1 Electromagnetism

Radio is a technology that is based upon electromagnetic phenomena and an understanding of electromagnetic theory is crucial to the understanding of radio. Ideas of electricity and magnetism have been in existence for many millennia, but the theory of electromagnetism was the result of a surge in activity over the last four centuries. The development of electromagnetic theory culminated in the Maxwell equations, equations that are crucial to our understanding of radio waves. Radio is an example of the triumph of theoretical science in that it was predicted through theory rather than being discovered by accident. It is the aim of the current chapter to describe electromagnetic theory through its historical development. The chapter has been written for those with very little knowledge in the area and so can be skipped by those who already possess a good knowledge of the subject. However, it is expected that some readers will be a little rusty on the theory and so this chapter will serve as revision for them.

1.1 Electricity

The first recorded observations of electrical effects go back to the Greeks. In the sixth century BC, Thales of Miletus observed that amber, when rubbed, would attract light objects. This phenomenon is exemplified by the old schoolboy trick of rubbing a comb on your trousers and then seeing it lift small scraps of paper. Today we know that matter is made up of atoms which contain particles with positive electric charge (protons), negative electric charge (electrons) and no charge (neutrons). Further, that like charges repel each other and that unlike charges attract. A simple model of a single atom consists of a number of electrons that orbit around a nucleus consisting of the same number of protons and possibly some neutrons (see Figure 1.1). The electrons are arranged in shells around the nucleus, each shell containing electrons of approximately the same energy (the energy increases with radius) and are designated, in order of energy, as K, L, M, N, O, P and Q (it should be noted that the energy gap between these shells is much larger than the range of energies within a shell). Due to quantum mechanical effects, the shells contain only a limited number of electrons (the K shell can contain a maximum of 2 electrons, the L shell 8 electrons, the M shell 18 electrons, the N shell 32, etc.). Matter will consist of a large collection of such atoms which, under normal circumstances, will be in overall electrical neutrality (the numbers of electrons and protons are equal). Under some circumstances, however, it is possible to increase, or decrease, the number

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

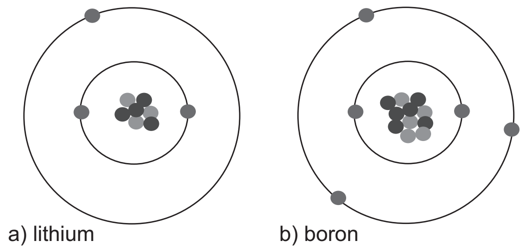

Fig. 1.1 Atomic structure consists of electrons orbiting an equal number of protons and possibly some neutrons.

of electrons and the material will become electrically charged. This is what is achieved in the above rubbing process, sometimes known as the triboelectric effect. The essential condition for the effect to exist is that the materials being rubbed together have different strengths of the force that bind their electrons to the nucleus (glass has a far stronger bond than rubber for example). When the materials are brought together, electrons in the material with the weaker force will be attracted to the material with the stronger force. When the materials are then separated, some of the transferred electrons will remain on the material with the stronger force and both materials will be charged, one positively (the one with the weaker force) and one negatively (the one with the stronger force).

Real matter can be quite complex in structure, with many materials composed of molecules that are complex combinations of different kinds of atoms. The heavier atoms (those with a large number of protons) can have many layers of electrons surrounding the nucleus and this means that the bond of the outer electrons can be relatively low. This can lead to high electron mobility in materials composed of such atoms. Materials for which the electrons are highly mobile, relative to the protons, are known as conductors and are exemplified by metals such as copper, silver and gold. Materials where the electrons are relatively immobile are known as insulators (glass and rubber being important examples). Insulators and conductors turned out to be of great importance in the development of electricity.

The seventeenth and eighteenth centuries were a period of great advances in our knowledge of electrical effects, much of it made possible by increasingly sophisticated machines for developing charged materials through the triboelectric effect. Figure 1.2 shows the basic mechanism of such machines. The rubber belt rolls over the glass cylinder and this causes electrical charge to build up on these components through the triboelectric effect. When the components separate, the belt will be negatively charged and the cylinder positively charged. The negative charges on the belt will eventually reach a conducting brush that sweeps them up onto a conducting metal wire along which they travel until reaching a conducting sphere on which they accumulate. In a similar fashion, the positive charge travels with the cylinder until it reaches a conducting brush. At this brush, the positive charge is neutralised by negative charge that has been drawn from the lower sphere along the conducting wire. In this fashion, positive charge accumulates on the lower sphere. As shown in Figure 1.2, the charge accumulates on opposing faces of the spheres. This occurs due to the mobility of electrons on conductors and the fact that

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

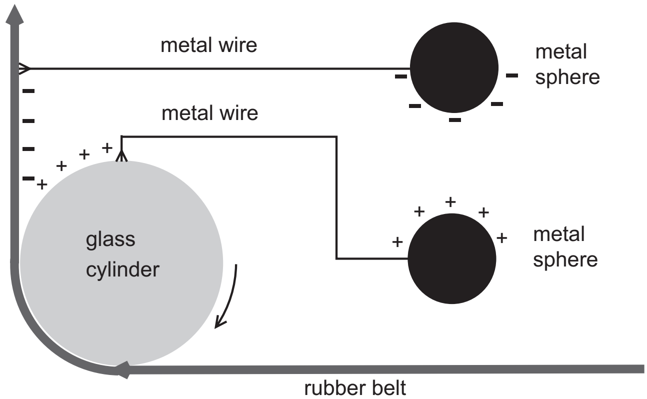

Fig. 1.2 A basic machine for creating positive and negative charge by the triboelectric effect.

opposing charges attract. The medium between the spheres is composed of air and this will tend to act as an insulator and so the charge will just accumulate on the spheres. Furthermore, the charges on the opposing spheres will balance each other out.

If a charged particle is placed between the spheres, it will be drawn towards the sphere with the opposing charge and repulsed by the sphere with the same charge. Consequently, if we want to increase the amount of negative charge on the upper sphere by directly moving positive charge to the lower sphere, this will require an external agency to do some work. This brings us to the important concept of *potential difference*. The potential difference between two points is defined to be the work done by an external force in moving positive charge between these points and is measured in terms of volts (1 volt is 1 joule per coulomb). In order to quantify this, we need to be able to calculate the force that one charge imposes upon another. The force F imposed on charge q by charge Q is given by *Coulomb's law*

$$F = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2}, \tag{1.1}$$

where r is the distance that separates the charges and ϵ_0 is known as the *permittivity* of free space (i.e. space that is devoid of matter). This force is repulsive if the charges have the same sign and attractive if the sign is different. The law was formulated by Charles Augustin de Coulomb in 1784 as the result of much experimental work. The units of charge are known as coulombs, with a proton having a charge 1.60219×10^{-19} coulombs and an electron minus that amount. If distances are measured in metres and the force in newtons, $\epsilon_0 = 8.85 \times 10^{-12}$.

Force is *vector* in nature, i.e. it has both magnitude and direction. Consequently, we need some understanding of vector quantities. Pictorially, we can represent a vector as an arrow that points in the direction of the vector with its length equal to the magnitude (Figure 1.3). Vectors are not only useful for describing quantities such as force, but can also be used for describing the geometrical concept of position. The position of a point can be described by the vector that joins some arbitrary origin to this point, the magnitude being the distance from the origin to the point. An important concept in vectors is that of

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

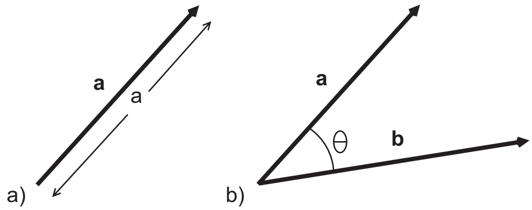

Fig. 1.3 a) Vector represented graphically as an arrow and b) angle between vectors for the vector dot product.

