Chapter One

ELECTRIC CHARGES AND FIELDS

1.1 Introduction

All of us have the experience of seeing a spark or hearing a crackle when we take off our synthetic clothes or sweater, particularly in dry weather. Have you ever tried to find any explanation for this phenomenon? Another common example of electric discharge is the lightning that we see in the sky during thunderstorms. We also experience a sensation of an electric shock either while opening the door of a car or holding the iron bar of a bus after sliding from our seat. The reason for these experiences is discharge of electric charges through our body, which were accumulated due to rubbing of insulating surfaces. You might have also heard that this is due to generation of static electricity. This is precisely the topic we are going to discuss in this and the next chapter. Static means anything that does not move or change with time. *Electrostatics deals with the study of forces, fields and potentials arising from static charges.*

1.2 Electric Charge

Historically the credit of discovery of the fact that amber rubbed with wool or silk cloth attracts light objects goes to Thales of Miletus, Greece, around 600 BC. The name electricity is coined from the Greek word

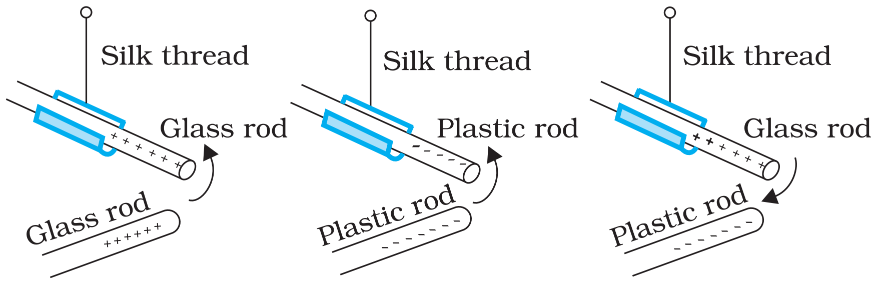


FIGURE 1.1 Rods: like charges repel and unlike charges attract each other.

elektron meaning *amber*. Many such pairs of materials were known which on rubbing could attract light objects like straw, pith balls and bits of papers.

It was observed that if two glass rods rubbed with wool or silk cloth are brought close to each other, they repel each other [Fig. 1.1(a)]. The two strands of wool or two pieces of silk cloth, with which the rods were rubbed, also repel each other. However, the glass rod and wool attracted each other. Similarly, two plastic rods rubbed with cat's fur repelled each other [Fig. 1.1(b)] but attracted the fur. On the other hand, the plastic rod attracts the glass rod [Fig. 1.1(c)] and repel the silk or wool with which the glass rod is rubbed. The glass rod repels the fur.

These seemingly simple facts were established from years of efforts and careful experiments and their analyses. It was concluded, after many careful studies by different scientists, that there were only two kinds of an entry which is called the *electric charge*. We say that the bodies like glass or plastic rods, silk, fur and pith balls are electrified. They acquire an electric charge on rubbing. There are two kinds of electrification and we find that (i) *like charges repel* and (ii) *unlike charges attract* each other. The property which differentiates the two kinds of charges is called the *polarity* of charge.

When a glass rod is rubbed with silk, the rod acquires one kind of charge and the silk acquires the second kind of charge. This is true for any pair of objects that are rubbed to be electrified. Now if the electrified glass rod is brought in contact with silk, with which it was rubbed, they no longer attract each other. They also do not attract or repel other light objects as they did on being electrified.

Thus, the charges acquired after rubbing are lost when the charged bodies are brought in contact. What can you conclude from these observations? It just tells us that unlike charges acquired by the objects neutralise or nullify each other's effect. Therefore, the charges were named as *positive* and *negative* by the American scientist Benjamin Franklin. By convention, the charge on glass rod or cat's fur is called positive and that on plastic rod or silk is termed negative. If an object possesses an electric charge, it is said to be electrified or charged. When it has no charge it is said to be electrically neutral.

Reprint 2025-26

A simple apparatus to detect charge on a body is the *gold-leaf electroscope* [Fig. 1.2(a)]. It consists of a vertical metal rod housed in a box, with two thin gold leaves attached to its bottom end. When a charged object touches the metal knob at the top of the rod, charge flows on to the leaves and they diverge. The degree of divergance is an indicator of the amount of charge.

Try to understand why material bodies acquire charge. You know that all matter is made up of atoms and/or molecules. Although normally the materials are electrically neutral, they do contain charges; but their charges are exactly balanced. Forces that hold the molecules together, forces that hold atoms together in a solid, the adhesive force of glue, forces associated with surface tension, all are basically electrical in nature, arising from the forces between charged particles. Thus the electric force is all pervasive and it encompasses almost each and every field associated with our life. It is therefore essential that we learn more about such a force.

To electrify a neutral body, we need to add or remove one kind of charge. When we say that a body is charged, we always refer to this excess charge or deficit of charge. In solids, some of the electrons, being less tightly bound in the atom, are the charges which are transferred from one body to the other. A body can thus be charged positively by losing some of its electrons. Similarly, a body can be charged negatively by gaining electrons. When we rub a glass rod with silk, some of the electrons from the rod are transferred to the silk cloth. Thus the rod gets positively charged and the silk gets negatively charged. No new charge is created in the process of rubbing. Also the number of electrons, that are transferred, is a very small fraction of the total number of electrons in the material body.

1.3 Conductors and Insulators

Some substances readily allow passage of electricity through them, others do not. Those which allow electricity to pass through them easily are called *conductors*. They have electric charges (electrons) that are comparatively free to move inside the material. Metals, human and animal bodies and earth are conductors. Most of the non-metals like glass, porcelain, plastic, nylon, wood offer high resistance to the passage of electricity through them. They are called *insulators*. Most substances fall into one of the two classes stated above*.

When some charge is transferred to a conductor, it readily gets distributed over the entire surface of the conductor. In contrast, if some charge is put on an insulator, it stays at the same place. You will learn why this happens in the next chapter.

This property of the materials tells you why a nylon or plastic comb gets electrified on combing dry hair or on rubbing, but a metal article

* There is a third category called *semiconductors*, which offer resistance to the movement of charges which is intermediate between the conductors and insulators.

Reprint 2025-26

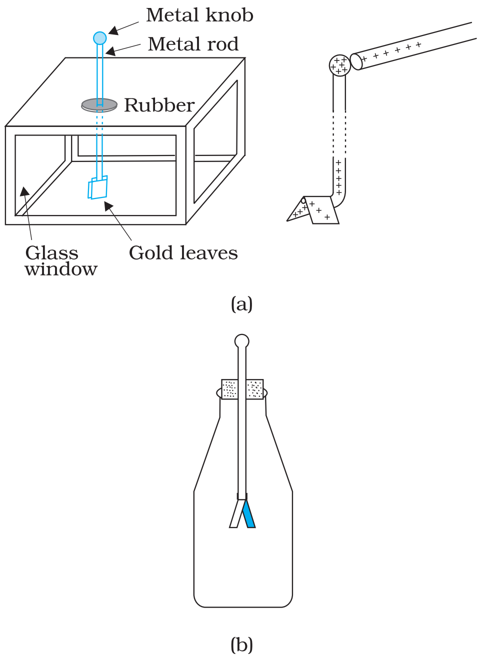


FIGURE 1.2 Electroscopes: (a) The gold leaf electroscope, (b) Schematics of a simple electroscope.

like spoon does not. The charges on metal leak through our body to the ground as both are conductors of electricity. However, if a metal rod with a wooden or plastic handle is rubbed without touching its metal part, it shows signs of charging.

1.4 Basic Properties of Electric Charge

We have seen that there are two types of charges, namely positive and negative and their effects tend to cancel each other. Here, we shall now describe some other properties of the electric charge.

If the sizes of charged bodies are very small as compared to the distances between them, we treat them as *point charges*. All the charge content of the body is assumed to be concentrated at one point in space.

1.4.1 Additivity of charges

We have not as yet given a quantitative definition of a charge; we shall follow it up in the next section. We shall tentatively assume that this can be done and proceed. If a system contains two point charges q_1 and q_2, the total charge of the system is obtained simply by adding algebraically q_1 and q_2 , i.e., charges add up like real numbers or they are scalars like the mass of a body. If a system contains n charges q_1, q_2, q_3, ..., q_n, then the total charge of the system is $q_1 + q_2 + q_3 + ... + q_n$. Charge has magnitude but no direction, similar to mass. However, there is one difference between mass and charge. Mass of a body is always positive whereas a charge can be either positive or negative. Proper signs have to be used while adding the charges in a system. For example, the total charge of a system containing five charges +1, +2, –3, +4 and –5, in some arbitrary unit, is (+1) + (+2) + (–3) + (+4) + (–5) = –1 in the same unit.

1.4.2 Charge is conserved

We have already hinted to the fact that when bodies are charged by rubbing, there is transfer of electrons from one body to the other; no new charges are either created or destroyed. A picture of particles of electric charge enables us to understand the idea of conservation of charge. When we rub two bodies, what one body gains in charge the other body loses. Within an isolated system consisting of many charged bodies, due to interactions among the bodies, charges may get redistributed but it is found that *the total charge of the isolated system is always conserved.* Conservation of charge has been established experimentally.

It is not possible to create or destroy net charge carried by any isolated system although the charge carrying particles may be created or destroyed in a process. Sometimes nature creates charged particles: a neutron turns into a proton and an electron. The proton and electron thus created have equal and opposite charges and the total charge is zero before and after the creation.

1.4.3 Quantisation of charge

Experimentally it is established that all free charges are integral multiples of a basic unit of charge denoted by e. Thus charge q on a body is always given by

$$q = ne$$

where n is any integer, positive or negative. This basic unit of charge is the charge that an electron or proton carries. By convention, the charge on an electron is taken to be negative; therefore charge on an electron is written as $-e$ and that on a proton as $+e$.

The fact that electric charge is always an integral multiple of e is termed as *quantisation of charge*. There are a large number of situations in physics where certain physical quantities are quantised. The quantisation of charge was first suggested by the experimental laws of electrolysis discovered by English experimentalist Faraday. It was experimentally demonstrated by Millikan in 1912.

In the International System (SI) of Units, a unit of charge is called a *coulomb* and is denoted by the symbol C. A coulomb is defined in terms the unit of the electric current which you are going to learn in a subsequent chapter. In terms of this definition, one coulomb is the charge flowing through a wire in 1 s if the current is 1 A (ampere), (see Chapter 1 of Class XI, Physics Textbook , Part I). In this system, the value of the basic unit of charge is

$$e = 1.602192 \times 10^{-19}\,\text{C}$$

Thus, there are about 6×10^{18} electrons in a charge of -1C. In electrostatics, charges of this large magnitude are seldom encountered and hence we use smaller units 1 μC (micro coulomb) = 10^{-6} C or 1 mC (milli coulomb) = 10^{-3} C.

If the protons and electrons are the only basic charges in the universe, all the observable charges have to be integral multiples of e. Thus, if a body contains n_1 electrons and n_2 protons, the total amount of charge on the body is $n_2 \times e + n_1 \times (-e) = (n_2 - n_1)\, e$. Since n_1 and n_2 are integers, their difference is also an integer. Thus the charge on any body is always an integral multiple of e and can be increased or decreased also in steps of e.

The step size e is, however, very small because at the macroscopic level, we deal with charges of a few μC. At this scale the fact that charge of a body can increase or decrease in units of e is not visible. In this respect, the grainy nature of the charge is lost and it appears to be continuous.

This situation can be compared with the geometrical concepts of points and lines. A dotted line viewed from a distance appears continuous to us but is not continuous in reality. As many points very close to

Reprint 2025-26

each other normally give an impression of a continuous line, many small charges taken together appear as a continuous charge distribution.

At the macroscopic level, one deals with charges that are enormous compared to the magnitude of charge *e*. Since $e = 1.6 \times 10^{-19}$ C, a charge of magnituOde, say 1 μC, contains something like 10^{13} times the electronic charge. At this scale, the fact that charge can increase or decrease only in units of *e* is not very different from saying that charge can take continuous values. Thus, at the macroscopic level, the quantisation of charge has no practical consequence and can be ignored. However, at the microscopic level, where the charges involved are of the order of a few tens or hundreds of *e*, i.e., they can be counted, they appear in discrete lumps and quantisation of charge cannot be ignored. It is the magnitude of scale involved that is very important.

Example 1.1 If 10^9 electrons move out of a body to another body every second, how much time is required to get a total charge of 1 C on the other body?

Solution In one second 10^9 electrons move out of the body. Therefore the charge given out in one second is $1.6 \times 10^{-19} \times 10^9$ C = 1.6×10^{-10} C. The time required to accumulate a charge of 1 C can then be estimated to be 1 C ÷ (1.6×10^{-10} C/s) = 6.25×10^9 s = 6.25×10^9 ÷ (365 × 24 × 3600) years = 198 years. Thus to collect a charge of one coulomb, from a body from which 10^9 electrons move out every second, we will need approximately 200 years. One coulomb is, therefore, a very large unit for many practical purposes.

It is, however, also important to know what is roughly the number of electrons contained in a piece of one cubic centimetre of a material. A cubic piece of copper of side 1 cm contains about 2.5×10^{24} electrons.

EXAMPLE 1.1

Example 1.2 How much positive and negative charge is there in a cup of water?

Solution Let us assume that the mass of one cup of water is 250 g. The molecular mass of water is 18g. Thus, one mole (= 6.02×10^{23} molecules) of water is 18 g. Therefore the number of molecules in one cup of water is (250/18) × 6.02×10^{23}.

Each molecule of water contains two hydrogen atoms and one oxygen atom, i.e., 10 electrons and 10 protons. Hence the total positive and total negative charge has the same magnitude. It is equal to (250/18) × $6.02 \times 10^{23} \times 10 \times 1.6 \times 10^{-19}$ C = 1.34×10^7 C.

EXAMPLE 1.2

1.5 Coulomb's Law

Coulomb's law is a quantitative statement about the force between two point charges. When the linear size of charged bodies are much smaller than the distance separating them, the size may be ignored and the charged bodies are treated as *point charges*. Coulomb measured the force between two point charges and found that *it varied inversely as the square of the distance between the charges and was directly proportional to the product of the magnitude of the two charges and* *acted along the line joining the two charges*. Thus, if two point charges q_1, q_2 are separated by a distance r in vacuum, the magnitude of the force (**F**) between them is given by

$$F = k \frac{|q_1 \; q_2|}{r^2} \tag{1.1}$$

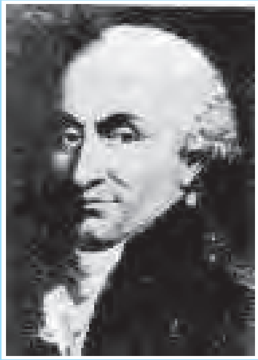
How did Coulomb arrive at this law from his experiments? Coulomb used a torsion balance* for measuring the force between two charged metallic spheres. When the separation between two spheres is much larger than the radius of each sphere, the charged spheres may be regarded as point charges. However, the charges on the spheres were unknown, to begin with. How then could he discover a relation like Eq. (1.1)? Coulomb thought of the following simple way: Suppose the charge on a metallic sphere is q. If the sphere is put in contact with an identical uncharged sphere, the charge will spread over the two spheres. By symmetry, the charge on each sphere will be $q/2$*. Repeating this process, we can get charges $q/2$, $q/4$, etc. Coulomb varied the distance for a fixed pair of charges and measured the force for different separations. He then varied the charges in pairs, keeping the distance fixed for each pair. Comparing forces for different pairs of charges at different distances, Coulomb arrived at the relation, Eq. (1.1).

Coulomb's law, a simple mathematical statement, was initially experimentally arrived at in the manner described above. While the original experiments established it at a macroscopic scale, it has also been established down to subatomic level ($r \sim 10^{-10}$ m).

Coulomb discovered his law without knowing the *explicit* magnitude of the charge. In fact, it is the other way round: Coulomb's law can *now* be employed to furnish a definition for a unit of charge. In the relation, Eq. (1.1), k is so far arbitrary. We can choose any positive value of k. The choice of k determines the size of the unit of charge. In SI units, the value of k is about $9 \times 10^9 \dfrac{\text{Nm}^2}{\text{C}^2}$. The unit of charge that results from this choice is called a coulomb which we defined earlier in Section 1.4. Putting this value of k in Eq. (1.1), we see that for $q_1 = q_2 = 1$ C, $r = 1$ m

$$F = 9 \times 10^9 \text{ N}$$

That is, 1 C is the charge that when placed at a distance of 1 m from another charge of the same magnitude *in vacuum* experiences an electrical force of repulsion of magnitude



Charles Augustin de Coulomb (1736 – 1806) Coulomb, a French physicist, began his career as a military engineer in the West Indies. In 1776, he returned to Paris and retired to a small estate to do his scientific research. He invented a torsion balance to measure the quantity of a force and used it for determination of forces of electric attraction or repulsion between small charged spheres. He thus arrived in 1785 at the inverse square law relation, now known as Coulomb's law. The law had been anticipated by Priestley and also by Cavendish earlier, though Cavendish never published his results. Coulomb also found the inverse square law of force between unlike and like magnetic poles.

CHARLES AUGUSTIN DE COULOMB (1736 –1806)

* A torsion balance is a sensitive device to measure force. It was also used later by Cavendish to measure the very feeble gravitational force between two objects, to verify Newton's Law of Gravitation.
* Implicit in this is the assumption of additivity of charges and conservation: two charges ($q/2$ each) add up to make a total charge q.

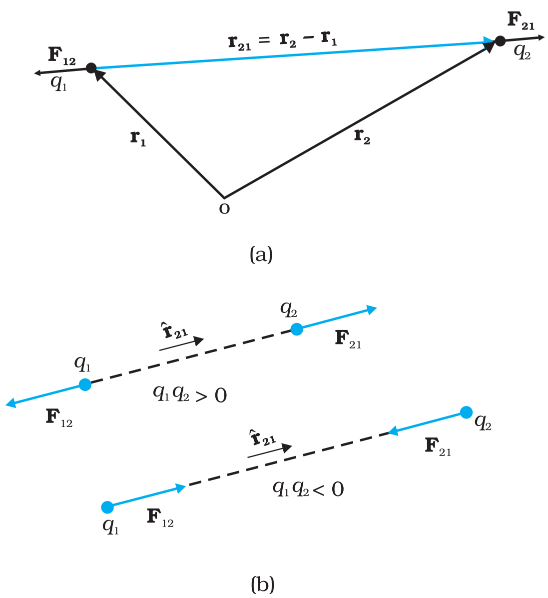


FIGURE 1.3 (a) Geometry and (b) Forces between charges.

9×10^9 N. One coulomb is evidently too big a unit to be used. In practice, in electrostatics, one uses smaller units like 1 mC or 1 μC.

The constant k in Eq. (1.1) is usually put as $k = 1/4\pi\varepsilon_0$ for later convenience, so that Coulomb's law is written as

$$F = \frac{1}{4\pi\varepsilon_0} \frac{|q_1 q_2|}{r^2} \tag{1.2}$$

ε_0 is called the *permittivity of free space* . The value of ε_0 in SI units is

$\varepsilon_0 = 8.854 \times 10^{-12}$ C^2 N^{-1}m^{-2}

Since force is a vector, it is better to write Coulomb's law in the vector notation. Let the position vectors of charges q_1 and q_2 be $\mathbf{r}_1$ and $\mathbf{r}_2$ respectively [see Fig. 1.3(a)]. We denote force on q_1 due to q_2 by $\mathbf{F}_{12}$ and force on q_2 due to q_1 by $\mathbf{F}_{21}$. The two point charges q_1 and q_2 have been numbered 1 and 2 for convenience and the vector leading from 1 to 2 is denoted by $\mathbf{r}_{21}$:

$\mathbf{r}_{21} = \mathbf{r}_2 - \mathbf{r}_1$

In the same way, the vector leading from 2 to 1 is denoted by $\mathbf{r}_{12}$:

$\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2 = - \mathbf{r}_{21}$

The magnitude of the vectors $\mathbf{r}_{21}$ and $\mathbf{r}_{12}$ is denoted by r_{21} and r_{12}, respectively ($r_{12} = r_{21}$). The direction of a vector is specified by a unit vector along the vector. To denote the direction from 1 to 2 (or from 2 to 1), we define the unit vectors:

$$\hat{\mathbf{r}}_{21} = \frac{\mathbf{r}_{21}}{r_{21}}, \quad \hat{\mathbf{r}}_{12} = \frac{\mathbf{r}_{12}}{r_{12}}, \quad \hat{\mathbf{r}}_{21} - \hat{\mathbf{r}}_{12}$$

Coulomb's force law between two point charges q_1 and q_2 located at $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively is then expressed as

$$\mathbf{F}_{21} = \frac{1}{4\pi\varepsilon_o} \frac{q_1 \; q_2}{r_{21}^2} \hat{\mathbf{r}}_{21} \tag{1.3}$$

Some remarks on Eq. (1.3) are relevant:

- Equation (1.3) is valid for any sign of q_1 and q_2 whether positive or negative. If q_1 and q_2 are of the same sign (either both positive or both negative), $\mathbf{F}_{21}$ is along $\hat{\mathbf{r}}_{21}$, which denotes repulsion, as it should be for like charges. If q_1 and q_2 are of opposite signs, $\mathbf{F}_{21}$ is along $-\hat{\mathbf{r}}_{21}(=\hat{\mathbf{r}}_{12})$, which denotes attraction, as expected for unlike charges. Thus, we do not have to write separate equations for the cases of like and unlike charges. Equation (1.3) takes care of both cases correctly [Fig. 1.3(b)].

- The force $\mathbf{F}_{12}$ on charge q_1 due to charge q_2, is obtained from Eq. (1.3), by simply interchanging 1 and 2, i.e.,

$$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} = -\mathbf{F}_{21}$$

Thus, Coulomb's law agrees with the Newton's third law.

- Coulomb's law [Eq. (1.3)] gives the force between two charges q_1 and q_2 in vacuum. If the charges are placed in matter or the intervening space has matter, the situation gets complicated due to the presence of charged constituents of matter. We shall consider electrostatics in matter in the next chapter.

Example 1.3 Coulomb's law for electrostatic force between two point charges and Newton's law for gravitational force between two stationary point masses, both have inverse-square dependence on the distance between the charges and masses respectively. (a) Compare the strength of these forces by determining the ratio of their magnitudes (i) for an electron and a proton and (ii) for two protons. (b) Estimate the accelerations of electron and proton due to the electrical force of their mutual attraction when they are 1 Å (= 10^{-10} m) apart? (m_p = 1.67 × 10^{-27} kg, m_e = 9.11 × 10^{-31} kg)

Solution

(a) (i) The electric force between an electron and a proton at a distance r apart is:

$$F_e = -\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2}$$

where the negative sign indicates that the force is attractive. The corresponding gravitational force (always attractive) is:

$$F_G = -G\frac{m_p\ m_e}{r^2}$$

where m_p and m_e are the masses of a proton and an electron respectively.

$$\left|\frac{F_e}{F_G}\right| = \frac{e^2}{4\pi\varepsilon_0 G m_p m_e} = 2.4 \times 10^{39}$$

(ii) On similar lines, the ratio of the magnitudes of electric force to the gravitational force between two protons at a distance r apart is:

$$\left|\frac{F_e}{F_G}\right| = \frac{e^2}{4\pi\varepsilon_0 G m_p m_p} = 1.3 \times 10^{36}$$

However, it may be mentioned here that the signs of the two forces are different. For two protons, the gravitational force is attractive in nature and the Coulomb force is repulsive. The actual values of these forces between two protons inside a nucleus (distance between two protons is ~ 10^{-15} m inside a nucleus) are F_e ~ 230 N, whereas, F_G ~ 1.9 × 10^{-34} N.

The (dimensionless) ratio of the two forces shows that electrical forces are enormously stronger than the gravitational forces.

EXAMPLE 1.3

Reprint 2025-26

(b) The electric force **F** exerted by a proton on an electron is same in magnitude to the force exerted by an electron on a proton; however, the masses of an electron and a proton are different. Thus, the magnitude of force is

$$|\mathbf{F}| = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} = 8.987 \times 10^9\ \mathrm{Nm^2/C^2} \times (1.6 \times 10^{-19}\mathrm{C})^2\ /\ (10^{-10}\mathrm{m})^2$$
$$= 2.3 \times 10^{-8}\ \mathrm{N}$$

Using Newton's second law of motion, $F = ma$, the acceleration that an electron will undergo is

$a = 2.3\times10^{-8}\ \mathrm{N}\ /\ 9.11 \times10^{-31}\ \mathrm{kg} = 2.5 \times 10^{22}\ \mathrm{m/s^2}$

Comparing this with the value of acceleration due to gravity, we can conclude that the effect of gravitational field is negligible on the motion of electron and it undergoes very large accelerations under the action of Coulomb force due to a proton.

The value for acceleration of the proton is

$2.3 \times 10^{-8}\ \mathrm{N}\ /\ 1.67 \times 10^{-27}\ \mathrm{kg} = 1.4 \times 10^{19}\ \mathrm{m/s^2}$

EXAMPLE 1.3

Example 1.4 A charged metallic sphere A is suspended by a nylon thread. Another charged metallic sphere B held by an insulating

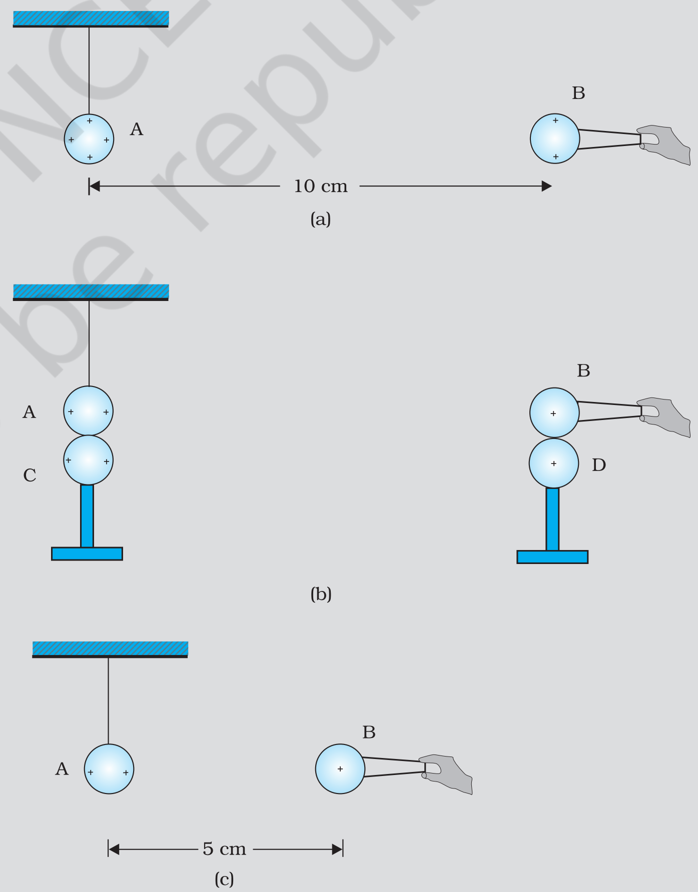


FIGURE 1.4

EXAMPLE 1.4

handle is brought close to A such that the distance between their centres is 10 cm, as shown in Fig. 1.4(a). The resulting repulsion of A is noted (for example, by shining a beam of light and measuring the deflection of its shadow on a screen). Spheres A and B are touched by uncharged spheres C and D respectively, as shown in Fig. 1.4(b). C and D are then removed and B is brought closer to A to a distance of 5.0 cm between their centres, as shown in Fig. 1.4(c). What is the expected repulsion of A on the basis of Coulomb's law? Spheres A and C and spheres B and D have identical sizes. Ignore the sizes of A and B in comparison to the separation between their centres.

Solution Let the original charge on sphere A be q and that on B be q'. At a distance r between their centres, the magnitude of the electrostatic force on each is given by

$$F = \frac{1}{4\pi\varepsilon_0}\frac{qq'}{r^2}$$

neglecting the sizes of spheres A and B in comparison to r. When an identical but uncharged sphere C touches A, the charges redistribute on A and C and, by symmetry, each sphere carries a charge $q/2$. Similarly, after D touches B, the redistributed charge on each is $q'/2$. Now, if the separation between A and B is halved, the magnitude of the electrostatic force on each is

$$F' = \frac{1}{4\pi\varepsilon_0}\frac{(q/2)(q'/2)}{(r/2)^2} = \frac{1}{4\pi\varepsilon_0}\frac{(qq')}{r^2} = F$$

Thus the electrostatic force on A, due to B, remains unaltered.

