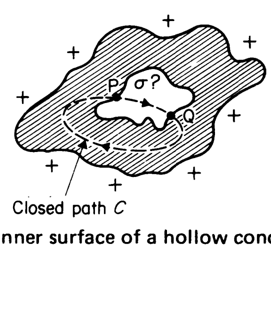

Fig. 2.15 Can the inner surface of a hollow conductor be charged?

$$\oint_C \mathbf{E}.\mathbf{ds} \neq 0.$$

But this would violate the circulation law (equation [2.10]) and so **E** inside a cavity must be zero and it is impossible to charge the inside of a hollow conductor. This important principle is the basis of electrostatic screening (the Faraday cage).

2.4 Electrostatic energy

Electric energy is stored in capacitors, for example in parallel plate capacitors (Fig. 2.14). From equation [2.11] we see that the potential difference V between the plates, for a uniform field $\mathbf{E}_i$ between the plates distance $\mathbf{d}$ apart is:

$$\mathbf{E}_i.\mathbf{d} = \phi_+ - \phi_- = V.$$

But $E_i = \sigma/\epsilon_0$ and for a uniform distribution of charge, $\sigma = Q/A$, where Q is the total charge on the plates of area A.

Hence $V = \left(\dfrac{d}{\epsilon_0 A}\right) Q$, or $V \propto Q$.

The proportionality between V and Q is always found for two oppositely charged conductors, since it arises from the principle of superposition: doubling the charges doubles the field and doubles the work done. By convention, we define *capacitance* by:

$$Q = CV$$

and so for the parallel plate capacitor the capacitance is

$$C = \epsilon_0 A/d. \qquad [2.19]$$

The unit of capacitance is the farad (after Faraday) and is a coulomb per volt. It is a very large unit and typical small capacitors are in microfarad (μF) for use at low frequencies and in picofarad (pF) for radio frequencies. When capacitors are in parallel they are all at the same potential (Fig. 2.16(a)) and so $C = \sum_i C_i$, but when they are in series they each carry the same charge (Fig. 2.16(b)) and then $1/C = \sum_i 1/C_i$.

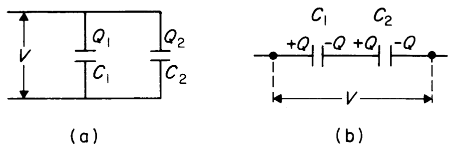

Fig. 2.16 Capacitors connected (a) in parallel and (b) in series.

The work done in charging a capacitor is equal to the *energy U* stored in it and so:

$$U = \int_0^Q V\mathrm{d}Q = \frac{1}{C}\int_0^Q Q\mathrm{d}Q = \frac{Q^2}{2C} = \frac{1}{2}CV^2. \qquad [2.20]$$

For the parallel plate capacitor, neglecting end-effects, the energy is:

$$U = \frac{Q^2}{2C} = \frac{(\sigma A)^2}{2(\epsilon_0 A/d)} = \frac{\sigma^2 Ad}{2\epsilon_0}.$$

The *energy density u* in this electrostatic field is the total energy U divided by the volume Ad and so

$$u = \frac{\sigma^2}{2\epsilon_0} = \frac{1}{2}\epsilon_0 E^2 \qquad [2.21]$$

since $E = \sigma/\epsilon_0$ between the plates. This expression for the energy density was derived from a specific example, the parallel plate capacitor, but it contains only the electric field E and the electric constant ϵ_0. It is, in fact, quite general, as can be seen from Fig. 2.17. There an electric field is described by a number of equipotentials $\phi_1, \phi_2, \ldots$ and we can imagine any small volume $\mathrm{d}\tau$ in that field as a tiny parallel plate capacitor ABCD, since a

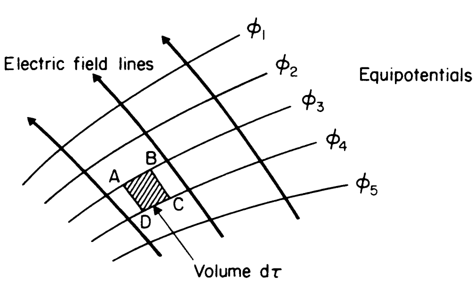

Fig. 2.17 Equipotentials in an electric field.

conducting plate is an equipotential surface. Therefore quite generally

$$U = \frac{1}{2}\epsilon_0 \int_T E^2\,d\tau. \qquad [2.22]$$

The idea that the electrostatic energy is stored ***in the electric field*** is an important one and enables the energy to be computed without knowing anything about the distribution of electric charge. It is even more important in discussing the energy of radio waves. Clearly radio stations transmit electromagnetic energy in the waves we receive at our aerials, but they do not transmit electric charges over long distances. The energy is stored, and travels, in the electromagnetic field of the wave.

There is one point of difficulty in calculating electric fields. The energy of a charged sphere of radius a (problem 2) is $\frac{1}{2}Q^2/(4\pi\epsilon_0 a)$ and so the self-energy of a point charge ($a\rightarrow 0$) would be infinite! Obviously the idea of stored energy in an electrostatic field is not consistent with the presence of point charges: either the electron has a finite size or we cannot extend our concept to elementary charged particles. To avoid this real difficulty we compute the energy only of the electrostatic fields between the charges, and omit their self-energy. Thus we can write equation [2.22] more generally as:

$$U = \frac{1}{2}\epsilon_0 \int_T \mathbf{E}.\mathbf{E}\,d\tau. \qquad [2.23]$$

Electric stress

The limit to the energy that can be stored in a particular capacitor depends on the maximum electric field E that its insulation will withstand before it breaks down under the electric stress. Typically the insulation strength is about $10^8\,\mathrm{V\,m^{-1}}$ so that a large capacitor of internal volume $0.1\,\mathrm{m^3}$ would hold a maximum of $\frac{1}{2}\epsilon_0.10^{16}.0.1 \simeq 5\,\mathrm{kJ}$. This is small compared with the 10 MJ of chemical energy in 1 kg of common salt and very small compared with the 50 TJ of nuclear energy in 1 kg of uranium.

Still larger capacitors cannot easily be built because of the enormous mechanical stresses they are subjected to when charged.

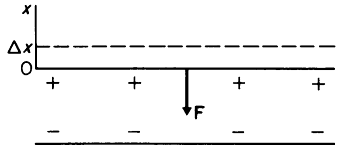

Fig. 2.18 Work is done in separating charged capacitor plates.

This can be seen by applying the principle of virtual work to charged capacitor plates (Fig. 2.18). If **F** is the attractive force between the plates, then the external work done ΔU increasing the separation by Δx must equal the change in electrostatic energy for constant Q. Using equations [2.19] and [2.20],

$$\Delta U = -F\,\Delta x = \tfrac{1}{2}Q^2\Delta\left(\frac{1}{C}\right) = \frac{Q^2\Delta x}{2\epsilon_0 A}$$

or $\quad F = -Q^2/2\,\epsilon_0 A.$

But the charge $Q = \sigma A$ and the electric field is $E = \sigma/\epsilon_0$, so that the stress is:

$$\frac{F}{A} = \frac{1}{2}\,\epsilon_0 E^2.$$

This is the same as the energy density (equation [2.21]) and so for $E = 10^8\,\mathrm{V m^{-1}}$, the mechanical stress is $\simeq 50\,\mathrm{kN\,m^{-2}}$ or about 5 tonne weight per square metre.

2.5 Dielectrics

Dielectric materials – like glass, paper and plastics – are electrical insulators. They have no free charges and do not conduct electricity, but they do influence electric fields. Faraday discovered that inserting an insulator between the plates of a parallel plate capacitor increased its capacitance. We define the *dielectric constant* (or relative permittivity) ϵ_r of an insulator from Faraday's experiment as the ratio of the capacitances when the capacitor is completely filled by the insulator to when it is empty:

$$\epsilon_r = C_{\text{full}}/C_{\text{empty}}. \qquad [2.24]$$

For commercial dielectrics the material is normally of uniform composition and when used in low electric fields is a linear, isotropic, homogeneous medium characterised by a single constant ϵ_r. Typical values of ϵ_r are air = 1.0006, polythene = 2.3, glass = 6, barium titanate ceramic = 3000. In designing capacitors, transformers, coaxial cables, etc., ϵ_r is an important factor influencing the design.

Physically there is a great deal to investigate in the dielectric behaviour of gases, liquids and, especially, solids, where ϵ_r can be an anisotropic parameter described by an appropriate tensor. The dielectric 'constant' will also vary with temperature, with the frequency of an electromagnetic field and will become nonlinear in high electric fields. It is only in its familiar usage in low-frequency, low-field capacitors that it can be treated as a simple constant.

Polarisation

What happens when a slab of dielectric is inserted into a parallel plate capacitor (Fig. 2.19)? We know that the capacitance

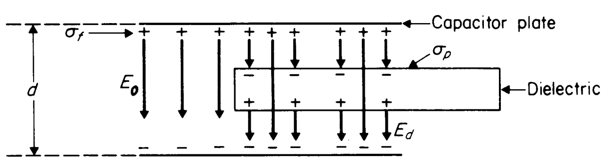

Fig. 2.19 Electric fields in a capacitor with and without a dielectric.

$C = Q/V$ increases from Faraday's experiment and so, if the charge Q on the plates has not leaked away, the potential V must have been reduced. Ignoring end-effects there is a uniform electric field $E = V/d$, so the electric field in the presence of the dielectric E_d must be less than that originally present, E_0. We can explain this by postulating an induced surface charge σ_p on each side of the dielectric slab, providing the charges are of opposite sign to those inducing them from the respective capacitor plates. If these free charges have surface density σ_f, then

$$E_d = \frac{\sigma_f - \sigma_p}{\epsilon_0} < \frac{\sigma_f}{\epsilon_0} = E_0 \qquad [2.25]$$

using the expression $E = \sigma/\epsilon_0$, which depends only on Gauss's law. This process is called *polarisation* of the dielectric and occurs only in the presence of the electric field (the part of the dielectric outside the capacitor plates is not polarised - Fig. 2.19).

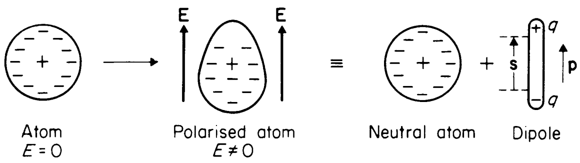

Fig. 2.20 Polarisation of an atom gives it a dipole moment **p**.

One way in which a neutral atom can acquire a dipole moment is shown in Fig. 2.20. The spherical electron cloud of the neutral atom is distorted by the applied electric field **E** and this distorted charge distribution is equivalent, by the principle of superposition, to the original spherical distribution plus a dipole distribution whose *dipole moment* is

$$\mathbf{p} = q\mathbf{s} \qquad [2.26]$$

where $\mathbf{s}$ is the distance vector from $-q$ to $+q$ of the dipole.

If a polarised dielectric consists of N such dipoles per unit volume, then we define its *polarisation* **P** as:

$$\mathbf{P} = N\mathbf{p} = \sum_i \mathbf{p}_i/\tau$$

where τ is the total volume. The SI unit of dipole moment is the coulomb-metre and that of polarisation coulomb per square metre. Going back to Fig. 2.19, we see that the surface charge on the dielectric is σ_p coulombs per square metre. If we assume this is due to N electrons per unit volume being displaced upwards a distance δ at each surface of area A', then the total charge is

$$\sigma_p A' = Ne\delta A'$$

or $\sigma_p = Ne\delta$.

But $Ne\delta$ is just Np and so $P = \sigma_p$ and equation [2.25] becomes

$$E_d = \frac{\sigma_f - P}{\epsilon_0}.$$

From this equation it is clear that $\epsilon_0 E_d$ and P have the same dimensions. We define *electric susceptibility* χ_e, a dimensionless quantity, as the ratio:

$$\chi_e = P/\epsilon_0 E_d.$$

Knowing χ_e for our dielectric material we therefore obtain the reduced electric field E_d as

$$E_d = \frac{\sigma_f}{\epsilon_0}\left(\frac{1}{1+\chi_e}\right). \qquad [2.27]$$

The capacitance of the capacitor is inversely proportional to the potential and so to the electric field. Using equations [2.24], [2.25] and [2.27] we therefore obtain

$$\epsilon_r = \frac{E_0}{E_d} = (1 + \chi_e). \qquad [2.28]$$

Measurement of the electric susceptibility χ_e of matter at low frequencies is thus a measurement of the dielectric constant ϵ_r, which at optical frequencies can be shown to be the square of the refractive index (see the sequel to this text, *Electromagnetism*).

Electric displacement

For all electrostatic systems we have the fundamental equation (equation [2.9]), Gauss's law:

$$\int_S \mathbf{E}.\mathbf{dS} = \frac{1}{\epsilon_0}\int_V \rho \, d\tau$$

where V is the volume enclosed by the surface S.