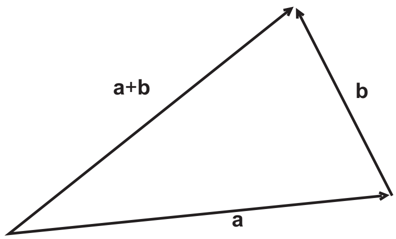

Fig. 1.4 The addition of vectors.

the *dot product* of two vectors $\mathbf{a}$ and $\mathbf{b}$, written as $\mathbf{a}\cdot\mathbf{b}$. If the two vectors have magnitudes a and b, respectively, the dot product is defined to be $ab\cos\theta$ where θ is the angle between these vectors (see Figure 1.3). It can now be seen that $a=\sqrt{\mathbf{a}\cdot\mathbf{a}}$ and $b=\sqrt{\mathbf{b}\cdot\mathbf{b}}$. (Note that we often use $|\mathbf{x}|$ as mathematical shorthand for magnitude $x=\sqrt{\mathbf{x}\cdot\mathbf{x}}$ of the vector.) The dot product can be used to find the component of a force $\mathbf{F}$ in a particular direction. Let $\hat{\mathbf{t}}$ be a unit vector ($|\hat{\mathbf{t}}|=1$) in the direction of interest, then $\hat{\mathbf{t}}\cdot\mathbf{F}$ is the component of force in that direction.

An important operation we can perform on a vector $\mathbf{p}$ is to multiply it by a scalar s to get a new vector $s\mathbf{p}$ that points in the same direction as $\mathbf{p}$ but now has the magnitude sp. Another important operation when we have multiple vectors is their addition. For the vectors $\mathbf{a}$ and $\mathbf{b}$, if we join the tip of the arrow representing $\mathbf{a}$ to the base of the arrow representing $\mathbf{b}$, the sum $\mathbf{a}+\mathbf{b}$ is represented by the arrow from the base of the arrow representing $\mathbf{a}$ to the tip of the arrow representing $\mathbf{b}$ (see Figure 1.4).

In terms of vectors, Coulomb's law can be rewritten as

$$\mathbf{F}=\frac{1}{4\pi\epsilon_0}\frac{qQ}{r^2}\hat{\mathbf{r}}, \tag{1.2}$$

where $\hat{\mathbf{r}}$ is a unit vector ($\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}=1$) in the direction from Q to q. An alternative way of looking at this is to regard charge Q as creating an *electric field* (sometimes known as the *electric intensity*)

$$\boldsymbol{\mathcal{E}}=\frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} \tag{1.3}$$

that pervades space. When a charge q is placed in this field, it is acted upon by a force $q\boldsymbol{\mathcal{E}}$ where $\boldsymbol{\mathcal{E}}$ is the value of the field at the position of charge q ($\boldsymbol{\mathcal{E}}$ will have units of volts

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

per metre). The concept of a field that exists at all points of space was a revolution in thinking and was an extremely important step in the development of electromagnetism.

One can now ask what the field will be when there are charges at a variety of locations. Fortunately, it turns out that this field will simply consist of the sum of the fields due to the individual charges. Consequently, at a position $\mathbf{r}$, a system of N charges has the electric field

$$\mathcal{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{Q_i}{|\mathbf{r}-\mathbf{r}_i|^3}(\mathbf{r}-\mathbf{r}_i), \tag{1.4}$$

where $\mathbf{r}_i$ is the position of the i^{th} charge Q_i and $|\mathbf{r}-\mathbf{r}_i|$ is the distance from $\mathbf{r}_i$ to $\mathbf{r}$.

We now return to the question of the potential difference between points $\mathbf{r}_A$ and $\mathbf{r}_B$. This is the work done in moving a unit charge from a point $\mathbf{r}_A$ to a point $\mathbf{r}_B$. If there is a constant electric field, the work done in moving from point $\mathbf{r}_A$ to $\mathbf{r}_B$ is $-(\mathbf{r}_B-\mathbf{r}_A)\cdot\mathcal{E}$ (i.e. minus the field in the direction of $\mathbf{r}_B$ from $\mathbf{r}_A$ multiplied by the distance in that direction). When moving through the field produced by a finite number of charges, however, the force will vary from point to point. Consequently, we will need to split the path over which the unit charge moves into a number of short segments on each of which the electric field can be regarded as constant (see Figure 1.5). The potential difference will now be approximated by

$$V = -\sum_{i=1}^{M}\mathcal{E}(\mathbf{r}_i)\cdot(\mathbf{r}_i-\mathbf{r}_{i-1}), \tag{1.5}$$

where M is the number of segments. Taking the limit where the segment lengths tend to zero, the above sum becomes the mathematical operation of integration along a line, that is

$$V = -\int_{\mathbf{r}_A}^{\mathbf{r}_B}\mathcal{E}(\mathbf{r})\cdot d\mathbf{r}. \tag{1.6}$$

In the case of our finite system of charge, we will define the *potential* V of the system to be the potential difference when point $\mathbf{r}_A$ is a point at infinity and $\mathbf{r}_B$ is the test point $\mathbf{r}$, then

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\frac{Q_i}{|\mathbf{r}-\mathbf{r}_i|}. \tag{1.7}$$

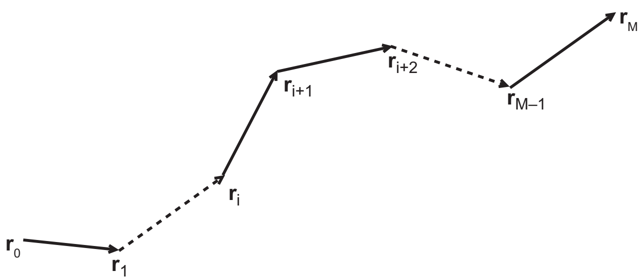

Fig. 1.5 Path for calculating work done when divided into segments ($\mathbf{r}_0=\mathbf{r}_A$ and $\mathbf{r}_M=\mathbf{r}_B$).

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

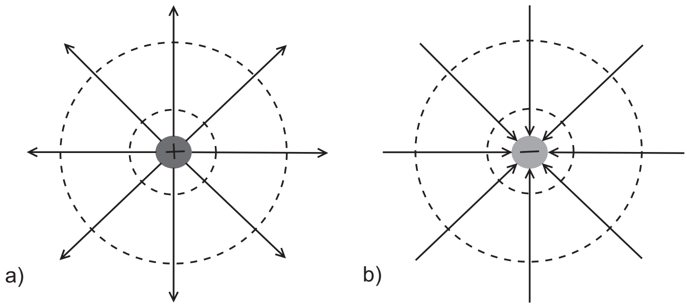

Fig. 1.6 Field lines and lines of constant potential for positive and negative charges.

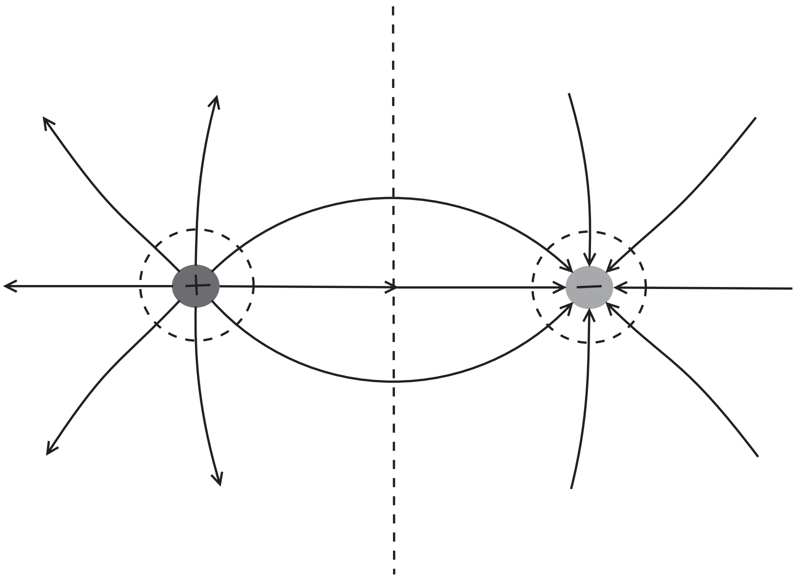

Fig. 1.7 Field lines and lines of constant potential for a dipole.