EXAMPLE 1.4

1.6 FORCES BETWEEN MULTIPLE CHARGES

The mutual electric force between two charges is given by Coulomb's law. How to calculate the force on a charge where there are not one but several charges around? Consider a system of n stationary charges q_1, q_2, q_3, ..., q_n in vacuum. What is the force on q_1 due to q_2, q_3, ..., q_n? Coulomb's law is not enough to answer this question. Recall that forces of mechanical origin add according to the parallelogram law of addition. Is the same true for forces of electrostatic origin?

Experimentally, it is verified that *force on any charge due to a number of other charges is the vector sum of all the forces on that charge due to the other charges, taken one at a time. The individual forces are unaffected due to the presence of other charges.* This is termed as the *principle of superposition.*

To better understand the concept, consider a system of three charges q_1, q_2 and q_3, as shown in Fig. 1.5(a). The force on one charge, say q_1, due to two other charges q_2, q_3 can therefore be obtained by performing a vector addition of the forces due to each one of these charges. Thus, if the force on q_1 due to q_2 is denoted by $\mathbf{F}_{12}$, $\mathbf{F}_{12}$ is given by Eq. (1.3) even though other charges are present.

Thus, $$\mathbf{F}_{12} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12}$$

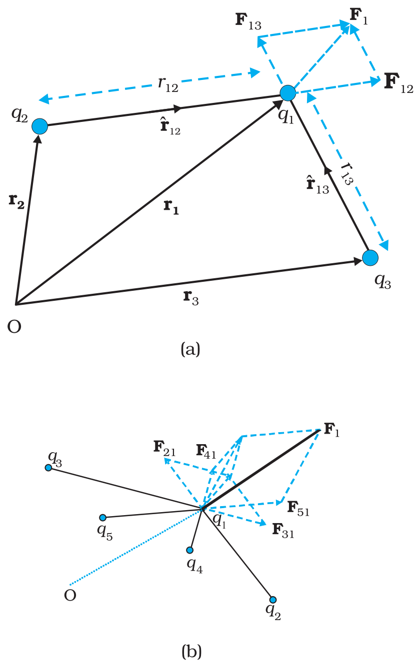


FIGURE 1.5 A system of (a) three charges (b) multiple charges.

In the same way, the force on q_1 due to q_3, denoted by $\mathbf{F}_{13}$, is given by

$$\mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13}$$

which again is the Coulomb force on q_1 due to q_3, even though other charge q_2 is present.

Thus the total force $\mathbf{F}_1$ on q_1 due to the two charges q_2 and q_3 is given as

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} + \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13} \tag{1.4}$$

The above calculation of force can be generalised to a system of charges more than three, as shown in Fig. 1.5(b).

The principle of superposition says that in a system of charges q_1, q_2, ..., q_n, the force on q_1 due to q_2 is the same as given by Coulomb's law, i.e., it is unaffected by the presence of the other charges q_3, q_4, ..., q_n. The total force $\mathbf{F}_1$ on the charge q_1, due to all other charges, is then given by the vector sum of the forces $\mathbf{F}_{12}$, $\mathbf{F}_{13}$, ..., $\mathbf{F}_{1n}$:

i.e.,

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} + ... + \mathbf{F}_{1n} = \frac{1}{4\pi\varepsilon_0}\left[\frac{q_1 q_2}{r_{12}^2}\hat{\mathbf{r}}_{12} + \frac{q_1 q_3}{r_{13}^2}\hat{\mathbf{r}}_{13} + ... + \frac{q_1 q_n}{r_{1n}^2}\hat{\mathbf{r}}_{1n}\right]$$

$$= \frac{q_1}{4\pi\varepsilon_0}\sum_{i=2}^{n}\frac{q_i}{r_{1i}^2}\hat{\mathbf{r}}_{1i} \tag{1.5}$$

The vector sum is obtained as usual by the parallelogram law of addition of vectors. All of electrostatics is basically a consequence of Coulomb's law and the superposition principle.

Example 1.5 Consider three charges q_1, q_2, q_3 each equal to q at the vertices of an equilateral triangle of side l. What is the force on a charge Q (with the same sign as q) placed at the centroid of the triangle, as shown in Fig. 1.6?

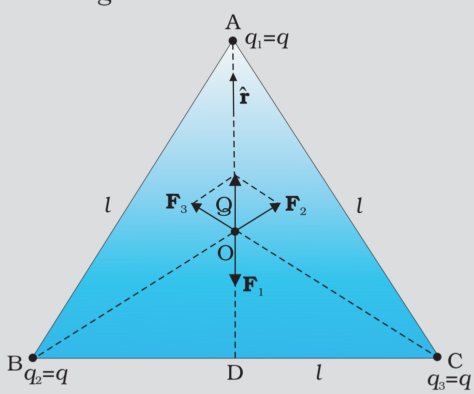


FIGURE 1.6

Solution In the given equilateral triangle ABC of sides of length l, if we draw a perpendicular AD to the side BC,

AD = AC cos 30° = $(\sqrt{3}/2)\, l$ and the distance AO of the centroid O from A is (2/3) AD = $(1/\sqrt{3})\, l$. By symmatry AO = BO = CO.

EXAMPLE 1.5

Thus,

Force $\mathbf{F_1}$ on Q due to charge q at A $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along AO

Force $\mathbf{F_2}$ on Q due to charge q at B $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along BO

Force $\mathbf{F_3}$ on Q due to charge q at C $= \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along CO

The resultant of forces $\mathbf{F_2}$ and $\mathbf{F_3}$ is $\dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}$ along OA, by the parallelogram law. Therefore, the total force on $Q = \dfrac{3}{4\pi\varepsilon_0}\dfrac{Qq}{l^2}(\hat{\mathbf{r}} - \hat{\mathbf{r}})$ = 0, where $\hat{\mathbf{r}}$ is the unit vector along OA.

It is clear also by symmetry that the three forces will sum to zero. Suppose that the resultant force was non-zero but in some direction. Consider what would happen if the system was rotated through 60° about O.

EXAMPLE 1.5

Example 1.6 Consider the charges q, q, and $-q$ placed at the vertices of an equilateral triangle, as shown in Fig. 1.7. What is the force on each charge?

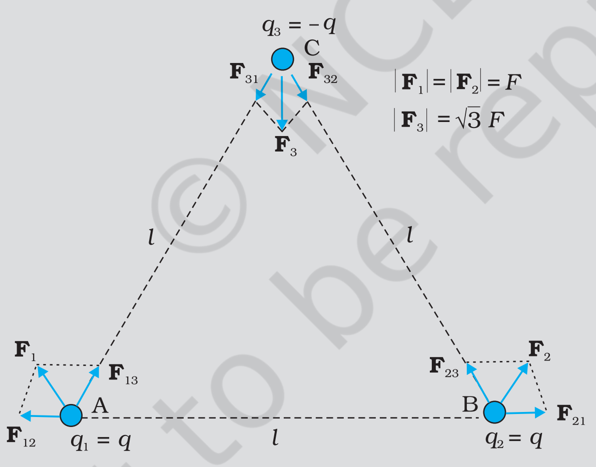


FIGURE 1.7

Solution The forces acting on charge q at A due to charges q at B and $-q$ at C are $\mathbf{F}_{12}$ along BA and $\mathbf{F}_{13}$ along AC respectively, as shown in Fig. 1.7. By the parallelogram law, the total force $\mathbf{F}_1$ on the charge q at A is given by

$\mathbf{F}_1 = F\,\hat{\mathbf{r}}_1$ where $\hat{\mathbf{r}}_1$ is a unit vector along BC.

The force of attraction or repulsion for each pair of charges has the same magnitude $F = \dfrac{q^2}{4\pi\varepsilon_0 l^2}$

The total force $\mathbf{F_2}$ on charge q at B is thus $\mathbf{F_2} = F\,\hat{\mathbf{r}}_2$, where $\hat{\mathbf{r}}_2$ is a unit vector along AC.

EXAMPLE 1.6

EXAMPLE 1.6

Similarly the total force on charge $-q$ at C is $\mathbf{F}_3 = \sqrt{3}\ F\ \hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the unit vector along the direction bisecting the $\angle$BCA.
It is interesting to see that the sum of the forces on the three charges is zero, i.e.,

$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = 0$

The result is not at all surprising. It follows straight from the fact that Coulomb's law is consistent with Newton's third law. The proof is left to you as an exercise.

1.7 Electric Field

Let us consider a point charge Q placed in vacuum, at the origin O. If we place another point charge q at a point P, where **OP** = **r**, then the charge Q will exert a force on q as per Coulomb's law. We may ask the question: If charge q is removed, then what is left in the surrounding? Is there nothing? If there is nothing at the point P, then how does a force act when we place the charge q at P. In order to answer such questions, the early scientists introduced the concept of *field*. According to this, we say that the charge Q produces an electric field everywhere in the surrounding. When another charge q is brought at some point P, the field there acts on it and produces a force. The electric field produced by the charge Q at a point **r** is given as

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}\hat{\mathbf{r}} \qquad (1.6)$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$, is a unit vector from the origin to the point **r**. Thus, Eq.(1.6) specifies the value of the electric field for each value of the position vector **r**. The word "field" signifies how some distributed quantity (which could be a scalar or a vector) varies with position. The effect of the charge has been incorporated in the existence of the electric field. We obtain the force **F** exerted by a charge Q on a charge q, as

$$\mathbf{F} = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2}\hat{\mathbf{r}} \qquad (1.7)$$

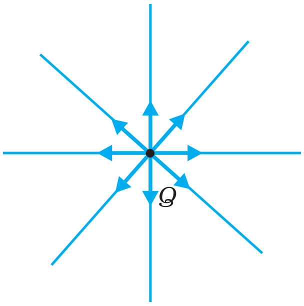
Note that the charge q also exerts an equal and opposite force on the charge Q. The electrostatic force between the charges Q and q can be looked upon as an interaction between charge q and the electric field of Q and *vice versa*. If we denote the position of charge q by the vector **r**, it experiences a force **F** equal to the charge q multiplied by the electric field **E** at the location of q. Thus,

$$\mathbf{F}(\mathbf{r}) = q\,\mathbf{E}(\mathbf{r}) \qquad (1.8)$$

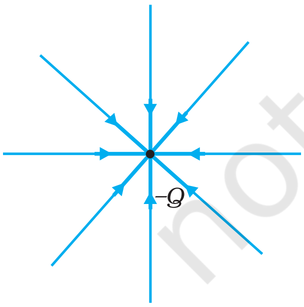
Equation (1.8) defines the SI unit of electric field as N/C*.

Some important remarks may be made here:

(i) From Eq. (1.8), we can infer that if q is unity, the electric field due to a charge Q is numerically equal to the force exerted by it. Thus, the *electric field due to a charge Q at a point in space may be defined as the force that a unit positive charge would experience if placed*



(a)



(b)

FIGURE 1.8 Electric field (a) due to a charge Q, (b) due to a charge $-Q$.

* An alternate unit V/m will be introduced in the next chapter.

at that point. The charge Q, which is producing the electric field, is called a *source charge* and the charge q, which tests the effect of a source charge, is called a *test charge*. Note that the source charge Q must remain at its original location. However, if a charge q is brought at any point around Q, Q itself is bound to experience an electrical force due to q and will tend to move. A way out of this difficulty is to make q negligibly small. The force $\mathbf{F}$ is then negligibly small but the ratio $\mathbf{F}/q$ is finite and defines the electric field:

$$\mathbf{E} = \lim_{q\to 0}\left(\frac{\mathbf{F}}{q}\right) \tag{1.9}$$

A practical way to get around the problem (of keeping Q undisturbed in the presence of q) is to hold Q to its location by unspecified forces! This may look strange but actually this is what happens in practice. When we are considering the electric force on a test charge q due to a charged planar sheet (Section 1.14), the charges on the sheet are held to their locations by the forces due to the unspecified charged constituents inside the sheet.

(ii) Note that the electric field $\mathbf{E}$ due to Q, though defined operationally in terms of some test charge q, is independent of q. This is because $\mathbf{F}$ is proportional to q, so the ratio $\mathbf{F}/q$ does not depend on q. The force $\mathbf{F}$ on the charge q due to the charge Q depends on the particular location of charge q which may take any value in the space around the charge Q. Thus, the electric field $\mathbf{E}$ due to Q is also dependent on the space coordinate $\mathbf{r}$. For different positions of the charge q all over the space, we get different values of electric field $\mathbf{E}$. The field exists at every point in three-dimensional space.

(iii) For a positive charge, the electric field will be directed radially outwards from the charge. On the other hand, if the source charge is negative, the electric field vector, at each point, points radially inwards.

(iv) Since the magnitude of the force $\mathbf{F}$ on charge q due to charge Q depends only on the distance r of the charge q from charge Q, the magnitude of the electric field $\mathbf{E}$ will also depend only on the distance r. Thus at equal distances from the charge Q, the magnitude of its electric field $\mathbf{E}$ is same. The magnitude of electric field $\mathbf{E}$ due to a point charge is thus same on a sphere with the point charge at its centre; in other words, it has a spherical symmetry.

1.7.1 Electric field due to a system of charges

Consider a system of charges q_1, q_2, ..., q_n with position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_n$ relative to some origin O. Like the electric field at a point in space due to a single charge, electric field at a point in space due to the system of charges is defined to be the force experienced by a unit test charge placed at that point, without disturbing the original positions of charges q_1, q_2, ..., q_n. We can use Coulomb's law and the superposition principle to determine this field at a point P denoted by position vector $\mathbf{r}$.

Reprint 2025-26

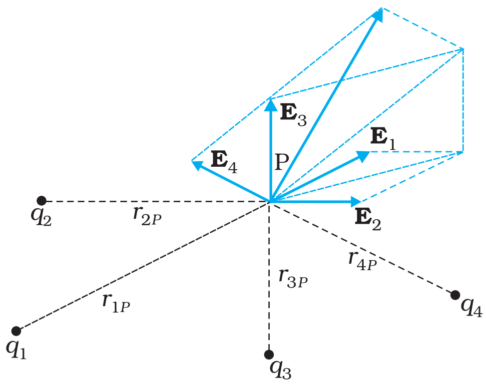


FIGURE 1.9 Electric field at a point due to a system of charges is the vector sum of the electric fields at the point due to individual charges.

Electric field $\mathbf{E}_1$ at $\mathbf{r}$ due to q_1 at $\mathbf{r}_1$ is given by

$$\mathbf{E}_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}^2}\hat{\mathbf{r}}_{1P}$$

where $\hat{\mathbf{r}}_{1P}$ is a unit vector in the direction from q_1 to P, and r_{1P} is the distance between q_1 and P.

In the same manner, electric field $\mathbf{E}_2$ at $\mathbf{r}$ due to q_2 at $\mathbf{r}_2$ is

$$\mathbf{E}_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}^2}\hat{\mathbf{r}}_{2P}$$

where $\hat{\mathbf{r}}_{2P}$ is a unit vector in the direction from q_2 to P and r_{2P} is the distance between q_2 and P. Similar expressions hold good for fields $\mathbf{E}_3$, $\mathbf{E}_4$, ..., $\mathbf{E}_n$ due to charges q_3, q_4, ..., q_n.

By the superposition principle, the electric field $\mathbf{E}$ at $\mathbf{r}$ due to the system of charges is (as shown in Fig. 1.9)

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_1(\mathbf{r}) + \mathbf{E}_2(\mathbf{r}) + \ldots + \mathbf{E}_n(\mathbf{r})$$

$$= \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}^2}\hat{\mathbf{r}}_{1P} + \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}^2}\hat{\mathbf{r}}_{2P} + \ldots + \frac{1}{4\pi\varepsilon_0}\frac{q_n}{r_{nP}^2}\hat{\mathbf{r}}_{nP}$$

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_{iP}^2}\hat{\mathbf{r}}_{iP} \qquad (1.10)$$

$\mathbf{E}$ is a vector quantity that varies from one point to another point in space and is determined from the positions of the source charges.

1.7.2 Physical significance of electric field

You may wonder why the notion of electric field has been introduced here at all. After all, for any system of charges, the measurable quantity is the force on a charge which can be directly determined using Coulomb's law and the superposition principle [Eq. (1.5)]. Why then introduce this intermediate quantity called the electric field?

For electrostatics, the concept of electric field is convenient, but not really necessary. Electric field is an elegant way of characterising the electrical environment of a system of charges. Electric field at a point in the space around a system of charges tells you the force a unit positive test charge would experience if placed at that point (without disturbing the system). Electric field is a characteristic of the system of charges and is independent of the test charge that you place at a point to determine the field. The term *field* in physics generally refers to a quantity that is defined at every point in space and may vary from point to point. Electric field is a vector field, since force is a vector quantity.

The true physical significance of the concept of electric field, however, emerges only when we go beyond electrostatics and deal with time-dependent electromagnetic phenomena. Suppose we consider the force between two distant charges q_1, q_2 in accelerated motion. Now the greatest speed with which a signal or information can go from one point to another is c, the speed of light. Thus, the effect of any motion of q_1 on q_2 cannot arise instantaneously. There will be some time delay between the effect (force on q_2) and the cause (motion of q_1). It is precisely here that the notion of electric field (strictly, electromagnetic field) is natural and very useful. *The field picture is this: the accelerated motion of charge q_1 produces electromagnetic waves, which then propagate with the speed c, reach q_2 and cause a force on q_2.* The notion of field elegantly accounts for the time delay. Thus, even though electric and magnetic fields can be detected only by their effects (forces) on charges, they are regarded as physical entities, not merely mathematical constructs. They have an *independent dynamics* of their own, i.e., they evolve according to laws of their *own*. They can also transport energy. Thus, a source of time-dependent electromagnetic fields, turned on for a short interval of time and then switched off, leaves behind propagating electromagnetic fields transporting energy. The concept of field was first introduced by Faraday and is now among the central concepts in physics.

Example 1.7 An electron falls through a distance of 1.5 cm in a uniform electric field of magnitude 2.0×10^4 N C^{-1} [Fig. 1.10(a)]. The direction of the field is reversed keeping its magnitude unchanged and a proton falls through the same distance [Fig. 1.10(b)]. Compute the time of fall in each case. Contrast the situation with that of 'free fall under gravity'.

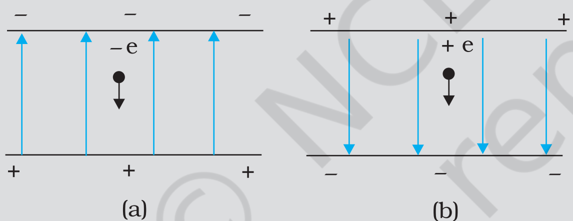


FIGURE 1.10

Solution In Fig. 1.10(a) the field is upward, so the negatively charged electron experiences a downward force of magnitude eE where E is the magnitude of the electric field. The acceleration of the electron is

$$a_e = eE/m_e$$

where m_e is the mass of the electron.

Starting from rest, the time required by the electron to fall through a distance h is given by $t_e = \sqrt{\dfrac{2h}{a_e}} = \sqrt{\dfrac{2h\,m_e}{e\,E}}$

For $e = 1.6 \times 10^{-19}$C, $m_e = 9.11 \times 10^{-31}$ kg,

$E = 2.0 \times 10^4$ N C^{-1}, $h = 1.5 \times 10^{-2}$ m,

$t_e = 2.9 \times 10^{-9}$s

In Fig. 1.10 (b), the field is downward, and the positively charged proton experiences a downward force of magnitude eE. The acceleration of the proton is

$$a_p = eE/m_p$$

where m_p is the mass of the proton; $m_p = 1.67 \times 10^{-27}$ kg. The time of fall for the proton is

EXAMPLE 1.7

$$t_p = \sqrt{\frac{2h}{a_p}} = \sqrt{\frac{2hm_p}{eE}} = 1.3\times10^{-7}\,\text{s}$$

Thus, the heavier particle (proton) takes a greater time to fall through the same distance. This is in basic contrast to the situation of 'free fall under gravity' where the time of fall is independent of the mass of the body. Note that in this example we have ignored the acceleration due to gravity in calculating the time of fall. To see if this is justified, let us calculate the acceleration of the proton in the given electric field:

$$a_p = \frac{eE}{m_p}$$

$$= \frac{(1.6\times10^{-19}\ \text{C})\times(2.0\times10^{4}\ \text{N C}^{-1})}{1.67\times10^{-27}\ \text{kg}}$$

$$= 1.9\times10^{12}\ \text{m s}^{-2}$$

which is enormous compared to the value of g (9.8 m s^{-2}), the acceleration due to gravity. The acceleration of the electron is even greater. Thus, the effect of acceleration due to gravity can be ignored in this example.

EXAMPLE 1.7

Example 1.8 Two point charges q_1 and q_2, of magnitude $+10^{-8}$ C and -10^{-8} C, respectively, are placed 0.1 m apart. Calculate the electric fields at points A, B and C shown in Fig. 1.11.

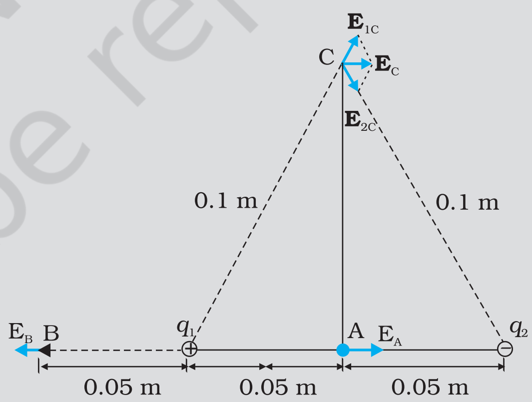


FIGURE 1.11

Solution The electric field vector $\mathbf{E}_{1A}$ at A due to the positive charge q_1 points towards the right and has a magnitude

$$E_{1A} = \frac{(9\times10^{9}\,\text{Nm}^2\text{C}^{-2})\times(10^{-8}\,\text{C})}{(0.05\,\text{m})^2} = 3.6\times10^{4}\ \text{N C}^{-1}$$

The electric field vector $\mathbf{E}_{2A}$ at A due to the negative charge q_2 points towards the right and has the same magnitude. Hence the magnitude of the total electric field E_A at A is

$E_A = E_{1A} + E_{2A} = 7.2\times10^{4}$ N C^{-1}

$\mathbf{E}_A$ is directed toward the right.

EXAMPLE 1.8

The electric field vector $\mathbf{E}_{1B}$ at B due to the positive charge q_1 points towards the left and has a magnitude

$$E_{1B} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.05\,\mathrm{m})^2} = 3.6 \times 10^4\ \mathrm{N\ C^{-1}}$$

The electric field vector $\mathbf{E}_{2B}$ at B due to the negative charge q_2 points towards the right and has a magnitude

$$E_{2B} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.15\,\mathrm{m})^2} = 4 \times 10^3\ \mathrm{N\ C^{-1}}$$

The magnitude of the total electric field at B is
$E_B = E_{1B} - E_{2B} = 3.2 \times 10^4\ \mathrm{N\ C^{-1}}$
$\mathbf{E}_B$ is directed towards the left.
The magnitude of each electric field vector at point C, due to charge q_1 and q_2 is

$$E_{1C} = E_{2C} = \frac{(9\times10^9\,\mathrm{Nm^2C^{-2}})\times(10^{-8}\,\mathrm{C})}{(0.10\,\mathrm{m})^2} = 9 \times 10^3\ \mathrm{N\ C^{-1}}$$

The directions in which these two vectors point are indicated in Fig. 1.11. The resultant of these two vectors is

$$E_C = E_{1c}\cos\frac{\pi}{3} + E_{2c}\cos\frac{\pi}{3} = 9 \times 10^3\ \mathrm{N\ C^{-1}}$$

$\mathbf{E}_C$ points towards the right.

EXAMPLE 1.8

1.8 Electric Field Lines

We have studied electric field in the last section. It is a vector quantity and can be represented as we represent vectors. Let us try to represent **E** due to a point charge pictorially. Let the point charge be placed at the origin. Draw vectors pointing along the direction of the electric field with their lengths proportional to the strength of the field at each point. Since the magnitude of electric field at a point decreases inversely as the square of the distance of that point from the charge, the vector gets shorter as one goes away from the origin, always pointing radially outward. Figure 1.12 shows such a picture. In this figure, each arrow indicates the electric field, i.e., the force acting on a unit positive charge, placed at the tail of that arrow. Connect the arrows pointing in one direction and the resulting figure represents a field line. We thus get many field lines, all pointing outwards from the point charge. Have we lost the information about the strength or magnitude of the field now, because it was contained in the length of the arrow? No. Now the magnitude of the field is represented by the density of field lines. **E** is strong near the charge, so the density of field lines is more near the charge and the lines are closer. Away from the charge, the field gets weaker and the density of field lines is less, resulting in well-separated lines.

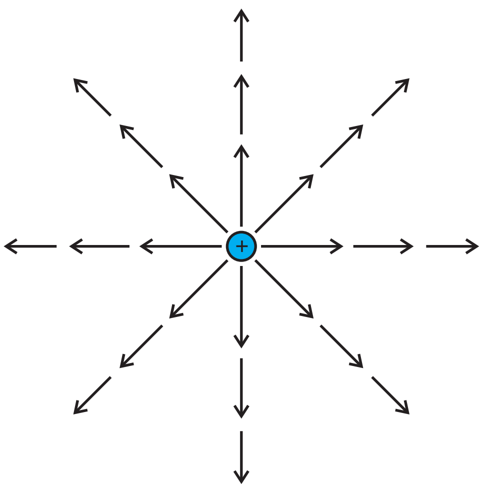


FIGURE 1.12 Field of a point charge.

Another person may draw more lines. But the number of lines is not important. In fact, an infinite number of lines can be drawn in any region.

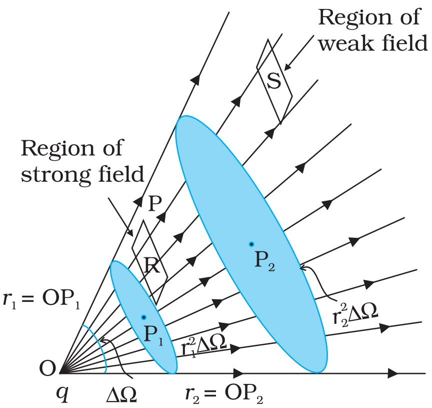


FIGURE 1.13 Dependence of electric field strength on the distance and its relation to the number of field lines.

It is the relative density of lines in different regions which is important.

We draw the figure on the plane of paper, *i.e.*, in two-dimensions but we live in three-dimensions. So if one wishes to estimate the density of field lines, one has to consider the number of lines per unit cross-sectional area, perpendicular to the lines. Since the electric field decreases as the square of the distance from a point charge and the area enclosing the charge increases as the square of the distance, the number of field lines crossing the enclosing area remains constant, whatever may be the distance of the area from the charge.

We started by saying that the field lines carry information about the direction of electric field at different points in space. Having drawn a certain set of field lines, the relative density (i.e., closeness) of the field lines at different points indicates the relative strength of electric field at those points. The field lines crowd where the field is strong and are spaced apart where it is weak. Figure 1.13 shows a set of field lines. We can imagine two equal and small elements of area placed at points R and S normal to the field lines there. The number of field lines in our picture cutting the area elements is proportional to the magnitude of field at these points. The picture shows that the field at R is stronger than at S.

To understand the dependence of the field lines on the area, or rather the *solid angle* subtended by an area element, let us try to relate the area with the solid angle, a generalisation of angle to three dimensions. Recall how a (plane) angle is defined in two-dimensions. Let a small transverse line element Δl be placed at a distance r from a point O. Then the angle subtended by Δl at O can be approximated as $\Delta\theta = \Delta l/r$. Likewise, in three-dimensions the solid angle* subtended by a small perpendicular plane area ΔS, at a distance r, can be written as $\Delta\Omega = \Delta S/r^2$. We know that in a given solid angle the number of radial field lines is the same. In Fig. 1.13, for two points P_1 and P_2 at distances r_1 and r_2 from the charge, the element of area subtending the solid angle $\Delta\Omega$ is $r_1^2\,\Delta\Omega$ at P_1 and an element of area $r_2^2\,\Delta\Omega$ at P_2, respectively. The number of lines (say n) cutting these area elements are the same. The number of field lines, cutting unit area element is therefore $n/(r_1^2\,\Delta\Omega)$ at P_1 and $n/(r_2^2\,\Delta\Omega)$ at P_2, respectively. Since n and $\Delta\Omega$ are common, the strength of the field clearly has a $1/r^2$ dependence.

The picture of field lines was invented by Faraday to develop an intuitive non-mathematical way of visualising electric fields around charged configurations. Faraday called them *lines of force. This term is somewhat misleading, especially in case of magnetic fields. The more appropriate term is field lines (electric or magnetic) that we have adopted in this book.*

Electric field lines are thus a way of pictorially mapping the electric field around a configuration of charges. An electric field line is, in general,

* Solid angle is a measure of a cone. Consider the intersection of the given cone with a sphere of radius R. The solid angle $\Delta\Omega$ of the cone is defined to be equal to $\Delta S/R^2$, where ΔS is the area on the sphere cut out by the cone.

a curve drawn in such a way that the tangent to it at each point is in the direction of the net field at that point. An arrow on the curve is obviously necessary to specify the direction of electric field from the two possible directions indicated by a tangent to the curve. A field line is a space curve, i.e., a curve in three dimensions.