When dielectrics are present, the charge density ρ will be the sum of any polarisation charges of density ρ_p and any free charges of density ρ_f. Therefore

$$\epsilon_0 \int_S \mathbf{E}.\mathbf{dS} = \int_V \rho_f \,\mathrm{d}\tau + \int_V \rho_p \,\mathrm{d}\tau. \qquad [2.29]$$

We have seen (Fig. 2.19) that in a parallel plate capacitor, when the polarisation **P** is normal to the surface of the dielectric, its magnitude is just the surface density of charge σ_p displaced from inside the dielectric. At the ends of the dielectric slab where **P** is tangential to the surface the surface density of charge is zero. It is the normal component of **P** that produces a surface charge, so that for an arbitrary surface S inside a dielectric (Fig. 2.21)

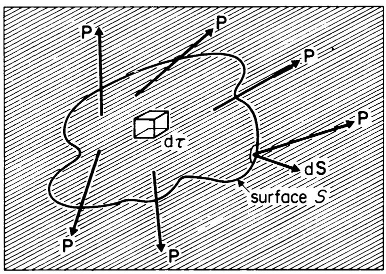

Fig. 2.21 Non-uniform polarisation of a dielectric.

the charge $\mathrm{d}q_p$ displaced across a surface element **dS** is **P.dS**. A non-uniform polarisation at the surface S therefore produces a total displacement of charge q_p across S given by:

$$q_p = \int_S \mathbf{P}.\mathbf{dS}$$

Since the dielectric is electrically neutral this is compensated by a volume density of charge $-\rho_p$ such that

$$\int_V -\rho_p \,\mathrm{d}\tau = -q_p.$$

Hence the flux of **P** is given by a type of Gauss's law for polarised dielectrics:

$$\int_S \mathbf{P}.\mathbf{dS} = -\int_V \rho_p \,\mathrm{d}\tau.$$

Combining this with equation [2.29], we have:

$$\int_S (\epsilon_0 \mathbf{E} + \mathbf{P}).\mathbf{dS} = \int_V \rho_f \,\mathrm{d}\tau$$

and define the *electric displacement*, **D**, as

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon_0(1 + \chi_e)\mathbf{E} \qquad [2.30]$$

so that Gauss's law can also be written:

$$\int_S \mathbf{D}.\mathrm{d}\mathbf{S} = \int_V \rho_f \,\mathrm{d}\tau. \qquad [2.31]$$

The flux of **D** thus depends solely on the free charges and this can be very useful, for example, in microwave physics.

However if we use this equation and, from [2.28] and [2.30], write

$$\mathbf{D} = \epsilon_r \epsilon_0 \mathbf{E} \qquad [2.32]$$

then we must remember that for many materials ϵ_r is not just a number. As we have emphasised before, **P** (and hence **D**) is not proportional to **E** for nonlinear materials and, in any case, χ_e (and so ϵ_r) can vary with frequency, temperature, crystal direction, etc. This is one reason why ϵ_r is often referred to as the relative permittivity rather than the dielectric constant of a dielectric.

Boundary conditions

What happens to the electric field when it crosses the boundary between two dielectrics of permittivities $\epsilon_1 = \epsilon_r(1)\epsilon_0$ and $\epsilon_2 = \epsilon_r(2)\epsilon_0$? To find out we apply Gauss's law as given in equation [2.31] and the circulation law, equation [2.10], to the electric vectors **D** and **E** shown in Fig. 2.22.

For the flux into and out of the Gaussian cylinder of cross-section dS and negligible height we have:

$$\mathbf{D}_1.\mathrm{d}\mathbf{S}_1 + \mathbf{D}_2.\mathrm{d}\mathbf{S}_2 = 0$$

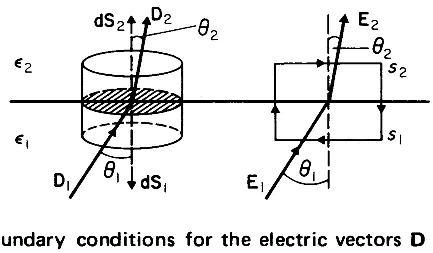

Fig. 2.22 Boundary conditions for the electric vectors **D** and **E** crossing between two dielectrics of permittivities ϵ_1 and ϵ_2.

since there are no free charges in dielectrics. Hence only the normal component D_n of each electric displacement contributes and

$$D_{1n} = D_{2n}. \qquad [2.33]$$

Applying the circulation law to the electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ crossing the closed loop of length $s_1 + s_2$, we have:

$$\oint_C \mathbf{E}.\mathbf{ds} = \mathbf{E}_1.\mathbf{s}_1 + \mathbf{E}_2.\mathbf{s}_2 = 0.$$

Since we can contract the loop to be as near as we wish to the surface, only the tangential component E_t of each electric field contributes and

$$E_{1t} = E_{2t}. \qquad [2.34]$$

At the boundary we therefore have continuity for D_n and E_t. When it is valid to use equation [2.32] we can write [2.33] and [2.34] as:

$$\epsilon_{r1} E_1 \cos\theta_1 = \epsilon_{r2} E_2 \cos\theta_2$$

$$E_1 \sin\theta_1 = E_2 \sin\theta_2.$$

We therefore get refraction of **D** and **E** at the boundary with the relation

$$\frac{\tan\theta_1}{\tan\theta_2} = \frac{\epsilon_{r1}}{\epsilon_{r2}}. \qquad [2.35]$$

Energy density

For a vacuum parallel plate capacitor we showed that the energy stored in it is (equation [2.19]):

$$U = \frac{1}{2} CV^2.$$

With a dielectric completely filling it, the capacitance is increased by a factor ϵ_r (equation [2.24]) and the electric field is reduced by ϵ_r (equation [2.28]). Hence the energy stored is:

$$U = \frac{1}{2}\left(\frac{\epsilon_r \epsilon_0 A}{d}\right)(Ed)^2$$

where E is the reduced field. Therefore the energy density is:

$$u = \frac{1}{2}\epsilon_r \epsilon_0 E^2 = \frac{1}{2} DE,$$

where equation [2.32] applies. As with the energy density in the vacuum (Fig. 2.17), the energy density still resides in the electric field. The difference is that equation [2.23] now becomes in general:

$$U = \frac{1}{2}\int_\tau \mathbf{D}.\mathbf{E}\, d\tau. \qquad [2.36]$$

We can apply the principle of virtual work to the force between two capacitance plates (charged conductors) in a dielectric liquid (Fig. 2.18, with a dielectric present) and find

$$F = \frac{-\delta U}{\partial x} = \frac{-Q^2}{2}\frac{\partial}{\partial x}\left(\frac{1}{C}\right).$$

The dielectric increases C by a factor ϵ_r and so decreases F by $1/\epsilon_r$. However this only leads to a revised Coulomb's law (compare equation [2.3]) in certain cases:

$$\mathbf{F}_1 = \frac{1}{4\pi\epsilon_r\epsilon_0}\sum_j \frac{q_1 q_j}{r_{1j}^2}\hat{\mathbf{r}}_{1j}.$$

It is limited to dielectrics which are isotropic, homogeneous, linear and have a constant relative permittivity ϵ_r. In practice this limits it to fluids over a narrow range of temperature and pressure, whereas the vacuum version of Coulomb's law is always true for stationary charges.

Chapter 5
Electromagnetism

So far we have been able to consider electric fields and magnetic fields separately, through imposing the conditions that these fields and any currents shall be in steady states with $\partial \mathbf{E}/\partial t$, $\partial \mathbf{B}/\partial t$ and $\partial \mathbf{j}/\partial t$ all zero. In this chapter we move from steady currents to varying currents, from steady electric fields to induced electric fields, from stationary to moving circuits. Electric and magnetic effects become intimately connected in the study of electromagnetism, which has already been introduced in the electromagnetic force equation [4.9]:

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}.$$

We first develop the concepts contained in Faraday's law and then apply it to a variety of examples.

5.1 Faraday's law

In a series of experiments in 1831–2 Faraday showed conclusively that electricity from batteries and magnetism from iron magnets were not separate phenomena but intimately related. He discovered that voltages can be generated in a circuit in three different ways:

1. by moving the circuit in a magnetic field;
2. by moving a magnet near the circuit; and
3. by changing the current in an adjacent circuit.

Consider the simple system shown in Fig. 5.1: a solenoid producing an axial field **B** which passes through a current loop connected to a galvanometer (current detector). The galvanometer needle kicks if:

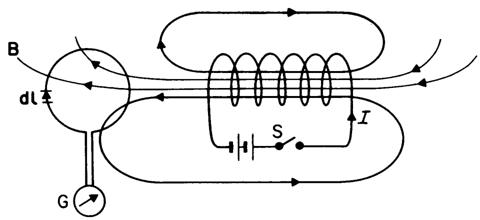

Fig. 5.1 A solenoid produces a magnetic field **B**, which is sensed by a current loop and a galvanometer G.

1. the solenoid is moved backwards and forwards;
2. the current loop is moved;
3. the current in the solenoid is switched on or off without moving any circuits.

The needle only moves when there is a current in the loop, i.e. when there is a net force on the electrons in the wire in one direction along it. There may be several different forces acting on different parts of the loop but what moves the needle is the net force integrated around the complete circuit. This is the *electromotive force* (e.m.f.)

$$\mathcal{E} = \oint \frac{\mathbf{F}}{q} \cdot \mathbf{dl} = \oint \mathbf{E} \cdot \mathbf{dl} \qquad [5.1]$$

where **F** is the force on charge q and the integral is taken round the loop. The definition of $\mathcal{E}$ is therefore the tangential force per unit charge in the wire integrated over the complete circuit.

We can illustrate this definition by considering the work done by a battery of e.m.f. $\mathcal{E}$ driving a charge q around a circuit (Fig. 5.2). The work done by the battery is $\mathcal{E}q$, while that done on the charge in moving a distance **dl** is **F.dl**. Therefore

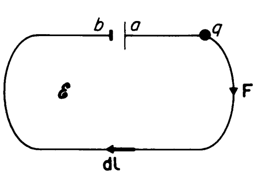

Fig. 5.2 A battery of e.m.f. $\mathcal{E}$ drives a charge q around a circuit.

$$\mathcal{E}q = \int_a^b \mathbf{F}.d\mathbf{l}$$

and $$\mathcal{E} = \int_a^b \frac{\mathbf{F}}{q}.d\mathbf{l} = \int_a^b \mathbf{E}.d\mathbf{l}$$

where $\mathbf{E} = \mathbf{F}/q$ is the electric field or force per unit charge. The SI units for $\mathcal{E}$ and $\mathbf{E}$ are obviously not the same: e.m.f. is measured in volts, electric field in volts per metre.

Motional e.m.f.s

The concept of electromotive force generated by the motion of circuits is best understood by considering some examples.

1. *Metal rod in a uniform field*

   In Fig. 5.3 a metal rod AB of length L is placed on the y-axis and moved along $0x$ in the uniform magnetic field $\mathbf{B}$ in the z-direction. By the Lorentz force law, each electron in the wire

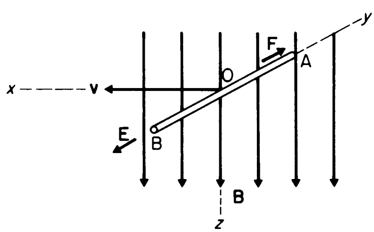

Fig. 5.3 A metal rod moves in a uniform magnetic field.

   experiences a force $\mathbf{F} = -e\mathbf{v} \times \mathbf{B}$ and so the free electrons tend to move towards A. This produces a distribution of excess negative charge, which by Gauss's law is equivalent to an electric field $\mathbf{E}$ and so a net force $-e\mathbf{E}$ on each electron. Therefore

$$\mathbf{E} = \mathbf{v} \times \mathbf{B}. \qquad [5.2]$$

   From equation [5.1] the total e.m.f. across the rod will be:

$$\mathcal{E}_{AB} = \int_B^A \mathbf{E}.d\mathbf{y} = v_x B_z L.$$

This e.m.f. is due to the electric field **E** induced in the rod by its motion through the operation of the Lorentz force law.

2. *Metal rod on rails in a uniform field*
In Fig. 5.4 the same metal rod AB is mounted on metal rails so that a circuit ABCD of variable size is formed as the rod moves.

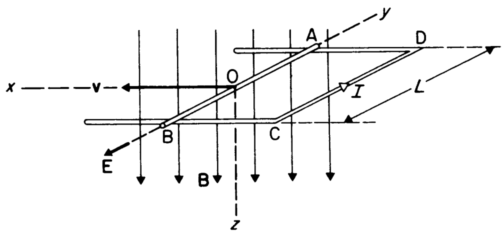

Fig. 5.4 A metal rod moving in a uniform magnetic field generates a current in a circuit.

The electrons now drift from B to A and go round DCB to form a conventional current I in the opposite direction. The e.m.f. generated by the rod moving is unchanged, but it now produces a current:

$$I = \frac{v_x B_z L}{R}$$

where R is the total resistance of the circuit ABCD.