We can visualise a field in terms of what are known as *field lines*. Such lines have the property that, at any point, their tangent is in the direction of the field at that point. Figure 1.6 shows the field lines for positive and negative charges, the fields run in the radial direction (outwards and inwards respectively). It will be noted that the field lines spread out as we move away from the sources and so the density of field lines at any point is an indication of the strength of the field at that point. Also shown are the surfaces of constant potential (spherical surfaces around the charge that are depicted as broken lines). Figure 1.7 shows the field lines for positive and negative charges of equal magnitude that are separated by a finite distance d. This combination is often known as a dipole and is important in the development of radio theory. At great distances from the dipole the effects of the charges will almost balance out and so the field will be much weaker than

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

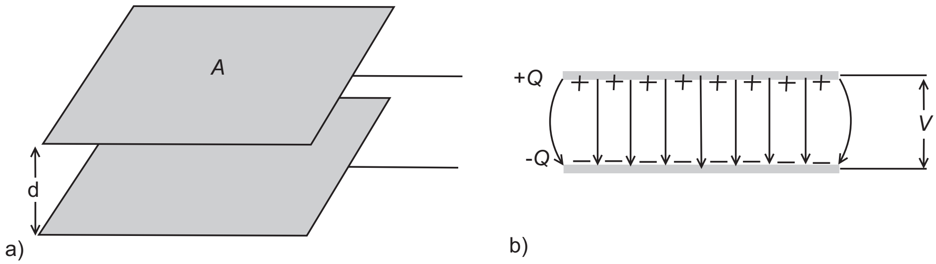

Fig. 1.8 a) Geometry of a parallel plate capacitor and b) field lines in a charged capacitor.

that of a single charge. At great distances, the field will have the form

$$\mathcal{E} = \frac{1}{4\pi\epsilon_0 r^3}(3\hat{\mathbf{r}}\mathbf{p}\cdot\hat{\mathbf{r}} - \mathbf{p}), \tag{1.8}$$

where $\mathbf{p} = Q(\mathbf{r}_+ - \mathbf{r}_-)$ is known as the dipole moment with $\mathbf{r}_+$ and $\mathbf{r}_-$ the positions of the positive and negative charges respectively.

We now return to the configuration of Figure 1.2 and note that the machine causes the accumulation equal numbers of opposite-signed charges, positive on the lower sphere and negative on the upper sphere. The spheres essentially store charge and are an example of an electrical device known as a *capacitor*. It will be noted that the potential on each sphere must be constant. This property follows from the fact that charges can move freely on a conducting sphere and so no further work is needed to move them around on the sphere. It turns out that the charge Q on the lower sphere is proportional to the potential difference $V = V_{+Q} - V_{-Q}$ between the spheres. The constant of proportionality C is known as the *capacitance* ($Q = CV$) and is measured in farads (coulombs per volt). Spheres are not the only capacitors and an important form of capacitor is known as the parallel plate capacitor (see Figure 1.8a). In this device the charge is accumulated on opposing faces of two parallel plates. The field between the plates is mainly constant (magnitude $\mathcal{E} = Q/\epsilon A$), except at the edges, where it adjusts to the zero field outside the capacitor. If the plates are distance d apart and have surface area A, the capacitance will be $C = \epsilon_0 A/d$. This value can be enhanced by inserting an insulating material between the plates. The capacitance will now given by $C = \epsilon A/d$ where ϵ is known as the *permittivity* of the insulator. When an insulator is added (see Figure 1.9a), the molecules become polarised (electrons are drawn towards the positive plate and protons towards the negative plate). The material will then consist of a collection of dipoles that are orientated along the original field line and this causes an additional field that partially counters the original field. The reduced field inside the dielectric will then result in an increased capacitance. The capacitor is an important component in electronic circuits and is represented by the symbol shown in Figure 1.9b.

If we connect the two sides of a capacitor by a conductor, electrons will flow from the negative side to the positive side until all the charge has been neutralised. For a perfect conductor, this will happen instantaneously. In reality, however, conductors are imperfect and there will be some resistance to the flow due to collisions on the molecular scale. The flow through an imperfect conductor is described by Ohm's law, according to which

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

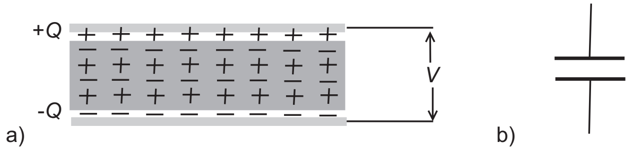

Fig. 1.9 a) Parallel capacitor with dielectric and b) symbol for capacitor.

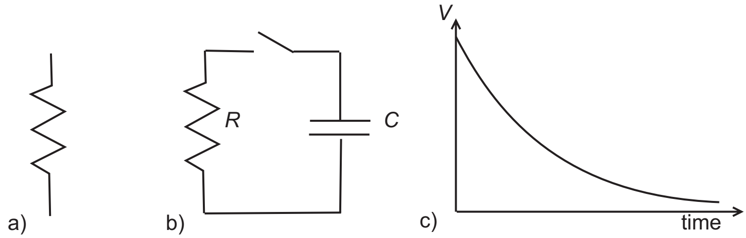

Fig. 1.10 Resistor and a capacitor drained by a resistor.

the potential drop V across the conductor is proportional to the current $\mathcal{I}$ through the conductor. Current is the rate at which charge flows in a conductor and is measured in amperes (1 ampere is 1 coulomb per second). Somewhat confusingly, current has always been taken to be flow of positive charge from higher to lower potential (the opposite direction to the reality of electron flow) and so is the rate of decrease of charge Q on the capacitor plate ($\mathcal{I} = -\mathrm{d}Q/\mathrm{d}t$ in the language of calculus). The constant of proportionality in Ohm's law is known as the resistance R ($V = R\mathcal{I}$) and has units of ohms (1 ohm is 1 amp per volt). George Ohm proposed his famous law in 1827 and it is an important relation in circuit theory. In the case of a wire of length L and cross-sectional area A, the resistance is given by $R = L/A\sigma$ where σ is a material property known as its conductivity.

An imperfect conductor is known as a resistor and is an important component in electronic circuits. A resistor is a lossy device and dissipates energy as heat at a rate $R\mathcal{I}^2$ (this is known as *Ohmic loss*). Figure 1.10b shows a simple circuit consisting of a capacitor and a resistor that dissipates the energy stored in the capacitor (Figure 1.10a shows the symbol used to represent the resistor). When the switch is thrown, a current $\mathcal{I}$ will flow through the resistor and the voltage drop across the capacitor will be given by $\mathcal{V} = R\mathcal{I}$. As the resistor drains the capacitor, the voltage across the capacitor will drop since the charge will be steadily depleted (see Figure 1.10c). Since $Q = C\mathcal{V}$ we will have $\mathcal{I} = -C\mathrm{d}\mathcal{V}/\mathrm{d}t$ and hence $\mathcal{V} = -RC\mathrm{d}\mathcal{V}/\mathrm{d}t$. This is an ordinary differential equation that has the solution $\mathcal{V} = V_0 \exp(-t/RC)$ where V_0 is the initial voltage difference between the capacitor plates and t is the time after switch on.