Figure 1.14 shows the field lines around some simple charge configurations. As mentioned earlier, the field lines are in 3-dimensional space, though the figure shows them only in a plane. The field lines of a single positive charge are radially outward while those of a single negative charge are radially inward. The field lines around a system of two positive charges (q, q) give a vivid pictorial description of their mutual repulsion, while those around the configuration of two equal and opposite charges (q, $-q$), a dipole, show clearly the mutual attraction between the charges. The field lines follow some important general properties:

(i) Field lines start from positive charges and end at negative charges. If there is a single charge, they may start or end at infinity.

(ii) In a charge-free region, electric field lines can be taken to be continuous curves without any breaks.

(iii) Two field lines can never cross each other. (If they did, the field at the point of intersection will not have a unique direction, which is absurd.)

(iv) Electrostatic field lines do not form any closed loops. This follows from the conservative nature of electric field (Chapter 2).

1.9 Electric Flux

Consider flow of a liquid with velocity **v**, through a small flat surface d*S*, in a direction normal to the surface. The rate of flow of liquid is given by the volume crossing the area per unit time $v\,dS$ and represents the flux of liquid flowing across the plane. If the normal to the surface is not parallel to the direction of flow of liquid, *i.e.*, to **v**, but makes an angle θ with it, the projected area in a plane perpendicular to **v** is $\delta\, dS \cos\theta$. Therefore, the flux going out of the surface d*S* is $\mathbf{v} \cdot \hat{\mathbf{n}}\, dS$. For the case of the electric field, we define an analogous quantity and call it *electric flux*. We should, however, note that there is no *flow* of a physically observable quantity unlike the case of liquid flow.

In the picture of electric field lines described above, we saw that the number of field lines crossing a unit area, placed normal to the field at a point is a measure of the strength of electric field at that point. This means that if

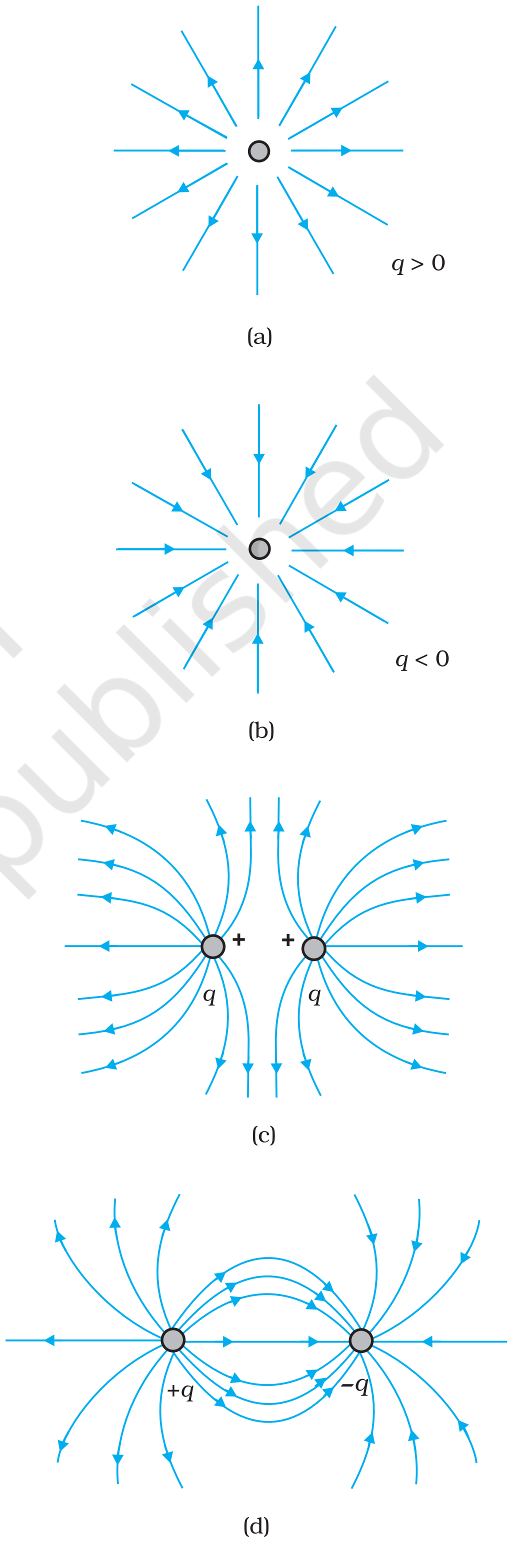


FIGURE 1.14 Field lines due to some simple charge configurations.

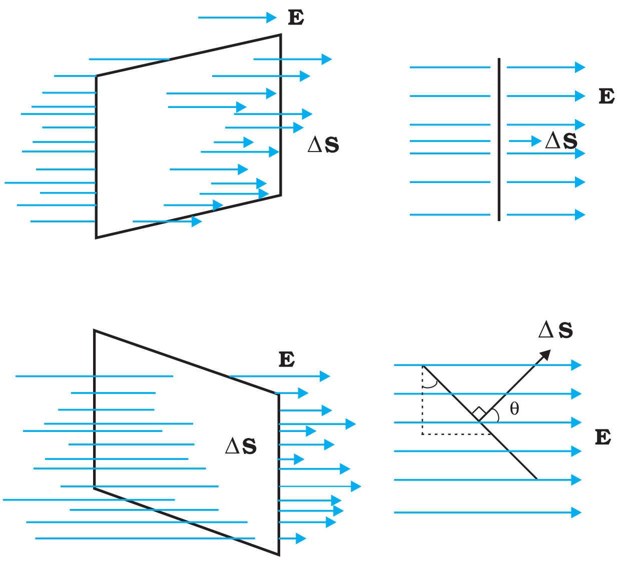


FIGURE 1.15 Dependence of flux on the inclination θ between **E** and $\hat{\mathbf{n}}$.

we place a small planar element of area ΔS normal to **E** at a point, the number of field lines crossing it is proportional* to $E\ \Delta S$. Now suppose we tilt the area element by angle θ. Clearly, the number of field lines crossing the area element will be smaller. The projection of the area element normal to E is $\Delta S \cos\theta$. Thus, the number of field lines crossing ΔS is proportional to $E\,\Delta S \cos\theta$. When $\theta = 90°$, field lines will be parallel to ΔS and will not cross it at all (Fig. 1.15).

The orientation of area element and not merely its magnitude is important in many contexts. For example, in a stream, the amount of water flowing through a ring will naturally depend on how you hold the ring. If you hold it normal to the flow, maximum water will flow through it than if you hold it with some other orientation. This shows that an area element should be treated as a vector. It has a magnitude and also a direction. How to specify the direction of a planar area? Clearly, the normal to the plane specifies the orientation of the plane. Thus the direction of a planar area vector is along its normal.

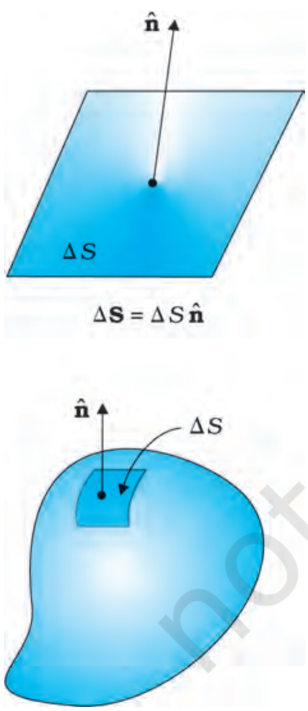


FIGURE 1.16 Convention for defining normal $\hat{\mathbf{n}}$ and ΔS.

How to associate a vector to the area of a curved surface? We imagine dividing the surface into a large number of very small area elements. Each small area element may be treated as planar and a vector associated with it, as explained before.

Notice one ambiguity here. The direction of an area element is along its normal. But a normal can point in two directions. Which direction do we choose as the direction of the vector associated with the area element? This problem is resolved by some convention appropriate to the given context. For the case of a closed surface, this convention is very simple. The vector associated with every area element of a closed surface is taken to be in the direction of the *outward* normal. This is the convention used in Fig. 1.16. Thus, the area element vector $\Delta\mathbf{S}$ at a point on a closed surface equals $\Delta S\ \hat{\mathbf{n}}$ where ΔS is the magnitude of the area element and $\hat{\mathbf{n}}$ is a unit vector in the direction of outward normal at that point.

We now come to the definition of electric flux. Electric flux $\Delta\phi$ through an area element $\Delta\mathbf{S}$ is defined by

$$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S} = E\,\Delta S\ \cos\theta \qquad (1.11)$$

which, as seen before, is proportional to the number of field lines cutting the area element. The angle θ here is the angle between **E** and $\Delta\mathbf{S}$. For a closed surface, with the convention stated already, θ is the angle between **E** and the outward normal to the area element. Notice we could look at the expression $E\ \Delta S \cos\theta$ in two ways: $E\ (\Delta S \cos\theta)$ i.e., E times the

* It will not be proper to say that the number of field lines is equal to $E\Delta S$. The number of field lines is after all, a matter of how many field lines we choose to draw. What is physically significant is the relative number of field lines crossing a given area at different points.

Reprint 2025-26

projection of area normal to **E**, or $E_{\perp}\,\Delta S$, i.e., component of **E** along the normal to the area element times the magnitude of the area element. The unit of electric flux is N C^{-1} m^2.

The basic definition of electric flux given by Eq. (1.11) can be used, in principle, to calculate the total flux through any given surface. All we have to do is to divide the surface into small area elements, calculate the flux at each element and add them up. Thus, the total flux ϕ through a surface S is

$$\phi \simeq \Sigma\, \mathbf{E} \cdot \Delta \mathbf{S} \tag{1.12}$$

The approximation sign is put because the electric field **E** is taken to be constant over the small area element. This is mathematically exact only when you take the limit $\Delta S \to 0$ and the sum in Eq. (1.12) is written as an integral.

1.10 Electric Dipole

An electric dipole is a pair of equal and opposite point charges q and $-q$, separated by a distance $2a$. The line connecting the two charges defines a direction in space. By convention, the direction from $-q$ to q is said to be the direction of the dipole. The mid-point of locations of $-q$ and q is called the centre of the dipole.

The total charge of the electric dipole is obviously zero. This does not mean that the field of the electric dipole is zero. Since the charge q and $-q$ are separated by some distance, the electric fields due to them, when added, do not exactly cancel out. However, at distances much larger than the separation of the two charges forming a dipole ($r >> 2a$), the fields due to q and $-q$ nearly cancel out. The electric field due to a dipole therefore falls off, at large distance, faster than like $1/r^2$ (the dependence on r of the field due to a single charge q). These qualitative ideas are borne out by the explicit calculation as follows:

1.10.1 The field of an electric dipole

The electric field of the pair of charges ($-q$ and q) at any point in space can be found out from Coulomb's law and the superposition principle. The results are simple for the following two cases: (i) when the point is on the dipole axis, and (ii) when it is in the *equatorial plane* of the dipole, i.e., on a plane perpendicular to the dipole axis through its centre. The electric field at any general point P is obtained by adding the electric fields $\mathbf{E}_{-q}$ due to the charge $-q$ and $\mathbf{E}_{+q}$ due to the charge q, by the parallelogram law of vectors.

(i) For points on the axis

Let the point P be at distance r from the centre of the dipole on the side of the charge q, as shown in Fig. 1.17(a). Then

$$\mathbf{E}_{-q} = -\frac{q}{4\pi\varepsilon_0 (r+a)^2}\,\mathbf{p} \tag{1.13(a)}$$

where $\hat{\mathbf{p}}$ is the unit vector along the dipole axis (from $-q$ to q). Also

$$\mathbf{E}_{+q} = \frac{q}{4\pi\varepsilon_0 (r-a)^2}\,\mathbf{p} \tag{1.13(b)}$$

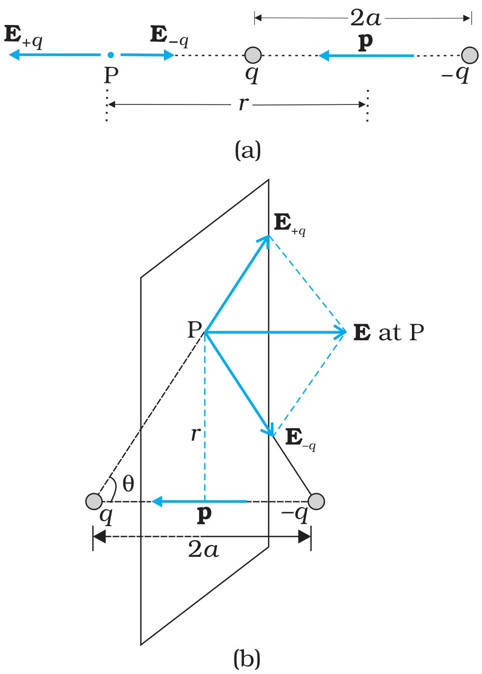


FIGURE 1.17 Electric field of a dipole at (a) a point on the axis, (b) a point on the equatorial plane of the dipole. **p** is the dipole moment vector of magnitude $p = q \times 2a$ and directed from $-q$ to q.

The total field at P is

$$\mathbf{E} = \mathbf{E}_{+q} + \mathbf{E}_{-q} = \frac{q}{4\pi\varepsilon_0}\left[\frac{1}{(r-a)^2} - \frac{1}{(r+a)^2}\right]\mathbf{p}$$

$$= \frac{q}{4\pi\varepsilon_o}\,\frac{4\,a\,r}{(r^2-a^2)^2}\,\mathbf{p} \qquad (1.14)$$

For $r >> a$

$$\mathbf{E} = \frac{4qa}{4\pi\varepsilon_0 r^3}\hat{\mathbf{p}} \qquad (r >> a) \qquad (1.15)$$

(ii) For points on the equatorial plane

The magnitudes of the electric fields due to the two charges $+q$ and $-q$ are given by

$$E_{+q} = \frac{q}{4\pi\varepsilon_0}\,\frac{1}{r^2+a^2} \qquad [1.16(a)]$$

$$E_{-q} = \frac{q}{4\pi\varepsilon_0}\,\frac{1}{r^2+a^2} \qquad [1.16(b)]$$

and are equal.

The directions of $\mathbf{E}_{+q}$ and $\mathbf{E}_{-q}$ are as shown in Fig. 1.17(b). Clearly, the components normal to the dipole axis cancel away. The components along the dipole axis add up. The total electric field is opposite to $\hat{\mathbf{p}}$. We have

$$\mathbf{E} = -\,(E_{+q} + E_{-q})\cos\theta\;\hat{\mathbf{p}}$$

$$= -\frac{2qa}{4\pi\varepsilon_o(r^2+a^2)^{3/2}}\mathbf{p} \qquad (1.17)$$

At large distances ($r >> a$), this reduces to

$$\mathbf{E} = -\frac{2qa}{4\pi\varepsilon_o r^3}\hat{\mathbf{p}} \qquad (r >> a) \qquad (1.18)$$

From Eqs. (1.15) and (1.18), it is clear that the dipole field at large distances does not involve q and a separately; it depends on the product qa. This suggests the definition of dipole moment. The *dipole moment vector* **p** of an electric dipole is defined by

$$\mathbf{p} = q \times 2a\;\hat{\mathbf{p}} \qquad (1.19)$$

that is, it is a vector whose magnitude is charge q times the separation $2a$ (between the pair of charges q, $-q$) and the direction is along the line from $-q$ to q. In terms of **p**, the electric field of a dipole at large distances takes simple forms:

At a point on the dipole axis

$$\mathbf{E} = \frac{2\mathbf{p}}{4\pi\varepsilon_o r^3} \qquad (r >> a) \qquad (1.20)$$

At a point on the equatorial plane

$$\mathbf{E} = -\frac{\mathbf{p}}{4\pi\varepsilon_o r^3} \qquad (r >> a) \qquad (1.21)$$

Notice the important point that the dipole field at large distances falls off not as $1/r^2$ but as $1/r^3$. Further, the magnitude and the direction of the dipole field depends not only on the distance r but also on the *angle* between the position vector **r** and the dipole moment **p**.

We can think of the limit when the dipole size $2a$ approaches zero, the charge q approaches infinity in such a way that the product $p = q \times 2a$ is finite. Such a dipole is referred to as a *point dipole*. For a point dipole, Eqs. (1.20) and (1.21) are exact, true for any r.

1.10.2 Physical significance of dipoles

In most molecules, the centres of positive charges and of negative charges* lie at the same place. Therefore, their dipole moment is zero. CO_2 and CH_4 are of this type of molecules. However, they develop a dipole moment when an electric field is applied. But in some molecules, the centres of negative charges and of positive charges do not coincide. Therefore they have a permanent electric dipole moment, even in the absence of an electric field. Such molecules are called polar molecules. Water molecules, H_2O, is an example of this type. Various materials give rise to interesting properties and important applications in the presence or absence of electric field.

Example 1.9 Two charges ±10 μC are placed 5.0 mm apart. Determine the electric field at (a) a point P on the axis of the dipole 15 cm away from its centre O on the side of the positive charge, as shown in Fig. 1.18(a), and (b) a point Q, 15 cm away from O on a line passing through O and normal to the axis of the dipole, as shown in Fig. 1.18(b).

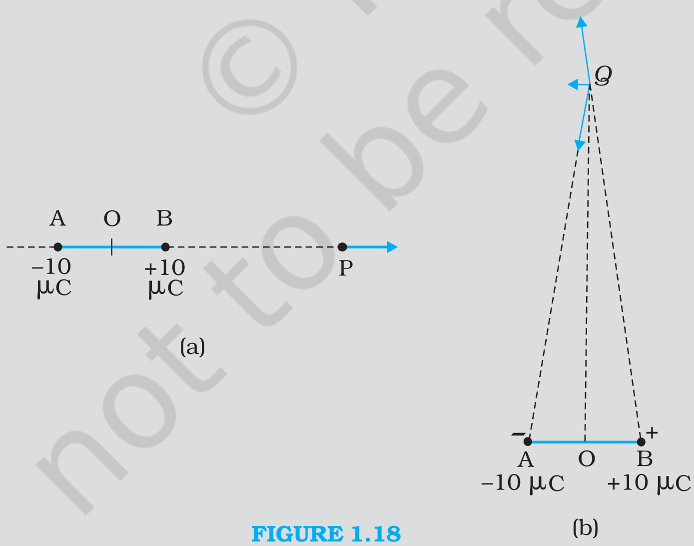


FIGURE 1.18

EXAMPLE 1.9

* Centre of a collection of positive point charges is defined much the same way as the centre of mass: $\mathbf{r}_{cm} = \dfrac{\sum_i q_i \mathbf{r}_i}{\sum_i q_i}$.

Reprint 2025-26

Solution (a) Field at P due to charge +10 μC

$$= \frac{10^{-5}\,\text{C}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15-0.25)^2\times10^{-4}\,\text{m}^2}$$

$= 4.13 \times 10^{6}\ \text{N C}^{-1}$ along BP

Field at P due to charge –10 μC

$$= \frac{10^{-5}\,\text{C}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15+0.25)^2\times10^{-4}\,\text{m}^2}$$

$= 3.86 \times 10^{6}\ \text{N C}^{-1}$ along PA

The resultant electric field at P due to the two charges at A and B is $= 2.7 \times 10^{5}\ \text{N C}^{-1}$ along BP.

In this example, the ratio OP/OB is quite large (= 60). Thus, we can expect to get approximately the same result as above by directly using the formula for electric field at a far-away point on the axis of a dipole. For a dipole consisting of charges $\pm q$, $2a$ distance apart, the electric field at a distance r from the centre on the axis of the dipole has a magnitude

$$E = \frac{2p}{4\pi\varepsilon_0 r^3} \qquad (r/a >> 1)$$

where $p = 2a\,q$ is the magnitude of the dipole moment.

The direction of electric field on the dipole axis is always along the direction of the dipole moment vector (i.e., from $-q$ to q). Here, $p = 10^{-5}\ \text{C} \times 5 \times 10^{-3}\ \text{m} = 5 \times 10^{-8}\ \text{C m}$

Therefore,

$$E = \frac{2\times5\times10^{-8}\,\text{C m}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15)^3\times10^{-6}\,\text{m}^3} = 2.6 \times 10^{5}\ \text{N C}^{-1}$$

along the dipole moment direction AB, which is close to the result obtained earlier.

(b) Field at Q due to charge + 10 μC at B

$$= \frac{10^{-5}\ \text{C}}{4\ \pi\ (8.854\times10^{-12}\ \text{C}^2\ \text{N}^{-1}\ \text{m}^{-2})} \times \frac{1}{[15^2 + (0.25)^2]\times10^{-4}\ \text{m}^2}$$

$= 3.99 \times 10^{6}\ \text{N C}^{-1}$ along BQ

Field at Q due to charge –10 μC at A

$$= \frac{10^{-5}\,\text{C}}{4\,\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{[15^2+(0.25)^2]\times10^{-4}\,\text{m}^2}$$

$= 3.99 \times 10^{6}\ \text{N C}^{-1}$ along QA.

Clearly, the components of these two forces with equal magnitudes cancel along the direction OQ but add up along the direction parallel to BA. Therefore, the resultant electric field at Q due to the two charges at A and B is

$$= 2 \times \frac{0.25}{\sqrt{15^2 + (0.25)^2}} \times 3.99 \times 10^{6}\,\text{N C}^{-1}\ \text{along BA}$$

$= 1.33 \times 10^{5}\ \text{N C}^{-1}$ along BA.

As in (a), we can expect to get approximately the same result by directly using the formula for dipole field at a point on the normal to the axis of the dipole:

EXAMPLE 1.9

$$E = \frac{p}{4\pi\varepsilon_0 r^3} \qquad (r/a >> 1)$$

$$= \frac{5\times10^{-8}\,\text{C m}}{4\pi(8.854\times10^{-12}\,\text{C}^2\,\text{N}^{-1}\,\text{m}^{-2})} \times \frac{1}{(15)^3\times10^{-6}\,\text{m}^3}$$

$= 1.33 \times 10^5\ \text{N C}^{-1}$.

The direction of electric field in this case is opposite to the direction of the dipole moment vector. Again, the result agrees with that obtained before.

EXAMPLE 1.9

1.11 Dipole in a Uniform External Field

Consider a permanent dipole of dipole moment **p** in a uniform external field **E**, as shown in Fig. 1.19. (By permanent dipole, we mean that **p** exists irrespective of **E**; it has not been induced by **E**.)

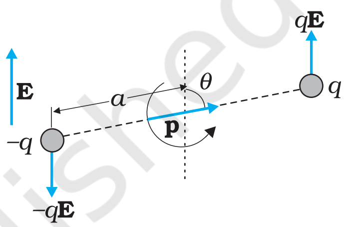


FIGURE 1.19 Dipole in a uniform electric field.

There is a force $q\mathbf{E}$ on q and a force $-q\mathbf{E}$ on $-q$. The net force on the dipole is zero, since **E** is uniform. However, the charges are separated, so the forces act at different points, resulting in a torque on the dipole. When the net force is zero, the torque (couple) is independent of the origin. Its magnitude equals the magnitude of each force multiplied by the arm of the couple (perpendicular distance between the two antiparallel forces).

Magnitude of torque = $q\,E \times 2\,a \sin\theta$

$= 2\,q\,a\,E \sin\theta$

Its direction is normal to the plane of the paper, coming out of it.

The magnitude of $\mathbf{p}\times\mathbf{E}$ is also $p\,E\sin\theta$ and its direction is normal to the paper, coming out of it. Thus,

$$\boldsymbol{\tau} = \mathbf{p}\times\mathbf{E} \qquad (1.22)$$

This torque will tend to align the dipole with the field **E**. When **p** is aligned with **E**, the torque is zero.

What happens if the field is not uniform? In that case, the net force will evidently be non-zero. In addition there will, in general, be a torque on the system as before. The general case is involved, so let us consider the simpler situations when **p** is parallel to **E** or antiparallel to **E**. In either case, the net torque is zero, but there is a net force on the dipole if **E** is not uniform.

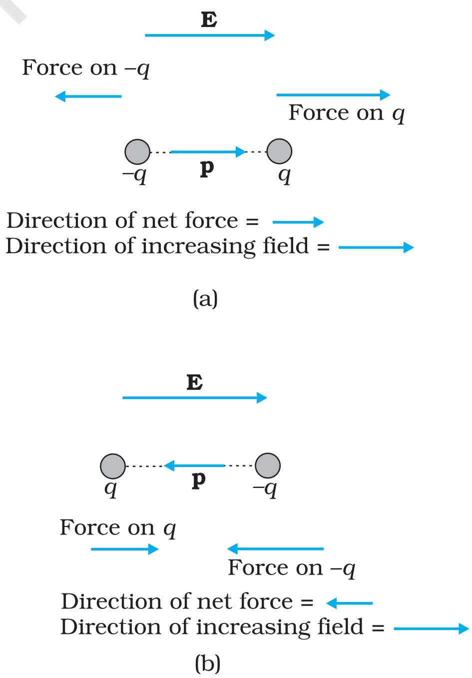


FIGURE 1.20 Electric force on a dipole: (a) **E** parallel to **p**, (b) **E** antiparallel to **p**.

Figure 1.20 is self-explanatory. It is easily seen that when **p** is parallel to **E**, the dipole has a net force in the direction of increasing field. When **p** is antiparallel to **E**, the net force on the dipole is in the direction of decreasing field. In general, the force depends on the orientation of **p** with respect to **E**.

This brings us to a common observation in frictional electricity. A comb run through dry hair attracts pieces of paper. The comb, as we know, acquires charge through friction. But the paper is not charged. What then explains the attractive force? Taking the clue from the preceding

Reprint 2025-26

discussion, the charged comb 'polarises' the piece of paper, i.e., induces a net dipole moment in the direction of field. Further, the electric field due to the comb is not uniform. This non-uniformity of the field makes a dipole to experience a net force on it. In this situation, it is easily seen that the paper should move in the direction of the comb!

1.12 Continuous Charge Distribution

We have so far dealt with charge configurations involving discrete charges $q_1, q_2, ..., q_n$. One reason why we restricted to discrete charges is that the mathematical treatment is simpler and does not involve calculus. For many purposes, however, it is impractical to work in terms of discrete charges and we need to work with continuous charge distributions. For example, on the surface of a charged conductor, it is impractical to specify the charge distribution in terms of the locations of the microscopic charged constituents. It is more feasible to consider an area element ΔS (Fig. 1.21) on the surface of the conductor (which is very small on the macroscopic scale but big enough to include a very large number of electrons) and specify the charge ΔQ on that element. We then define a *surface charge density* σ at the area element by

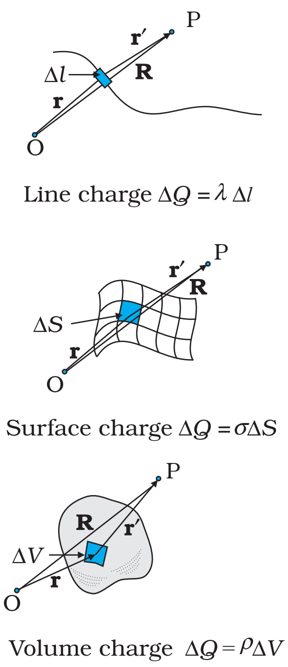


FIGURE 1.21 Definition of linear, surface and volume charge densities. In each case, the element (Δl, ΔS, ΔV) chosen is small on the macroscopic scale but contains a very large number of microscopic constituents.

$$\sigma = \frac{\Delta Q}{\Delta S} \tag{1.23}$$

We can do this at different points on the conductor and thus arrive at a continuous function σ, called the surface charge density. The surface charge density σ so defined ignores the quantisation of charge and the discontinuity in charge distribution at the microscopic level*. σ represents macroscopic surface charge density, which in a sense, is a smoothed out average of the microscopic charge density over an area element ΔS which, as said before, is large microscopically but small macroscopically. The units for σ are C/m^2.

Similar considerations apply for a line charge distribution and a volume charge distribution. The *linear charge density* λ of a wire is defined by

$$\lambda = \frac{\Delta Q}{\Delta l} \tag{1.24}$$

where Δl is a small line element of wire on the macroscopic scale that, however, includes a large number of microscopic charged constituents, and ΔQ is the charge contained in that line element. The units for λ are C/m. The *volume charge density* (sometimes simply called charge density) is defined in a similar manner:

$$\rho = \frac{\Delta Q}{\Delta V} \tag{1.25}$$

where ΔQ is the charge included in the macroscopically small volume element ΔV that includes a large number of microscopic charged constituents. The units for ρ are C/m^3.