3. *Current loop in a uniform field*
In Fig. 5.5 a square current loop abcd is moving into and out of a uniform magnetic field **B**. If the sides of the loop are parallel to Ox and Oy and **B** is in the z direction, then motion

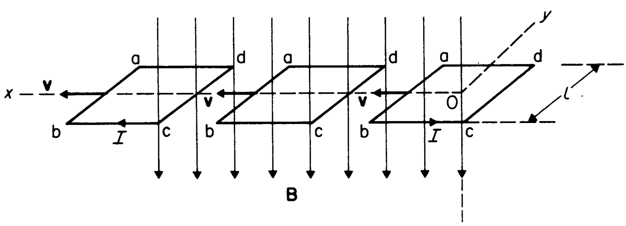

Fig. 5.5 A current loop moves in a uniform magnetic field.

at velocity **v** along Ox will produce an e.m.f. $\mathcal{E}$ along ab as it enters the field. This will generate a current $I = v_x B_z l/r$, where l is the length ab and r is the resistance of the loop, until the side cd enters the field. Then an e.m.f. $\mathcal{E}$ will be generated in dc and this will exactly cancel that in ab giving zero current. Finally as it emerges from the field there will only be the e.m.f. $\mathcal{E}$ in dc and this will generate I in the opposite direction in the loop.

The source of this electrical energy is the mechanical work done in moving the coil. It is dissipated in the loop as heat, which by Joule's law is $I^2 r$ watts. The mechanical work can equally well be done by moving the magnetic field across a stationary loop: it is essentially a relative motion effect. An observer on the coil in this case would see a moving magnetic field $\mathbf{B}(t)$ and would ascribe the current to a moving electric field $\mathbf{E}(t)$. In terms of magnetic flux, at any instant the flux through the coil

$$\Phi = B_z lx$$

and so:

$$|\mathcal{E}| = v_x B_z l = \frac{\mathrm{d}x}{\mathrm{d}t} B_z l = \frac{\mathrm{d}\Phi}{\mathrm{d}t}. \qquad [5.3]$$

This equation is an expression of the *'flux rule'* found experimentally by Faraday: an e.m.f. is induced in a circuit whenever the flux through it changes from any cause. The *direction* of the e.m.f. was established by Lenz and is known as *Lenz's law*: the current induced tends to oppose the change of flux through the circuit. The combination of the flux rule and Lenz's law is known as *Faraday's law*: the e.m.f. induced in a circuit is equal to the negative rate of change of the magnetic flux through that circuit. That is:

$$\mathcal{E} = \oint \mathbf{E}.\mathrm{d}\mathbf{l} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t}. \qquad [5.4]$$

From the definition of magnetic flux this can be written:

$$\oint \mathbf{E}.\mathrm{d}\mathbf{l} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}.\mathrm{d}\mathbf{S}. \qquad [5.5]$$

Here the direction of the vectors **dS** are given by the right-hand screw rule for the circulation around S in the line integral.

Clearly the induced electric field in equation [5.5] is not an electrostatic field, for which the circulation law gives $\oint \mathbf{E}.\mathbf{dl} = 0$, but arises from the Lorentz force $q\mathbf{v} \times \mathbf{B}$ in the case of motional e.m.f.s and from $\mathrm{d}B/\mathrm{d}t$ when the magnetic field is varying. These can be different phenomena, although both are represented by Faraday's law. It is therefore important to distinguish between them.

Motional and transformer e.m.f.s

We can summarise the results on *motional e.m.f.s* by

$$\oint \mathbf{E}.\mathbf{dl} = \oint (\mathbf{v} \times \mathbf{B}).\mathbf{dl} \qquad [5.6]$$

where **v** is the relative motion of a circuit with respect to the frame (usually the laboratory frame) in which **B** is fixed. We say that the circuit, which must not be changing in its shape or composition, is cutting the magnetic flux. This is true whether **B** is a steady magnetic field or a time-varying magnetic field.

The *transformer e.m.f.s* arise when there is no motion and **E** and **B** are fixed in the same coordinate system, so that:

$$\oint \mathbf{E}.\mathbf{dl} = -\frac{\mathrm{d}}{\mathrm{d}t}\int_S \mathbf{B}.\mathbf{dS} = -\int_S \frac{\partial \mathbf{B}}{\partial t}.\mathbf{dS} \qquad [5.7]$$

where the induced field **E** is due solely to the time variation of the magnetic field **B** and is therefore zero for a steady field. The time derivative now refers to each elementary area separately and so is a partial derivative. Now there is no longer any flux cutting and so there is no need to restrict the line integral to a circuit. It can be any contour in space and equation [5.7] becomes:

$$\oint_C \mathbf{E}.\mathbf{ds} = -\int_S \frac{\partial \mathbf{B}}{\partial t}.\mathbf{dS}. \qquad [5.8]$$

Combining this with Stokes's theorem (equation [3.6]), we obtain the differential form of Faraday's law:

$$\mathrm{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \qquad [5.9]$$

This is always true, giving as its time-independent limit the circulation law of the electrostatic field, curl **E** = 0. Equation [5.9] and the electromagnetic force equation [4.9] are the two fundamental equations of electromagnetism.

5.2 Applications of Faraday's law

We will now illustrate the usefulness of Faraday's law, as expressed in equations [5.4], [5.6] and [5.8].

Betatron

The betatron is a circular electron accelerator (Fig. 5.6) with the electrons circulating in a vacuum chamber placed in a powerful,

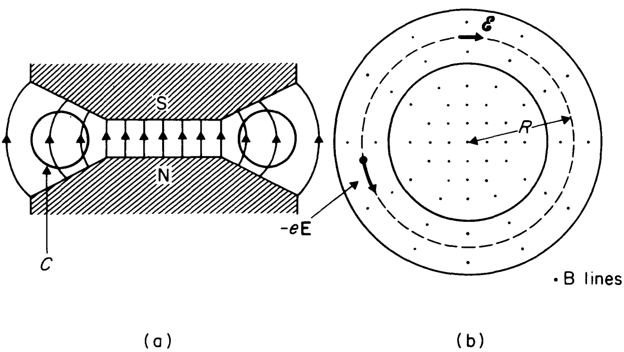

Fig. 5.6 The betatron. (a) Vertical section through the magnet NS and vacuum chamber C. (b) Top view of central section of the vacuum chamber.

non-uniform magnetic field produced by shaped pole-pieces. The electrons are accelerated by increasing the magnetic field, which generates an e.m.f. in the vacuum given by:

$$\mathcal{E} = \oint_C \mathbf{E}.\mathbf{ds} = -\int_S \frac{\partial \mathbf{B}}{\partial t}.\mathbf{dS}. \qquad [5.10]$$

If we assume the electrons are injected into an orbit radius R for which the mean field is $\bar{B}$, then equation [5.10] becomes:

$$2\pi RE = -\frac{d\bar{B}}{dt}.\pi R^2$$

or $$E = -\frac{R}{2}\frac{d\bar{B}}{dt}.$$

The e.m.f. generated opposes the increasing magnetic flux and so its direction is clockwise when viewed from above the magnet (Fig. 5.6(b)) and therefore the electron motion is anticlockwise when driven by the force $-e\mathbf{E}$. The rate of change of momentum p of the electron is therefore:

$$\frac{dp}{dt} = -eE = \frac{eR}{2}\frac{d\bar{B}}{dt}.$$

But we have already seen (equation [4.10]) that the momentum of an electron in a circular orbit is:

$$p = eRB_R$$

where B_R in this case is the magnetic field at radius R. Clearly B_R must vary in time so that:

$$\frac{dp}{dt} = \frac{eR\,dB_R}{dt} = \frac{eR}{2}\frac{d\bar{B}}{dt}$$

that is B_R must increase so that it is always equal to $\frac{1}{2}\bar{B}$ if the electrons are to be confined to their orbit as they accelerate. This is the principle of the betatron, which can accelerate electrons up to energies of many MeV, when they begin to radiate significantly.

Faraday's disc

A homopolar generator can be made from a disc rotating in a steady magnetic field, as first shown by Faraday. In Fig. 5.7 the circular disc of radius a rotates at a steady angular velocity ω in a uniform field $\mathbf{B}$. The simplest type of disc is an insulating one with a conducting ring round its circumference, a conducting axle RO and a radial conducting wire OP embedded in it. The circuit QROPQ is then completed by the brushes on the moving parts at Q and R. Since there is a steady field the only source of e.m.f. must be a motional e.m.f. given by equation [5.6]:

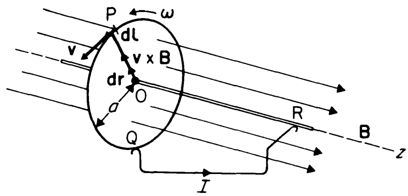

Fig. 5.7 Faraday's disc is a homopolar generator.

$$\mathcal{E} = \oint \mathbf{E}.\mathbf{dl} = \oint (\mathbf{v} \times \mathbf{B}).\mathbf{dl}.$$

The circuit QRO is stationary in the frame of **B** and so the only contributions to the e.m.f. come from OP and PQ. For OP, $\mathbf{v} \times \mathbf{B}$ is along **r** and $\mathbf{dl} = \mathbf{dr}$, while $v = r\omega$, so that

$$\oint (\mathbf{v} \times \mathbf{B}).\mathbf{dl} = \int_0^a r\omega B \mathrm{d}r = \frac{1}{2}a^2 \omega B.$$

And for PQ, $\mathbf{v} \times \mathbf{B}$ is normal to **dl** at all points so that the contribution to the integral is zero. Therefore $\mathcal{E} = \frac{1}{2}a^2 \omega B$ is the e.m.f. driving a current along the circuit in the direction QR.

In terms of Faraday's law (equation [5.4]), we must be careful to specify the flux being cut by the circuit. For the circuit in Fig. 5.7, the flux is cut as the radial wire OP sweeps round through angle POQ. If this angle is θ radians, then the flux cut $= B\pi a^2 (\theta/2\pi)$ and Faraday's law gives:

$$\mathcal{E} = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = -\frac{Ba^2}{2}\frac{\mathrm{d}\theta}{\mathrm{d}t} = -\tfrac{1}{2}a^2 \omega B$$

as before.

When the whole disc is a conductor (exercise 4), very high currents can be generated in the small resistance of the disc and so several brushes are joined together to conduct the radial currents back to the axle.

Mutual inductance

A typical example of a transformer e.m.f. (equation [5.8]) is provided by a coil 1 producing a time-varying magnetic field within the turns of a second coil 2 fixed to it (Fig. 5.8(a)). When the same **B** is parallel to the area S of each of N turns of a coil, the e.m.f. becomes:

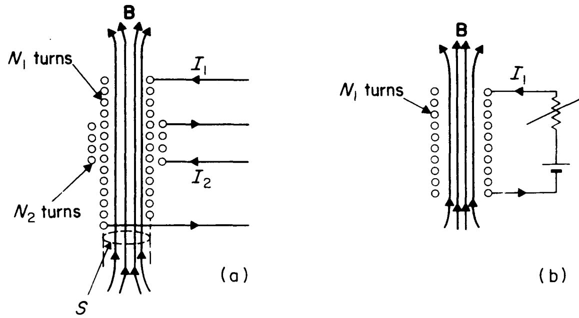

Fig. 5.8 (a) A mutual inductance. (b) A self-inductance.

$$\mathcal{E} = -\frac{d}{dt}(NBS)$$

that is, it is the negative of the rate of change of the flux linkage $N\Phi$.

Applying Ampère's law, $\oint \mathbf{B.ds} = \mu_0 I$, to each turn of coil 1 in Fig. 5.8(a), the field inside is:

$$B_1 = \mu_0 N_1 I_1/l$$

and for a long solenoid, B_0 outside is negligible. Therefore the flux linking coil 2 from coil 1 is:

$$N_2\Phi_{21} = N_2 B_1 S = \mu_0 N_1 N_2 S I_1/l$$

where I_1 is the only quantity varying with time. Hence the e.m.f. induced in coil 2 is:

$$\mathcal{E}_{21} = -\left(\frac{\mu_0 N_1 N_2 S}{l}\right)\frac{dI_1}{dt}.$$

Clearly $\mathcal{E}_{21}$ is proportional to dI_1/dt. The constant of proportionality, which depends only on the geometry of the coils, is called the *mutual inductance, M.* In particular M_{21} is the flux linking coil 2 due to *unit* current in coil 1, that is:

$$M_{21} = \frac{N_2\Phi_{21}}{I_1} \qquad [5.11]$$

and $\mathcal{E}_{21} = -M_{21}\dfrac{dI_1}{dt}$.

For two overwound coils there is clearly a reciprocal relationship and the e.m.f. $\mathcal{E}_{12}$ induced in coil 1 due to current I_2 varying is:

$$\mathcal{E}_{12} = -M_{12}\frac{dI_2}{dt} \qquad [5.12]$$

where

$$M_{12} = \frac{N_1\Phi_{12}}{I_2}. \qquad [5.13]$$

For this case it is obvious that $M_{12} = M_{21} = M$, a result which can be proved for any pair of coupled circuits (Neumann's theorem).