Much of the early development of the science of electricity was hindered by the need to use machines, such as that shown in Figure 1.2, to generate electric charge. In 1794, however, this process was revolutionised through the invention of the *battery* by Alessandro Volta, a device that creates charge through a chemical process rather than a mechanical process. Figure 1.11 shows a single-cell version of Volta's battery

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

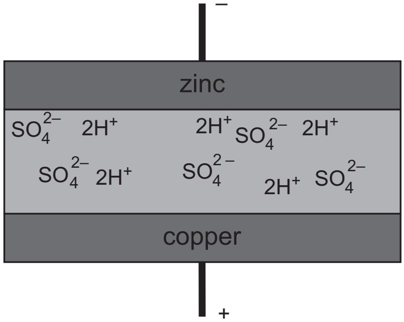

Fig. 1.11 Volta's battery.

(Volta in fact made a stack of these in order to produce large potential differences). It consists of a layer of copper (the anode), a layer of felt that is soaked in a mixture of water and sulphuric acid (the electrolyte) and a layer of zinc (the cathode). Within the electrolyte, the sulphuric acid will disassociate into SO_4^{2-} and H^{2+} ions. At the copper plate electrons are drawn into the electrolyte to combine with hydrogen ions and form hydrogen gas, hence causing an accumulation of positive charge. Meanwhile, at the zinc plate, this is counterbalanced by zinc ions dissolving into the electrolyte, hence causing an accumulation of negative charge. The chemistry can be summarised as

$$Cu + Zn + 3H_2SO_4 = Zn^{2+} + 2SO_4^{2-} + Cu^{2+} + SO_2 + H_2 + 2H_2O. \qquad (1.9)$$

An important concept in electromagnetic theory (and many other field theories) is the concept of *flux*. Consider a flat surface with area A and unit normal $\mathbf{n}$. If $\mathbf{G}$ is a constant vector field, it will have a flux $\mathbf{n} \cdot \mathbf{G}A$ across the surface (i.e. the normal component of the field multiplied by the area of the surface). A good illustration of the notion of flux comes from the study of fluid flow. Such a medium is usually described in terms of its velocity field, a vector field that gives the magnitude and direction of the fluid velocity at a given point. The flux is then the total volume of fluid crossing the surface in a unit time. For a general surface surface S with unit normal $\mathbf{n}$, the *flux* through S is defined by the integral over the surface of the normal component of the vector field, i.e. $\int_S \mathbf{G}(\mathbf{r}) \cdot \mathbf{n} dS$. The surface integral is a calculus concept that can be understood by approximating the surface by a set of small flat surface elements on each of which $\mathbf{n}$ and $\mathbf{G}$ can be approximated by constant values. If the ith element has area ΔS_i, we approximate $\mathbf{n}$ by a constant vector $\mathbf{n}_i$ and $\mathbf{G}$ by a constant vector $\mathbf{G}_i$. The total flux through S is then approximated by the sum of the fluxes $\mathbf{G}(\mathbf{r}_i) \cdot \mathbf{n}_i \Delta S_i$ through these smaller elements, i.e.

$$\text{total flux through } S \approx \sum_{i=1}^{N} \mathbf{G}_i \cdot \mathbf{n}_i \Delta S_i. \qquad (1.10)$$

In the limit of this sum as the areas of the surface elements tend to zero, the above sum then becomes the surface integral $\int_S \mathbf{G}(\mathbf{r}) \cdot \mathbf{n} dS$.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

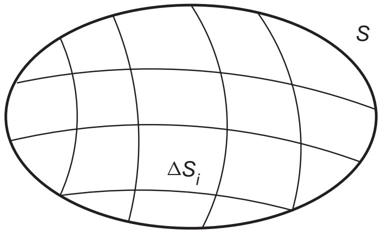

Fig. 1.12 Flux surface integral.

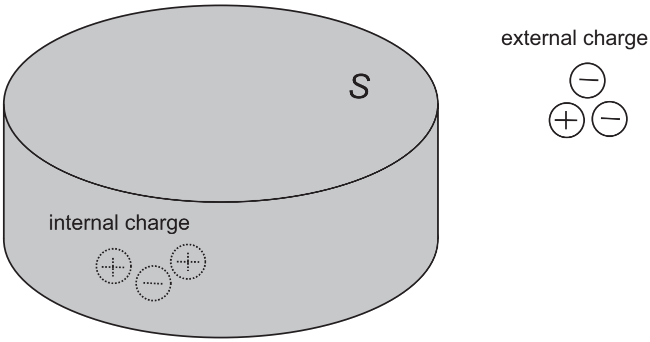

Fig. 1.13 Gauss' law.

An important property of the electric field is that the total flux through a closed surface S is proportional to the charge contained within that surface. This is known as *Gauss' Law* which, in mathematical terms, is given by

$$\int_S \epsilon \boldsymbol{\mathcal{E}}(\mathbf{r}) \cdot \mathbf{n} dS = \text{total charge within } S, \tag{1.11}$$

where S is an arbitrary closed surface in space and $\mathbf{n}$ is unit normal on this surface. Gauss' law is one of the fundamental laws of electromagnetism. A simple example is given by a single charge located at the origin and a surface S that consists of a sphere of radius a with centre at the charge. The field is given by Eq. 1.3 and from which $\boldsymbol{\mathcal{E}} \cdot \mathbf{n} = Q/4\pi\epsilon_0 a^2$ since $\mathbf{n}$ is a unit vector in the radial direction (i.e. the field direction). Since $\boldsymbol{\mathcal{E}} \cdot \mathbf{n}$ is constant, we simply multiply by the area of the sphere ($4\pi a^2$) to get the integral over the sphere. As a consequence $\int_S \boldsymbol{\mathcal{E}}(\mathbf{r}) \cdot \mathbf{n} dS = q/\epsilon_0$, which is Gauss' law.

1.2 Magnetism

At the time of their discovery of electrostatic attraction, the Greeks were also aware that the mineral magnetite (the oxide of iron Fe_3O_4) could attract pieces of non-oxide iron. Further, that the iron itself could be magnetised by stoking with the magnetite. The Chinese were also aware that magnetite (also known as lodestone) was a naturally occurring *magnet* that could attract iron. Indeed, the Chinese also discovered the effect

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

of a magnet orientating itself with respect to Earth. By the twelfth century, both the Chinese and Europeans were using compasses in the form of lodestones for navigation. However, it took until 1600 for the Earth itself to be recognised as having the property of a magnet. This was recognised by William Gilbert in his book 'De Magnete', one of the first works on magnetism. The recognition of the Earth's magnetic properties led to the designation of the two ends of a magnet as North and South. However, unlike electric sources where positive and negative charge can have separate existence, the sources of magnetic fields are always found in North/South pairs. Because of this, the flux of a magnetic field through a closed surface S is zero, i.e.

$$\int_S \mathcal{B}(\mathbf{r}) \cdot \mathbf{n} \mathrm{d}S = 0, \tag{1.12}$$

where $\mathcal{B}$ is the magnetic field (sometimes known as the *magnetic flux density*).

Since the magnetic poles always appear in North and South pairs, the basic source of magnetism is the magnetic dipole. This has a field

$$\mathcal{B} = \frac{\mu_0}{4\pi r^3}(3\hat{\mathbf{r}} \cdot \mathbf{M}\hat{\mathbf{r}} - \mathbf{M}), \tag{1.13}$$

where $\mathbf{M}$ is known as the dipole moment and μ_0 is a constant that is known as the permeability of free space. The dipole will have the field lines shown in Figure 1.14 (also shown is the field lines of Earth's magnetic field for which north is at 79° latitude).