The notion of continuous charge distribution is similar to that we adopt for continuous mass distribution in mechanics. When we refer to

* At the microscopic level, charge distribution is discontinuous, because they are discrete charges separated by intervening space where there is no charge.

Reprint 2025-26

the density of a liquid, we are referring to its macroscopic density. We regard it as a continuous fluid and ignore its discrete molecular constitution.

The field due to a continuous charge distribution can be obtained in much the same way as for a system of discrete charges, Eq. (1.10). Suppose a continuous charge distribution in space has a charge density ρ. Choose any convenient origin O and let the position vector of any point in the charge distribution be **r**. The charge density ρ may vary from point to point, i.e., it is a function of **r**. Divide the charge distribution into small volume elements of size ΔV. The charge in a volume element ΔV is $\rho \Delta V$.

Now, consider any general point P (inside or outside the distribution) with position vector **R** (Fig. 1.21). Electric field due to the charge $\rho \Delta V$ is given by Coulomb's law:

$$\Delta \mathbf{E} = \frac{1}{4\pi\varepsilon_0} \frac{\rho\, \Delta V}{r'^2} \hat{\mathbf{r}}' \tag{1.26}$$

where r' is the distance between the charge element and P, and $\hat{\mathbf{r}}'$ is a unit vector in the direction from the charge element to P. By the superposition principle, the total electric field due to the charge distribution is obtained by summing over electric fields due to different volume elements:

$$\mathbf{E} \cong \frac{1}{4\pi\varepsilon_0} \sum_{all\ \Delta V} \frac{\rho\, \Delta V}{r'^2} \hat{\mathbf{r}}' \tag{1.27}$$

Note that ρ, r', $\hat{\mathbf{r}}'$ all can vary from point to point. In a strict mathematical method, we should let $\Delta V \to 0$ and the sum then becomes an integral; but we omit that discussion here, for simplicity. In short, using Coulomb's law and the superposition principle, electric field can be determined for any charge distribution, discrete or continuous or part discrete and part continuous.

1.13 Gauss's Law

As a simple application of the notion of electric flux, let us consider the total flux through a sphere of radius r, which encloses a point charge q at its centre. Divide the sphere into small area elements, as shown in Fig. 1.22.

The flux through an area element $\Delta \mathbf{S}$ is

$$\Delta \phi = \mathbf{E} \cdot \Delta \mathbf{S} = \frac{q}{4\pi\varepsilon_0\, r^2}\, \hat{\mathbf{r}} \cdot \Delta \mathbf{S} \tag{1.28}$$

where we have used Coulomb's law for the electric field due to a single charge q. The unit vector $\hat{\mathbf{r}}$ is along the radius vector from the centre to the area element. Now, since the normal to a sphere at every point is along the radius vector at that point, the area element $\Delta \mathbf{S}$ and $\hat{\mathbf{r}}$ have the same direction. Therefore,

$$\Delta \phi = \frac{q}{4\pi\varepsilon_0\, r^2} \Delta S \tag{1.29}$$

since the magnitude of a unit vector is 1.

The total flux through the sphere is obtained by adding up flux through all the different area elements:

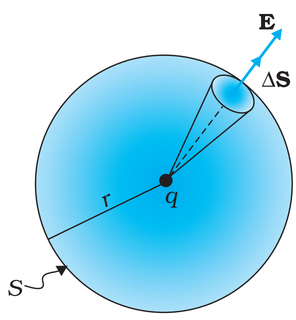


FIGURE 1.22 Flux through a sphere enclosing a point charge q at its centre.

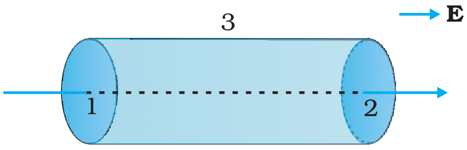


FIGURE 1.23 Calculation of the flux of uniform electric field through the surface of a cylinder.

$$\phi = \sum_{\text{all } \Delta S} \frac{q}{4\pi\varepsilon_0 r^2} \Delta S$$

Since each area element of the sphere is at the same distance r from the charge,

$$\phi = \frac{q}{4\pi\varepsilon_0 r^2} \sum_{\text{all } \Delta S} \Delta S = \frac{q}{4\pi\varepsilon_0 r^2} S$$

Now S, the total area of the sphere, equals $4\pi r^2$. Thus,

$$\phi = \frac{q}{4\pi\varepsilon_0 r^2} \times 4\pi r^2 = \frac{q}{\varepsilon_0} \tag{1.30}$$

Equation (1.30) is a simple illustration of a general result of electrostatics called Gauss's law.

We state *Gauss's law* without proof:

Electric flux through a closed surface S

$$= q/\varepsilon_0 \tag{1.31}$$

q = total charge enclosed by S.

The law implies that the total electric flux through a closed surface is zero if no charge is enclosed by the surface. We can see that explicitly in the simple situation of Fig. 1.23.

Here the electric field is uniform and we are considering a closed cylindrical surface, with its axis parallel to the uniform field **E**. The total flux ϕ through the surface is $\phi = \phi_1 + \phi_2 + \phi_3$, where ϕ_1 and ϕ_2 represent the flux through the surfaces 1 and 2 (of circular cross-section) of the cylinder and ϕ_3 is the flux through the curved cylindrical part of the closed surface. Now the normal to the surface 3 at every point is perpendicular to **E**, so by definition of flux, $\phi_3 = 0$. Further, the outward normal to 2 is along **E** while the outward normal to 1 is opposite to **E**. Therefore,

$$\phi_1 = -E\,S_1, \qquad \phi_2 = +E\,S_2$$

$$S_1 = S_2 = S$$

where S is the area of circular cross-section. Thus, the total flux is zero, as expected by Gauss's law. Thus, whenever you find that the net electric flux through a closed surface is zero, we conclude that the total charge contained in the closed surface is zero.

The great significance of Gauss's law Eq. (1.31), is that it is true in general, and not only for the simple cases we have considered above. Let us note some important points regarding this law:

(i) Gauss's law is true for any closed surface, no matter what its shape or size.

(ii) The term q on the right side of Gauss's law, Eq. (1.31), includes the sum of all charges enclosed by the surface. The charges may be located anywhere inside the surface.

(iii) In the situation when the surface is so chosen that there are some charges inside and some outside, the electric field [whose flux appears on the left side of Eq. (1.31)] is due to all the charges, both inside and outside S. The term q on the right side of Gauss's law, however, represents only the total charge inside S.

Reprint 2025-26

(iv) The surface that we choose for the application of Gauss's law is called the Gaussian surface. You may choose any Gaussian surface and apply Gauss's law. However, take care not to let the Gaussian surface pass through any discrete charge. This is because electric field due to a system of discrete charges is not well defined at the location of any charge. (As you go close to the charge, the field grows without any bound.) However, the Gaussian surface can pass through a continuous charge distribution.

(v) Gauss's law is often useful towards a much easier calculation of the electrostatic field *when the system has some symmetry*. This is facilitated by the choice of a suitable Gaussian surface.

(vi) Finally, Gauss's law is based on the inverse square dependence on distance contained in the Coulomb's law. Any violation of Gauss's law will indicate departure from the inverse square law.

Example 1.10 The electric field components in Fig. 1.24 are $E_x = \alpha x^{1/2}$, $E_y = E_z = 0$, in which $\alpha = 800$ N/C m$^{1/2}$. Calculate (a) the flux through the cube, and (b) the charge within the cube. Assume that $a = 0.1$ m.

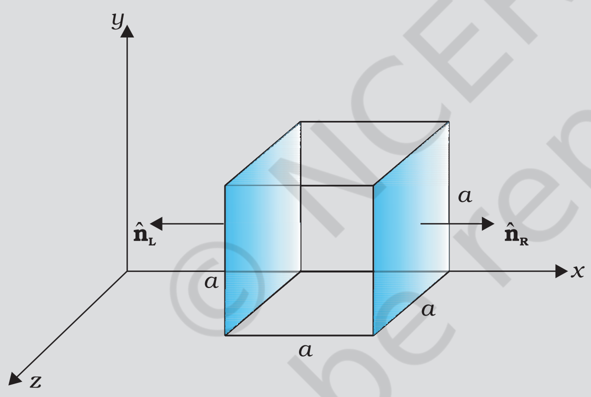


FIGURE 1.24

Solution

(a) Since the electric field has only an x component, for faces perpendicular to x direction, the angle between **E** and Δ**S** is $\pm \pi/2$. Therefore, the flux $\phi = \mathbf{E}.\Delta\mathbf{S}$ is separately zero for each face of the cube except the two shaded ones. Now the magnitude of the electric field at the left face is

$E_L = \alpha x^{1/2} = \alpha a^{1/2}$

($x = a$ at the left face).

The magnitude of electric field at the right face is

$E_R = \alpha\, x^{1/2} = \alpha\,(2a)^{1/2}$

($x = 2a$ at the right face).

The corresponding fluxes are

$$\phi_L = \mathbf{E}_L \cdot \Delta\mathbf{S} = \Delta S\, \mathbf{E}_L \cdot \hat{\mathbf{n}}_L = E_L\, \Delta S \cos\theta = -E_L\, \Delta S, \text{ since } \theta = 180°$$
$$= -E_L a^2$$
$$\phi_R = \mathbf{E}_R \cdot \Delta\mathbf{S} = E_R\, \Delta S \cos\theta = E_R\, \Delta S, \quad \text{since } \theta = 0°$$
$$= E_R a^2$$

Net flux through the cube

EXAMPLE 1.10

$= \phi_R + \phi_L = E_R a^2 - E_L a^2 = a^2\,(E_R - E_L) = \alpha a^2\,[(2a)^{1/2} - a^{1/2}]$

$= \alpha a^{5/2}\left(\sqrt{2}-1\right)$

$= 800\,(0.1)^{5/2}\left(\sqrt{2}-1\right)$

$= 1.05$ N m^2 C^{-1}

(b) We can use Gauss's law to find the total charge q inside the cube. We have $\phi = q/\varepsilon_0$ or $q = \phi\varepsilon_0$. Therefore,

$q = 1.05 \times 8.854 \times 10^{-12}$ C $= 9.27 \times 10^{-12}$ C.

EXAMPLE 1.10

Example 1.11 An electric field is uniform, and in the positive x direction for positive x, and uniform with the same magnitude but in the negative x direction for negative x. It is given that $\mathbf{E} = 200\ \hat{\mathbf{i}}$ N/C for $x > 0$ and $\mathbf{E} = -200\ \hat{\mathbf{i}}$ N/C for $x < 0$. A right circular cylinder of length 20 cm and radius 5 cm has its centre at the origin and its axis along the x-axis so that one face is at $x = +10$ cm and the other is at $x = -10$ cm (Fig. 1.25). (a) What is the net outward flux through each flat face? (b) What is the flux through the side of the cylinder? (c) What is the net outward flux through the cylinder? (d) What is the net charge inside the cylinder?

Solution

(a) We can see from the figure that on the left face $\mathbf{E}$ and $\Delta\mathbf{S}$ are parallel. Therefore, the outward flux is

$\phi_L = \mathbf{E}\cdot\Delta\mathbf{S} = -\,200\ \hat{\mathbf{i}}\cdot\Delta\mathbf{S}$

$= +\,200\ \Delta S$, since $\hat{\mathbf{i}}\cdot\Delta\mathbf{S} = -\,\Delta S$

$= +\,200 \times \pi\,(0.05)^2 = +\,1.57$ N m^2 C^{-1}

On the right face, $\mathbf{E}$ and $\Delta\mathbf{S}$ are parallel and therefore

$\phi_R = \mathbf{E}\cdot\Delta\mathbf{S} = +\,1.57$ N m^2 C^{-1}.

(b) For any point on the side of the cylinder $\mathbf{E}$ is perpendicular to $\Delta\mathbf{S}$ and hence $\mathbf{E}\cdot\Delta\mathbf{S} = 0$. Therefore, the flux out of the side of the cylinder is zero.

(c) Net outward flux through the cylinder

$\phi = 1.57 + 1.57 + 0 = 3.14$ N m^2 C^{-1}

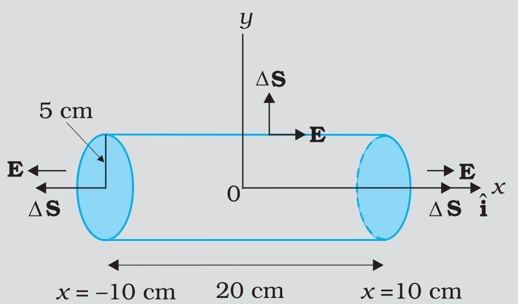


FIGURE 1.25

(d) The net charge within the cylinder can be found by using Gauss's law which gives

$q = \varepsilon_0 \phi$

$= 3.14 \times 8.854 \times 10^{-12}$ C

$= 2.78 \times 10^{-11}$ C

EXAMPLE 1.11

1.14 Applications of Gauss's Law

The electric field due to a general charge distribution is, as seen above, given by Eq. (1.27). In practice, except for some special cases, the summation (or integration) involved in this equation cannot be carried out to give electric field at every point in space. For some symmetric charge configurations, however, it is possible to obtain the electric field in a simple way using the Gauss's law. This is best understood by some examples.

1.14.1 Field due to an infinitely long straight uniformly charged wire

Consider an infinitely long thin straight wire with uniform linear charge density λ. The wire is obviously an axis of symmetry. Suppose we take the radial vector from O to P and rotate it around the wire. The points P, P′, P″ so obtained are completely equivalent with respect to the charged wire. This implies that the electric field must have the same magnitude at these points. The direction of electric field at every point must be radial (outward if $\lambda > 0$, inward if $\lambda < 0$). This is clear from Fig. 1.26.

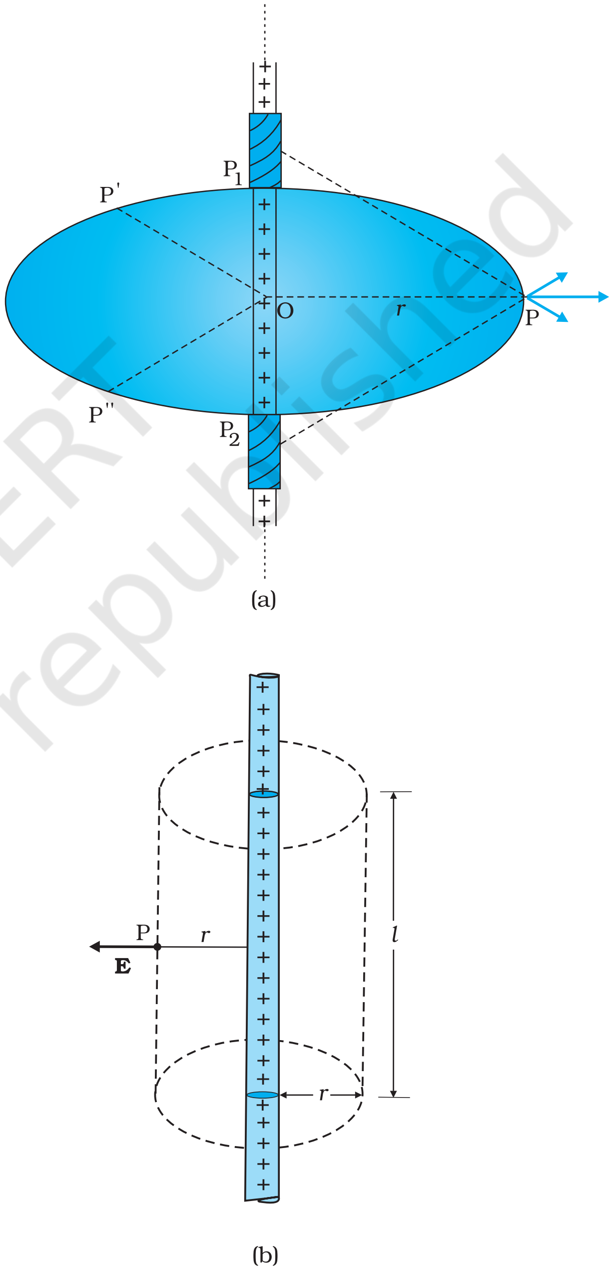


FIGURE 1.26 (a) Electric field due to an infinitely long thin straight wire is radial, (b) The Gaussian surface for a long thin wire of uniform linear charge density.

Consider a pair of line elements P_1 and P_2 of the wire, as shown. The electric fields produced by the two elements of the pair when summed give a resultant electric field which is radial (the components normal to the radial vector cancel). This is true for any such pair and hence the total field at any point P is radial. Finally, since the wire is infinite, electric field does not depend on the position of P along the length of the wire. In short, the electric field is everywhere radial in the plane cutting the wire normally, and its magnitude depends only on the radial distance *r*.

To calculate the field, imagine a cylindrical Gaussian surface, as shown in the Fig. 1.26(b). Since the field is everywhere radial, flux through the two ends of the cylindrical Gaussian surface is zero. At the cylindrical part of the surface, **E** is normal to the surface at every point, and its magnitude is constant, since it depends only on *r*. The surface area of the curved part is $2\pi rl$, where l is the length of the cylinder.

Flux through the Gaussian surface

= flux through the curved cylindrical part of the surface

= $E \times 2\pi rl$

The surface includes charge equal to $\lambda\, l$. Gauss's law then gives
$E \times 2\pi rl = \lambda l/\varepsilon_0$

i.e., $E = \dfrac{\lambda}{2\pi\varepsilon_0 r}$

Vectorially, **E** at any point is given by

$$\mathbf{E} = \frac{\lambda}{2\pi\varepsilon_0 r}\,\hat{\mathbf{n}} \qquad (1.32)$$

where $\hat{\mathbf{n}}$ is the radial unit vector in the plane normal to the wire passing through the point. **E** is directed outward if λ is positive and inward if λ is negative.

Note that when we write a vector **A** as a scalar multiplied by a unit vector, i.e., as $\mathbf{A} = A\,\hat{\mathbf{a}}$, the scalar A is an algebraic number. It can be negative or positive. The direction of **A** will be the same as that of the unit vector $\hat{\mathbf{a}}$ if $A > 0$ and opposite to $\hat{\mathbf{a}}$ if $A < 0$. When we want to restrict to non-negative values, we use the symbol $|\mathbf{A}|$ and call it the modulus of **A**. Thus, $|\mathbf{A}| \geq 0$.

Also note that though only the charge enclosed by the surface (λl) was included above, the electric field **E** is due to the charge on the entire wire. Further, the assumption that the wire is infinitely long is crucial. Without this assumption, we cannot take **E** to be normal to the curved part of the cylindrical Gaussian surface. However, Eq. (1.32) is approximately true for electric field around the central portions of a long wire, where the end effects may be ignored.

1.14.2 Field due to a uniformly charged infinite plane sheet

Let σ be the uniform surface charge density of an infinite plane sheet (Fig. 1.27). We take the x-axis normal to the given plane. By symmetry, the electric field will not depend on y and z coordinates and its direction at every point must be parallel to the x-direction.

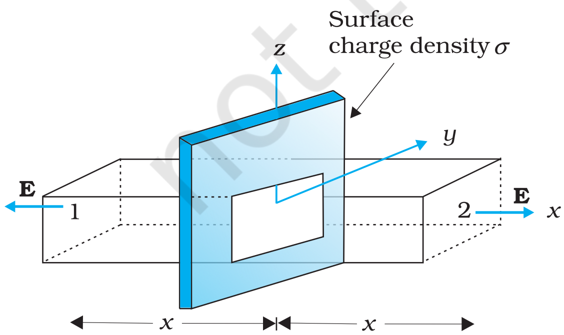


FIGURE 1.27 Gaussian surface for a uniformly charged infinite plane sheet.

We can take the Gaussian surface to be a rectangular parallelepiped of cross-sectional area A, as shown. (A cylindrical surface will also do.) As seen from the figure, only the two faces 1 and 2 will contribute to the flux; electric field lines are parallel to the other faces and they, therefore, do not contribute to the total flux.

The unit vector normal to surface 1 is in $-x$ direction while the unit vector normal to surface 2 is in the $+x$ direction. Therefore, flux $\mathbf{E}.\Delta\mathbf{S}$ through both the surfaces are equal and add up. Therefore the net flux through the Gaussian surface is $2\,EA$. The charge enclosed by the closed surface is σA. Therefore by Gauss's law,

Reprint 2025-26

$2\,EA = \sigma A/\varepsilon_0$
or, $E = \sigma/2\varepsilon_0$
Vectorially,

$$\mathbf{E} = \frac{\sigma}{2\varepsilon_0}\hat{\mathbf{n}} \tag{1.33}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane and going away from it.

E is directed away from the plate if σ is positive and toward the plate if σ is negative. Note that the above application of the Gauss' law has brought out an additional fact: *E* is independent of *x* also.

For a finite large planar sheet, Eq. (1.33) is approximately true in the middle regions of the planar sheet, away from the ends.

1.14.3 Field due to a uniformly charged thin spherical shell

Let σ be the uniform surface charge density of a thin spherical shell of radius *R* (Fig. 1.28). The situation has obvious spherical symmetry. The field at any point P, outside or inside, can depend only on *r* (the radial distance from the centre of the shell to the point) and must be radial (i.e., along the radius vector).

(i) Field outside the shell: Consider a point P outside the shell with radius vector **r**. To calculate **E** at P, we take the Gaussian surface to be a sphere of radius *r* and with centre O, passing through P. All points on this sphere are equivalent relative to the given charged configuration. (That is what we mean by spherical symmetry.) The electric field at each point of the Gaussian surface, therefore, has the same magnitude *E* and is along the radius vector at each point. Thus, **E** and $\Delta\mathbf{S}$ at every point are parallel and the flux through each element is $E\,\Delta S$. Summing over all ΔS, the flux through the Gaussian surface is $E \times 4\,\pi\,r^2$. The charge enclosed is $\sigma \times 4\,\pi\,R^2$. By Gauss's law

$$E \times 4\,\pi\,r^2 = \frac{\sigma}{\varepsilon_0}\,4\,\pi\,R^2$$

$$\text{Or,}\quad E = \frac{\sigma R^2}{\varepsilon_0\,r^2} = \frac{q}{4\pi\varepsilon_0 r^2}$$

where $q = 4\,\pi R^2\,\sigma$ is the total charge on the spherical shell.
Vectorially,

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0\,r^2}\hat{\mathbf{r}} \tag{1.34}$$

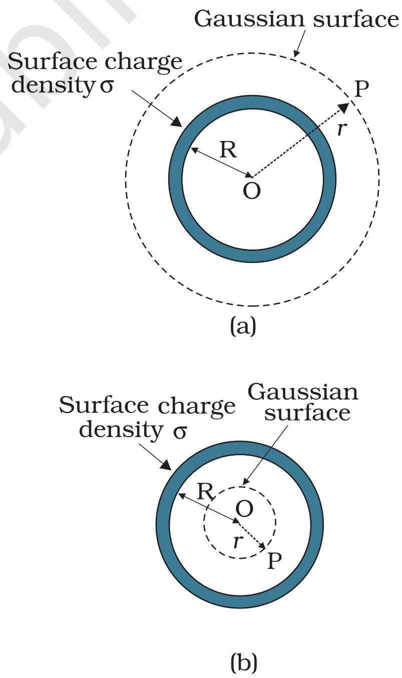


FIGURE 1.28 Gaussian surfaces for a point with (a) $r > R$, (b) $r < R$.

The electric field is directed outward if $q > 0$ and inward if $q < 0$. This, however, is exactly the field produced by a charge *q* placed at the centre O. Thus for points outside the shell, the field due to a uniformly charged shell is as if the entire charge of the shell is concentrated at its centre.

(ii) Field inside the shell: In Fig. 1.28(b), the point P is inside the shell. The Gaussian surface is again a sphere through P centred at O.

The flux through the Gaussian surface, calculated as before, is $E \times 4\pi r^2$. However, in this case, the Gaussian surface encloses no charge. Gauss's law then gives

$$E \times 4\pi r^2 = 0$$

i.e., $E = 0 \qquad (r < R)$ (1.35)

that is, the field due to a uniformly charged thin shell is zero at all points inside the shell*. This important result is a direct consequence of Gauss's law which follows from Coulomb's law. The experimental verification of this result confirms the $1/r^2$ dependence in Coulomb's law.

Example 1.12 An early model for an atom considered it to have a positively charged point nucleus of charge *Ze*, surrounded by a uniform density of negative charge up to a radius *R*. The atom as a whole is neutral. For this model, what is the electric field at a distance *r* from the nucleus?

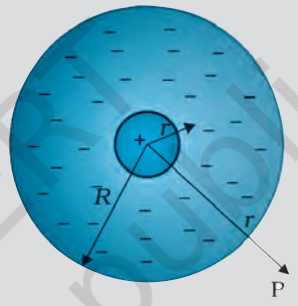


FIGURE 1.29

Solution The charge distribution for this model of the atom is as shown in Fig. 1.29. The total negative charge in the uniform spherical charge distribution of radius *R* must be $-Z e$, since the atom (nucleus of charge $Z e$ + negative charge) is neutral. This immediately gives us the negative charge density ρ, since we must have

$$\frac{4\pi R^3}{3}\rho = 0 - Ze$$

or $$\rho = -\frac{3\,Ze}{4\pi R^3}$$

To find the electric field $\mathbf{E}(\mathbf{r})$ at a point P which is a distance *r* away from the nucleus, we use Gauss's law. Because of the spherical symmetry of the charge distribution, the magnitude of the electric field $\mathbf{E}(\mathbf{r})$ depends only on the radial distance, no matter what the direction of $\mathbf{r}$. Its direction is along (or opposite to) the radius vector $\mathbf{r}$ from the origin to the point P. The obvious Gaussian surface is a spherical surface centred at the nucleus. We consider two situations, namely, $r < R$ and $r > R$.

(i) $r < R$: The electric flux ϕ enclosed by the spherical surface is

$$\phi = E(r) \times 4\pi r^2$$

EXAMPLE 1.12

* Compare this with a uniform mass shell discussed in Section 7.5 of Class XI Textbook of Physics.

where $E(r)$ is the magnitude of the electric field at r. This is because the field at any point on the spherical Gaussian surface has the same direction as the normal to the surface there, and has the same magnitude at all points on the surface.

The charge q enclosed by the Gaussian surface is the positive nuclear charge and the negative charge within the sphere of radius r,

i.e., $q = Ze + \dfrac{4\pi r^3}{3}\rho$

Substituting for the charge density ρ obtained earlier, we have

$$q = Ze - Ze\frac{r^3}{R^3}$$

Gauss's law then gives,

$$E(r) = \frac{Ze}{4\pi\varepsilon_0}\left(\frac{1}{r^2} - \frac{r}{R^3}\right); \quad r < R$$

The electric field is directed radially outward.

(ii) $r > R$: In this case, the total charge enclosed by the Gaussian spherical surface is zero since the atom is neutral. Thus, from Gauss's law,

$E(r) \times 4\pi r^2 = 0$ or $E(r) = 0$; $\quad r > R$

At $r = R$, both cases give the same result: $E = 0$.

EXAMPLE 1.12

SUMMARY

1. Electric and magnetic forces determine the properties of atoms, molecules and bulk matter.
2. From simple experiments on frictional electricity, one can infer that there are two types of charges in nature; and that like charges repel and unlike charges attract. By convention, the charge on a glass rod rubbed with silk is positive; that on a plastic rod rubbed with fur is then negative.
3. Conductors allow movement of electric charge through them, insulators do not. In metals, the mobile charges are electrons; in electrolytes both positive and negative ions are mobile.
4. Electric charge has three basic properties: quantisation, additivity and conservation.

   Quantisation of electric charge means that total charge (q) of a body is always an integral multiple of a basic quantum of charge (e) i.e., $q = ne$, where $n = 0, \pm1, \pm2, \pm3, ...$ Proton and electron have charges $+e$, $-e$, respectively. For macroscopic charges for which n is a very large number, quantisation of charge can be ignored.

   Additivity of electric charges means that the total charge of a system is the algebraic sum (i.e., the sum taking into account proper signs) of all individual charges in the system.

   Conservation of electric charges means that the total charge of an isolated system remains unchanged with time. This means that when bodies are charged through friction, there is a transfer of electric charge from one body to another, but no creation or destruction of charge.

5. *Coulomb's Law*: The mutual electrostatic force between two point charges q_1 and q_2 is proportional to the product $q_1 q_2$ and inversely proportional to the square of the distance r_{21} separating them. Mathematically,

$$\mathbf{F}_{21} = \text{force on } q_2 \text{ due to } q_1 = \frac{k\ (q_1 q_2)}{r_{21}^2}\ \hat{\mathbf{r}}_{21}$$

where $\hat{\mathbf{r}}_{21}$ is a unit vector in the direction from q_1 to q_2 and $k = \dfrac{1}{4\pi\varepsilon_0}$ is the constant of proportionality.