Self-inductance

Faraday induction is not confined to pairs of circuits: we can have self-induction in a single coil. In Fig. 5.8(b) if we vary R then I_1 changes and so the flux Φ_{11} changes. This generates a self-e.m.f. given by:

$$\mathcal{E}_{11} = -\frac{d}{dt}(N_1\Phi_{11}).$$

The *self-inductance* of a coil L is therefore the flux linkage in the coil per unit current in the coil, or

$$L = \frac{N_1\Phi_{11}}{I_1} \qquad [5.14]$$

and $$\mathcal{E} = -L\frac{dI_1}{dt}. \qquad [5.15]$$

The SI unit for both mutual and self-inductance is the *henry*, equivalent to 1 volt-second per ampere. Since $M_{21} = \mu_0 N_1 N_2 S/l$, we note that an alternative, and commonly used, unit for the magnetic constant μ_0 is henry per metre, so that:

$$\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H\,m^{-1}} \qquad [5.16]$$

an exact relationship. The self-inductance of a coil is

$$L = M_{11} = \mu_0 N^2 S/l \qquad [5.17]$$

and an inductance of 1 H requires a very large, air-cored solenoid (e.g. 7000 turns of diameter 0.1 m over a length 0.5 m). Commonly air-cored coils are in the range μH to mH and larger ones have cores of high permeability to enhance their value. At very high frequencies the 'skin effect' causes a small change in L, but for most purposes the self-inductance of an air-cored coil given by equation [5.15] is independent of frequency.

Coupled circuits

When two circuits are coupled, as in Fig. 5.9(a), the current in 1 depends on both the battery V_1 and the e.m.f. induced in it by variations in the current in 2. From equation [5.12],

$$\mathcal{E}_{12} = -M_{12}\frac{dI_2}{dt} = -M\frac{dI_2}{dt}$$

where the negative sign means that if both I_1 and I_2 are positive (anticlockwise) currents, then $\mathcal{E}_{12}$ will be opposed to V_1 when dI_2/dt is positive (increasing). If the polarity of V_1 is reversed, then I_1 is negative and so $\mathcal{E}_{12}$ and V_1 act in the same direction for increasing dI_2/dt.

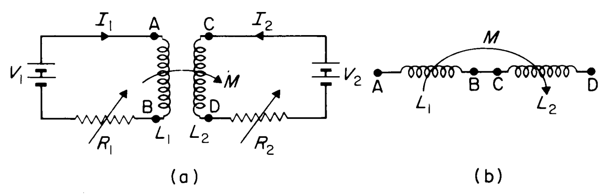

Fig. 5.9 (a) Coupled circuits. (b) Series coupled inductances.

The mutual inductance is reversed in sign if one of the coils is reversed, as can be seen more clearly in Fig. 5.9(b), where the total inductance will be:

$$L_{ABCD} = L_1 + L_2 + 2M \quad [5.18]$$

or $\quad L_{ABCD} = L_1 + L_2 - 2M$

according to whether the currents in L_1 and L_2 are both of the same, or of opposite, signs. So, in general, the e.m.f. in circuit 1 of Fig. 5.9(a) is:

$$V_1 \pm M \frac{dI_2}{dt} - L_1 \frac{dI_1}{dt} = I_1 R$$

where the e.m.f. $\mathcal{E}_{11}$ from equation [5.15] has also been included.

When the coils L_1 and L_2 are coupled tightly together, for example by winding one on top of the other in a toroid, there is no leakage of magnetic flux and so:

$$\Phi_{12} = \Phi_{22} \quad \text{and} \quad \Phi_{21} = \Phi_{11}.$$

But $$M_{12} = \frac{N_1 \Phi_{12}}{I} \quad \text{and} \quad M_{21} = \frac{N_2 \Phi_{21}}{I}.$$

Therefore

$$M_{12} M_{21} = M^2 = N_1 N_2 \Phi_{22} \Phi_{11} / I_1 I_2$$

and, from equation [5.14] defining self-inductance,

$$M^2 = L_1 L_2$$

or $$M = \sqrt{L_1 L_2}. \qquad [5.19]$$

This is the maximum value of M for tight coupling. In general there is some flux leakage and

$$M = k\sqrt{L_1 L_2} \qquad [5.20]$$

where $0 \leqslant k \leqslant 1$ and k is the coupling coefficient.

Magnetic energy

A coil with a current flowing in it stores magnetic energy, an outstanding example being a superconducting coil in its persistent mode. This energy is due to the rate of doing electrical work W by the induced e.m.f. $\mathcal{E}$ in the coil against the induced current I:

$$\frac{dW}{dt} = \mathcal{E} I = -LI \frac{dI}{dt}.$$

For a perfect (loss-less) coil, the total energy stored is therefore:

$$U = -W = \tfrac{1}{2} L I^2. \qquad [5.21]$$

1

Introduction

The theory of electrical networks or circuits has a very specific and useful purpose: it is to allow the calculation of the currents which will flow in the different components or branches of a particular network when one or more electrical signals, or sources of electrical energy, are applied to it. Of course, the motion of the electrical charges which make up the electrical currents takes place under the influence of either electrostatic or magnetic forces and will, therefore, strictly be determined by the basic laws of electricity and magnetism. Similarly, the effects which the moving charges cause as they pass along or are stored in conductors in different configurations will be governed by those same laws. However, for the purpose of electrical network theory, whether for constant (dc) or alternating (ac) currents, an attempt is made to simplify the analysis by expressing the various effects in terms of the properties of specific circuit elements such as resistors, capacitors or inductors. At the same time, the forces acting on the charges are expressed in terms of electromotive forces or electrical potentials while the movement of charges is described by reference to the electric currents.

A summary of the main features of the currents, the potentials and the circuit components utilized in network theory are given below. A fuller treatment of the concepts will be found in E. R. Dobbs, *Electricity and Magnetism*, Routledge, London, 1984.

1.1 ELECTRIC CURRENT

When a charge $\mathrm{d}Q$ passes a given point in time $\mathrm{d}t$, a current will flow in the direction of the charge given by

$$I = \frac{dQ}{dt}$$

the convention being that the direction of current flow is the same as that of the motion of the positive charge.

The SI unit of current—the ampere, A—is the fundamental electrical unit.

The charge which passes the given point in a given time will be

$$Q = \int I \, dt$$

for which the SI unit is the *coulomb*, C, so that 1 C = 1 A s.

1.2 ELECTRICAL POTENTIAL AND EMF

When an electrical current flows, charge moves from one position to another position where it has a lower energy. If the energy of one coulomb of charge flowing between the two positions changes by one joule, there is said to be a *potential difference* (or pd) between the two positions equal to one *volt* (1 V). One volt is thus $1\,J\,C^{-1}$. The pd between two points may be considered to be due to an electric field $\boldsymbol{E}$ acting between the points, $\boldsymbol{E}$ being the force, expressed in newtons per coulomb. The situation is illustrated in Fig. 1.1 where the field $\boldsymbol{E}$ acts at the position of the element of length ds, although not necessarily parallel to it. The pd between A and B will thus be the work done by the field acting on unit charge in moving it between those points, i.e.

$$V_{AB} = \int \boldsymbol{E} \cdot d\boldsymbol{s}$$

Clearly the unit of the field will be $V\,m^{-1}$.

If the potential of the earth is taken as zero and the *potentials* of the points A and B are taken to be V_A and V_B, i.e. the pds of those

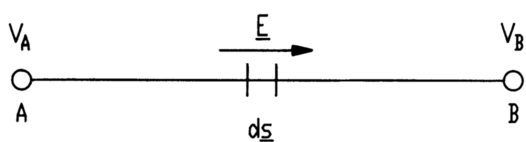

Fig. 1.1 Electric field acting between two points.

two points relative to the earth then, clearly,

$$V_{AB} = V_A - V_B$$

In a circuit which forms a closed loop, the pd which drives the current will have to be supplied to the circuit by some *electromotive force* (or emf) which is given by the integral round the loop

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s}$$

The source of this emf may be, for example, a thermocouple or a dynamo or a voltaic cell. Such a source is said to be an *active* component in a circuit while elements such as resistors, capacitors etc., are said to be passive.

1.3 RESISTANCE AND CONDUCTANCE

Ohm's law, which is obeyed in a wide range of situations and by most, though not all, materials states that the voltage or pd across a conductor (or resistor) is proportional to the current flowing through the conductor provided its temperature remains constant. If the pd is V and the current is I then

$$V = IR$$

where the constant of proportionality, R, is called the *resistance* of the conductor for which the SI unit is the ohm, Ω, and $1\ \Omega = 1\ \mathrm{V\,A^{-1}}$. If Ohm's law had been written instead as $I = GV$, then G would have been the conductance, equal to R^{-1}, and its unit would have been the siemen, S, where $1\ \mathrm{S} = 1\ \Omega^{-1}$.

When current I flows through the resistance R, each coulomb of charge will do work $V = IR$ joules, and heating, known as *Joule heating*, will occur at the rate of VI watts. The power lost as heat is thus

$$VI = I^2R = V^2/R \text{ watts}$$

If the temperature of the resistor changes from θ_0, at which the resistance is R_0, to a new value θ, the resistance will change, approximately, to

$$R = R_0[1 + \alpha(\theta - \theta_0)]$$

where α, the *temperature coefficient of resistance*, is generally positive. For some special alloys, e.g. manganin and constantan, $\alpha \approx 0$. For large temperature changes, a polynomial in $\theta - \theta_0$ will generally be needed to give an accurate description. In the case of semiconductors, e.g. germanium and silicon, the resistance usually falls with temperature.

1.4 INTERNAL RESISTANCE

Most voltage generators or sources of emf will have an intrinsic or internal resistance. This situation is illustrated in Fig. 1.2 where the emf of the generator is V_0 and the internal resistance is R_0. Clearly, the pd between A and B is

$$V_{AB} = V_0 - IR_0$$

and $V_{AB} = IR$. If $R \to \infty$, $I \to 0$ so that V_0 is clearly given by V_{AB} in the open-circuit condition.

1.5 INDUCTANCE

For two coils in proximity, a current flowing in one coil will cause a linkage of magnetic flux lines to the other. If a current I_1 in the coil 1 causes a flux-linkage MI_1 to coil 2 then, by Neumann's theorem, a current I_2 in coil 2 will cause a flux-linkage MI_2 to coil 1. The quantity M is called the *mutual inductance* of the two coils. If the current in either coil changes at a rate $\mathrm{d}I/\mathrm{d}t$ then, by Faraday's law, there will be an emf generated in the other coil of $-M\,\mathrm{d}I/\mathrm{d}t$. The SI unit of mutual inductance is the *henry*, H and $1\ \mathrm{H} = 1\ \mathrm{V\,s\,A^{-1}}$ (or $1\ \mathrm{Wb\,A^{-1}}$).

For a single coil, the passage of a current will cause a magnetic flux to link to the coil itself. If unit current causes a flux linkage L,

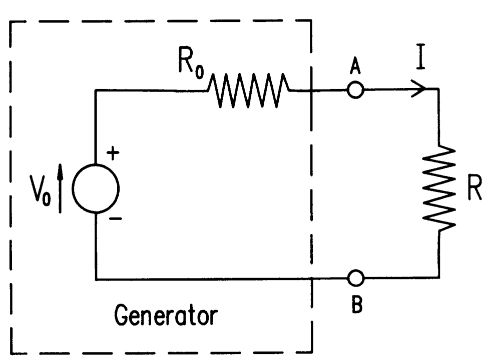

Fig. 1.2 Generator with internal resistance R_0.

then a change in the current at a rate $\mathrm{d}I/\mathrm{d}t$ will generate a back emf in the coil given by

$$V_{\mathrm{back}} = -L\frac{\mathrm{d}I}{\mathrm{d}t}$$

L is called the self-inductance of the coil (or *inductor*) and will also have the same unit, the henry, as M above.

Since a potential $V = -V_{\mathrm{back}}$ will be needed to drive the current I through the coil, work will be done by the potential, equal to $I(L\,\mathrm{d}I/\mathrm{d}t)\,\mathrm{d}t$ in time $\mathrm{d}t$. As the current increases from zero to I, the work done will build up stored magnetic energy in the coil equal to

$$W = \int_0^I IL\,\mathrm{d}I = \tfrac{1}{2}LI^2$$

When two coils of self-inductance L_1 and L_2 are coupled tightly so that all the flux from one circuit links with the other, it can be shown that the mutual inductance between them is

$$M = (L_1L_2)^{1/2} \tag{1.1}$$

In general there will be flux leakage and then $M = k\,(L_1L_2)^{1/2}$ where $0 \leqslant k \leqslant 1$ and k is called the *coefficient of coupling.*

1.6 CAPACITANCE

Any conductor raised to a potential V will store a charge Q which will increase proportionately with the potential. Thus it is possible to write $Q = CV$ where C is the *capacitance* of the conductor. The SI unit for capacitance is the farad, F, where $1\ \mathrm{F} = 1\ \mathrm{C\,V^{-1}}$. Similarly, if a *capacitor* is formed by placing two conductors in proximity, then the charge stored in the capacitor will be given by the same relation where V is the pd between the conductors and C is the capacitance of the capacitor so formed. Because the farad is a large unit, sub-multiples such as the microfarad, μF ($1\ \mu\mathrm{F} = 10^{-6}\ \mathrm{F}$), or the nanofarad, nF ($1\ \mathrm{nF} = 10^{-9}\ \mathrm{F}$), or the picofarad, pF ($1\ \mathrm{pF} = 10^{-12}\ \mathrm{F}$), are generally used.