If the basic magnetic source is the dipole, how do we interpret the dipole moment? It turns out that, if we suspend a magnetic dipole of moment $\mathbf{m}$ in field $\mathcal{B}$, the dipole will experience a torque

$$\tau = \mathbf{m} \times \mathcal{B}. \tag{1.14}$$

This is a more complex behaviour than the interaction of an electric charge with an electric field. In particular, it involves a *vector product*, defined by $\mathbf{a} \times \mathbf{b} = ab\sin(\theta)\hat{\mathbf{n}}$ where θ is the angle between the vectors and $\hat{\mathbf{n}}$ is a unit vector that is perpendicular to

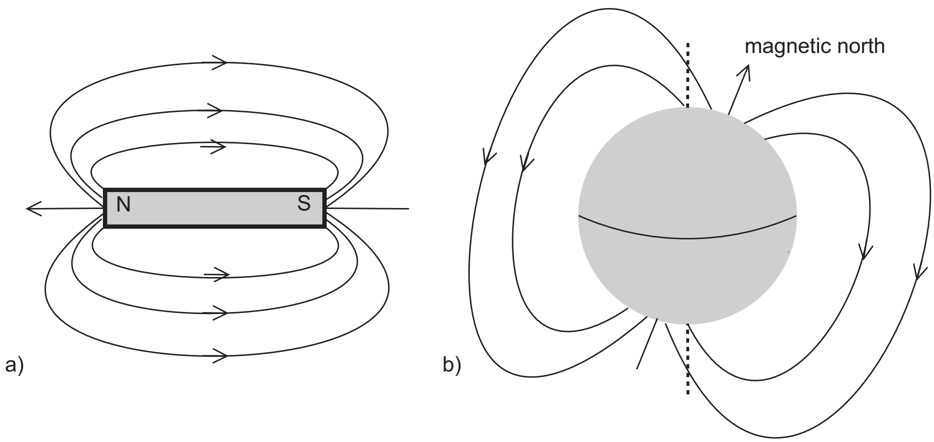

Fig. 1.14 Magnetic field lines.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

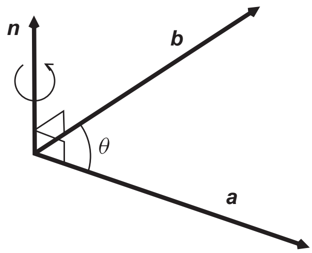

Fig. 1.15 Vector product.

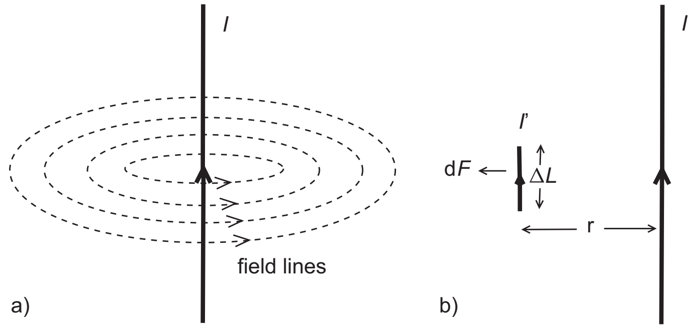

Fig. 1.16 The magnetic field of a current-carrying wire.

both **a** and **b** (direction defined by the right-hand screw rule as shown in Figure 1.15). An important consequence of 1.14 is that a magnet, freely suspended in the magnetic field of Earth, will rotate until it aligns with Earth's field lines (i.e. until the torque becomes zero), an effect that is used in navigation in the guise of a compass.

In 1820, Hans Christian Oersted discovered the magnetic effect of current. By observing the deflections of a compass, he showed that a long straight wire carrying a steady current $\mathcal{I}$ caused a magnetic field that had circular field lines centred on the wire (see Figure 1.16a) and a magnitude $\mathcal{B}$ that depended upon the distance r from the wire

$$\mathcal{B} = \frac{\mu_0 \mathcal{I}}{2\pi r}, \tag{1.15}$$

where $\mathcal{I}$ is the current in the wire. The unit for the magnetic field is usually the tesla, a quantity that is one newton per ampere per metre. Further, in such units, the permeability has the value $4\pi \times 10^{-7}$.

From the work of Oersted, it became clear that the moving charge had the ability to cause a magnetic field. Further, that a current carrying wire could experience the force of a magnetic field. According to Oersted, a wire element of length ΔL carrying a current $\mathcal{I}'$ will suffer a force

$$\Delta \mathbf{F} = \mathcal{I}' \Delta L \mathbf{t} \times \boldsymbol{\mathcal{B}}, \tag{1.16}$$

where $\mathbf{t}$ is a unit vector in the direction of the current.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

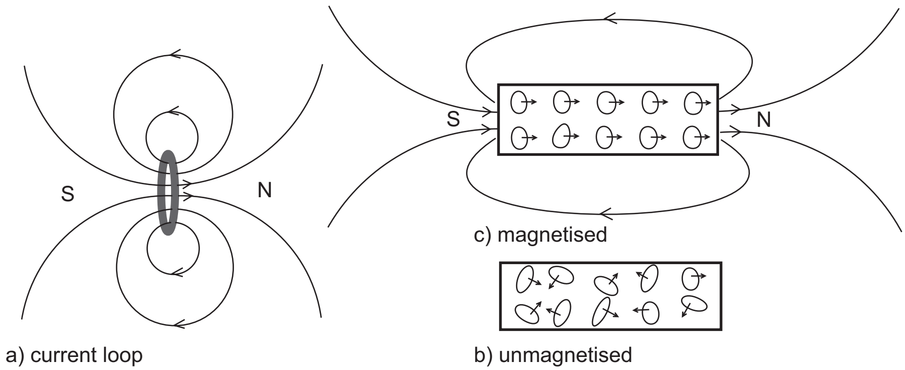

Fig. 1.17 The current loop and the current loop model of magnetism.

The fact that moving charge could produce a magnetic field led to further illumination of the concept of a magnetic dipole. Around 1820, the formula of Oersted was further generalised to allow for an arbitrary circuit C by the work of Jean-Baptiste Biot and Felix Savart. The Biot–Savart formula is

$$\boldsymbol{\mathcal{B}}(\mathbf{r}) = \frac{\mu_0 I}{4\pi r} \int_C \frac{\mathrm{d}\mathbf{r}' \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}. \tag{1.17}$$

For points at a large distance from a planar current loop, this expression reduces to (1.13) with $\mathbf{M} = \mathcal{I}A\mathbf{n}$ where $\mathbf{n}$ is a unit vector perpendicular to the plane of the loop and A is the area of the loop. The magnetic dipole can thus be pictured as a loop of current. In fact, at a molecular level, we can interpret this as orbiting electrons or spinning charge. All matter will consist of many such dipoles, but in most matter these will be in a random configuration and hence have no net effect. However, for materials such as magnetite, these dipoles are aligned with each other and hence the material will exhibit magnetic properties. Materials such as iron can be magnetised when external fields align their dipoles and materials such as steel can retain this magnetism.

Current can be regarded as a stream of charge travelling down a wire and the above considerations suggest that the force $\mathbf{F}$ that acts upon a charge q will be

$$\mathbf{F} = q\boldsymbol{\mathcal{E}} + q\mathbf{v} \times \boldsymbol{\mathcal{B}}, \tag{1.18}$$

where $\mathbf{v}$ is the velocity vector of the charge. This force is often known as the *Lorentz force* and is important for understanding the interaction of matter with the electromagnetic field.