In SI units, the unit of charge is coulomb. The experimental value of the constant ε_0 is

$$\varepsilon_0 = 8.854 \times 10^{-12}\ \mathrm{C^2\ N^{-1}\ m^{-2}}$$

The approximate value of k is

$$k = 9 \times 10^9\ \mathrm{N\ m^2\ C^{-2}}$$

6. The ratio of electric force and gravitational force between a proton and an electron is

$$\frac{k\ e^2}{G\ \ m_e m_p} \cong 2.4\ \times\ 10^{39}$$

7. *Superposition Principle*: The principle is based on the property that the forces with which two charges attract or repel each other are not affected by the presence of a third (or more) additional charge(s). For an assembly of charges q_1, q_2, q_3, ..., the force on any charge, say q_1, is the vector sum of the force on q_1 due to q_2, the force on q_1 due to q_3, and so on. For each pair, the force is given by the Coulomb's law for two charges stated earlier.

8. The electric field **E** at a point due to a charge configuration is the force on a small positive test charge q placed at the point divided by the magnitude of the charge. Electric field due to a point charge q has a magnitude $|q|/4\pi\varepsilon_0 r^2$; it is radially outwards from q, if q is positive, and radially inwards if q is negative. Like Coulomb force, electric field also satisfies superposition principle.

9. An electric field line is a curve drawn in such a way that the tangent at each point on the curve gives the direction of electric field at that point. The relative closeness of field lines indicates the relative strength of electric field at different points; they crowd near each other in regions of strong electric field and are far apart where the electric field is weak. In regions of constant electric field, the field lines are uniformly spaced parallel straight lines.

10. Some of the important properties of field lines are: (i) Field lines are continuous curves without any breaks. (ii) Two field lines cannot cross each other. (iii) Electrostatic field lines start at positive charges and end at negative charges —they cannot form closed loops.

11. An electric dipole is a pair of equal and opposite charges q and $-q$ separated by some distance $2a$. Its dipole moment vector $\mathbf{p}$ has magnitude $2qa$ and is in the direction of the dipole axis from $-q$ to q.

12. Field of an electric dipole in its equatorial plane (i.e., the plane perpendicular to its axis and passing through its centre) at a distance r from the centre:

$$\mathbf{E} = \frac{-\mathbf{p}}{4\,\pi\varepsilon_o} \frac{1}{(a^2 + r^2)^{3/2}}$$

$$\cong \frac{-\mathbf{p}}{4\,\pi\varepsilon_o\, r^3}, \qquad \textit{for } r >> a$$

Dipole electric field on the axis at a distance r from the centre:

$$\mathbf{E} = \frac{2\,\mathbf{p}r}{4\,\pi\varepsilon_0 (r^2 - a^2)^2}$$

$$\cong \frac{2\mathbf{p}}{4\,\pi\varepsilon_0 r^3} \qquad \textit{for} \quad r >> a$$

The $1/r^3$ dependence of dipole electric fields should be noted in contrast to the $1/r^2$ dependence of electric field due to a point charge.

13. In a uniform electric field **E**, a dipole experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E}$$

but experiences no net force.

14. The flux $\Delta\phi$ of electric field **E** through a small area element $\Delta\mathbf{S}$ is given by

$$\Delta\phi = \mathbf{E} \cdot \Delta\mathbf{S}$$

The vector area element $\Delta\mathbf{S}$ is

$$\Delta\mathbf{S} = \Delta S\ \hat{\mathbf{n}}$$

where ΔS is the magnitude of the area element and $\hat{\mathbf{n}}$ is normal to the area element, which can be considered planar for sufficiently small ΔS.

For an area element of a closed surface, $\hat{\mathbf{n}}$ is taken to be the direction of *outward* normal, by convention.

15. *Gauss's law*: The flux of electric field through any closed surface S is $1/\varepsilon_0$ times the total charge enclosed by S. The law is especially useful in determining electric field **E**, when the source distribution has simple symmetry:

(i) *Thin infinitely long straight wire of uniform linear charge density* λ

$$\mathbf{E} = \frac{\lambda}{2\,\pi\varepsilon_0\, r}\ \hat{\mathbf{n}}$$

where r is the perpendicular distance of the point from the wire and $\hat{\mathbf{n}}$ is the radial unit vector in the plane normal to the wire passing through the point.

(ii) Infinite thin plane sheet of uniform surface charge density σ

$$\mathbf{E} = \frac{\sigma}{2\,\varepsilon_0}\ \hat{\mathbf{n}}$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane, outward on either side.

Reprint 2025-26

(iii) *Thin spherical shell of uniform surface charge density* σ

$$\mathbf{E} = \frac{q}{4\pi\varepsilon_0 r^2}\hat{\mathbf{r}} \qquad (r \geq R)$$

$$\mathbf{E} = 0 \qquad (r < R)$$

where r is the distance of the point from the centre of the shell and R the radius of the shell. q is the total charge of the shell: $q = 4\pi R^2\sigma$.

The electric field outside the shell is as though the total charge is concentrated at the centre. The same result is true for a solid sphere of uniform volume charge density. The field is zero at all points inside the shell.

Physical quantity	Symbol	Dimensions	Unit	Remarks
Vector area element	$\Delta\mathbf{S}$	$[L^2]$	m^2	$\Delta\mathbf{S} = \Delta S\,\hat{\mathbf{n}}$
Electric field	$\mathbf{E}$	$[MLT^{-3}A^{-1}]$	$V\,m^{-1}$	
Electric flux	ϕ	$[ML^3\,T^{-3}A^{-1}]$	V m	$\Delta\phi = \mathbf{E}\cdot\Delta\mathbf{S}$
Dipole moment	$\mathbf{p}$	[LTA]	C m	Vector directed from negative to positive charge
Charge density:				
linear	λ	$[L^{-1}\,TA]$	$C\,m^{-1}$	Charge/length
surface	σ	$[L^{-2}\,TA]$	$C\,m^{-2}$	Charge/area
volume	ρ	$[L^{-3}\,TA]$	$C\,m^{-3}$	Charge/volume

POINTS TO PONDER

1. You might wonder why the protons, all carrying positive charges, are compactly residing inside the nucleus. Why do they not fly away? You will learn that there is a third kind of a fundamental force, called the strong force which holds them together. The range of distance where this force is effective is, however, very small ~10^{-14} m. This is precisely the size of the nucleus. Also the electrons are not allowed to sit on top of the protons, i.e. inside the nucleus, due to the laws of quantum mechanics. This gives the atoms their structure as they exist in nature.
2. Coulomb force and gravitational force follow the same inverse-square law. But gravitational force has only one sign (always attractive), while

Reprint 2025-26

Coulomb force can be of both signs (attractive and repulsive), allowing possibility of cancellation of electric forces. This is how gravity, despite being a much weaker force, can be a dominating and more pervasive force in nature.

3. The constant of proportionality k in Coulomb's law is a matter of choice if the unit of charge is to be defined using Coulomb's law. In SI units, however, what is defined is the unit of current (A) via its magnetic effect (Ampere's law) and the unit of charge (coulomb) is simply defined by (1C = 1 A s). In this case, the value of k is no longer arbitrary; it is approximately 9×10^9 N m^2 C^{-2}.
4. The rather large value of k, i.e., the large size of the unit of charge (1C) from the point of view of electric effects arises because (as mentioned in point 3 already) the unit of charge is defined in terms of magnetic forces (forces on current–carrying wires) which are generally much weaker than the electric forces. Thus while 1 ampere is a unit of reasonable size for magnetic effects, 1 C = 1 A s, is too big a unit for electric effects.
5. The additive property of charge is not an 'obvious' property. It is related to the fact that electric charge has no direction associated with it; charge is a scalar.
6. Charge is not only a scalar (or invariant) under rotation; it is also invariant for frames of reference in relative motion. This is not always true for every scalar. For example, kinetic energy is a scalar under rotation, but is not invariant for frames of reference in relative motion.
7. Conservation of total charge of an isolated system is a property independent of the scalar nature of charge noted in point 6. Conservation refers to invariance in time in a given frame of reference. A quantity may be scalar but not conserved (like kinetic energy in an inelastic collision). On the other hand, one can have conserved vector quantity (e.g., angular momentum of an isolated system).
8. Quantisation of electric charge is a basic (unexplained) law of nature; interestingly, there is no analogous law on quantisation of mass.
9. Superposition principle should not be regarded as 'obvious', or equated with the law of addition of vectors. It says two things: force on one charge due to another charge is unaffected by the presence of other charges, and there are no additional three-body, four-body, etc., forces which arise only when there are more than two charges.
10. The electric field due to a discrete charge configuration is not defined at the locations of the discrete charges. For continuous volume charge distribution, it is defined at any point in the distribution. For a surface charge distribution, electric field is discontinuous across the surface.
11. The electric field due to a charge configuration with total charge zero is not zero; but for distances large compared to the size of the configuration, its field falls off faster than $1/r^2$, typical of field due to a single charge. An electric dipole is the simplest example of this fact.

Reprint 2025-26

EXERCISES

1.1 What is the force between two small charged spheres having charges of 2×10^{-7}C and 3×10^{-7}C placed 30 cm apart in air?

1.2 The electrostatic force on a small sphere of charge 0.4 μC due to another small sphere of charge –0.8 μC in air is 0.2 N. (a) What is the distance between the two spheres? (b) What is the force on the second sphere due to the first?

1.3 Check that the ratio $ke^2/G\ m_e m_p$ is dimensionless. Look up a Table of Physical Constants and determine the value of this ratio. What does the ratio signify?

1.4 (a) Explain the meaning of the statement 'electric charge of a body is quantised'.
(b) Why can one ignore quantisation of electric charge when dealing with macroscopic i.e., large scale charges?

1.5 When a glass rod is rubbed with a silk cloth, charges appear on both. A similar phenomenon is observed with many other pairs of bodies. Explain how this observation is consistent with the law of conservation of charge.

1.6 Four point charges q_A = 2 μC, q_B = –5 μC, q_C = 2 μC, and q_D = –5 μC are located at the corners of a square ABCD of side 10 cm. What is the force on a charge of 1 μC placed at the centre of the square?

1.7 (a) An electrostatic field line is a continuous curve. That is, a field line cannot have sudden breaks. Why not?
(b) Explain why two field lines never cross each other at any point?

1.8 Two point charges q_A = 3 μC and q_B = –3 μC are located 20 cm apart in vacuum.
(a) What is the electric field at the midpoint O of the line AB joining the two charges?
(b) If a negative test charge of magnitude 1.5×10^{-9} C is placed at this point, what is the force experienced by the test charge?

1.9 A system has two charges $q_A = 2.5 \times 10^{-7}$ C and $q_B = -2.5 \times 10^{-7}$ C located at points A: (0, 0, –15 cm) and B: (0,0, +15 cm), respectively. What are the total charge and electric dipole moment of the system?

1.10 An electric dipole with dipole moment 4×10^{-9} C m is aligned at 30° with the direction of a uniform electric field of magnitude 5×10^4 NC^{-1}. Calculate the magnitude of the torque acting on the dipole.

1.11 A polythene piece rubbed with wool is found to have a negative charge of 3×10^{-7} C.
(a) Estimate the number of electrons transferred (from which to which?)
(b) Is there a transfer of mass from wool to polythene?

1.12 (a) Two insulated charged copper spheres A and B have their centres separated by a distance of 50 cm. What is the mutual force of electrostatic repulsion if the charge on each is 6.5×10^{-7} C? The radii of A and B are negligible compared to the distance of separation.
(b) What is the force of repulsion if each sphere is charged double the above amount, and the distance between them is halved?

1.13 Figure 1.30 shows tracks of three charged particles in a uniform electrostatic field. Give the signs of the three charges. Which particle has the highest charge to mass ratio?

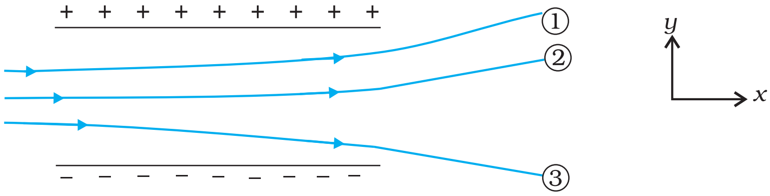


FIGURE 1.30

1.14 Consider a uniform electric field $\mathbf{E} = 3 \times 10^3\ \hat{\mathbf{i}}$ N/C. (a) What is the flux of this field through a square of 10 cm on a side whose plane is parallel to the *yz* plane? (b) What is the flux through the same square if the normal to its plane makes a 60° angle with the *x*-axis?

1.15 What is the net flux of the uniform electric field of Exercise 1.14 through a cube of side 20 cm oriented so that its faces are parallel to the coordinate planes?

1.16 Careful measurement of the electric field at the surface of a black box indicates that the net outward flux through the surface of the box is 8.0×10^3 Nm^2/C. (a) What is the net charge inside the box? (b) If the net outward flux through the surface of the box were zero, could you conclude that there were no charges inside the box? Why or Why not?

1.17 A point charge +10 μC is a distance 5 cm directly above the centre of a square of side 10 cm, as shown in Fig. 1.31. What is the magnitude of the electric flux through the square? (*Hint:* Think of the square as one face of a cube with edge 10 cm.)

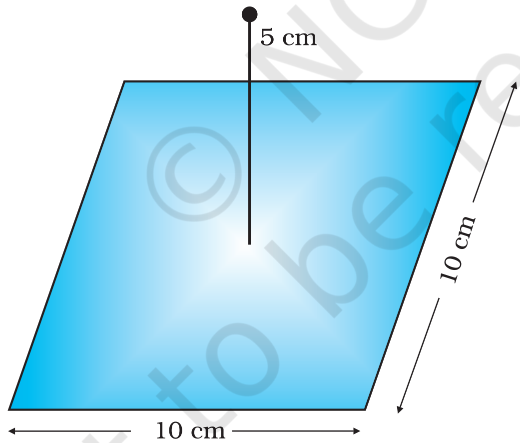


FIGURE 1.31

1.18 A point charge of 2.0 μC is at the centre of a cubic Gaussian surface 9.0 cm on edge. What is the net electric flux through the surface?

1.19 A point charge causes an electric flux of -1.0×10^3 Nm^2/C to pass through a spherical Gaussian surface of 10.0 cm radius centred on the charge. (a) If the radius of the Gaussian surface were doubled, how much flux would pass through the surface? (b) What is the value of the point charge?

1.20 A conducting sphere of radius 10 cm has an unknown charge. If the electric field 20 cm from the centre of the sphere is 1.5×10^3 N/C and points radially inward, what is the net charge on the sphere?

Reprint 2025-26

1.21 A uniformly charged conducting sphere of 2.4 m diameter has a surface charge density of 80.0 $\mu C/m^2$. (a) Find the charge on the sphere. (b) What is the total electric flux leaving the surface of the sphere?

1.22 An infinite line charge produces a field of 9×10^4 N/C at a distance of 2 cm. Calculate the linear charge density.

1.23 Two large, thin metal plates are parallel and close to each other. On their inner faces, the plates have surface charge densities of opposite signs and of magnitude 17.0×10^{-22} C/m^2. What is **E**: (a) in the outer region of the first plate, (b) in the outer region of the second plate, and (c) between the plates?

12089CH02

Chapter Two

ELECTROSTATIC POTENTIAL AND CAPACITANCE

2.1 Introduction

In Chapters 5 and 7 (Class XI), the notion of potential energy was introduced. When an external force does work in taking a body from a point to another against a force like spring force or gravitational force, that work gets stored as potential energy of the body. When the external force is removed, the body moves, gaining kinetic energy and losing an equal amount of potential energy. The sum of kinetic and potential energies is thus conserved. Forces of this kind are called conservative forces. Spring force and gravitational force are examples of conservative forces.

Coulomb force between two (stationary) charges is also a conservative force. This is not surprising, since both have inverse-square dependence on distance and differ mainly in the proportionality constants – the masses in the gravitational law are replaced by charges in Coulomb's law. Thus, like the potential energy of a mass in a gravitational field, we can define electrostatic potential energy of a charge in an electrostatic field.

Consider an electrostatic field **E** due to some charge configuration. First, for simplicity, consider the field **E** due to a charge Q placed at the origin. Now, imagine that we bring a test charge q from a point R to a point P against the repulsive force on it due to the charge Q. With reference

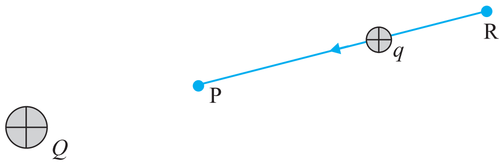


FIGURE 2.1 A test charge q (> 0) is moved from the point R to the point P against the repulsive force on it by the charge Q (> 0) placed at the origin.

to Fig. 2.1, this will happen if Q and q are both positive or both negative. For definiteness, let us take Q, $q > 0$.

Two remarks may be made here. First, we assume that the test charge q is so small that it does not disturb the original configuration, namely the charge Q at the origin (or else, we keep Q fixed at the origin by some unspecified force). Second, in bringing the charge q from R to P, we apply an external force $\mathbf{F}_{ext}$ just enough to counter the repulsive electric force $\mathbf{F}_E$ (i.e, $\mathbf{F}_{ext} = -\mathbf{F}_E$). This means there is no net force on or acceleration of the charge q when it is brought from R to P, i.e., it is brought with infinitesimally slow constant speed. In this situation, work done by the external force is the negative of the work done by the electric force, and gets fully stored in the form of potential energy of the charge q. If the external force is removed on reaching P, the electric force will take the charge away from Q – the stored energy (potential energy) at P is used to provide kinetic energy to the charge q in such a way that the sum of the kinetic and potential energies is conserved.

Thus, work done by external forces in moving a charge q from R to P is

$$W_{RP} = \int_R^P \mathbf{F}_{ext} \cdot d\mathbf{r}$$

$$= -\int_R^P \mathbf{F}_{ext} \cdot d\mathbf{r} \qquad (2.1)$$

This work done is against electrostatic repulsive force and gets stored as potential energy.

At every point in electric field, a particle with charge q possesses a certain electrostatic potential energy, this work done increases its potential energy by an amount equal to potential energy difference between points R and P.

Thus, potential energy difference

$$\Delta U = U_P - U_R = W_{RP} \qquad (2.2)$$

(Note here that this displacement is in an opposite sense to the electric force and hence work done by electric field is negative, i.e., $-W_{RP}$.)

Therefore, we can define electric potential energy difference between two points as the work required to be done by an external force in moving (without accelerating) charge q from one point to another for electric field of any arbitrary charge configuration.

Two important comments may be made at this stage:

(i) The right side of Eq. (2.2) depends only on the initial and final positions of the charge. It means that the work done by an electrostatic field in moving a charge from one point to another depends only on the initial and the final points and is independent of the path taken to go from one point to the other. This is the fundamental characteristic of a conservative force. The concept of the potential energy would not be meaningful if the work depended on the path. The path-independence of work done by an electrostatic field can be proved using the Coulomb's law. We omit this proof here.

Reprint 2025-26

(ii) Equation (2.2) defines *potential energy difference* in terms of the physically meaningful quantity *work*. Clearly, potential energy so defined is undetermined to within an additive constant. What this means is that the actual value of potential energy is not physically significant; it is only the difference of potential energy that is significant. We can always add an arbitrary constant α to potential energy at every point, since this will not change the potential energy difference:

$$(U_P + \alpha) - (U_R + \alpha) = U_P - U_R$$

Put it differently, there is a freedom in choosing the point where potential energy is zero. A convenient choice is to have electrostatic potential energy zero at infinity. With this choice, if we take the point R at infinity, we get from Eq. (2.2)

$$W_{\infty P} = U_P - U_\infty = U_P \tag{2.3}$$

Since the point P is arbitrary, Eq. (2.3) provides us with a definition of potential energy of a charge q at any point. *Potential energy of charge q at a point* (in the presence of field due to any charge configuration) *is the work done by the external force* (equal and opposite to the electric force) *in bringing the charge q from infinity to that point.*

Count Alessandro Volta (1745 – 1827) Italian physicist, professor at Pavia. Volta established that the *animal electricity* observed by Luigi Galvani, 1737–1798, in experiments with frog muscle tissue placed in contact with dissimilar metals, was not due to any exceptional property of animal tissues but was also generated whenever any wet body was sandwiched between dissimilar metals. This led him to develop the first *voltaic pile*, or battery, consisting of a large stack of moist disks of cardboard (electrolyte) sandwiched between disks of metal (electrodes).

COUNT ALESSANDRO VOLTA (1745 – 1827)

2.2 Electrostatic Potential

Consider any general static charge configuration. We define potential energy of a test charge q in terms of the work done on the charge q. This work is obviously proportional to q, since the force at any point is $q\mathbf{E}$, where $\mathbf{E}$ is the electric field at that point due to the given charge configuration. It is, therefore, convenient to divide the work by the amount of charge q, so that the resulting quantity is independent of q. In other words, work done per unit test charge is characteristic of the electric field associated with the charge configuration. This leads to the idea of electrostatic potential V due to a given charge configuration. From Eq. (2.1), we get:

Work done by external force in bringing a unit positive charge from point R to P

$$= V_P - V_R \left(= \frac{U_P - U_R}{q} \right) \tag{2.4}$$

where V_P and V_R are the electrostatic potentials at P and R, respectively. Note, as before, that it is not the actual value of potential but the potential difference that is physically significant. If, as before, we choose the potential to be zero at infinity, Eq. (2.4) implies:

Work done by an external force in bringing a unit positive charge from infinity to a point = electrostatic potential (V) at that point.

Reprint 2025-26

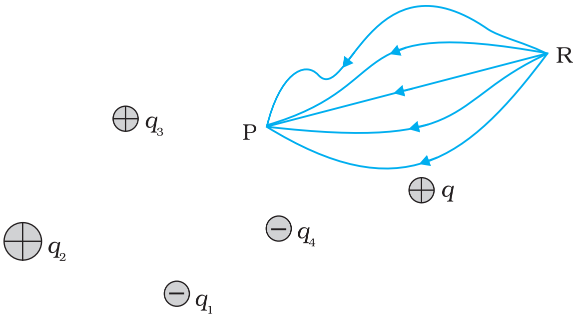


FIGURE 2.2 Work done on a test charge q by the electrostatic field due to any given charge configuration is independent of the path, and depends only on its initial and final positions.

In other words, the electrostatic potential (V) at any point in a region with electrostatic field is the work done in bringing a unit positive charge (without acceleration) from infinity to that point.

The qualifying remarks made earlier regarding potential energy also apply to the definition of potential. To obtain the work done per unit test charge, we should take an infinitesimal test charge δq, obtain the work done δW in bringing it from infinity to the point and determine the ratio $\delta W/\delta q$. Also, the external force at every point of the path is to be equal and opposite to the electrostatic force on the test charge at that point.

2.3 Potential due to a Point Charge

Consider a point charge Q at the origin (Fig. 2.3). For definiteness, take Q to be positive. We wish to determine the potential at any point P with position vector **r** from the origin. For that we must calculate the work done in bringing a unit positive test charge from infinity to the point P. For $Q > 0$, the work done against the repulsive force on the test charge is positive. Since work done is independent of the path, we choose a convenient path – along the radial direction from infinity to the point P.

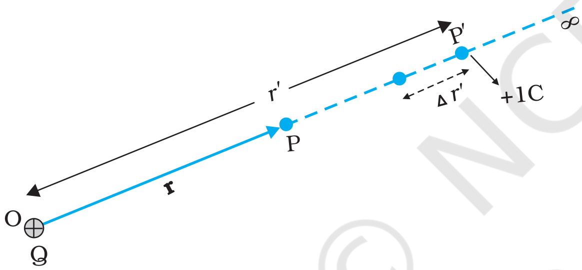


FIGURE 2.3 Work done in bringing a unit positive test charge from infinity to the point P, against the repulsive force of charge Q ($Q > 0$), is the potential at P due to the charge Q.

At some intermediate point P′ on the path, the electrostatic force on a unit positive charge is

$$\frac{Q \times 1}{4\pi\varepsilon_0 r'^2}\hat{\mathbf{r}}' \qquad (2.5)$$

where $\hat{\mathbf{r}}'$ is the unit vector along OP′. Work done against this force from $\mathbf{r}'$ to $\mathbf{r}' + \Delta\mathbf{r}'$ is

$$\Delta W = -\frac{Q}{4\pi\varepsilon_0 r'^2}\Delta r' \qquad (2.6)$$

The negative sign appears because for $\Delta r' < 0$, ΔW is positive. Total work done (W) by the external force is obtained by integrating Eq. (2.6) from $r' = \infty$ to $r' = r$,

$$W = -\int_{\infty}^{r}\frac{Q}{4\pi\varepsilon_0 r'^2}dr' = \left.\frac{Q}{4\pi\varepsilon_0 r'}\right|_{\infty}^{r} = \frac{Q}{4\pi\varepsilon_0 r} \qquad (2.7)$$

This, by definition is the potential at P due to the charge Q

$$V(r) = \frac{Q}{4\pi\varepsilon_0 r} \qquad (2.8)$$

Reprint 2025-26

Equation (2.8) is true for any sign of the charge Q, though we considered $Q > 0$ in its derivation. For $Q < 0$, $V < 0$, i.e., work done (by the external force) per unit positive test charge in bringing it from infinity to the point is negative. This is equivalent to saying that work done by the electrostatic force in bringing the unit positive charge form infinity to the point P is positive. [This is as it should be, since for $Q < 0$, the force on a unit positive test charge is attractive, so that the electrostatic force and the displacement (from infinity to P) are in the same direction.] Finally, we note that Eq. (2.8) is consistent with the choice that potential at infinity be zero.

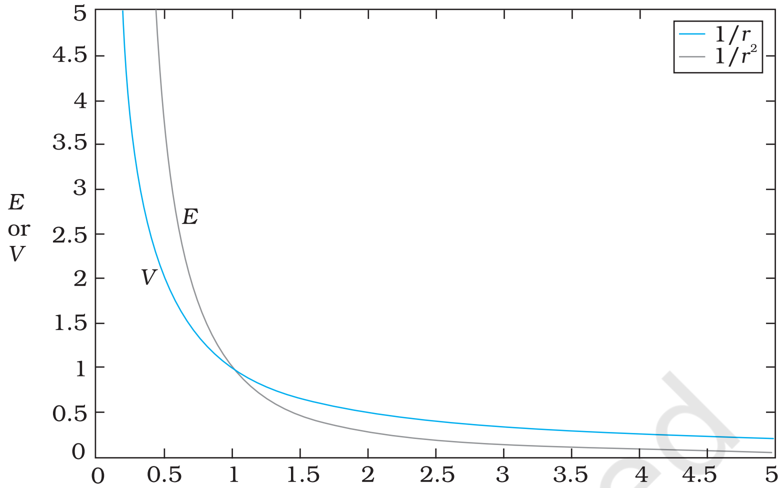


FIGURE 2.4 Variation of potential V with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-1}] (blue curve) and field with r [in units of $(Q/4\pi\varepsilon_0)$ m^{-2}] (black curve) for a point charge Q.

Figure (2.4) shows how the electrostatic potential ($\propto 1/r$) and the electrostatic field ($\propto 1/r^2$) varies with r.

Example 2.1

(a) Calculate the potential at a point P due to a charge of 4×10^{-7}C located 9 cm away.

(b) Hence obtain the work done in bringing a charge of 2×10^{-9} C from infinity to the point P. Does the answer depend on the path along which the charge is brought?

Solution

(a) $V = \dfrac{1}{4\pi\varepsilon_0}\dfrac{Q}{r} = 9 \times 10^9 \text{ Nm}^2 \text{ C}^{-2} \times \dfrac{4 \times 10^{-7}\text{ C}}{0.09\text{ m}}$

$= 4 \times 10^4$ V

(b) $W = qV = 2 \times 10^{-9}\text{C} \times 4 \times 10^4\text{V}$

$= 8 \times 10^{-5}$ J

No, work done will be path independent. Any arbitrary infinitesimal path can be resolved into two perpendicular displacements: One along **r** and another perpendicular to **r**. The work done corresponding to the later will be zero.