Since the current is $\mathrm{d}Q/\mathrm{d}t$ and $Q = CV$, the work done in time $\mathrm{d}t$ in charging a capacitor will be $VI\,\mathrm{d}t = V(\mathrm{d}Q/\mathrm{d}t)\,\mathrm{d}t = V\,\mathrm{d}Q = VC\,\mathrm{d}V$ and the energy stored in the capacitor at the potential V will be

$$W = \int_0^V VC\,\mathrm{d}V = \tfrac{1}{2}CV^2$$

4

Alternating current theory

Alternating current, or ac, theory is concerned with the mathematical analysis of the steady-state behaviour of electrical circuits in which the currents and voltages vary periodically with time. The analysis is simplified by considering only sinusoidal variations, an approach which is not restrictive since any general periodic waveform can be represented as a sum of such quantities, i.e. a Fourier series. In this chapter, it is shown how the sinusoidal waveforms can be represented both graphically and mathematically and how, in consequence, the effect of various circuit elements can be expressed in terms of generalized impedances.

4.1 GRAPHICAL REPRESENTATION OF AC VOLTAGES

An ac voltage which varies sinusoidally with time can be represented by the equation

$$V = V_0 \sin \omega t \tag{4.1}$$

and can be represented graphically (Fig. 4.1(a)), where V is the *instantaneous voltage* or *potential* at time t, V_0 is the *peak* or *maximum value* or *amplitude* of the voltage, and ω is the *angular frequency* or *pulsatance* of the wave.

The *period* T, the time for one complete cycle such that $\omega T = 2\pi$, is given by

$$T = 2\pi/\omega \tag{4.2}$$

The number of cycles or periods occurring in unit time is the

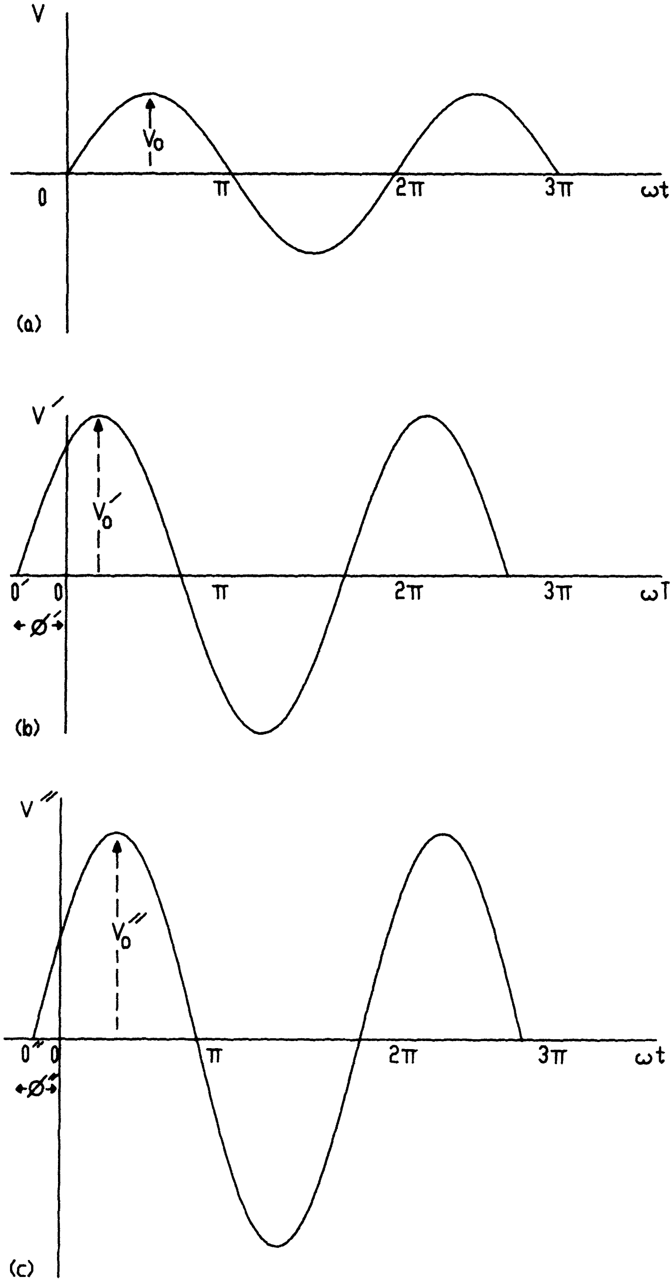

Fig. 4.1 (a) Representation of $V = V_0 \sin \omega t$. (b) Representation of $V' = V'_0 \sin(\omega t + \phi')$. (c) Representation of $V'' = V + V' = V''_0 \sin(\omega t + \phi'')$.

frequency of the wave, f, and is given by

$$f = T^{-1} = \omega/2\pi \tag{4.3}$$

In SI units, pulsatance ω is measured in *radians per second* or *rad* s^{-1} and the period T is quoted in *seconds* or, if appropriate, submultiples of a second, e.g. milliseconds (ms) $\equiv 10^{-3}$ s, microsecond (μs) $\equiv 10^{-6}$ s and nanoseconds (ns) $\equiv 10^{-9}$ s. Frequency f, which has dimensions s^{-1}, is quoted in *hertz* (Hz) or in multiples such as kilohertz (kHz) $\equiv 10^3$ Hz, megahertz (MHz) $\equiv 10^6$ Hz and gigahertz (GHz) $\equiv 10^9$ Hz.

A second voltage V', having the same frequency as V but a different phase and a peak value V'_0 (Fig. 4.1(b)), can be represented mathematically by the equation

$$V' = V'_0 \sin(\omega t + \phi') \tag{4.4}$$

where ϕ' is the *phase difference* between V and V'. The curve for V' is said to *lead* that for V by an angle ϕ', the lead being expressed by the positive value of ϕ' in (4.4). Conversely, V is said to *lag behind*, or simple *lags* V' by ϕ'. The phase angle is normally expressed in degrees, or multiples or fractions of π.

4.2 GRAPHICAL ADDITION OF AC WAVEFORMS

If the two voltages V and V' are applied simultaneously and in series to a circuit, the resultant voltage V'' is the instantaneous sum of V and V'. The result of adding the two curves represented in Figs 4.1(a) and 4.1(b) is shown in Fig. 4.1(c) and it can be seen that V'' (i) has a peak value V''_0, (ii) has the same frequency as its two components, V and V', and (iii) leads V by the angle ϕ'' corresponding to the separation 0″0. Consequently, the resultant voltage V'' can be expressed as

$$V'' = V' \sin(\omega t + \phi'') \tag{4.5}$$

This method of obtaining the resultant V'' of two voltages V and V' is clearly tedious, and is likely to be inaccurate unless the addition is made from digitized waveforms on a computer.

4.3 ALGEBRAIC ADDITION OF AC VOLTAGES AND CURRENTS

The instantaneous values of V, V' and V'' in Fig. 4.1 at a time t are related by $V'' = V + V'$ or

$$V'' = V_0 \sin \omega t + V'_0 \sin(\omega t + \phi')$$

The expansion $\sin(\omega t + \phi') = \sin \omega t \cos \phi' + \cos \omega t \sin \phi'$ gives

$$V'' = (V_0 + V'_0 \cos \phi') \sin \omega t + (V'_0 \sin \phi') \cos \omega t$$

Using the well-known result that $(a \sin \theta + b \cos \theta) = (a^2 + b^2)^{1/2} \times \sin(\theta + \delta)$ where $\delta = \tan^{-1}(b/a)$, it is possible to write

$$V'' = V''_0 \sin(\omega t + \phi'') \tag{4.6}$$

where

$$V''_0 = (V_0^2 + V_0'^2 + 2V_0 V'_0 \cos \phi')^{1/2} \tag{4.7}$$

and

$$\phi'' = \tan^{-1}\left(\frac{V'_0 \sin \phi'}{V_0 + V'_0 \cos \phi'}\right) \tag{4.8}$$

This method, although more accurate and simpler than the graphical summation, is only true when the waveforms are precisely sinusoidal—a rare occurrence in practice.

If two sinusoidal currents $I_0 \sin \omega t$ and $I'_0 \sin(\omega t + \phi')$ were to be added together, the resultant I'' would have a form corresponding to (4.6) to (4.8) with the peak currents I_0 and I'_0 replacing V_0 and V'_0.

4.4 PHASOR REPRESENTATION AND THE ADDITION OF AC VOLTAGES AND CURRENTS

It is well known that a sinusoidal motion will be generated by the projection of a vector rotating about a point with constant angular frequency. Such a vector 0A, of length V_0, is shown in Fig. 4.2 rotating at angular frequency ω in an anticlockwise sense: the projection OP can be taken to represent a voltage $V = V_0 \sin \omega t$. Expressed in this way, the vector 0A is referred to as a *phasor*.

If the two voltages $V = V_0 \sin \omega t$ and $V' = V'_0 \sin(\omega t + \phi')$ are both represented by phasors (Fig. 4.3), then the voltage V'' is represented by the phasor which is the instantaneous vector sum of V and V' and is obtained by completing the parallelogram 0ACB. Since V and V' have the same angular frequency, ω, so will their resultant V''. The *rotating phasor diagram* (Fig. 4.3) can be applied to both voltages and currents.

When each phasor has the same frequency, only the resultant amplitude and phase are of interest, and these may be obtained from

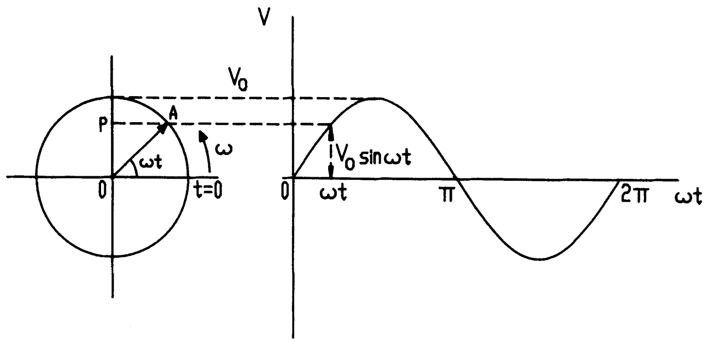

Fig. 4.2 Sine wave generated by the projection of a rotating phasor OA. (Conventionally, anti-clockwise rotation is assumed.)

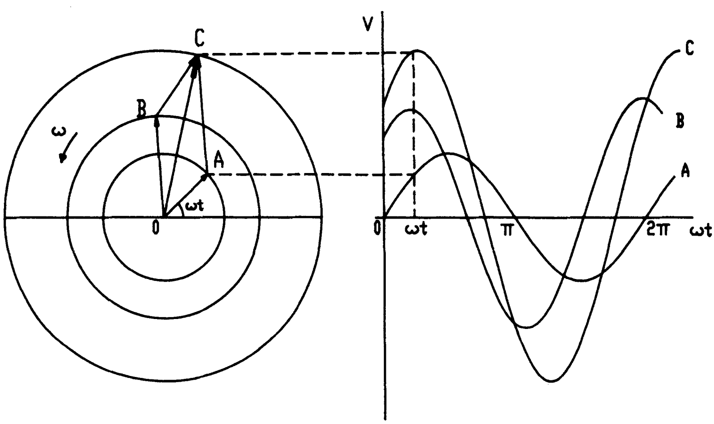

Fig. 4.3 Summation of phasors with $OA \equiv V$, $OB \equiv V'$ and $OC \equiv V''$ such that $\mathbf{OC} = \mathbf{OA} + \mathbf{OB}$. In the curves which are generated, curve A represents $V = V_0 \sin \omega t$, curve B, $V' = V'_0 \sin(\omega t + \phi')$, and curve C, $V'' = V''_0 \sin(\omega t + \phi'')$.

the phasor diagram taken at an instant in time, i.e. the *stationary phasor diagram.* The simplest such diagram is shown in Fig. 4.4, from which the resultant phasors may be derived geometrically in terms of

$$V''_0 = (V_0^2 + V_0'^2 + 2V_0 V'_0 \cos \phi')^{1/2}$$

and

$$\phi'' = \tan^{-1}\left(\frac{V'_0 \sin \phi'}{V_0 + V'_0 \cos \phi'}\right)$$

corresponding to (4.7) and (4.8).

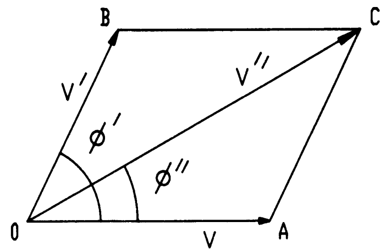

Fig. 4.4 Stationary phasor diagram.