A general law connecting magnetic fields and current was discovered by Andre-Marie Ampère in 1823. Consider a surface S through which current passes and which is bounded by a curve C. *Ampère's law*, in its mathematical form, then states that

$$\int_C \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \mu_0 \mathcal{I}, \tag{1.19}$$

where $\mathcal{I}$ is the total current passing through the surface S. If we consider the case of a long straight wire carrying current $\mathcal{I}$, we could take the curve C to be a circle of radius a that is centred on the wire. In this case, the magnetic field B will be constant on C

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

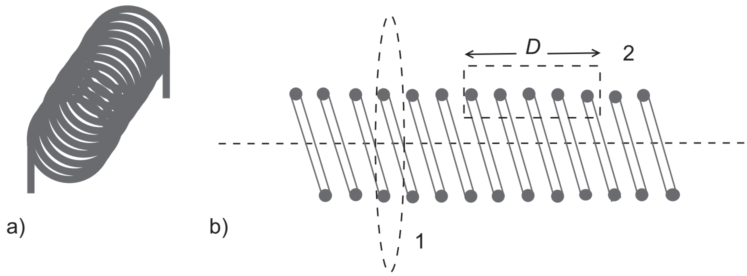

Fig. 1.18 The magnetic field of a long solenoid.

and we will have $\int_C \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = 2\pi \mu_0 a\mathcal{B}$. Substituting from (1.15), the right-hand side becomes $\mu_0 \mathcal{I}$, i.e. we have Ampère's law.

Ampère's law is a useful result for determining complex magnetic fields. Consider the example of an infinitely long solenoid (a good approximation to a long solenoid, as in Figure 1.18a). By symmetry, the only dependence of the magnetic field is the radial distance r from the axis. Further, by analogy with the limiting case of an infinite wire, the field lines outside the wire will be circular and centred on the solenoid axis, i.e. the field will point in the rotational direction. We first apply Ampère's law on a circular curve C with radius r and centred on the axis (the curve 1 in Figure 1.18b). On curve C the magnetic field will be constant and so $\int_C \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = 2\pi r\mathcal{B}$, where $\mathcal{B}$ is the component magnetic field that is tangent to C. Consequently, if current $\mathcal{I}$ flows through the solenoid, Ampère's law will imply $\mathcal{B} = \mu_0 \mathcal{I}/2\pi r$. Inside the solenoid, Ampère's law will imply that the rotational component of the field is zero and so, as a consequence, we take the field to be parallel to the solenoid axis. If we apply Ampère's law on the rectangular curve in Figure 1.18 (curve 2), we find that $\int_C \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = D\mathcal{B}$ where $\mathcal{B}$ is the magnitude of the magnetic field parallel to the axis. The current through the loop will be $nD\mathcal{I}$ where n is the number of turns per unit length on the solenoid. As a consequence, Ampère's law will imply that $\mathcal{B} = \mu_0 n\mathcal{I}$ inside the solenoid. Up to now, we have implicitly assumed that the radial component of the magnetic field is zero, but we can verify this using (1.12). We take the surface S to be a cylinder of radius r, and length D, with the same axis as the solenoid. The magnetic flux through the cylinder ends will cancel, but the contribution from the curved surface will be $2\pi rD\mathcal{B}$, where $\mathcal{B}$ is now the radial component of the magnetic field. Equation 1.12 will then imply that this radial component is zero.

What emerges from our considerations is that the various integral results, such as Ampère's and Gauss' laws, constitute a powerful and self-contained description of electromagnetism. In order to complete this description, however, we need to introduce one further integral law and this is the subject of the next section.

1.3 Electromagnetism

We now consider the consequences of the variation of fields with time. This brings us to Faraday's law, one of the key discoveries in the development of electromagnetic theory.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

In 1830, Michael Faraday discovered magnetic induction when he noted that, by moving a loop of wire in and out of a magnetic field, he could cause a current to flow in the loop. This was an important discovery as, hitherto, the battery and the charged capacitor had been the only means of driving a current through an electrical circuit. Somewhat confusingly, the effective potential of this new sort of generator came to be known as the *electromotive force* (or EMF for short). Faraday concluded that the EMF induced in a circuit was proportional to the rate of change of magnetic flux through that circuit, a result that is known as *Faraday's law*.. For a surface S with bounding curve C (see Figure 1.19), the flux is given by $\Phi = \int_S \mathcal{B}(\mathbf{r},t)\cdot\mathbf{n}dS$ and the law has the mathematical form

$$\int_C \mathcal{E}(\mathbf{r},t)\cdot \mathrm{d}\mathbf{r} = -\frac{d}{dt}\int_S \mathcal{B}(\mathbf{r},t)\cdot\mathbf{n}\mathrm{d}S. \tag{1.20}$$

Figure 1.20 shows two different ways in which magnetic flux can vary in a circuit. If we consider the loop in Figure 1.20a to be rotating at angular speed ω, the flux Φ through the loop will be $\Phi = A\mathcal{B}\sin(\omega t)$ where A is the area of the loop and $\mathcal{B}$ is the magnetic field (we assume the plane of the loop is parallel to the field when $t = 0$). As a consequence, an EMF $A\mathcal{B}\cos(\omega t)$ will be generated and this causes an *alternating current* to flow in the load. In Figure 1.20b a conducting bar with load moves over a rectangular circuit at speed v and so causes the total area of the circuit to change at rate vd. As a consequence, the magnetic flux will increase at a rate $vd\mathcal{B}$ and so, by Faraday's law, an EMF of $-vdB$ will be generated in the circuit ($\mathcal{B}$ is a magnetic field orthogonal to the loop). We can view this last example from the viewpoint of the Lorentz force. A unit charge, located on the conducting bar, will suffer a Lorentz force $v\mathcal{B}$ in the clockwise direction due to

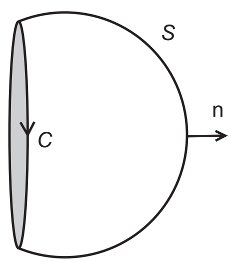

Fig. 1.19 Geometry for Faraday's law.

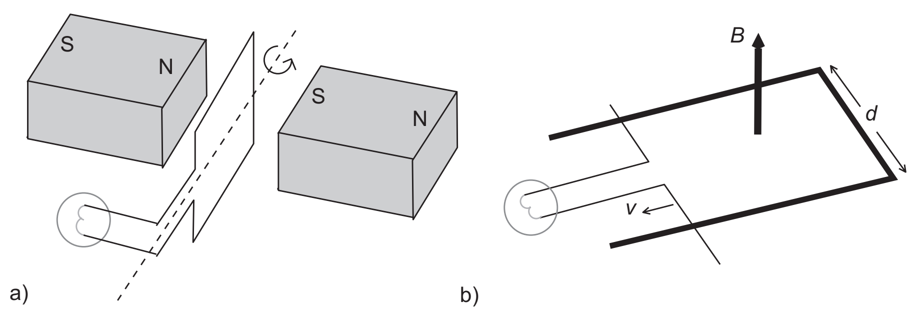

Fig. 1.20 Magnetic induction.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

the imposed motion transverse to the bar. Consequently, integrating along the bar, we obtain an EMF of $-vd\mathcal{B}$.