EXAMPLE 2.1

2.4 Potential due to an Electric Dipole

As we learnt in the last chapter, an electric dipole consists of two charges q and $-q$ separated by a (small) distance $2a$. Its total charge is zero. It is characterised by a dipole moment vector **p** whose magnitude is $q \times 2a$ and which points in the direction from $-q$ to q (Fig. 2.5). We also saw that the electric field of a dipole at a point with position vector **r** depends not just on the magnitude r, but also on the angle between **r** and **p**. Further,

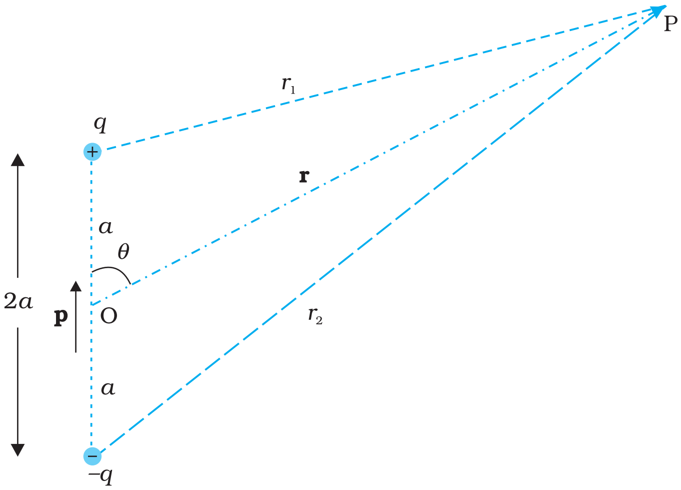


FIGURE 2.5 Quantities involved in the calculation of potential due to a dipole.

the field falls off, at large distance, not as $1/r^2$ (typical of field due to a single charge) but as $1/r^3$. We, now, determine the electric potential due to a dipole and contrast it with the potential due to a single charge.

As before, we take the origin at the centre of the dipole. Now we know that the electric field obeys the superposition principle. Since potential is related to the work done by the field, electrostatic potential also follows the superposition principle. Thus, the potential due to the dipole is the sum of potentials due to the charges q and $-q$

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r_1} - \frac{q}{r_2}\right) \tag{2.9}$$

where r_1 and r_2 are the distances of the point P from q and $-q$, respectively.

Now, by geometry,

$$r_1^2 = r^2 + a^2 - 2ar\cos\theta$$

$$r_2^2 = r^2 + a^2 + 2ar\cos\theta \tag{2.10}$$

We take r much greater than a ($r >> a$) and retain terms only upto the first order in a/r

$$r_1^2 = r^2\left(1 - \frac{2a\cos\theta}{r} + \frac{a^2}{r^2}\right)$$

$$\cong r^2\left(1 - \frac{2a\cos\theta}{r}\right) \tag{2.11}$$

Similarly,

$$r_2^2 \cong r^2\left(1 + \frac{2a\cos\theta}{r}\right) \tag{2.12}$$

Using the Binomial theorem and retaining terms upto the first order in a/r; we obtain,

$$\frac{1}{r_1} \cong \frac{1}{r}\left(1 - \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 + \frac{a}{r}\cos\theta\right) \tag{2.13(a)}$$

$$\frac{1}{r_2} \cong \frac{1}{r}\left(1 + \frac{2a\cos\theta}{r}\right)^{-1/2} \cong \frac{1}{r}\left(1 - \frac{a}{r}\cos\theta\right) \tag{2.13(b)}$$

Using Eqs. (2.9) and (2.13) and $p = 2qa$, we get

$$V = \frac{q}{4\pi\varepsilon_0}\frac{2a\cos\theta}{r^2} = \frac{p\cos\theta}{4\pi\varepsilon_0 r^2} \tag{2.14}$$

Now, $p\cos\theta = \mathbf{p}.\hat{\mathbf{r}}$

Reprint 2025-26

where $\hat{\mathbf{r}}$ is the unit vector along the position vector **OP**.

The electric potential of a dipole is then given by

$$V = \frac{1}{4\pi\varepsilon_0}\frac{\mathbf{p}.\hat{\mathbf{r}}}{r^2}\,; \qquad (r >> a) \tag{2.15}$$

Equation (2.15) is, as indicated, approximately true only for distances large compared to the size of the dipole, so that higher order terms in a/r are negligible. For a point dipole **p** at the origin, Eq. (2.15) is, however, exact.

From Eq. (2.15), potential on the dipole axis ($\theta = 0, \pi$) is given by

$$V = \pm\frac{1}{4\pi\varepsilon_0}\frac{p}{r^2} \tag{2.16}$$

(Positive sign for $\theta = 0$, negative sign for $\theta = \pi$.) The potential in the equatorial plane ($\theta = \pi/2$) is zero.

The important contrasting features of electric potential of a dipole from that due to a single charge are clear from Eqs. (2.8) and (2.15):

(i) The potential due to a dipole depends not just on r but also on the angle between the position vector **r** and the dipole moment vector **p**. (It is, however, axially symmetric about **p**. That is, if you rotate the position vector **r** about **p**, keeping θ fixed, the points corresponding to P on the cone so generated will have the same potential as at P.)

(ii) The electric dipole potential falls off, at large distance, as $1/r^2$, not as $1/r$, characteristic of the potential due to a single charge. (You can refer to the Fig. 2.5 for graphs of $1/r^2$ versus r and $1/r$ versus r, drawn there in another context.)

2.5 Potential due to a System of Charges

Consider a system of charges $q_1, q_2,\ldots, q_n$ with position vectors $\mathbf{r}_1, \mathbf{r}_2,\ldots, \mathbf{r}_n$ relative to some origin (Fig. 2.6). The potential V_1 at P due to the charge q_1 is

$$V_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}}$$

where r_{1P} is the distance between q_1 and P.

Similarly, the potential V_2 at P due to q_2 and V_3 due to q_3 are given by

$$V_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_2}{r_{2P}}, \quad V_3 = \frac{1}{4\pi\varepsilon_0}\frac{q_3}{r_{3P}}$$

where r_{2P} and r_{3P} are the distances of P from charges q_2 and q_3, respectively; and so on for the potential due to other charges. By the superposition principle, the potential V at P due to the total charge configuration is the algebraic sum of the potentials due to the individual charges

$$V = V_1 + V_2 + \ldots + V_n \tag{2.17}$$

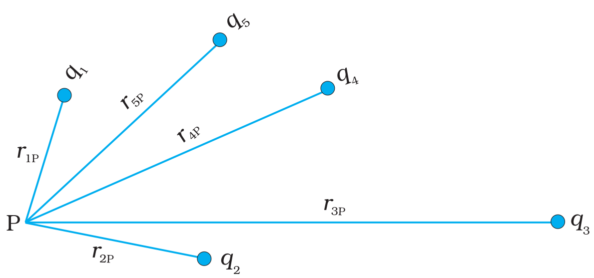


FIGURE 2.6 Potential at a point due to a system of charges is the sum of potentials due to individual charges.

Reprint 2025-26

$$= \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}} + ... + \frac{q_n}{r_{nP}}\right) \tag{2.18}$$

If we have a continuous charge distribution characterised by a charge density ρ (**r**), we divide it, as before, into small volume elements each of size Δv and carrying a charge $\rho\Delta v$. We then determine the potential due to each volume element and sum (strictly speaking , integrate) over all such contributions, and thus determine the potential due to the entire distribution.

We have seen in Chapter 1 that for a uniformly charged spherical shell, the electric field outside the shell is as if the entire charge is concentrated at the centre. Thus, the potential outside the shell is given by

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{r} \quad (r \geq R) \qquad [2.19(a)]$$

where q is the total charge on the shell and R its radius. The electric field inside the shell is zero. This implies (Section 2.6) that potential is constant inside the shell (as no work is done in moving a charge inside the shell), and, therefore, equals its value at the surface, which is

$$V = \frac{1}{4\pi\varepsilon_0}\frac{q}{R} \qquad [2.19(b)]$$

Example 2.2 Two charges 3×10^{-8} C and -2×10^{-8} C are located 15 cm apart. At what point on the line joining the two charges is the electric potential zero? Take the potential at infinity to be zero.

Solution Let us take the origin O at the location of the positive charge. The line joining the two charges is taken to be the x-axis; the negative charge is taken to be on the right side of the origin (Fig. 2.7).

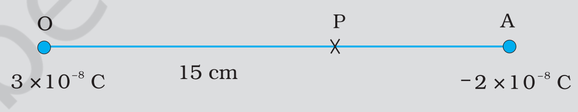


FIGURE 2.7

Let P be the required point on the x-axis where the potential is zero. If x is the x-coordinate of P, obviously x must be positive. (There is no possibility of potentials due to the two charges adding up to zero for $x < 0$.) If x lies between O and A, we have

$$\frac{1}{4\pi\varepsilon_0}\left[\frac{3\times10^{-8}}{x\times10^{-2}} - \frac{2\times10^{-8}}{(15-x)\times10^{-2}}\right] = 0$$

where x is in cm. That is,

$$\frac{3}{x} - \frac{2}{15-x} = 0$$

which gives x = 9 cm.

If x lies on the extended line OA, the required condition is

$$\frac{3}{x} - \frac{2}{x-15} = 0$$

EXAMPLE 2.2

which gives

$x = 45$ cm

Thus, electric potential is zero at 9 cm and 45 cm away from the positive charge on the side of the negative charge. Note that the formula for potential used in the calculation required choosing potential to be zero at infinity.

EXAMPLE 2.2

Example 2.3 Figures 2.8 (a) and (b) show the field lines of a positive and negative point charge respectively.

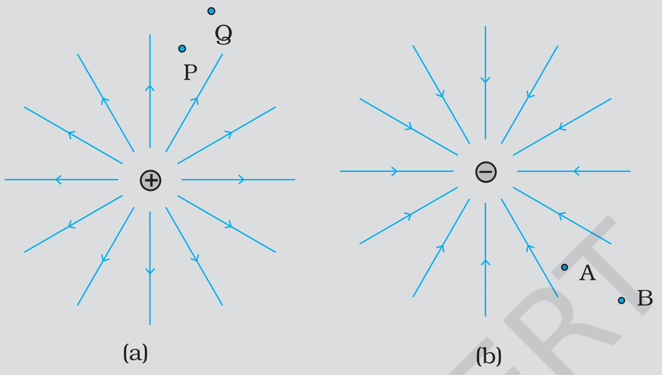


FIGURE 2.8

(a) Give the signs of the potential difference $V_P - V_Q$; $V_B - V_A$.
(b) Give the sign of the potential energy difference of a small negative charge between the points Q and P; A and B.
(c) Give the sign of the work done by the field in moving a small positive charge from Q to P.
(d) Give the sign of the work done by the external agency in moving a small negative charge from B to A.
(e) Does the kinetic energy of a small negative charge increase or decrease in going from B to A?

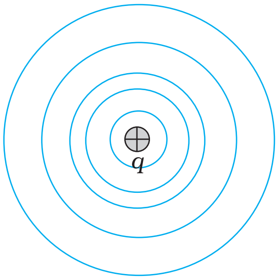
Solution

(a) As $V \propto \frac{1}{r}$, $V_P > V_Q$. Thus, $(V_P - V_Q)$ is positive. Also V_B is less negative than V_A. Thus, $V_B > V_A$ or $(V_B - V_A)$ is positive.
(b) A small negative charge will be attracted towards positive charge. The negative charge moves from higher potential energy to lower potential energy. Therefore the sign of potential energy difference of a small negative charge between Q and P is positive.
Similarly, $(P.E.)_A > (P.E.)_B$ and hence sign of potential energy differences is positive.
(c) In moving a small positive charge from Q to P, work has to be done by an external agency against the electric field. Therefore, work done by the field is negative.
(d) In moving a small negative charge from B to A work has to be done by the external agency. It is positive.
(e) Due to force of repulsion on the negative charge, velocity decreases and hence the kinetic energy decreases in going from B to A.

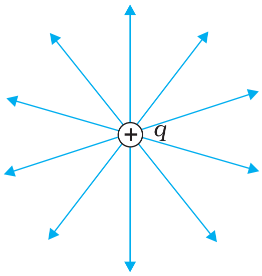
EXAMPLE 2.3

PHYSICS

Electric potential, equipotential surfaces:
http://video.mit.edu/watch/4-electrostatic-potential-electric-energy-ev-conservative-field-equipotential-sufaces-12584/



(a)



(b)

FIGURE 2.9 For a single charge q (a) equipotential surfaces are spherical surfaces centred at the charge, and (b) electric field lines are radial, starting from the charge if $q > 0$.

2.6 Equipotential Surfaces

An equipotential surface is a surface with a constant value of potential at all points on the surface. For a single charge q, the potential is given by Eq. (2.8):

$$V = \frac{1}{4\pi\varepsilon_o}\frac{q}{r}$$

This shows that V is a constant if r is constant. Thus, equipotential surfaces of a single point charge are concentric spherical surfaces centred at the charge.

Now the electric field lines for a single charge q are radial lines starting from or ending at the charge, depending on whether q is positive or negative. Clearly, the electric field at every point is normal to the equipotential surface passing through that point. This is true in general: *for any charge configuration, equipotential surface through a point is normal to the electric field at that point*. The proof of this statement is simple.

If the field were not normal to the equipotential surface, it would have non-zero component along the surface. To move a unit test charge against the direction of the component of the field, work would have to be done. But this is in contradiction to the definition of an equipotential surface: there is no potential difference between any two points on the surface and no work is required to move a test charge on the surface. The electric field must, therefore, be normal to the equipotential surface at every point. Equipotential surfaces offer an alternative visual picture in addition to the picture of electric field lines around a charge configuration.

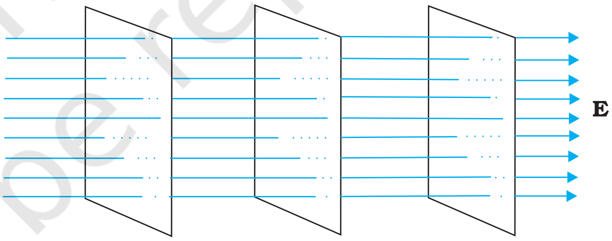


FIGURE 2.10 Equipotential surfaces for a uniform electric field.

For a uniform electric field **E**, say, along the x-axis, the equipotential surfaces are planes normal to the x-axis, i.e., planes parallel to the y-z plane (Fig. 2.10). Equipotential surfaces for (a) a dipole and (b) two identical positive charges are shown in Fig. 2.11.

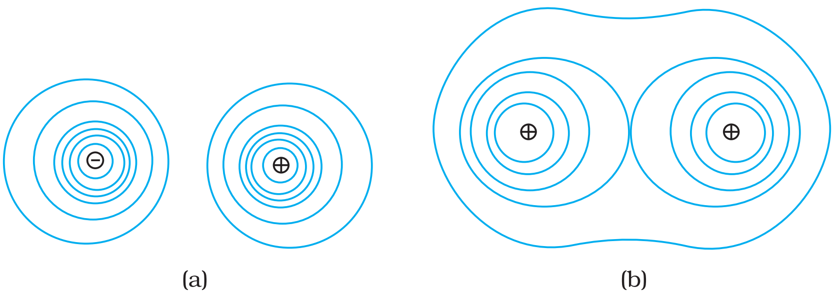


FIGURE 2.11 Some equipotential surfaces for (a) a dipole, (b) two identical positive charges.

Reprint 2025-26

2.6.1 Relation between field and potential

Consider two closely spaced equipotential surfaces A and B (Fig. 2.12) with potential values V and $V + \delta V$, where δV is the change in V in the direction of the electric field **E**. Let P be a point on the surface B. δl is the perpendicular distance of the surface A from P. Imagine that a unit positive charge is moved along this perpendicular from the surface B to surface A against the electric field. The work done in this process is $|\mathbf{E}|\,\delta l$.

This work equals the potential difference $V_A - V_B$.

Thus,

$$|\mathbf{E}|\,\delta l = V - (V + \delta V) = -\delta V$$

i.e.,
$$|\mathbf{E}| = -\frac{\delta V}{\delta l} \qquad (2.20)$$

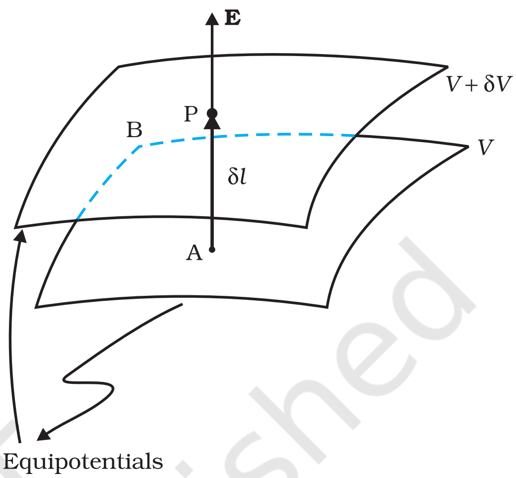


FIGURE 2.12 From the potential to the field.

Since δV is negative, $\delta V = -|\delta V|$. we can rewrite Eq (2.20) as

$$|\mathbf{E}| = -\frac{\delta V}{\delta l} = +\frac{|\delta V|}{\delta l} \qquad (2.21)$$

We thus arrive at two important conclusions concerning the relation between electric field and potential:

(i) *Electric field is in the direction in which the potential decreases steepest.*

(ii) *Its magnitude is given by the change in the magnitude of potential per unit displacement normal to the equipotential surface at the point.*

2.7 Potential Energy of a System of Charges

Consider first the simple case of two charges q_1 and q_2 with position vector $\mathbf{r}_1$ and $\mathbf{r}_2$ relative to some origin. Let us calculate the work done (externally) in building up this configuration. This means that we consider the charges q_1 and q_2 initially at infinity and determine the work done by an external agency to bring the charges to the given locations. Suppose, first the charge q_1 is brought from infinity to the point $\mathbf{r}_1$. There is no external field against which work needs to be done, so work done in bringing q_1 from infinity to $\mathbf{r}_1$ is zero. This charge produces a potential in space given by

$$V_1 = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r_{1P}}$$

where r_{1P} is the distance of a point P in space from the location of q_1. From the definition of potential, work done in bringing charge q_2 from infinity to the point $\mathbf{r}_2$ is q_2 times the potential at $\mathbf{r}_2$ due to q_1:

$$\text{work done on } q_2 = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}}$$

Reprint 2025-26

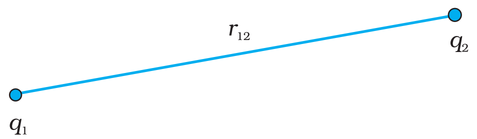


FIGURE 2.13 Potential energy of a system of charges q_1 and q_2 is directly proportional to the product of charges and inversely to the distance between them.

where r_{12} is the distance between points 1 and 2.

Since electrostatic force is conservative, this work gets stored in the form of potential energy of the system. Thus, the potential energy of a system of two charges q_1 and q_2 is

$$U = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \tag{2.22}$$

Obviously, if q_2 was brought first to its present location and q_1 brought later, the potential energy U would be the same. More generally, the potential energy expression, Eq. (2.22), is unaltered whatever way the charges are brought to the specified locations, because of path-independence of work for electrostatic force.

Equation (2.22) is true for any sign of q_1 and q_2. If $q_1q_2 > 0$, potential energy is positive. This is as expected, since for like charges ($q_1q_2 > 0$), electrostatic force is repulsive and a positive amount of work is needed to be done against this force to bring the charges from infinity to a finite distance apart. For unlike charges ($q_1 q_2 < 0$), the electrostatic force is attractive. In that case, a positive amount of work is needed against this force to take the charges from the given location to infinity. In other words, a negative amount of work is needed for the reverse path (from infinity to the present locations), so the potential energy is negative.

Equation (2.22) is easily generalised for a system of any number of point charges. Let us calculate the potential energy of a system of three charges q_1, q_2 and q_3 located at $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, respectively. To bring q_1 first from infinity to $\mathbf{r}_1$, no work is required. Next we bring q_2 from infinity to $\mathbf{r}_2$. As before, work done in this step is

$$q_2 V_1(\mathbf{r}_2) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}} \tag{2.23}$$

The charges q_1 and q_2 produce a potential, which at any point P is given by

$$V_{1,2} = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}}\right) \tag{2.24}$$

Work done next in bringing q_3 from infinity to the point $\mathbf{r}_3$ is q_3 times $V_{1,\,2}$ at $\mathbf{r}_3$

$$q_3 V_{1,2}(\mathbf{r}_3) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \tag{2.25}$$

The total work done in assembling the charges at the given locations is obtained by adding the work done in different steps [Eq. (2.23) and Eq. (2.25)],

$$U = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_2 q_3}{r_{23}}\right) \tag{2.26}$$

Again, because of the conservative nature of the electrostatic force (or equivalently, the path independence of work done), the final expression for U, Eq. (2.26), is independent of the manner in which the configuration is assembled. *The potential energy*

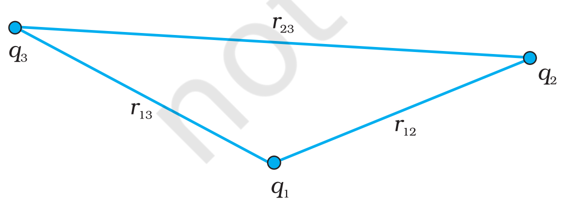


FIGURE 2.14 Potential energy of a system of three charges is given by Eq. (2.26), with the notation given in the figure.

is characteristic of the present state of configuration, and not the way the state is achieved.

Example 2.4 Four charges are arranged at the corners of a square ABCD of side d, as shown in Fig. 2.15.(a) Find the work required to put together this arrangement. (b) A charge q_0 is brought to the centre E of the square, the four charges being held fixed at its corners. How much extra work is needed to do this?

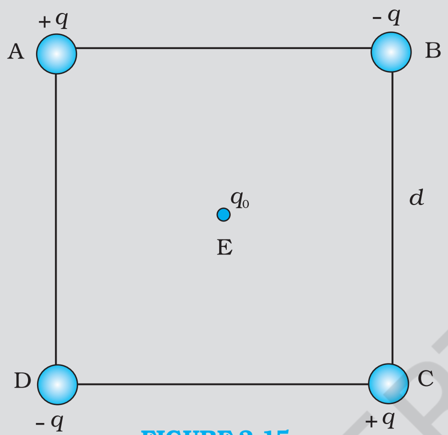


FIGURE 2.15

Solution

(a) Since the work done depends on the final arrangement of the charges, and not on how they are put together, we calculate work needed for one way of putting the charges at A, B, C and D. Suppose, first the charge $+q$ is brought to A, and then the charges $-q$, $+q$, and $-q$ are brought to B, C and D, respectively. The total work needed can be calculated in steps:

(i) Work needed to bring charge $+q$ to A when no charge is present elsewhere: this is zero.

(ii) Work needed to bring $-q$ to B when $+q$ is at A. This is given by (charge at B) × (electrostatic potential at B due to charge $+q$ at A)

$$= -q \times \left(\frac{q}{4\pi\varepsilon_0 d}\right) = -\frac{q^2}{4\pi\varepsilon_0 d}$$

(iii) Work needed to bring charge $+q$ to C when $+q$ is at A and $-q$ is at B. This is given by (charge at C) × (potential at C due to charges at A and B)

$$= +q\left(\frac{+q}{4\pi\varepsilon_0 d\sqrt{2}} + \frac{-q}{4\pi\varepsilon_0 d}\right)$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(1 - \frac{1}{\sqrt{2}}\right)$$

(iv) Work needed to bring $-q$ to D when $+q$ at A,$-q$ at B, and $+q$ at C. This is given by (charge at D) × (potential at D due to charges at A, B and C)

$$= -q\left(\frac{+q}{4\pi\varepsilon_0 d} + \frac{-q}{4\pi\varepsilon_0 d\sqrt{2}} + \frac{q}{4\pi\varepsilon_0 d}\right)$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(2 - \frac{1}{\sqrt{2}}\right)$$

EXAMPLE 2.4

Add the work done in steps (i), (ii), (iii) and (iv). The total work required is

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left\{(0)+(1)+\left(1-\frac{1}{\sqrt{2}}\right)+\left(2-\frac{1}{\sqrt{2}}\right)\right\}$$

$$= \frac{-q^2}{4\pi\varepsilon_0 d}\left(4-\sqrt{2}\right)$$

The work done depends only on the arrangement of the charges, and not how they are assembled. By definition, this is the total electrostatic energy of the charges.

(Students may try calculating same work/energy by taking charges in any other order they desire and convince themselves that the energy will remain the same.)

(b) The extra work necessary to bring a charge q_0 to the point E when the four charges are at A, B, C and D is q_0 × (electrostatic potential at E due to the charges at A, B, C and D). The electrostatic potential at E is clearly zero since potential due to A and C is cancelled by that due to B and D. Hence, no work is required to bring any charge to point E.

EXAMPLE 2.4

2.8 Potential Energy in an External Field

2.8.1 Potential energy of a single charge

In Section 2.7, the source of the electric field was specified – the charges and their locations - and the potential energy of the system of those charges was determined. In this section, we ask a related but a distinct question. What is the potential energy of a charge q in a given field? This question was, in fact, the starting point that led us to the notion of the electrostatic potential (Sections 2.1 and 2.2). But here we address this question again to clarify in what way it is different from the discussion in Section 2.7.

The main difference is that we are now concerned with the potential energy of a charge (or charges) in an *external* field. The external field **E** is *not* produced by the given charge(s) whose potential energy we wish to calculate. **E** is produced by sources external to the given charge(s).The external sources may be known, but often they are unknown or unspecified; what is specified is the electric field **E** or the electrostatic potential V due to the external sources. We assume that the charge q does not significantly affect the sources producing the external field. This is true if q is very small, or the external sources are held fixed by other unspecified forces. Even if q is finite, its influence on the external sources may still be ignored in the situation when very strong sources far away at infinity produce a finite field **E** in the region of interest. Note again that we are interested in determining the potential energy of a given charge q (and later, a system of charges) in the external field; we are not interested in the potential energy of the sources producing the external electric field.

The external electric field **E** and the corresponding external potential V may vary from point to point. By definition, V at a point P is the work done in bringing a unit positive charge from infinity to the point P.

Reprint 2025-26

(We continue to take potential at infinity to be zero.) Thus, work done in bringing a charge q from infinity to the point P in the external field is qV. This work is stored in the form of potential energy of q. If the point P has position vector **r** relative to some origin, we can write:

Potential energy of q at **r** in an external field

$$= qV(\mathbf{r}) \tag{2.27}$$

where $V(\mathbf{r})$ is the external potential at the point **r**.

Thus, if an electron with charge $q = e = 1.6\times10^{-19}$ C is accelerated by a potential difference of ΔV = 1 volt, it would gain energy of $q\Delta V = 1.6 \times 10^{-19}$J. This unit of energy is defined as 1 *electron volt* or 1eV, i.e., 1 eV=1.6×10^{-19}J. The units based on eV are most commonly used in atomic, nuclear and particle physics, (1 keV = 10^3eV = 1.6×10^{-16}J, 1 MeV = 10^6eV = 1.6×10^{-13}J, 1 GeV = 10^9eV = 1.6×10^{-10}J and 1 TeV = 10^{12}eV = 1.6×10^{-7}J). [This has already been defined on Page 117, XI Physics Part I, Table 6.1.]

2.8.2 Potential energy of a system of two charges in an external field

Next, we ask: what is the potential energy of a system of two charges q_1 and q_2 located at $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively, in an external field? First, we calculate the work done in bringing the charge q_1 from infinity to $\mathbf{r}_1$. Work done in this step is $q_1\, V(\mathbf{r}_1)$, using Eq. (2.27). Next, we consider the work done in bringing q_2 to $\mathbf{r}_2$. In this step, work is done not only against the external field **E** but also against the field due to q_1.

Work done on q_2 against the external field

$= q_2\, V(\mathbf{r}_2)$

Work done on q_2 against the field due to q_1

$$= \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}}$$

where r_{12} is the distance between q_1 and q_2. We have made use of Eqs. (2.27) and (2.22). By the superposition principle for fields, we add up the work done on q_2 against the two fields (**E** and that due to q_1):

Work done in bringing q_2 to $\mathbf{r}_2$

$$= q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}} \tag{2.28}$$

Thus,

Potential energy of the system

= the total work done in assembling the configuration

$$= q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}} \tag{2.29}$$

Example 2.5

(a) Determine the electrostatic potential energy of a system consisting of two charges 7 μC and –2 μC (and with no external field) placed at (–9 cm, 0, 0) and (9 cm, 0, 0) respectively.

(b) How much work is required to separate the two charges infinitely away from each other?

EXAMPLE 2.5

EXAMPLE 2.5

(c) Suppose that the same system of charges is now placed in an external electric field $E = A\,(1/r^2)$; $A = 9 \times 10^5$ NC^{-1} m^2. What would the electrostatic energy of the configuration be?

Solution

(a) $U = \dfrac{1}{4\pi\varepsilon_0}\dfrac{q_1 q_2}{r} = 9 \times 10^9 \times \dfrac{7 \times (-2) \times 10^{-12}}{0.18} = -0.7$ J.

(b) $W = U_2 - U_1 = 0 - U = 0 - (-0.7) = 0.7$ J.