The phasor method can be readily extended to the summation of several voltages or currents. Consider the voltages

$$V_i = V_{i0} \sin(\omega t + \phi_i); \quad i = 1, 2, 3, 4, \ldots$$

each with the same frequency but different amplitudes and phase. If the horizontal axis is taken to correspond to $\phi = 0$ (Fig. 4.5) then the horizontal component of the resultant will be given by

$$V_{\mathrm{HO}} = \sum_i V_{i0} \cos \phi_i$$

where the summation includes all of the terms. The vertical component corresponding to $\phi = \pi/2$ will be

$$V_{\mathrm{VO}} = \sum_i V_{i0} \sin \phi_i$$

The amplitude of the resultant will be

$$V_{\mathrm{RO}} = (V_{\mathrm{HO}}^2 + V_{\mathrm{VO}}^2)^{1/2} \tag{4.9}$$

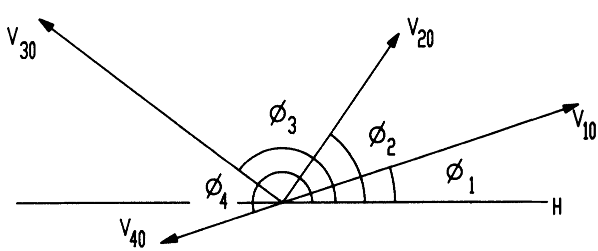

Fig. 4.5 Representation of several phasors.

while its phase angle will be

$$\phi_R = \tan^{-1}\left(\frac{V_{VO}}{V_{HO}}\right) \tag{4.10}$$

Inspection will show that (4.6)–(4.8) are a special case of this result.

4.5 RESISTANCE, SELF-INDUCTANCE AND CAPACITANCE

Consider a sinusoidally varying current $I = I_0 \sin \omega t$ which flows in turn through a resistor, an inductor and a capacitor as shown in Fig. 4.6. The instantaneous voltages required to cause the current to flow through each element are given by the following.

1. For the *resistor* (resistance R)

$$V_R = IR = I_0 R \sin \omega t = V_{R0} \sin \omega t \tag{4.11}$$

2. For the *inductor* (self-inductance L)

$$\begin{aligned} V_L &= L\frac{dI}{dt} \\ &= L\frac{d}{dt}(I_0 \sin \omega t) \\ &= I_0 \omega L \cos \omega t = I_0 \omega L \sin\left(\omega t + \frac{\pi}{2}\right) \\ &= V_{L0} \sin\left(\omega t + \frac{\pi}{2}\right) \end{aligned} \tag{4.12}$$

3. For the *capacitor* (capacitance C)

$$\begin{aligned} V_C &= Q/C \\ &= \frac{\int I\,dt}{C} = \frac{\int I_0 \sin \omega t\,dt}{C} = -\frac{I_0 \cos \omega t}{\omega C} = \frac{I_0}{\omega C}\sin\left(\omega t - \frac{\pi}{2}\right) \\ &= V_{C0} \sin\left(\omega t - \frac{\pi}{2}\right) \end{aligned} \tag{4.13}$$

From (4.11) it is seen that V_R and I are in phase and are simply related by Ohm's law, $V_R = IR$. Equation (4.12) shows that the

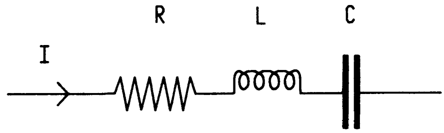

Fig. 4.6 Series R, C, L circuit with $I = I_0 \sin \omega t$.

voltage across the inductor leads the current by $\pi/2$ or 90° and that peak voltage is given by

$$V_{L0} = I_0 \omega L = I_0 X_L \tag{4.14}$$

where $X_L = \omega L$ is called the inductive reactance. Equation (4.13) shows that the voltage V_C across the capacitor lags the current by $\pi/2$ or 90° and the peak voltage is given by

$$V_{C0} = \frac{I_0}{\omega C} = I_0 X_C \tag{4.15}$$

where $X_C = 1/\omega C$ is called the capacitive reactance.

The various voltages for the series circuit can be represented on a voltage phasor diagram (Fig. 4.7(a)), in which the peak current is represented by the horizontal phasor I_0. The peak voltage across the resistor is in phase with the current and the phasor which represents it will therefore be parallel to I_0 and have magnitude $V_{R0} = I_0 R$. However, the peak voltage across the inductor leads the current by 90° and the phasor which represents it will be 90° 'ahead' of I_0 in the phasor diagram and will have a magnitude of $V_{L0} = I_0 \omega L$. Similarly, the phasor V_{C0} of magnitude $I_0/\omega C$ which is in the negative sense relative to V_{L0} will completely represent the peak voltage across the capacitor, since V_{C0} lags I_0 by 90°. Figure 4.7(b) illustrates how

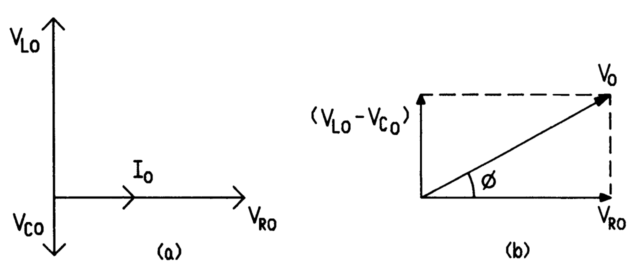

Fig. 4.7 Phasor diagram for R, C, L circuit.

the resultant peak voltage across the series combination of R, L and C can be determined by combining the total vector normal to I_0 namely $V_{L0} - V_{C0}$ with the phasor parallel to I_0 namely V_{R0}: it is seen that the resultant V_0 leads the current by the angle ϕ. Geometrically,

$$V_0 = [V_{R0}^2 + (V_{L0} - V_{C0})^2]^{1/2} = I_0[R^2 + (\omega L - 1/\omega C)^2]^{1/2} \quad (4.16(a))$$

and

$$\phi = \tan^{-1}\left(\frac{\omega L - 1/\omega C}{R}\right) \quad (4.16(b))$$

whence the voltage V across the series combination is

$$\begin{aligned} V &= V_0 \sin(\omega t + \phi) \\ &= I_0 |Z| \sin(\omega t + \phi) \end{aligned} \quad (4.17)$$

where

$$|Z| = \left[R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2\right]^{1/2} \quad (4.18)$$

$|Z|$ is called the *magnitude of the impedance* of the series circuit and represents the ratio of the peak voltage to the peak current, i.e. $|Z| = V_0/I_0$. The reason for writing $|Z|$ will become apparent in section 4.6 below.

Remembering that $X_L = \omega L$ and $X_C = 1/\omega C$, (4.18) can be written as

$$|Z| = (R^2 + X^2)^{1/2} \quad (4.19)$$

while (4.16(b)) becomes

$$\phi = \tan^{-1}\left(\frac{X}{R}\right) \quad (4.20)$$

Equations (4.19) and (4.20) are general expressions for the impedance and phase angle of an ac circuit and in both equations X represents the total reactance of the circuit, i.e. $X = X_L - X_C$.

More complicated circuits can also be treated by phasor methods. As a general rule, however, the phasor diagram is best set up relative to that quantity—current or voltage—which is common to the elements of the circuit, i.e. the current for a series circuit, the voltage for a parallel circuit: in the latter case a *current* phasor diagram will result.

4.6 COMPLEX REPRESENTATION OF AC VOLTAGES AND CURRENTS—THE j NOTATION

Treatment of ac circuits is greatly simplified by the use of complex numbers. Use is made of the two well-known results

$$e^{j\theta} = \cos\theta + j\sin\theta$$

and

$$e^{\pm j\pi/2} = \cos\left(\pm\frac{\pi}{2}\right) + j\sin\left(\pm\frac{\pi}{2}\right) = \pm j$$

so that the operator $j = \sqrt{-1}$ is equivalent to a positive or anticlockwise rotation of 90° and $-j$ represents a negative or clockwise rotation of 90°. The usual notation for $\sqrt{-1}$, namely i, was replaced in network theory by j, to avoid confusion with currents.

The function $\sin\omega t$ can be replaced by the imaginary part of $e^{j\omega t}$ and $\cos\omega t$ by the real part of $e^{j\omega t}$. The mathematical treatment then proceeds as if $e^{j\omega t}$ were the actual time variation.

The method is best illustrated by applying it to the series R, L, C circuit treated in section 4.5. Let the current in Fig. 4.6 be written as

$$I = I_0 e^{j\omega t} = I_0(\cos\omega t + j\sin\omega t) \tag{4.21}$$

The instantaneous total voltage will be

$$V = V_R + V_L + V_C = IR + \frac{L dI}{dt} + \frac{\int I\,dt}{C} \tag{4.22}$$

where $\int I\,dt = Q$, the charge. Substituting for I from (4.21) gives

$$\begin{aligned} V &= I_0 e^{j\omega t} R + L I_0 j\omega e^{j\omega t} + \frac{I_0 e^{j\omega t}}{j\omega C} \\ &= \left(R + j\omega L + \frac{1}{j\omega C}\right) I \\ &= ZI \end{aligned}$$

where Z is the *complex impedance* given by

$$Z = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{4.23}$$

and V and I are the complex voltage and current respectively.

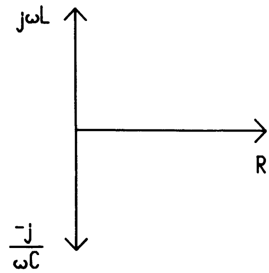

Fig. 4.8 Argand diagram showing the 'rotation' of the impedances due to L and C.

In complex notation, the impedance due to the inductance is $j\omega L$ and that due to the capacitance is $1/j\omega C$ or $-j/\omega C$. On an Argand diagram, in which imaginary quantities are represented by coordinates at right angles to the real axis (Fig. 4.8), the inductive impedance is shown to lead R by 90° and the capacitive impedance to lag by the same amount. The lead or lag of 90° is represented in the corresponding impedance equation by the positive and negative sign of j, respectively.

The total complex impedance given by (4.23) can be written as

$$Z = |Z| e^{j\phi}$$

whence

$$|Z| = \left[R^2 + \left(\omega L - \frac{1}{\omega L} \right)^2 \right]^{1/2} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{\omega L - 1/\omega C}{R} \right)$$

which correspond to the values obtained in (4.18) and (4.16(b)) by phasor methods.

In a general circuit the impedance can be written

$$Z = R + jX \tag{4.24}$$

so that $|Z| = (R^2 + X^2)^{1/2}$ and $\phi = \tan^{-1}(X/R)$ as in (4.19) and (4.20), and X is the reactance of the circuit. Thus the time dependence of the voltage V is given by

$$V = ZI = |Z| e^{j\phi} I_0 e^{j\omega t} = |Z| I_0 e^{j(\omega t + \phi)} \tag{4.25}$$

and the phase angle ϕ is derived from the impedance Z. If the actual current flowing is $I_0 \sin \omega t$ = imaginary part$(I_0 e^{j\omega t}) = \text{Im}(I_0 e^{j\omega t})$ then V is the imaginary part of (4.25) and

$$V = |Z| I_0 \sin(\omega t + \phi) \tag{4.26}$$

and $V_0 = I_0|Z|$ as in (4.17). In (4.26), V represents the instantaneous voltage across the circuit. Similarly, if $I = I_0 \cos \omega t = \text{real part}(I_0 e^{j\omega t}) = \text{Re}(I_0 e^{j\omega t})$ then (4.25) gives

$$V = |Z| I_0 \cos(\omega t + \phi) \tag{4.27}$$

with the same values for $|Z|$ and ϕ as above.

4.7 MANIPULATION OF COMPLEX IMPEDANCES

If there are two complex quantities such as $Z_1 = |Z_1|e^{j\phi_1}$ and $Z_2 = |Z_2|e^{j\phi_2}$ then

$$Z_1 Z_2 = (|Z_1||Z_2|)\,e^{j(\phi_1+\phi_2)} \tag{4.28}$$

and

$$\frac{Z_1}{Z_2} = \left(\frac{|Z_1|}{|Z_2|}\right) e^{j(\phi_1-\phi_2)} \tag{4.29}$$

(Thus $|Z_1Z_2| = |Z_1||Z_2, |Z_1/Z_2| = |Z_1|/|Z_2|$ since $|e^{j\phi}| = 1$ always.)

For convenience, the complex expression can be abbreviated so that

$$Z = |Z|\,e^{j\phi} = |Z|/\phi \tag{4.30}$$

Suppose that a voltage $V = V_0 e^{j(\omega t+\phi_1)}$, where V_0 is a real quantity, is applied across an impedance $Z = |Z|\,e^{j\phi_2}$. The current through the impedance is given by

$$I = \frac{V}{Z} = \frac{V_0 e^{j(\omega t+\phi_1)}}{|Z|\,e^{j\phi_2}} = \frac{V_0}{|Z|}\,e^{j(\omega t+\phi_1-\phi_2)} \tag{4.31}$$

or

$$I = I_0 e^{j(\omega t+\phi)} \tag{4.32}$$

with $I_0 = V_0/|Z|$, a real quantity, and $\phi = \phi_1 - \phi_2$.