A time-varying current brings us to the concept of mutual impedance. Consider a solenoid with a wire loop wrapped around it. If we now drive the solenoid by a alternating current $\mathcal{I}_1(t)$, there will be a magnetic field $\mathcal{B}(t) = \mu_0 n \mathcal{I}_1(t)$ through the loop and hence a flux $\Phi = A\mu_0 n \mathcal{I}_1(t)$ where A is the area of the loop. According to Faraday's law, this will generate an EMF of

$$\mathcal{E}_2 = -L_{21}\frac{\mathrm{d}\mathcal{I}_1}{\mathrm{d}t} \tag{1.21}$$

in the wire, where $L_{21} = A\mu_0 n$ is known as the *mutual inductance* of the wire loop and solenoid. If the solenoid has a finite length l with N_1 turns then $n = N_1/l$ and the mutual inductance will be given by $L_{12} = \mu_0 N_1 A/l$. Further, and if the loop has N_2 turns, the mutual inductance will now be given by $L_{12} = \mu_0 N_1 N_2 A/l$. A device with mutually interacting windings is known as a *transformer* and is represented by the symbol shown in Figure 1.21b. The longer solenoid is often known as the primary and the loop winding as the secondary. If current flows through the secondary, it is clear the secondary itself will cause additional flux and so

$$\mathcal{E}_2 = -L_{21}\frac{\mathrm{d}\mathcal{I}_1}{\mathrm{d}t} - L_{22}\frac{\mathrm{d}\mathcal{I}_2}{\mathrm{d}t}, \tag{1.22}$$

where L_{22} is known as the self inductance of the secondary winding. It is clear that the primary will also experience self inductance and that the EMF generated in the primary will take the form

$$\mathcal{E}_1 = -L_{11}\frac{\mathrm{d}\mathcal{I}_1}{\mathrm{d}t} - L_{12}\frac{\mathrm{d}\mathcal{I}_2}{\mathrm{d}t}, \tag{1.23}$$

where it should be noted that $L_{12} = L_{21}$. For the solenoid, it is obvious that it will induce a flux $\Phi = A\mu_0 \mathcal{I}_1(t) A\mu_0 N_1/l$ in itself and so it will have a self inductance $L_{11} = \mu_0 N_1^2 A/l$. Likewise, the secondary will have a self inductance $L_{22} = \mu_0 N_2^2 A/l$. The unit of inductance is the henry and is named after Joseph Henry who discovered magnetic induction independently of Faraday and at about the same time. An inductance of 1 henry will result in an EMF of 1 volt in a closed loop for a change of 1 amp in the current over a period of 1 second.

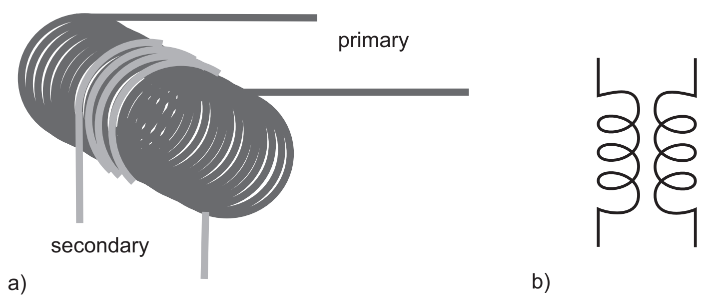

Fig. 1.21 Transformer and transformer symbol.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

From the above considerations, it will be noted that a solenoid that carries a current $\mathcal{I}$ will always have a self-induced EMF, even if the secondary winding does not exist. For this reason, such a device is known as an *inductor* and is an important component in radio technology (it is represented by the symbol shown in Figure 1.21a). For such a component we will have the self-induced EMF

$$\mathcal{E} = -L\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t}, \tag{1.24}$$

where $\mathcal{I}$ is the current in the inductor and L is the self inductance ($L = \mu_0 N^2 A/l$ for a solenoid with length l and cross-sectional area A). We can enhance the inductance of the solenoid by winding it around a core made up of ferromagnetic material (iron and cobalt for example). As can be seen in Figure 1.22, the solenoid field causes the current loops within the core to align and this causes an increase in the magnetic flux density. Consequently, the inductance in the solenoid will now become $L = \mu N^2 A/l$, where μ is known as the *permeability* of the core. Permeability is a material property of the solenoid core and has units of henries per metre.

We now consider the circuit shown in Figure 1.23b, consisting of a series capacitor, resistor, inductor and switch. Before the switch is closed we assume the capacitor to be charged to a voltage V_0. After the switch is closed, however, the capacitor discharges and the potential difference across the capacitor will decay.

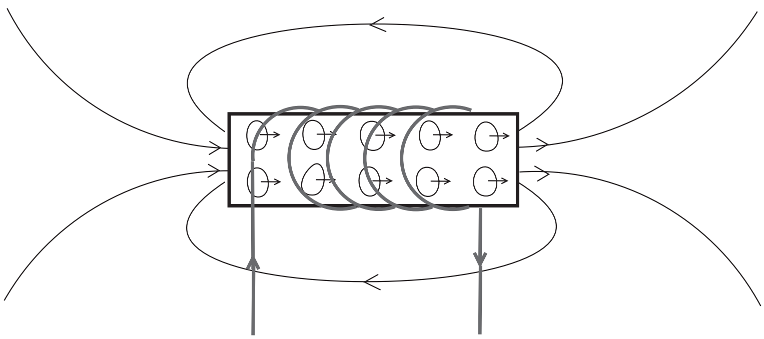

Fig. 1.22 Magnetic flux enhanced by a ferromagnetic material.

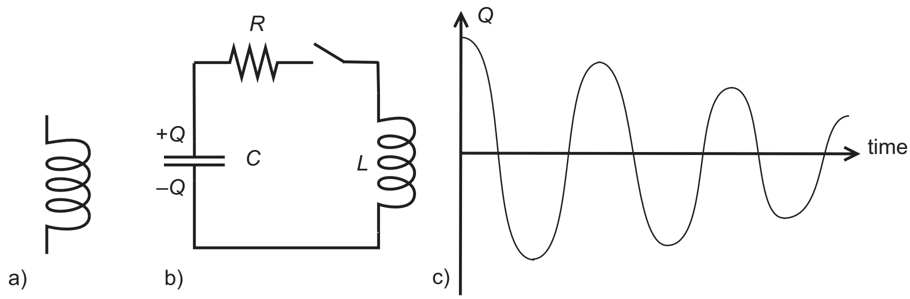

Fig. 1.23 Inductor and oscillating circuit.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

We can analyse the circuit by noting the famous Kirchhoff circuit laws:

1. The total current into any junction is equal to the total current out.
2. The total voltage drop around any circuit loop is zero.

(Since they both have units of volts, voltage is a terminology often used for both potential difference and EMF.) From the first law we obtain that the same current $\mathcal{I}$ flows in and out of all components and from the second law we obtain that

$$-L\frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} - \mathcal{I}R + \frac{Q}{C} = 0. \tag{1.25}$$

In terms of the charge Q on the upper capacitor plate, we have the ordinary differential equation (ODE)

$$L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C} = 0, \tag{1.26}$$

which can be solved to yield

$$Q = V_0 C \exp(-\zeta\omega_0 t)\cos(\sqrt{1-\zeta^2}\omega_0 t), \tag{1.27}$$

where $\omega_0 = 1/\sqrt{LC}$ and $\zeta = R\sqrt{C/L}/2$. It will be noted that when the capacitor is discharged through an inductor, it will *ring*, i.e. there will be oscillations in the circuit at an angular frequency of $\sqrt{1-\zeta^2}\omega_0$. Further, the oscillations will decay at a rate that is dependent upon the amount of resistance R in the circuit.