(c) The mutual interaction energy of the two charges remains unchanged. In addition, there is the energy of interaction of the two charges with the external electric field. We find,

$$q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) = A\frac{7\mu\text{C}}{0.09\text{m}} + A\frac{-2\mu\text{C}}{0.09\text{m}}$$

and the net electrostatic energy is

$$q_1 V(\mathbf{r}_1) + q_2 V(\mathbf{r}_2) + \frac{q_1 q_2}{4\pi\varepsilon_0 r_{12}} = A\frac{7\,\mu\text{C}}{0.09\,\text{m}} + A\frac{-2\,\mu\text{C}}{0.09\,\text{m}} - 0.7\,\text{J}$$

$$= 70 - 20 - 0.7 = 49.3\,\text{J}$$

2.8.3 Potential energy of a dipole in an external field

Consider a dipole with charges $q_1 = +q$ and $q_2 = -q$ placed in a uniform electric field **E**, as shown in Fig. 2.16.

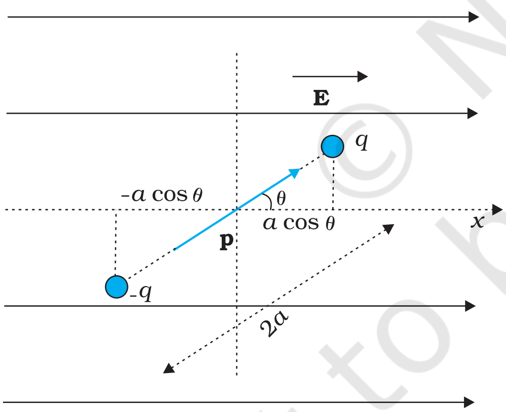


FIGURE 2.16 Potential energy of a dipole in a uniform external field.

As seen in the last chapter, in a uniform electric field, the dipole experiences no net force; but experiences a torque $\boldsymbol{\tau}$ given by

$$\boldsymbol{\tau} = \mathbf{p} \times \mathbf{E} \qquad (2.30)$$

which will tend to rotate it (unless **p** is parallel or antiparallel to **E**). Suppose an external torque $\boldsymbol{\tau}_{ext}$ is applied in such a manner that it just neutralises this torque and rotates it in the plane of paper from angle θ_0 to angle θ_1 at an infinitesimal angular speed and *without angular acceleration.* The amount of work done by the external torque will be given by

$$W = \int_{\theta_0}^{\theta_1} t_{ext}(\theta)d\theta = \int_{\theta_0}^{\theta_1} pE\sin\theta\, d\theta$$

$$= pE(\cos\theta_0 - \cos\theta_1) \qquad (2.31)$$

This work is stored as the potential energy of the system. We can then associate potential energy $U(\theta)$ with an inclination θ of the dipole. Similar to other potential energies, there is a freedom in choosing the angle where the potential energy U is taken to be zero. A natural choice is to take $\theta_0 = \pi / 2$. (An explanation for it is provided towards the end of discussion.) We can then write,

$$U(\theta) = pE\left(\cos\frac{\pi}{2} - \cos\theta\right) = pE\cos\theta = -\mathbf{p.E} \qquad (2.32)$$

This expression can alternately be understood also from Eq. (2.29). We apply Eq. (2.29) to the present system of two charges $+q$ and $-q$. The potential energy expression then reads

$$U'(\theta) = q[V(\mathbf{r}_1) - V(\mathbf{r}_2)] - \frac{q^2}{4\pi\varepsilon_0 \times 2a} \tag{2.33}$$

Here, $\mathbf{r_1}$ and $\mathbf{r_2}$ denote the position vectors of $+q$ and $-q$. Now, the potential difference between positions $\mathbf{r_1}$ and $\mathbf{r_2}$ equals the work done in bringing a unit positive charge against field from $\mathbf{r_2}$ to $\mathbf{r_1}$. The displacement parallel to the force is $2a\cos\theta$. Thus, $[V(\mathbf{r}_1) - V(\mathbf{r}_2)] = -E \times 2a\cos\theta$. We thus obtain,

$$U'(\theta) = -pE\cos\theta - \frac{q^2}{4\pi\varepsilon_0 \times 2a} = -\mathbf{p.E} - \frac{q^2}{4\pi\varepsilon_0 \times 2a} \tag{2.34}$$

We note that $U'(\theta)$ differs from $U(\theta)$ by a quantity which is just a constant for a given dipole. Since a constant is insignificant for potential energy, we can drop the second term in Eq. (2.34) and it then reduces to Eq. (2.32).

We can now understand why we took $\theta_0 = \pi/2$. In this case, the work done against the *external* field **E** in bringing $+q$ and $-q$ are equal and opposite and cancel out, i.e., $q[V(\mathbf{r}_1) - V(\mathbf{r}_2)] = 0$.

Example 2.6 A molecule of a substance has a permanent electric dipole moment of magnitude 10^{-29} C m. A mole of this substance is polarised (at low temperature) by applying a strong electrostatic field of magnitude 10^6 V m^{-1}. The direction of the field is suddenly changed by an angle of 60°. Estimate the heat released by the substance in aligning its dipoles along the new direction of the field. For simplicity, assume 100% polarisation of the sample.

Solution Here, dipole moment of each molecules = 10^{-29} C m
As 1 mole of the substance contains 6×10^{23} molecules,
total dipole moment of all the molecules, $p = 6 \times 10^{23} \times 10^{-29}$ C m
$= 6 \times 10^{-6}$ C m
Initial potential energy, $U_i = -pE\cos\theta = -6\times10^{-6}\times10^{6}\cos 0° = -6$ J
Final potential energy (when $\theta = 60°$), $U_f = -6 \times 10^{-6} \times 10^{6} \cos 60° = -3$ J
Change in potential energy = –3 J – (–6J) = 3 J
So, there is loss in potential energy. This must be the energy released by the substance in the form of heat in aligning its dipoles.

EXAMPLE 2.6

2.9 Electrostatics of Conductors

Conductors and insulators were described briefly in Chapter 1. Conductors contain mobile charge carriers. In metallic conductors, these charge carriers are electrons. In a metal, the outer (valence) electrons part away from their atoms and are free to move. These electrons are free within the metal but not free to leave the metal. The free electrons form a kind of 'gas'; they collide with each other and with the ions, and move randomly in different directions. In an external electric field, they drift against the direction of the field. The positive ions made up of the nuclei and the bound electrons remain held in their fixed positions. In electrolytic conductors, the charge carriers are both positive and negative ions; but the situation in this case is more involved – the movement of the charge carriers is affected both by the external electric field as also by the so-called chemical forces (see Chapter 3). We shall restrict our discussion to metallic solid conductors. Let us note important results regarding electrostatics of conductors.

1. Inside a conductor, electrostatic field is zero

Consider a conductor, neutral or charged. There may also be an external electrostatic field. In the static situation, when there is no current inside or on the surface of the conductor, the electric field is zero everywhere inside the conductor. This fact can be taken as the defining property of a conductor. A conductor has free electrons. As long as electric field is not zero, the free charge carriers would experience force and drift. In the static situation, the free charges have so distributed themselves that the electric field is zero everywhere inside. *Electrostatic field is zero inside a conductor.*

2. At the surface of a charged conductor, electrostatic field must be normal to the surface at every point

If **E** were not normal to the surface, it would have some non-zero component along the surface. Free charges on the surface of the conductor would then experience force and move. In the static situation, therefore, **E** should have no tangential component. Thus *electrostatic field at the surface of a charged conductor must be normal to the surface at every point.* (For a conductor without any surface charge density, field is zero even at the surface.) See result 5.

3. The interior of a conductor can have no excess charge in the static situation

A neutral conductor has equal amounts of positive and negative charges in every small volume or surface element. When the conductor is charged, the excess charge can reside only on the surface in the static situation. This follows from the Gauss's law. Consider any arbitrary volume element v inside a conductor. On the closed surface S bounding the volume element v, electrostatic field is zero. Thus the total electric flux through S is zero. Hence, by Gauss's law, there is no net charge enclosed by S. But the surface S can be made as small as you like, i.e., the volume v can be made vanishingly small. This means *there is no net charge at any point inside the conductor*, and *any excess charge must reside at the surface.*

4. Electrostatic potential is constant throughout the volume of the conductor and has the same value (as inside) on its surface

This follows from results 1 and 2 above. Since **E** = 0 inside the conductor and has no tangential component on the surface, no work is done in moving a small test charge within the conductor and on its surface. That is, there is no potential difference between any two points inside or on the surface of the conductor. Hence, the result. If the conductor is charged, electric field normal to the surface exists; this means potential will be different for the surface and a point just outside the surface.

In a system of conductors of arbitrary size, shape and charge configuration, each conductor is characterised by a constant value of potential, but this constant may differ from one conductor to the other.

5. Electric field at the surface of a charged conductor

$$\mathbf{E} = \frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}} \tag{2.35}$$

where σ is the surface charge density and $\hat{\mathbf{n}}$ is a unit vector normal to the surface in the outward direction.

To derive the result, choose a pill box (a short cylinder) as the Gaussian surface about any point P on the surface, as shown in Fig. 2.17. The pill box is partly inside and partly outside the surface of the conductor. It has a small area of cross section δS and negligible height.

Just inside the surface, the electrostatic field is zero; just outside, the field is normal to the surface with magnitude E. Thus, the contribution to the total flux through the pill box comes only from the outside (circular) cross-section of the pill box. This equals $\pm E\delta S$ (positive for $\sigma > 0$, negative for $\sigma < 0$), since over the small area δS, **E** may be considered constant and **E** and δS are parallel or antiparallel. The charge enclosed by the pill box is $\sigma\delta S$.

By Gauss's law

$$E\delta S = \frac{|\sigma|\,\delta S}{\varepsilon_0}$$

$$E = \frac{|\sigma|}{\varepsilon_0} \tag{2.36}$$

Including the fact that electric field is normal to the surface, we get the vector relation, Eq. (2.35), which is true for both signs of σ. For $\sigma > 0$, electric field is normal to the surface outward; for $\sigma < 0$, electric field is normal to the surface inward.

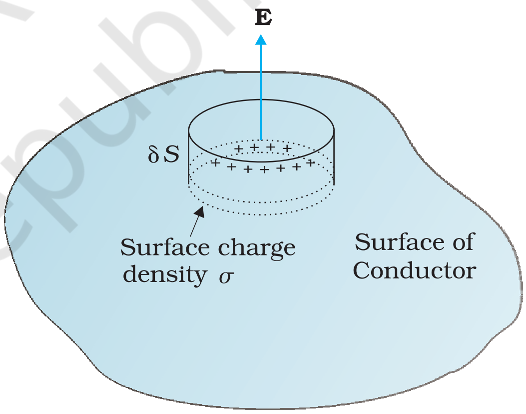


FIGURE 2.17 The Gaussian surface (a pill box) chosen to derive Eq. (2.35) for electric field at the surface of a charged conductor.

6. Electrostatic shielding

Consider a conductor with a cavity, with no charges inside the cavity. A remarkable result is that the electric field inside the cavity is zero, whatever be the size and shape of the cavity and whatever be the charge on the conductor and the external fields in which it might be placed. We have proved a simple case of this result already: the electric field inside a charged spherical shell is zero. The proof of the result for the shell makes use of the spherical symmetry of the shell (see Chapter 1). But the vanishing of electric field in the (charge-free) cavity of a conductor is, as mentioned above, a very general result. A related result is that even if the conductor

Reprint 2025-26

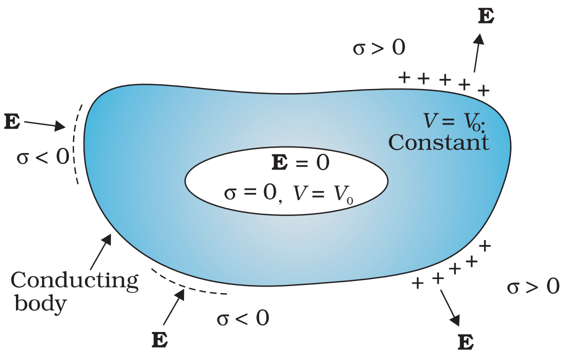


FIGURE 2.18 The electric field inside a cavity of any conductor is zero. All charges reside only on the outer surface of a conductor with cavity. (There are no charges placed in the cavity.)

is charged or charges are induced on a neutral conductor by an external field, all charges reside only on the outer surface of a conductor with cavity.

The proofs of the results noted in Fig. 2.18 are omitted here, but we note their important implication. Whatever be the charge and field configuration outside, any cavity in a conductor remains shielded from outside electric influence: *the field inside the cavity is always zero*. This is known as *electrostatic shielding*. The effect can be made use of in protecting sensitive instruments from outside electrical influence. Figure 2.19 gives a summary of the important electrostatic properties of a conductor.

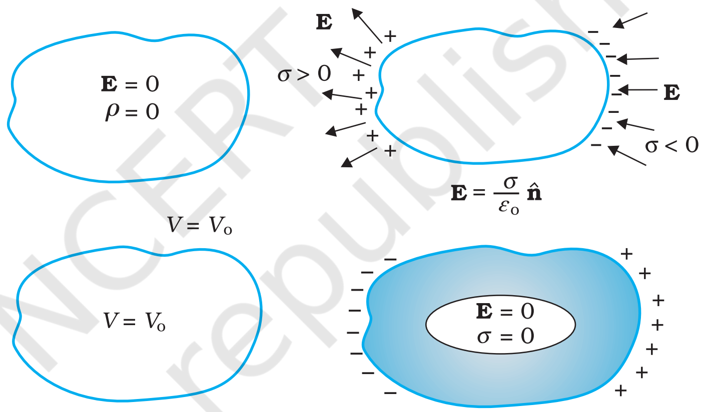


FIGURE 2.19 Some important electrostatic properties of a conductor.

Example 2.7

(a) A comb run through one's dry hair attracts small bits of paper. Why?
What happens if the hair is wet or if it is a rainy day? (Remember, a paper does not conduct electricity.)

(b) Ordinary rubber is an insulator. But special rubber tyres of aircraft are made slightly conducting. Why is this necessary?

(c) Vehicles carrying inflammable materials usually have metallic ropes touching the ground during motion. Why?

(d) A bird perches on a bare high power line, and nothing happens to the bird. A man standing on the ground touches the same line and gets a fatal shock. Why?

Solution

(a) This is because the comb gets charged by friction. The molecules in the paper gets polarised by the charged comb, resulting in a net force of attraction. If the hair is wet, or if it is rainy day, friction between hair and the comb reduces. The comb does not get charged and thus it will not attract small bits of paper.

EXAMPLE 2.7

(b) To enable them to conduct charge (produced by friction) to the ground; as too much of static electricity accumulated may result in spark and result in fire.

(c) Reason similar to (b).

(d) Current passes only when there is difference in potential.

EXAMPLE 2.7

2.10 Dielectrics and Polarisation

Dielectrics are non-conducting substances. In contrast to conductors, they have no (or negligible number of) charge carriers. Recall from Section 2.9 what happens when a conductor is placed in an external electric field. The free charge carriers move and charge distribution in the conductor adjusts itself in such a way that the electric field due to induced charges opposes the external field within the conductor. This happens until, in the static situation, the two fields cancel each other and the net electrostatic field in the conductor is zero. In a dielectric, this free movement of charges is not possible. It turns out that the external field induces dipole moment by stretching or re-orienting molecules of the dielectric. The collective effect of all the molecular dipole moments is net charges on the surface of the dielectric which produce a field that opposes the external field. Unlike in a conductor, however, the opposing field so induced does not exactly cancel the external field. It only reduces it. The extent of the effect depends on the nature of the dielectric. To understand the effect, we need to look at the charge distribution of a dielectric at the molecular level.

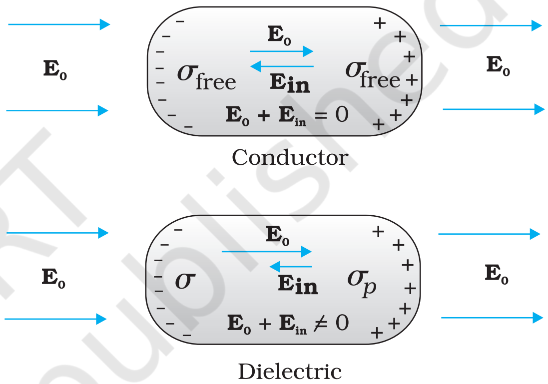


FIGURE 2.20 Difference in behaviour of a conductor and a dielectric in an external electric field.

The molecules of a substance may be polar or non-polar. In a non-polar molecule, the centres of positive and negative charges coincide. The molecule then has no permanent (or intrinsic) dipole moment. Examples of non-polar molecules are oxygen (O_2) and hydrogen (H_2) molecules which, because of their symmetry, have no dipole moment. On the other hand, a polar molecule is one in which the centres of positive and negative charges are separated (even when there is no external field). Such molecules have a permanent dipole moment. An ionic molecule such as HCl or a molecule of water (H_2O) are examples of polar molecules.

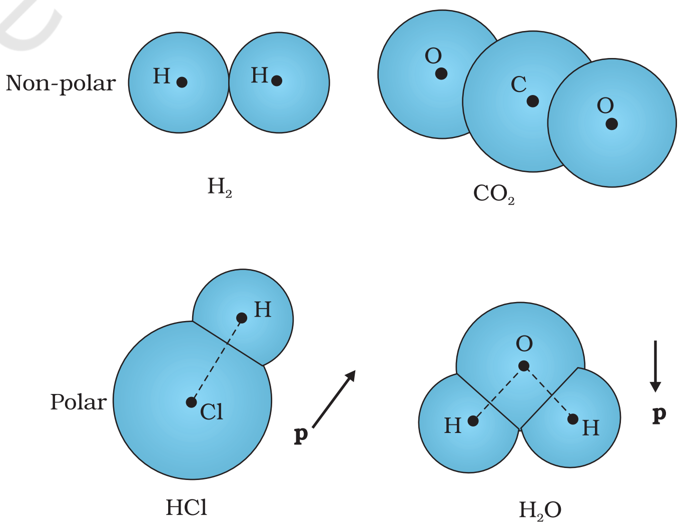


FIGURE 2.21 Some examples of polar and non-polar molecules.

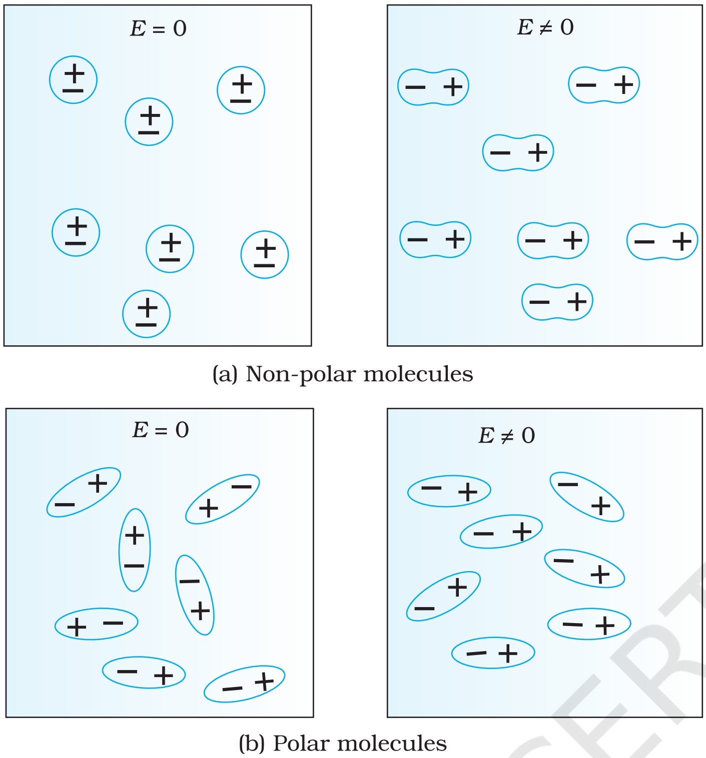


FIGURE 2.22 A dielectric develops a net dipole moment in an external electric field. (a) Non-polar molecules, (b) Polar molecules.

In an external electric field, the positive and negative charges of a non-polar molecule are displaced in opposite directions. The displacement stops when the external force on the constituent charges of the molecule is balanced by the restoring force (due to internal fields in the molecule). The non-polar molecule thus develops an induced dipole moment. The dielectric is said to be polarised by the external field. We consider only the simple situation when the induced dipole moment is in the direction of the field and is proportional to the field strength. (Substances for which this assumption is true are called *linear isotropic dielectrics.*) The induced dipole moments of different molecules add up giving a net dipole moment of the dielectric in the presence of the external field.

A dielectric with polar molecules also develops a net dipole moment in an external field, but for a different reason. In the absence of any external field, the different permanent dipoles are oriented randomly due to thermal agitation; so the total dipole moment is zero. When an external field is applied, the individual dipole moments tend to align with the field. When summed overall the molecules, there is then a net dipole moment in the direction of the external field, i.e., the dielectric is polarised. The extent of polarisation depends on the relative strength of two mutually opposite factors: the dipole potential energy in the external field tending to align the dipoles with the field and thermal energy tending to disrupt the alignment. There may be, in addition, the 'induced dipole moment' effect as for non-polar molecules, but generally the alignment effect is more important for polar molecules.

Thus in either case, whether polar or non-polar, a dielectric develops a net dipole moment in the presence of an external field. The dipole moment per unit volume is called *polarisation* and is denoted by **P**. For linear isotropic dielectrics,

$$\mathbf{P} = \varepsilon_0 \chi_e \mathbf{E} \tag{2.37}$$

where χ_e is a constant characteristic of the dielectric and is known as the *electric susceptibility* of the dielectric medium.

It is possible to relate χ_e to the molecular properties of the substance, but we shall not pursue that here.

The question is: how does the polarised dielectric modify the original external field inside it? Let us consider, for simplicity, a rectangular dielectric slab placed in a uniform external field $\mathbf{E}_0$ parallel to two of its faces. The field causes a uniform polarisation **P** of the dielectric. Thus

Reprint 2025-26

every volume element Δv of the slab has a dipole moment $\mathbf{P}\,\Delta v$ in the direction of the field. The volume element Δv is macroscopically small but contains a very large number of molecular dipoles. Anywhere inside the dielectric, the volume element Δv has no net charge (though it has net dipole moment). This is, because, the positive charge of one dipole sits close to the negative charge of the adjacent dipole. However, at the surfaces of the dielectric normal to the electric field, there is evidently a net charge density. As seen in Fig 2.23, the positive ends of the dipoles remain unneutralised at the right surface and the negative ends at the left surface. The unbalanced charges are the induced charges due to the external field.

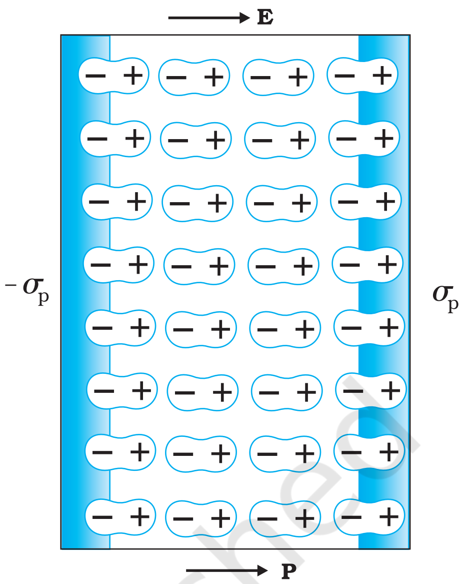


FIGURE 2.23 A uniformly polarised dielectric amounts to induced surface charge density, but no volume charge density.

Thus, the polarised dielectric is equivalent to two charged surfaces with induced surface charge densities, say σ_p and $-\sigma_p$. Clearly, the field produced by these surface charges opposes the external field. The total field in the dielectric is, thereby, reduced from the case when no dielectric is present. We should note that the surface charge density $\pm\sigma_p$ arises from bound (not free charges) in the dielectric.

2.11 Capacitors and Capacitance

A capacitor is a system of two conductors separated by an insulator (Fig. 2.24). The conductors have charges, say Q_1 and Q_2, and potentials V_1 and V_2. Usually, in practice, the two conductors have charges Q and $-Q$, with potential difference $V = V_1 - V_2$ between them. We shall consider only this kind of charge configuration of the capacitor. (Even a single conductor can be used as a capacitor by assuming the other at infinity.) The conductors may be so charged by connecting them to the two terminals of a battery. Q is called the charge of the capacitor, though this, in fact, is the charge on one of the conductors – the total charge of the capacitor is zero.

The electric field in the region between the conductors is proportional to the charge Q. That is, if the charge on the capacitor is, say doubled, the electric field will also be doubled at every point. (This follows from the direct proportionality between field and charge implied by Coulomb's law and the superposition principle.) Now, potential difference V is the work done per unit positive charge in taking a small test charge from the conductor 2 to 1 against the field. Consequently, V is also proportional to Q, and the ratio Q/V is a constant:

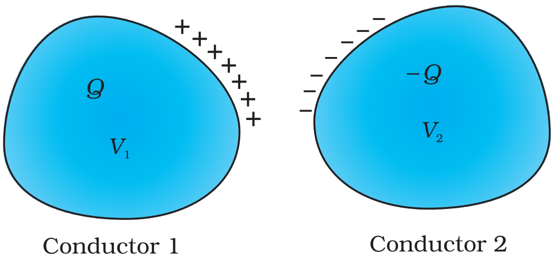


FIGURE 2.24 A system of two conductors separated by an insulator forms a capacitor.

$$C = \frac{Q}{V} \tag{2.38}$$

The constant C is called the *capacitance* of the capacitor. C is independent of Q or V, as stated above. The capacitance C depends only on the geometrical configuration (shape, size, separation) of the system of two conductors. [As we shall see later, it also depends on the nature of the insulator (dielectric) separating the two conductors.] The SI unit of capacitance is 1 farad (=1 coulomb volt^{-1}) or 1 F = 1 C V^{-1}. A capacitor with fixed capacitance is symbolically shown as ─┤├─, while the one with variable capacitance is shown as ⫲.

Equation (2.38) shows that for large C, V is small for a given Q. This means a capacitor with large capacitance can hold large amount of charge Q at a relatively small V. This is of practical importance. High potential difference implies strong electric field around the conductors. A strong electric field can ionise the surrounding air and accelerate the charges so produced to the oppositely charged plates, thereby neutralising the charge on the capacitor plates, at least partly. In other words, the charge of the capacitor leaks away due to the reduction in insulating power of the intervening medium.

The maximum electric field that a dielectric medium can withstand without break-down (of its insulating property) is called its *dielectric strength*; for air it is about 3×10^6 Vm^{-1}. For a separation between conductors of the order of 1 cm or so, this field corresponds to a potential difference of 3×10^4 V between the conductors. Thus, for a capacitor to store a large amount of charge without leaking, its capacitance should be high enough so that the potential difference and hence the electric field do not exceed the break-down limits. Put differently, there is a limit to the amount of charge that can be stored on a given capacitor without significant leaking. In practice, a farad is a very big unit; the most common units are its sub-multiples 1 μF = 10^{-6} F, 1 nF = 10^{-9} F, 1 pF = 10^{-12} F, etc. Besides its use in storing charge, a capacitor is a key element of most ac circuits with important functions, as described in Chapter 7.

2.12 The Parallel Plate Capacitor

A parallel plate capacitor consists of two large plane parallel conducting plates separated by a small distance (Fig. 2.25). We first take the intervening medium between the plates to be vacuum. The effect of a dielectric medium between the plates is discussed in the next section. Let A be the area of each plate and d the separation between them. The two plates have charges Q and $-Q$. Since d is much smaller than the linear dimension of the plates ($d^2 << A$), we can use the result on electric field by an infinite plane sheet of uniform surface charge density (Section 1.15). Plate 1 has surface charge density $\sigma = Q/A$ and plate 2 has a surface charge density $-\sigma$. Using Eq. (1.33), the electric field in different regions is:

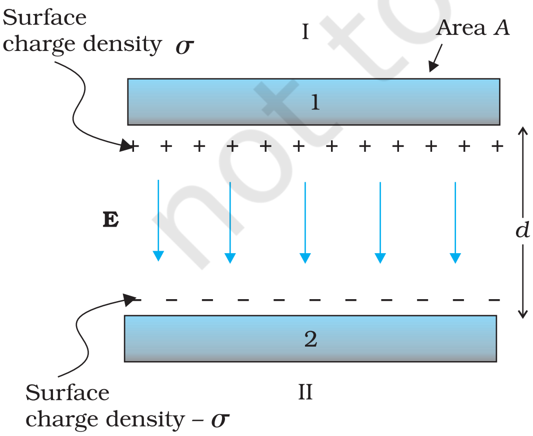


FIGURE 2.25 The parallel plate capacitor.

Outer region I (region above the plate 1),

$$E = \frac{\sigma}{2\varepsilon_0} - \frac{\sigma}{2\varepsilon_0} = 0 \quad (2.39)$$

Outer region II (region below the plate 2),

$$E = \frac{\sigma}{2\varepsilon_0} - \frac{\sigma}{2\varepsilon_0} = 0 \tag{2.40}$$

In the inner region between the plates 1 and 2, the electric fields due to the two charged plates add up, giving

$$E = \frac{\sigma}{2\varepsilon_0} + \frac{\sigma}{2\varepsilon_0} = \frac{\sigma}{\varepsilon_0} = \frac{Q}{\varepsilon_0 A} \tag{2.41}$$

The direction of electric field is from the positive to the negative plate.