Similarly, if the current through a circuit of impedance $|Z|\,e^{j\phi_2}$ is $I_0 e^{j(\omega t+\phi_1)}$, then the voltage across the impedance is

$$V = I_0|Z|\,e^{j(\omega t+\phi_1+\phi_2)} = V_0 e^{j(\omega t+\phi)} \tag{4.33}$$

Since the time variation $e^{j\omega t}$ is common to all terms, the current and voltage are often abbreviated into the form of (4.30) so that

$$V = V_0 e^{j\phi_1} = V_0/\phi_1$$

and

$$Z = |Z|\,e^{j\phi_2} = |Z|/\phi_2$$

and

$$I = \frac{V}{Z} = \frac{V_0\underline{/\phi_1}}{|Z|\underline{/\phi_2}} = \frac{V_0}{|Z|}\underline{/\phi_1 - \phi_2} \tag{4.34}$$

4.8 MUTUAL INDUCTANCE IN AN AC CIRCUIT

When the magnetic flux from a coil links with a second coil, a change of current in the first, or primary, coil will cause an emf to be induced in the secondary coil; the value of the induced emf will be determined by the laws of electromagnetic induction. For the present purposes, it is possible to consider two coils, 1° and 2° in Fig. 4.9, as forming a mutual inductance M. The emf generated in the secondary 2° is given by

$$V_2 = -M\frac{dI_1}{dt} \tag{4.35}$$

where I_1 is the current in the primary 1°. If $I_1 = I_{10}e^{j\omega t}$ then

$$V_2 = -j\omega M I_{10}e^{j\omega t} = -j\omega M I_1 \tag{4.36}$$

Since the coupling of the magnetic field between the two coils is reciprocal, the emf V_1 generated in the primary by an alternating current in the secondary will be given by $V_1 = -j\omega M I_2$.

Clearly the action of the mutual inductance can be represented by an operator $-j\omega M$, where $-j$ indicates a voltage lag of 90° behind the current I_1. The peak secondary emf is $\omega M I_{10} = V_{20}$.

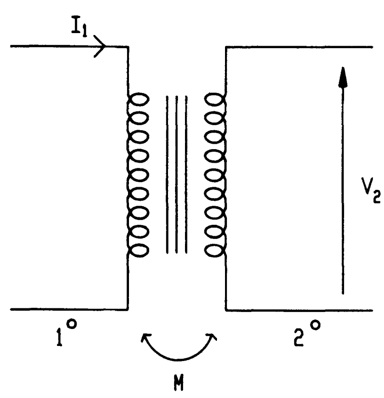

Fig. 4.9 Schematic diagram of a mutual inductance.

4.9 KIRCHHOFF'S LAWS IN AC CIRCUITS

In ac circuits it is necessary to express voltages and currents in terms of their magnitude and phase difference and this can be achieved by expressing them as complex quantities. Thus, in order that the use of Kirchhoff's laws may be extended to ac circuits, it follows that they need to be stated in terms of complex quantities:

1. *current law*—the algebraic sum of the complex currents at any node in a circuit is zero;
2. *voltage law*—in any loop the algebraic sum of the complex potential differences across the impedances in the loop is equal to the algebraic sum of the complex emfs acting in that loop.

4.10 COMPLEX IMPEDANCES IN SERIES AND PARALLEL

The derivation of the total impedance due to a combination of complex impedances is entirely analogous to the dc case treated in section 2.3.

4.10.1 Series circuit

Suppose that a current I flows through three impedances Z_1, Z_2 and Z_3 as shown in Fig. 4.10. The total voltage V will be the sum of the individual voltages across each impedance so that, by Kirchhoff's laws,

$$V = IZ_1 + IZ_2 + IZ_3 \tag{4.37}$$

If the combination of impedances is to be equivalent to a single impedance Z then

$$V = IZ \tag{4.38}$$

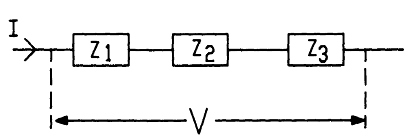

Fig. 4.10 Current I flowing in turn through Z_1, Z_2 and Z_3.

Thus from (4.37) and (4.38)

$$Z = Z_1 + Z_2 + Z_3 \tag{4.39}$$

The general case will be

$$Z = \sum_n Z_n. \tag{4.40}$$

4.10.2 In parallel

If three impedances are connected in parallel (Fig. 4.11), a voltage V placed across them will generate currents I_1, I_2 and I_3 which, by Kirchhoff's laws, must in sum equal the total current I flowing into the circuit and must also satisfy the condition that $V = I_1 Z_1 = I_2 Z_2 = I_3 Z_3$. Thus

$$I = I_1 + I_2 + I_3 = \frac{V}{Z_1} + \frac{V}{Z_2} + \frac{V}{Z_3} \tag{4.41}$$

If the parallel combination of impedances is to be equivalent to a single impedance Z then

$$I = V/Z \tag{4.42}$$

and, from (4.41) and (4.42),

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3} \tag{4.43}$$

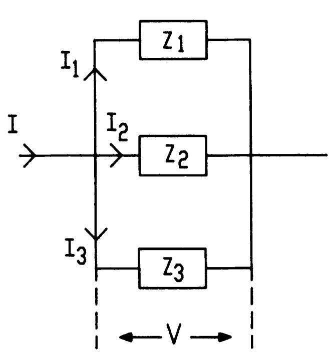

Fig. 4.11 Current divided between three parallel impedances.

The general case will, therefore, be

$$\frac{1}{Z}=\sum_n \frac{1}{Z_n} \tag{4.44}$$

4.11 ADMITTANCE OF AN AC CIRCUIT

The reciprocal of the impedance Z of a circuit is defined as the *admittance* Y of the circuit; thus

$$Y=1/Z \tag{4.45}$$

and, by definition, the current and voltage are related by

$$I=YV \tag{4.46}$$

Since $Z=R+\mathrm{j}X$

$$Y=\frac{1}{R+\mathrm{j}X}=\frac{R-\mathrm{j}X}{R^2+X^2} \tag{4.47}$$

The admittance Y can be expressed in terms of a *conductance* G and a *susceptance* B so that

$$Y=G+\mathrm{j}B \tag{4.48}$$

Hence, from (4.47) and (4.48)

$$G=\frac{R}{R^2+X^2} \quad \text{and} \quad B=\frac{-X}{R^2+X^2} \tag{4.49}$$

(Note that only if $X=0$ will $G=1/R$ and only if $R=0$ will $B=-1/X$.)

It is generally more convenient to work in terms of admittances when a number of elements are connected in parallel and also when applying nodal analysis to a circuit.

4.12 ROOT MEAN SQUARE VALUES OF AC QUANTITIES

If, at a given instance, a current I flows in a resistor of resistance R, the instantaneous rate of energy loss due to Joule heating is

$$p=I^2R$$

If the current varies over a period of time, the *mean power*, $\bar{p}$, dissipated can be expressed as

$$\bar{p}=\overline{I^2R}=\overline{I^2}R$$

where the bar indicates the mean value over the given period of time and $\overline{I^2}$ is called the mean square value of I.

For the purpose of considering power dissipation, an effective current may be defined as I_{rms}, the *root mean square* or *rms* current, where

$$I_{rms} = \sqrt{\overline{I^2}} \tag{4.50}$$

Thus

$$\bar{p} = I_{rms}^2 R \tag{4.51}$$

I_{rms} is the magnitude of the dc current which would produce the same average Joule heating as the varying current I. It is usual, therefore, to quote ac voltages and currents in terms of their rms values; so that when the emf of the mains supply is quoted as 240 V, it is understood that this is actually the rms value.

For a current $I = I_0 \sin \omega t$

$$I_{rms}^2 = \overline{I^2} = \overline{I_0^2 \sin^2 \omega t} = \overline{\tfrac{1}{2} I_0^2 (1 - \cos 2\omega t)}$$

Over a complete cycle, $\cos 2\omega t$ averages to zero so that

$$I_{rms}^2 = \tfrac{1}{2} I_0^2$$

or

$$I_{rms} = I_0/\sqrt{2} \quad \text{and} \quad I_0 = \sqrt{2} I_{rms} \tag{4.52}$$

Similarly, it follows that

$$V_{rms} = V_0/\sqrt{2} \quad \text{and} \quad V_0 = \sqrt{2} V_{rms} \tag{4.53}$$

Thus a voltage of 240 V (rms) will have a peak value of about 340 V.

From (4.52) and (4.53), it follows that any relationship established between I_0 and V_0 will equally apply between I_{rms} and V_{rms}; for example, if $V_0 = I_0 |Z|$ then $V_{rms} = I_{rms} |Z|$.

4.13 POWER IN AC CIRCUITS

The current I which flows through an impedance, or a network of impedances, will generally not be in phase with the voltage V across it. This fact can be expressed by writing $V = V_0 \sin \omega t$ and $I = I_0 \sin(\omega t - \phi)$ where $-\pi/2 \leqslant \phi \leqslant \pi/2$.

The instantaneous power generated in the impedance will then be

$$\begin{aligned} p = VI &= V_0 I_0 \sin \omega t \sin(\omega t - \phi) \\ &= V_0 I_0 (\sin^2 \omega t \cos \phi - \sin \omega t \cos \omega t \sin \phi) \\ &= \tfrac{1}{2} V_0 I_0 [(1 - \cos 2\omega t) \cos \phi - \sin 2\omega t \sin \phi] \end{aligned} \tag{4.54}$$

The component $\frac{1}{2}V_0I_0(1 - \cos 2\omega t)\cos\phi$ is never less than zero and is called the *instantaneous active power*: it represents the power transferred from the generator to the impedance or network at the given time t. The component $-\frac{1}{2}V_0I_0 \sin 2\omega t \sin\phi$ can have either sign and is called the *instantaneous reactive power* and represents the continual interchange of power between the generator and the reactive part of the impedance or network; the time average of this term is zero.

The *average power* supplied to the network or impedance over one cycle will be

$$P = \tfrac{1}{2} V_0 I_0 \cos\phi = V_{rms} I_{rms} \cos\phi \tag{4.55}$$

since

$$\overline{(1 - \cos 2\omega t)} = 1$$

The product $V_{rms}I_{rms}$ is called the *apparent power* S, and is measured in *voltamperes* (VA) to distinguish it from real power which is measured in watts. The quantity $\cos\phi$ is known as the *power factor* of the impedance or network and $0 \leqslant \cos\phi \leqslant 1$ giving $\cos\phi = 1$ for a pure resistance, $\cos\phi = 0$ for a pure reactance.

For completeness, since the power factor does not indicate the sign of ϕ, it is usual to state the power factor as either a lagging ($\phi > 0$) or a leading ($\phi < 0$) power factor.

The *reactive power* Q, which is said to be in quadrature with the real or average power, is given by

$$Q = \tfrac{1}{2} V_0 I_0 \sin\phi = V_{rms} I_{rms} \sin\phi \tag{4.56}$$

and is measured in *var* (voltamperes reactive) or, in practical use, kvar. Since $\sin\phi$ can be positive or negative, it is usual to quote Q as a positive quantity and indicate its value in var (inductive) for $\phi > 0$ and var (capacitive) for $\phi < 0$.

To summarize, the apparent power is

$$S = V_{rms} I_{rms} = I_{rms}^2 |Z| \text{ VA}$$

the average power is

$$P = V_{rms} I_{rms} \cos\phi = I_{rms}^2 R \text{ W}$$

the reactive power is

$$Q = V_{rms}I_{rms}\sin\phi = I_{rms}^2 X \text{ var}$$

and

$$S^2 = P^2 + Q^2 \tag{4.57}$$

From (4.54) the power at an instant t can be written

$$p = P(1 - \cos 2\omega t) - Q\sin 2\omega t \tag{4.58}$$

and the frequency of p is twice that of V or I.

Figure 4.12(a) represents the variation of the current and voltage in a circuit such that the current lags the voltage by angle ϕ. The corresponding instantaneous power p is shown in Fig. 4.12(b) and it should be noted that, when p is positive, energy is being transferred from the generator to the circuit and, when p is negative, energy is being returned from the circuit to the generator.

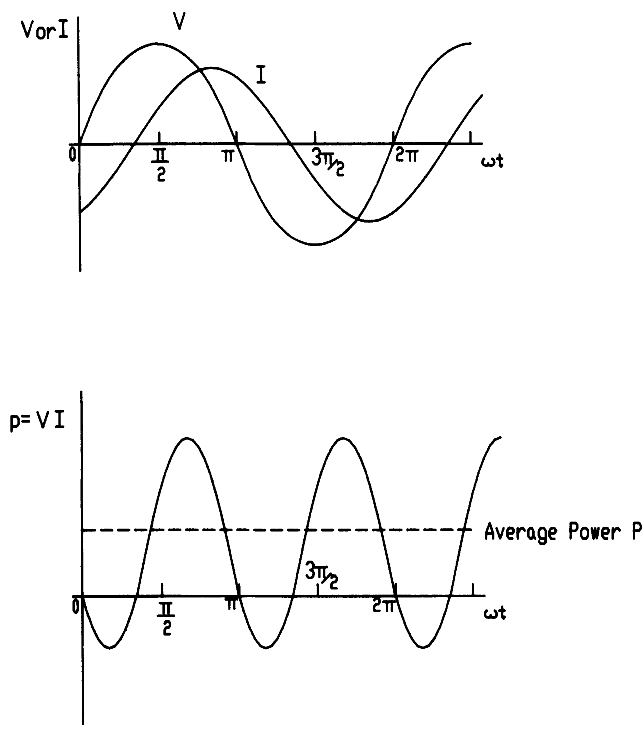

Fig. 4.12 Instantaneous power p for voltage $V = V_0 \sin\omega t$ and current $I = I_0 \sin(\omega t - \phi)$.