The frequency ω at which a circuit rings is of great important to us in radio. Consider the parallel combination of a capacitor C, an inductor L and a load resistance R with the inductor driven by harmonic voltage source $V_S\cos(\omega t)$ (see Figure 1.24). By the Kirchhoff current law, we have that

$$\mathcal{I} = C\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} + \frac{\mathcal{V}}{R} \tag{1.28}$$

and from the Kirchoff voltage law

$$\mathcal{V} = V_S\cos(\omega t) - L\frac{d\mathcal{I}}{dt}. \tag{1.29}$$

As a consequence

$$V_S\cos(\omega t) - \mathcal{V} = LC\frac{d^2\mathcal{V}}{dt^2} + \frac{L}{R}\frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} \tag{1.30}$$

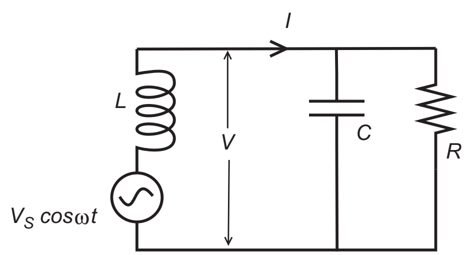

Fig. 1.24 A sinusoidally driven circuit.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

and, dividing by LC, we obtain

$$\frac{d^2\mathcal{V}}{dt^2}+\frac{1}{RC}\frac{d\mathcal{V}}{dt}+\frac{\mathcal{V}}{LC}=\frac{V_S}{LC}\cos(\omega t). \tag{1.31}$$

After the source is switched on, the solution will settle down to the steady state

$$\mathcal{V}(t)=\frac{V_S}{LC}\frac{(\omega_0^2-\omega^2)\cos(\omega t)+\frac{\omega}{RC}\sin(\omega t)}{(\omega_0^2-\omega^2)^2+\frac{\omega^2}{R^2C^2}}, \tag{1.32}$$

where $\omega_0=1/\sqrt{LC}$, i.e. the voltage in the circuit oscillates at the forcing frequency ω. It will be noted, however, that as frequency ω approaches ω_0, the amplitude of oscillations will increase, reaching a peak value of QV_S where $Q=R/\omega L$. The circuit is said to *resonate* at frequency ω_0 and Q is a measure of the strength of this resonance.

1.4 Maxwell's Equations

Until about 1860, Eqs. (1.11), (1.12), (1.19) and (1.20) were presumed to correctly reflect the content of electromagnetic theory. Whilst these equations imply that time-varying magnetic flux will cause an electric field, they do not imply that time-varying electric flux will cause a magnetic field. Around 1860 the physicist James Clerk Maxwell became convinced that time-varying electric flux should cause a magnetic field. Indeed, there are good reasons for believing that Ampère's law needs some form of modification. Consider Ampère's law in form

$$\int_C \boldsymbol{\mathcal{B}}(\mathbf{r})\cdot d\mathbf{r}=\mu_0\mathcal{I}. \tag{1.33}$$

We consider the case of two charged spheres that are made to discharge by connecting them through a conducting wire that, as a result, carries a current $\mathcal{I}$. Referring to Figure 1.25, if we apply Eq. (1.33) using surface S_1 we obtain that $\int_{C_1}\boldsymbol{\mathcal{B}}(\mathbf{r})\cdot d\mathbf{r}=\mu_0\mathcal{I}$ and using surface S_2 we obtain that $\int_{C_2}\boldsymbol{\mathcal{B}}(\mathbf{r})\cdot d\mathbf{r}=0$. If we add these two results, the path integrals will cancel (the curves are identical but the integrals are evaluated in opposite directions) and this will imply that $\mathcal{I}=0$. This clearly poses a problem for electromagnetic theory. Maxwell's solution was to add another term to Eq. (1.33), which he called the *displacement current*. The resulting equation is

$$\int_C \boldsymbol{\mathcal{B}}(\mathbf{r},t)\cdot d\mathbf{r}=\mu_0\frac{d}{dt}\int_S \epsilon\boldsymbol{\mathcal{E}}(\mathbf{r},t)\cdot\mathbf{n}dS+\mu_0\mathcal{I}. \tag{1.34}$$

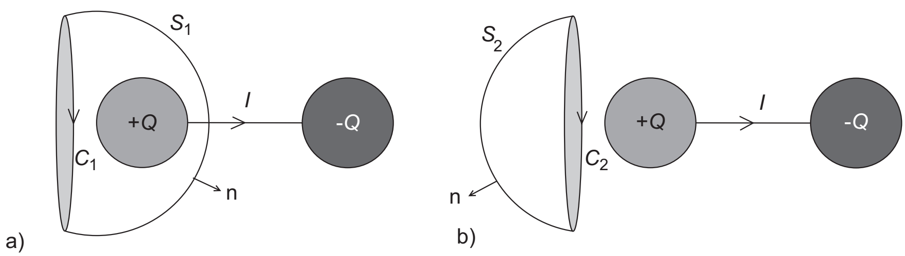

Fig. 1.25 Configuration for testing Ampère's law.

https://doi.org/10.1017/9781108684514.002 Published online by Cambridge University Press

If we now consider this modified Ampère's law on the surfaces S_1 and S_2 we now obtain $\int_{C_1} \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \mu_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_{S_1} \epsilon \boldsymbol{\mathcal{E}}(\mathbf{r}) \cdot \mathbf{n} \mathrm{d}S + \mu_0 \mathcal{I}$ and $\int_{C_2} \boldsymbol{\mathcal{B}}(\mathbf{r}) \cdot \mathrm{d}\mathbf{r} = \mu_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_{S_2} \epsilon \boldsymbol{\mathcal{E}}(\mathbf{r}) \cdot \mathbf{n} \mathrm{d}S$. If we now add these two results, we obtain that $\mu_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_{S} \epsilon \boldsymbol{\mathcal{E}}(\mathbf{r}) \cdot \mathbf{n} \mathrm{d}S + \mu_0 \mathcal{I}$ where $S = S_1 + S_2$. Since S will be a closed surface, Gauss' law will therefore imply that $\mathrm{d}Q/\mathrm{d}t + \mathcal{I} = 0$, the definition of the current, i.e. we now have consistency.

We can now state the integral form of what are known as *Maxwell's equations*, equations that fully describe an electromagnetic (EM) field. The first pair of these four vector equations are:

$$\int_C \frac{\boldsymbol{\mathcal{B}}}{\mu} \cdot \mathrm{d}\mathbf{r} = \frac{\mathrm{d}}{\mathrm{d}t} \int_S \epsilon \boldsymbol{\mathcal{E}} \cdot \mathbf{n} \mathrm{d}S + \mathcal{I} \tag{1.35}$$

and

$$\int_C \boldsymbol{\mathcal{E}} \cdot \mathrm{d}\mathbf{r} = -\frac{\mathrm{d}}{\mathrm{d}t} \int_S \boldsymbol{\mathcal{B}} \cdot \mathbf{n} \mathrm{d}S, \tag{1.36}$$

where S is a surface through which a total current $\mathcal{I}$ flows and C is a contour that is the boundary of S (see Figure 1.19). The second pair are

$$\int_S \boldsymbol{\mathcal{B}} \cdot \mathbf{n} \mathrm{d}S = 0 \tag{1.37}$$

and

$$\int_S \epsilon \boldsymbol{\mathcal{E}} \cdot \mathbf{n} \mathrm{d}S = Q, \tag{1.38}$$

where S is now a closed surface that contains an amount of charge Q. It should be noted that we have included the possibility of a magnetic medium (i.e. $\mu \neq \mu_0$) and, for convenience, have written $\boldsymbol{\mathcal{E}}(\mathbf{r}, t)$ as $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{B}}(\mathbf{r}, t)$ as $\boldsymbol{\mathcal{B}}$.

1.5 Conclusion

In the present chapter we have developed electromagnetic theory up to the integral form of Maxwell's equations. Sources apart (current and charge), these equations possess a remarkable symmetry, something we have come to expect from the laws of physics. Indeed, Maxwell's equations are the only laws of physics to emerge unscathed from the relativity revolution that Einstein unleashed in 1905. It is fair to say that Maxwell's equations were the first truly relativistic theory of physics.

Maxwell showed that his equations predicted wavelike fields and, around 1886, this prediction led the German physicist Heinrich Hertz to construct apparatus that could transmit and receive electromagnetic waves. The wavelengths involved were those that we now commonly associate with radio and so it was with Hertz that radio was born. In the next chapter we will develop the basic theory behind radio and describe the experiments that led to its practical realisation. Further, we will describe some of the early work that helped turn this scientific discovery into one of the most dynamic technologies of the twentieth century.