4.14 COMPLEX POWER AND THE POWER TRIANGLE

In order to express the power in a circuit in complex form, the dc equivalent voltage is written as

$$V = V_{\text{rms}}\,e^{j\phi_1}$$

and the current as

$$I = I_{\text{rms}}\,e^{j\phi_2}$$

where both V_{rms} and I_{rms} are taken to be real.

Because it is the difference in phase angle which determines the power factor, the *complex power* S' is obtained by using the complex conjugate of I, i.e.

$$S' = VI^* = V_{\text{rms}}e^{j\phi_1}I_{\text{rms}}e^{-j\phi_2} = V_{\text{rms}}I_{\text{rms}}\,e^{j\phi} \tag{4.59}$$

where ϕ is the phase difference. From (4.59), the complex power S' becomes

$$S' = V_{\text{rms}}I_{\text{rms}}\cos\phi + jV_{\text{rms}}I_{\text{rms}}\sin\phi = P + jQ \tag{4.60}$$

where P and Q are, from (4.55) and (4.56), the average and reactive power respectively. Since $|S'| = (P^2 + Q^2)^{1/2} = S$, the complex power may be written as

$$S' = S\underline{/\phi} = S\underline{/\phi_1 - \phi_2} \tag{4.61}$$

The components of the power may thus be represented diagrammatically as in Fig. 4.13, where the figures ABC are said to be the *power triangles.*

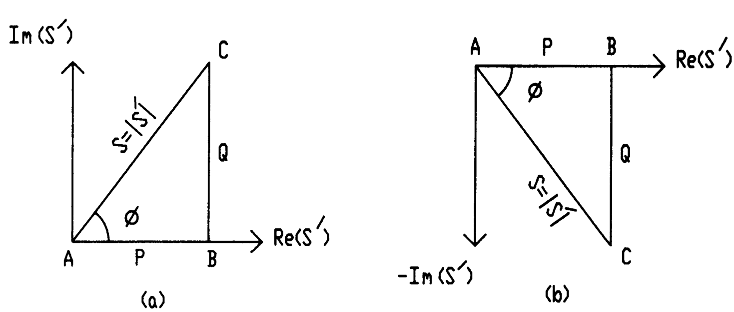

Fig. 4.13 Representation of the complex power: (a) $\cos\phi$ lagging, Q inductive; (b) $\cos\phi$ leading, Q capacitive.

4.15 POWER USAGE: IMPROVEMENT OF THE POWER FACTOR

Consider a supply voltage V_{rms} connected to a number of impedances connected in parallel as in Fig. 4.14. Each impedance can be regarded as representing, for example, the electrical network of a private or commercial consumer.

The total complex power S_T supplied by the source is

$$S'_T = V_{rms} I_{1rms} + V_{rms} I_{2rms} + \cdots + V_{rms} I_{nrms}$$

which, from (4.59), becomes

$$S'_T = P_1 + jQ_1 + P_2 + jQ_2 + \cdots + P_n + jQ_n = P_T + jQ_T \quad (4.62)$$

where P_T is the total average power and Q_T is the total reactive power.

The total apparent power $S_T = |S'_T|$ is given by

$$S_T = (P_T^2 + Q_T^2)^{1/2} \quad (4.63)$$

while the total power factor is

$$\cos \phi_T = P_T / Q_T \quad (4.64)$$

(It is a simple matter to show that (4.62), (4.63) and (4.64) apply equally to impedances connected in series.)

Equation (4.62) shows that the power triangles for the individual impedances or networks can be connected in sequence, as shown in Fig. 4.15, to form a total power triangle, ABC. The figure has been drawn for Q_1 and Q_2 inductive, Q_4 capacitive, whilst Q_3 is zero indicating a purely resistive load. The total power triangle has $AB = P_T$, $BC = Q_T$ and $AC = S_T$ whilst $\angle BAC = \phi_T$ corresponds to a total power factor $\cos \phi_T$ lagging.

The suppliers of mains electricity rate their alternators, transformers and power lines in terms of apparent power, kVA or MVA,

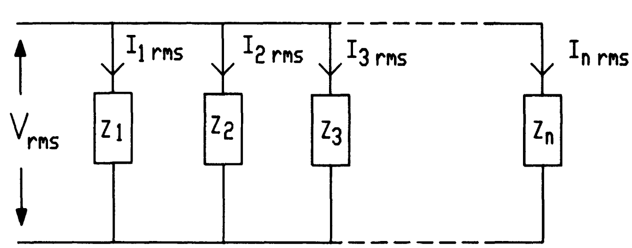

Fig. 4.14 Power supplied to networks in parallel.

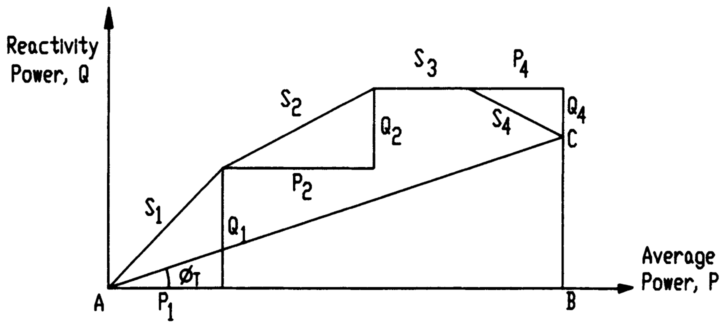

Fig. 4.15 The total power triangle.

rather than average power (kW or MW) and charge accordingly. Any consumer, therefore, who operates equipment which has a complex impedance as opposed to a pure resistance, is charged for more than the real energy consumed. Thus, from the point of view of both the supplier and the consumer, energy is wasted if the power factor differs from unity.

Much electrical equipment has a reactive power component and, for most, this component is inductive. For large industrial users, there may be advantages in 'improving the power factor' by seeking to cancel out the reactive power by introducing a network with a Q of the opposite sign: in most cases this would mean introducing a bank of capacitors in parallel with the source of supply to reduce the inductive component. If Q_T could be made zero, the running costs to the consumer would be reduced, thus enabling some of the generating capabilities of the supplier to be made available to other customers. However, the installation of such a capacitor bank rests solely on the economic decision of the consumer, since he has to compare the cost of installation with the savings to be realized.

4.16 IMPEDANCE MATCHING

When a voltage generator which is connected to a load of impedance Z_L has an internal impedance Z_G, the power delivered to the load will depend on the relationship of Z_G to Z_L. When the power delivered is a maximum, the load is said to be *matched*.

Consider the simplest case shown as in Fig. 4.16 and let

$$Z_G = R_G + jX_G \quad \text{and} \quad Z_L = R_L + jX_L$$

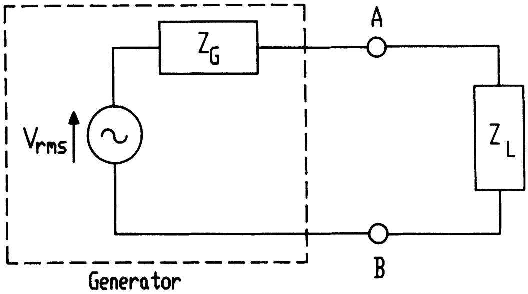

Fig. 4.16 Generator with internal impedance Z_G connected to a load Z_L.

The rms current will be

$$I_{rms} = \frac{V_{rms}}{|Z_G + Z_L|} = V_{rms}[(R_G + R_L)^2 + (X_G + X_L)^2]^{-1/2}$$

The average power dissipated in the load will be

$$P = I_{rms}^2 R_L = V_{rms}^2 R_L[(R_G + R_L)^2 + (X_G + X_L)^2]^{-1} \qquad (4.65)$$

It is possible to think of maximizing the power under three practical conditions: either X_L is fixed and R_L is variable, or R_L and X_L are both variable but their ratio is fixed, i.e. the phase angle ϕ_L of the load is fixed, or R_L and X_L are both independently variable.

1. X_L is fixed and R_L is variable.
   From (4.65)

$$\frac{dP}{dR_L} = \frac{V_{rms}^2[(R_G + R_L)^2 + (X_G + X_L)^2] - V_{rms}^2 R_L \times 2(R_G + R_L)}{[(R_G + R_L)^2 + (X_G + X_L)^2]^2}$$

   For maximum transfer of power to the load $dP/dR_L = 0$. Therefore

$$(R_G + R_L)^2 + (X_G + X_L)^2 - 2R_L(R_G + R_L) = 0$$

   which is satisfied if

$$R_L = [R_G^2 + (X_G + X_L)^2]^{1/2} \qquad (4.66)$$

   From (4.66), if $X_L = 0$ then

$$R_L = |Z_G| \qquad (4.67)$$

2. R_L and X_L are both variable but their ratio X_L/R_L is a constant; $k = \tan \phi_L$, where ϕ_L is the phase angle of the load.

Since $X_L = kR_L$, (4.65) becomes

$$P = V_{rms}^2 R_L[(R_G + R_L)^2 + (X_G + kR_L)^2]^{-1}$$

For maximum power $dP/dR_L = 0$ which requires that

$$(R_G + R_L)^2 + (X_G + kR_L)^2 - R_L[2(R_G + R_L) + 2k(X_G + kR_L)] = 0$$

and this condition is satisfied if

$$R_L = \left(\frac{R_G^2 + X_G^2}{1 + k^2}\right)^{1/2} = |Z_G| \cos \phi_L \qquad (4.68)$$

3. R_L and X_L are independently variable.

It is obvious from (4.65) that, in terms of X_L, the condition for maximum P is $X_L = -X_G$. This may be verified by putting $(\partial P/\partial X_L)_{R_L} = 0$.

The expression for the power reduces to an equivalent of the dc expression, (2.14).

$$P = \frac{V_{rms}^2 R_L}{(R_G + R_L)^2}$$

which gives a maximum when $R_L = R_G$.

It is clear that the maximum power transferable from the generator under these conditions occurs when the load impedance satisfies the conditions $R_L = R_G$ and $X_L = X_G$ or

$$Z_L = Z_G^* \qquad (4.69)$$

i.e. the impedance of the load is equal to the complex conjugate of the impedance of the source.

4.17 COMPLEX FREQUENCY—*s* NOTATION

The analysis of electrical circuits can usefully be extended to the case where the voltage or current variation $e^{j\omega t}$ is replaced by e^{st} where

$$s = j\omega + \sigma = j(\omega - j\sigma)$$

which corresponds to a *complex frequency*, $\omega - j\sigma$.

σ is the real part of s and is called the *neper frequency*: it is measured in *nepers per second* and corresponds to the exponential growth or decay in a signal such that, when $\sigma t = 1$ neper, the signal amplitude changes by a factor e. (Normally, $\sigma < 0$, corresponding to an exponential decay.)

Consider the voltage

$$V = V_{co} e^{st} \tag{4.70}$$

where V_{co} is the complex amplitude of the voltage such that $V_{co} = V_0 e^{j\phi}$ and ϕ corresponds to a phase difference. Then

$$\begin{aligned} V &= V_0 e^{j\phi} e^{(j\omega+\sigma)t} = V_0 e^{\sigma t} e^{j(\omega t+\phi)} \\ &= V_0 e^{\sigma t}[\cos(\omega t+\phi) + j\sin(\omega t+\phi)] \end{aligned} \tag{4.71}$$

Thus

$$\mathrm{Re}(V) = V_0 e^{\sigma t} \cos(\omega t + \phi) \tag{4.72}$$

and

$$\mathrm{Im}(V) = V_0 e^{\sigma t} \sin(\omega t + \phi) \tag{4.73}$$

If $\phi = 0$ (4.70) becomes

$$V = V_0 e^{st} \tag{4.74}$$

Hence

for $s = 0, V = V_0$ i.e. a dc voltage

for $s = \sigma, V = V_0 e^{\sigma t}$ i.e. an exponential voltage (usually $\sigma < 0$ corresponding to a decay)

for $s = j\omega, V = V_0 e^{j\omega t}$ i.e. a sinusoidal voltage

for $s = j\omega + \sigma, V = V_0 e^{(j\omega+\sigma)t}$ i.e. an exponentially varying sinusoidal voltage.

Obviously it is also possible to express current in the same form, and the various ac equations may be generalized as follows:

1. $$V = V_{co} e^{st} \quad \text{and} \quad I = I_{co} e^{st} \tag{4.75}$$

2. $$V = IR \quad \text{and} \quad I = GV \tag{4.76}$$

3. $$V = L\frac{dI}{dt} = L\frac{d}{dt}(I_{co} e^{st}) = sLI \quad \text{and} \quad I = \frac{V}{sL} \tag{4.77}$$

4. $$V = \frac{Q}{C} = \frac{\int I\,dt}{C} = \frac{\int I_{co} e^{st}\,dt}{C} = \frac{I}{sC} \quad \text{and} \quad I = sCV \tag{4.78}$$

5. $$Z(s) = \frac{V(s)}{I(s)} \quad \text{and} \quad Y(s) = \frac{I(s)}{V(s)} \tag{4.79}$$

where $Z(s)$ and $Y(s)$ are the generalized impedance and admittance respectively.