Thus, the electric field is localised between the two plates and is uniform throughout. For plates with finite area, this will not be true near the outer boundaries of the plates. The field lines bend outward at the edges — an effect called 'fringing of the field'. By the same token, σ will not be strictly uniform on the entire plate. [E and σ are related by Eq. (2.35).] However, for $d^2 << A$, these effects can be ignored in the regions sufficiently far from the edges, and the field there is given by Eq. (2.41). Now for uniform electric field, potential difference is simply the electric field times the distance between the plates, that is,

$$V = E\,d = \frac{1}{\varepsilon_0}\frac{Qd}{A} \tag{2.42}$$

The capacitance C of the parallel plate capacitor is then

$$C = \frac{Q}{V} = = \frac{\varepsilon_0 A}{d} \tag{2.43}$$

which, as expected, depends only on the geometry of the system. For typical values like $A = 1\ \text{m}^2$, $d = 1$ mm, we get

$$C = \frac{8.85 \times 10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2} \times 1\text{m}^2}{10^{-3}\text{m}} = 8.85 \times 10^{-9}\,\text{F} \tag{2.44}$$

(You can check that if 1F= 1C V^{-1} = 1C $(\text{NC}^{-1}\text{m})^{-1}$ = 1 $\text{C}^2\,\text{N}^{-1}\text{m}^{-1}$.) This shows that 1F is too big a unit in practice, as remarked earlier. Another way of seeing the 'bigness' of 1F is to calculate the area of the plates needed to have C = 1F for a separation of, say 1 cm:

$$A = \frac{Cd}{\varepsilon_0} = \frac{1\text{F} \times 10^{-2}\,\text{m}}{8.85 \times 10^{-12}\,\text{C}^2\text{N}^{-1}\text{m}^{-2}} = 10^9\,\text{m}^2 \tag{2.45}$$

which is a plate about 30 km in length and breadth!

2.13 Effect of Dielectric on Capacitance

With the understanding of the behaviour of dielectrics in an external field developed in Section 2.10, let us see how the capacitance of a parallel plate capacitor is modified when a dielectric is present. As before, we have two large plates, each of area A, separated by a distance d. The charge on the plates is $\pm Q$, corresponding to the charge density $\pm\sigma$ (with $\sigma = Q/A$). When there is vacuum between the plates,

$$E_0 = \frac{\sigma}{\varepsilon_0}$$

PHYSICS

Factors affecting capacitance, capacitors in action
Interactive Java tutorial
http://micro.magnet.fsu.edu/electromag/java/capacitance/

and the potential difference V_0 is

$V_0 = E_0 d$

The capacitance C_0 in this case is

$$C_0 = \frac{Q}{V_0} = \varepsilon_0 \frac{A}{d} \tag{2.46}$$

Consider next a dielectric inserted between the plates fully occupying the intervening region. The dielectric is polarised by the field and, as explained in Section 2.10, the effect is equivalent to two charged sheets (at the surfaces of the dielectric normal to the field) with surface charge densities σ_p and $-\sigma_p$. The electric field in the dielectric then corresponds to the case when the net surface charge density on the plates is $\pm(\sigma - \sigma_p)$. That is,

$$E = \frac{\sigma - \sigma_P}{\varepsilon_0} \tag{2.47}$$

so that the potential difference across the plates is

$$V = E\,d = \frac{\sigma - \sigma_P}{\varepsilon_0} d \tag{2.48}$$

For linear dielectrics, we expect σ_p to be proportional to E_0, i.e., to σ. Thus, $(\sigma - \sigma_p)$ is proportional to σ and we can write

$$\sigma - \sigma_P = \frac{\sigma}{K} \tag{2.49}$$

where K is a constant characteristic of the dielectric. Clearly, $K > 1$. We then have

$$V = \frac{\sigma d}{\varepsilon_0 K} = \frac{Q d}{A \varepsilon_0 K} \tag{2.50}$$

The capacitance C, with dielectric between the plates, is then

$$C = \frac{Q}{V} = \frac{\varepsilon_0 K A}{d} \tag{2.51}$$

The product $\varepsilon_0 K$ is called the *permittivity* of the medium and is denoted by ε

$$\varepsilon = \varepsilon_0 K \tag{2.52}$$

For vacuum $K = 1$ and $\varepsilon = \varepsilon_0$; ε_0 is called the *permittivity of the vacuum*. The dimensionless ratio

$$K = \frac{\varepsilon}{\varepsilon_0} \tag{2.53}$$

is called the *dielectric constant* of the substance. As remarked before, from Eq. (2.49), it is clear that K is greater than 1. From Eqs. (2.46) and (2. 51)

$$K = \frac{C}{C_0} \tag{2.54}$$

Thus, the dielectric constant of a substance is the factor (>1) by which the capacitance increases from its vacuum value, when the dielectric is inserted fully between the plates of a capacitor. Though we arrived at

Eq. (2.54) for the case of a parallel plate capacitor, it holds good for any type of capacitor and can, in fact, be viewed in general as a definition of the dielectric constant of a substance.

Example 2.8 A slab of material of dielectric constant K has the same area as the plates of a parallel-plate capacitor but has a thickness $(3/4)d$, where d is the separation of the plates. How is the capacitance changed when the slab is inserted between the plates?

Solution Let $E_0 = V_0/d$ be the electric field between the plates when there is no dielectric and the potential difference is V_0. If the dielectric is now inserted, the electric field in the dielectric will be $E = E_0/K$. The potential difference will then be

$$V = E_0\left(\frac{1}{4}d\right) + \frac{E_0}{K}\left(\frac{3}{4}d\right)$$

$$= E_0 d\left(\frac{1}{4} + \frac{3}{4K}\right) = V_0 \frac{K+3}{4K}$$

The potential difference decreases by the factor $(K + 3)/4K$ while the free charge Q_0 on the plates remains unchanged. The capacitance thus increases

$$C = \frac{Q_0}{V} = \frac{4K}{K+3}\frac{Q_0}{V_0} = \frac{4K}{K+3}C_0$$

EXAMPLE 2.8

2.14 Combination of Capacitors

We can combine several capacitors of capacitance $C_1, C_2,\ldots, C_n$ to obtain a system with some effective capacitance C. The effective capacitance depends on the way the individual capacitors are combined. Two simple possibilities are discussed below.

2.14.1 Capacitors in series

Figure 2.26 shows capacitors C_1 and C_2 combined in series.

The left plate of C_1 and the right plate of C_2 are connected to two terminals of a battery and have charges Q and $-Q$, respectively. It then follows that the right plate of C_1 has charge $-Q$ and the left plate of C_2 has charge Q. If this was not so, the net charge on each capacitor would not be zero. This would result in an electric field in the conductor connecting $C_1$and C_2. Charge would flow until the net charge on both C_1 and C_2 is zero and there is no electric field in the conductor connecting C_1 and C_2. Thus, in the series combination, charges on the two plates ($\pm Q$) are the same on each capacitor. The total

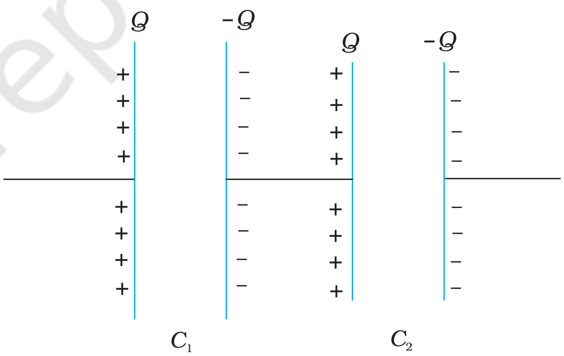


FIGURE 2.26 Combination of two capacitors in series.

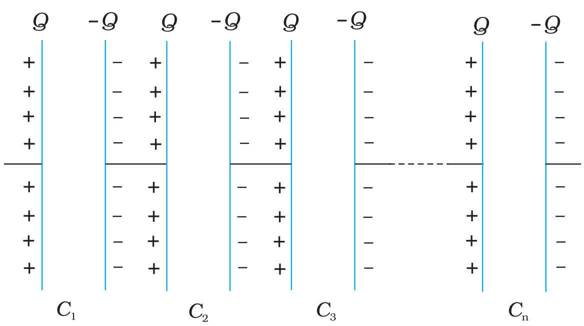


FIGURE 2.27 Combination of n capacitors in series.

potential drop V across the combination is the sum of the potential drops V_1 and V_2 across C_1 and C_2, respectively.

$$V = V_1 + V_2 = \frac{Q}{C_1} + \frac{Q}{C_2} \tag{2.55}$$

$$\text{i.e.,} \quad \frac{V}{Q} = \frac{1}{C_1} + \frac{1}{C_2}, \tag{2.56}$$

Now we can regard the combination as an effective capacitor with charge Q and potential difference V. The *effective capacitance* of the combination is

$$C = \frac{Q}{V} \tag{2.57}$$

We compare Eq. (2.57) with Eq. (2.56), and obtain

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} \tag{2.58}$$

The proof clearly goes through for any number of capacitors arranged in a similar way. Equation (2.55), for n capacitors arranged in series, generalises to

$$V = V_1 + V_2 + \ldots + V_n = \frac{Q}{C_1} + \frac{Q}{C_2} + \ldots + \frac{Q}{C_n} \tag{2.59}$$

Following the same steps as for the case of two capacitors, we get the general formula for effective capacitance of a series combination of n capacitors:

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots + \frac{1}{C_n} \tag{2.60}$$

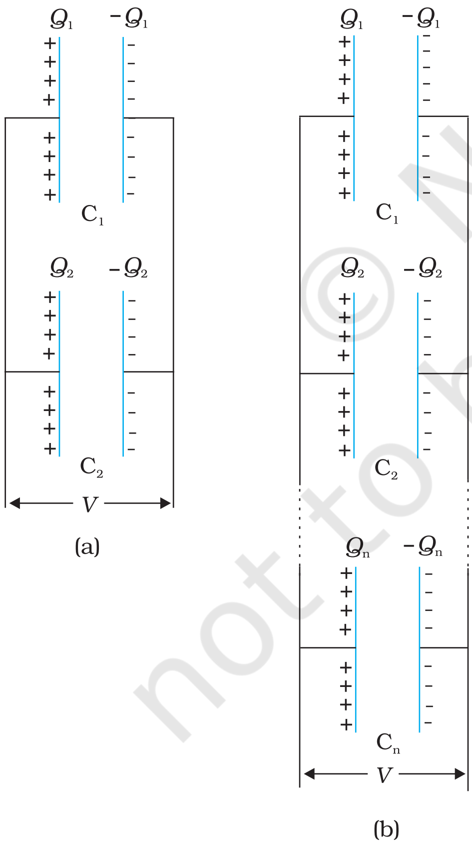


FIGURE 2.28 Parallel combination of (a) two capacitors, (b) n capacitors.

2.14.2 Capacitors in parallel

Figure 2.28 (a) shows two capacitors arranged in parallel. In this case, the same potential difference is applied across both the capacitors. But the plate charges ($\pm Q_1$) on capacitor 1 and the plate charges ($\pm Q_2$) on the capacitor 2 are not necessarily the same:

$$Q_1 = C_1 V, \; Q_2 = C_2 V \tag{2.61}$$

The equivalent capacitor is one with charge

$$Q = Q_1 + Q_2 \tag{2.62}$$

and potential difference V.

$$Q = CV = C_1 V + C_2 V \tag{2.63}$$

The effective capacitance C is, from Eq. (2.63),

$$C = C_1 + C_2 \tag{2.64}$$

The general formula for effective capacitance C for parallel combination of n capacitors [Fig. 2.28 (b)] follows similarly,

$$Q = Q_1 + Q_2 + \ldots + Q_n \tag{2.65}$$

$$\text{i.e.,} \quad CV = C_1 V + C_2 V + \ldots C_n V \tag{2.66}$$

which gives

$$C = C_1 + C_2 + \ldots C_n \tag{2.67}$$

Reprint 2025-26

Example 2.9 A network of four 10 μF capacitors is connected to a 500 V supply, as shown in Fig. 2.29. Determine (a) the equivalent capacitance of the network and (b) the charge on each capacitor. (Note, the *charge on a capacitor* is the charge on the plate with higher potential, equal and opposite to the charge on the plate with lower potential.)

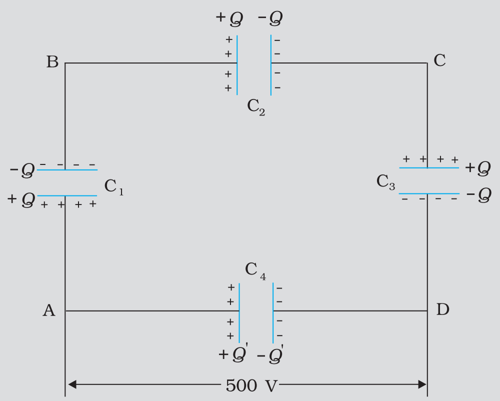


FIGURE 2.29

Solution

(a) In the given network, C_1, C_2 and C_3 are connected in series. The effective capacitance C′ of these three capacitors is given by

$$\frac{1}{C'} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}$$

For $C_1 = C_2 = C_3 = 10$ μF, $C' = (10/3)$ μF. The network has C' and C_4 connected in parallel. Thus, the equivalent capacitance C of the network is

$$C = C' + C_4 = \left(\frac{10}{3} + 10\right) \mu\text{F} = 13.3 \mu\text{F}$$

(b) Clearly, from the figure, the charge on each of the capacitors, C_1, C_2 and C_3 is the same, say Q. Let the charge on C_4 be Q'. Now, since the potential difference across AB is Q/C_1, across BC is Q/C_2, across CD is Q/C_3, we have

$$\frac{Q}{C_1} + \frac{Q}{C_2} + \frac{Q}{C_3} = 500\,\text{V}.$$

Also, $Q'/C_4 = 500$ V.

This gives for the given value of the capacitances,

$$Q = 500\,V \times \frac{10}{3} \mu\text{F} = 1.7 \times 10^{-3}\,\text{C} \text{ and}$$

$$Q' = 500\,V \times 10\,\mu\text{F} = 5.0 \times 10^{-3}\,\text{C}$$

EXAMPLE 2.9

2.15 Energy Stored in a Capacitor

A capacitor, as we have seen above, is a system of two conductors with charge Q and $-Q$. To determine the energy stored in this configuration, consider initially two uncharged conductors 1 and 2. Imagine next a process of transferring charge from conductor 2 to conductor 1 bit by

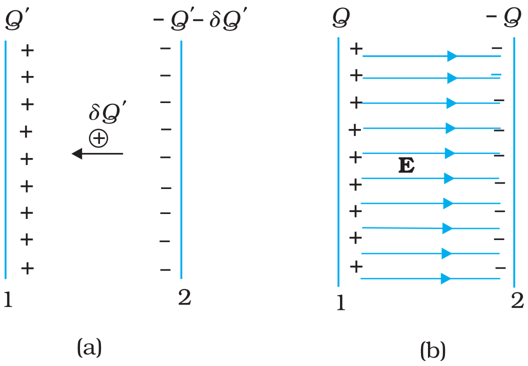


FIGURE 2.30 (a) Work done in a small step of building charge on conductor 1 from Q' to $Q' + \delta Q'$. (b) Total work done in charging the capacitor may be viewed as stored in the energy of electric field between the plates.

bit, so that at the end, conductor 1 gets charge Q. By charge conservation, conductor 2 has charge $-Q$ at the end (Fig 2.30).

In transferring positive charge from conductor 2 to conductor 1, work will be done externally, since at any stage conductor 1 is at a higher potential than conductor 2. To calculate the total work done, we first calculate the work done in a small step involving transfer of an infinitesimal (i.e., vanishingly small) amount of charge. Consider the intermediate situation when the conductors 1 and 2 have charges Q' and $-Q'$ respectively. At this stage, the potential difference V' between conductors 1 to 2 is Q'/C, where C is the capacitance of the system. Next imagine that a small charge $\delta Q'$ is transferred from conductor 2 to 1. Work done in this step (δW), resulting in charge Q' on conductor 1 increasing to $Q' + \delta Q'$, is given by

$$\delta W = V'\delta Q' = \frac{Q'}{C}\delta Q' \tag{2.68}$$

Integrating eq. (2.68)

$$W = \int_0^Q \frac{Q'}{C}\delta Q' = \frac{1}{C}\frac{Q'^2}{2}\Big|_0^Q = \frac{Q^2}{2C}$$

We can write the final result, in different ways

$$W = \frac{Q^2}{2C} = \frac{1}{2}CV^2 = \frac{1}{2}QV \tag{2.69}$$

Since electrostatic force is conservative, this work is stored in the form of potential energy of the system. For the same reason, the final result for potential energy [Eq. (2.69)] is independent of the manner in which the charge configuration of the capacitor is built up. When the capacitor discharges, this stored-up energy is released. It is possible to view the potential energy of the capacitor as 'stored' in the electric field between the plates. To see this, consider for simplicity, a parallel plate capacitor [of area A (of each plate) and separation d between the plates].

Energy stored in the capacitor

$$= \frac{1}{2}\frac{Q^2}{C} = \frac{(A\sigma)^2}{2} \times \frac{d}{\varepsilon_0 A} \tag{2.70}$$

The surface charge density σ is related to the electric field E between the plates,

$$E = \frac{\sigma}{\varepsilon_0} \tag{2.71}$$

From Eqs. (2.70) and (2.71) , we get

Energy stored in the capacitor

$$U = (1/2)\,\varepsilon_0 E^2 \times A d \tag{2.72}$$

Note that *Ad* is the volume of the region between the plates (where electric field alone exists). If we define *energy density as energy stored per unit volume of space*, Eq (2.72) shows that

Energy density of electric field,

$$u = (1/2)\varepsilon_0 E^2 \tag{2.73}$$

Though we derived Eq. (2.73) for the case of a parallel plate capacitor, the result on energy density of an electric field is, in fact, very general and holds true for electric field due to any configuration of charges.

Example 2.10 (a) A 900 pF capacitor is charged by 100 V battery [Fig. 2.31(a)]. How much electrostatic energy is stored by the capacitor? (b) The capacitor is disconnected from the battery and connected to another 900 pF capacitor [Fig. 2.31(b)]. What is the electrostatic energy stored by the system?

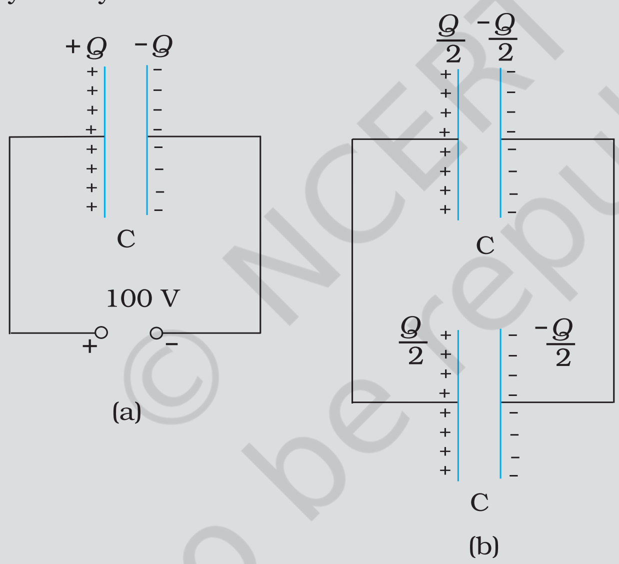


FIGURE 2.31

Solution

(a) The charge on the capacitor is

$Q = CV = 900 \times 10^{-12}\text{ F} \times 100\text{ V} = 9 \times 10^{-8}\text{ C}$

The energy stored by the capacitor is

$= (1/2)\ CV^2 = (1/2)\ QV$

$= (1/2) \times 9 \times 10^{-8}\text{C} \times 100\text{ V} = 4.5 \times 10^{-6}\text{ J}$

(b) In the steady situation, the two capacitors have their positive plates at the same potential, and their negative plates at the same potential. Let the common potential difference be V′. The

EXAMPLE 2.10

charge on each capacitor is then $Q' = CV'$. By charge conservation, $Q' = Q/2$. This implies $V' = V/2$. The total energy of the system is

$$= 2 \times \frac{1}{2} Q'V' = \frac{1}{4} QV = 2.25 \times 10^{-6} \text{ J}$$

Thus in going from (a) to (b), though no charge is lost; the final energy is only half the initial energy. *Where has the remaining energy gone*?

There is a transient period before the system settles to the situation (b). During this period, a transient current flows from the first capacitor to the second. Energy is lost during this time in the form of heat and electromagnetic radiation.

EXAMPLE 2.10

SUMMARY

1. Electrostatic force is a conservative force. Work done by an external force (equal and opposite to the electrostatic force) in bringing a charge q from a point R to a point P is $q(V_P–V_R)$, which is the difference in potential energy of charge q between the final and initial points.
2. Potential at a point is the work done per unit charge (by an external agency) in bringing a charge from infinity to that point. Potential at a point is arbitrary to within an additive constant, since it is the potential difference between two points which is physically significant. If potential at infinity is chosen to be zero; potential at a point with position vector **r** due to a point charge Q placed at the origin is given is given by

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o} \frac{Q}{r}$$

3. The electrostatic potential at a point with position vector **r** due to a point dipole of dipole moment **p** placed at the origin is

$$V(\mathbf{r}) = \frac{1}{4\pi\varepsilon_o} \frac{\mathbf{p}.\hat{\mathbf{r}}}{r^2}$$

The result is true also for a dipole (with charges $–q$ and q separated by $2a$) for $r >> a$.

4. For a charge configuration q_1, q_2, ..., q_n with position vectors $\mathbf{r}_1$, $\mathbf{r}_2$, ... $\mathbf{r}_n$, the potential at a point P is given by the superposition principle

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}} + \frac{q_2}{r_{2P}} + ... + \frac{q_n}{r_{nP}}\right)$$

where r_{1P} is the distance between q_1 and P, as and so on.

5. An equipotential surface is a surface over which potential has a constant value. For a point charge, concentric spheres centred at a location of the charge are equipotential surfaces. The electric field **E** at a point is perpendicular to the equipotential surface through the point. **E** is in the direction of the steepest decrease of potential.

Reprint 2025-26

6. Potential energy stored in a system of charges is the work done (by an external agency) in assembling the charges at their locations. Potential energy of two charges q_1, q_2 at $\mathbf{r}_1$, $\mathbf{r}_2$ is given by

$$U = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r_{12}}$$

where r_{12} is distance between q_1 and q_2.

7. The potential energy of a charge q in an external potential $V(\mathbf{r})$ is $qV(\mathbf{r})$. The potential energy of a dipole moment $\mathbf{p}$ in a uniform electric field $\mathbf{E}$ is $-\mathbf{p}\cdot\mathbf{E}$.

8. Electrostatics field $\mathbf{E}$ is zero in the interior of a conductor; just outside the surface of a charged conductor, $\mathbf{E}$ is normal to the surface given by

$\mathbf{E} = \dfrac{\sigma}{\varepsilon_0}\hat{\mathbf{n}}$ where $\hat{\mathbf{n}}$ is the unit vector along the outward normal to the surface and σ is the surface charge density. Charges in a conductor can reside only at its surface. Potential is constant within and on the surface of a conductor. In a cavity within a conductor (with no charges), the electric field is zero.

9. A capacitor is a system of two conductors separated by an insulator. Its capacitance is defined by $C = Q/V$, where Q and $-Q$ are the charges on the two conductors and V is the potential difference between them. C is determined purely geometrically, by the shapes, sizes and relative positions of the two conductors. The unit of capacitance is farad:, 1 F = 1 C V^{-1}. For a parallel plate capacitor (with vacuum between the plates),

$$C = \varepsilon_0 \frac{A}{d}$$

where A is the area of each plate and d the separation between them.

10. If the medium between the plates of a capacitor is filled with an insulating substance (dielectric), the electric field due to the charged plates induces a net dipole moment in the dielectric. This effect, called polarisation, gives rise to a field in the opposite direction. The net electric field inside the dielectric and hence the potential difference between the plates is thus reduced. Consequently, the capacitance C increases from its value C_0 when there is no medium (vacuum),

$$C = KC_0$$

where K is the dielectric constant of the insulating substance.

11. For capacitors in the series combination, the total capacitance C is given by

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots$$

In the parallel combination, the total capacitance C is:

$$C = C_1 + C_2 + C_3 + \ldots$$

where C_1, C_2, C_3... are individual capacitances.

Reprint 2025-26

12. The energy U stored in a capacitor of capacitance C, with charge Q and voltage V is

$$U = \frac{1}{2}QV = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q^2}{C}$$

The electric energy density (energy per unit volume) in a region with electric field is $(1/2)\varepsilon_0 E^2$.

Physical quantity	Symbol	Dimensions	Unit	Remark
Potential	ϕ or V	$[M^1 L^2 T^{-3} A^{-1}]$	V	Potential difference is physically significant
Capacitance	C	$[M^{-1} L^{-2} T^{-4} A^2]$	F	
Polarisation	**P**	$[L^{-2} AT]$	C m^{-2}	Dipole moment per unit volume
Dielectric constant	K	[Dimensionless]		

POINTS TO PONDER

1. Electrostatics deals with forces between charges at rest. But if there is a force on a charge, how can it be at rest? Thus, when we are talking of electrostatic force between charges, it should be understood that each charge is being kept at rest by some unspecified force that opposes the net Coulomb force on the charge.
2. A capacitor is so configured that it confines the electric field lines within a small region of space. Thus, even though field may have considerable strength, the potential difference between the two conductors of a capacitor is small.
3. Electric field is discontinuous across the surface of a spherical charged shell. It is zero inside and $\frac{\sigma}{\varepsilon_0}\hat{\mathbf{n}}$ outside. Electric potential is, however continuous across the surface, equal to $q/4\pi\varepsilon_0 R$ at the surface.
4. The torque $\mathbf{p} \times \mathbf{E}$ on a dipole causes it to oscillate about $\mathbf{E}$. Only if there is a dissipative mechanism, the oscillations are damped and the dipole eventually aligns with $\mathbf{E}$.
5. Potential due to a charge q at its own location is not defined – it is infinite.
6. In the expression $qV(\mathbf{r})$ for potential energy of a charge q, $V(\mathbf{r})$ is the potential due to external charges and not the potential due to q. As seen in point 5, this expression will be ill-defined if $V(\mathbf{r})$ includes potential due to a charge q itself.

Reprint 2025-26

7. A cavity inside a conductor is shielded from outside electrical influences. It is worth noting that electrostatic shielding does not work the other way round; that is, if you put charges inside the cavity, the exterior of the conductor is not shielded from the fields by the inside charges.

EXERCISES

2.1 Two charges 5×10^{-8} C and -3×10^{-8} C are located 16 cm apart. At what point(s) on the line joining the two charges is the electric potential zero? Take the potential at infinity to be zero.

2.2 A regular hexagon of side 10 cm has a charge 5 µC at each of its vertices. Calculate the potential at the centre of the hexagon.

2.3 Two charges 2 µC and –2 µC are placed at points A and B 6 cm apart.

(a) Identify an equipotential surface of the system.

(b) What is the direction of the electric field at every point on this surface?

2.4 A spherical conductor of radius 12 cm has a charge of 1.6×10^{-7}C distributed uniformly on its surface. What is the electric field

(a) inside the sphere

(b) just outside the sphere

(c) at a point 18 cm from the centre of the sphere?

2.5 A parallel plate capacitor with air between the plates has a capacitance of 8 pF (1pF = 10^{-12} F). What will be the capacitance if the distance between the plates is reduced by half, and the space between them is filled with a substance of dielectric constant 6?

2.6 Three capacitors each of capacitance 9 pF are connected in series.

(a) What is the total capacitance of the combination?

(b) What is the potential difference across each capacitor if the combination is connected to a 120 V supply?

2.7 Three capacitors of capacitances 2 pF, 3 pF and 4 pF are connected in parallel.

(a) What is the total capacitance of the combination?

(b) Determine the charge on each capacitor if the combination is connected to a 100 V supply.

2.8 In a parallel plate capacitor with air between the plates, each plate has an area of 6×10^{-3} m^2 and the distance between the plates is 3 mm. Calculate the capacitance of the capacitor. If this capacitor is connected to a 100 V supply, what is the charge on each plate of the capacitor?

Reprint 2025-26

2.9 Explain what would happen if in the capacitor given in Exercise 2.8, a 3 mm thick mica sheet (of dielectric constant = 6) were inserted between the plates,

(a) while the voltage supply remained connected.

(b) after the supply was disconnected.

2.10 A 12pF capacitor is connected to a 50V battery. How much electrostatic energy is stored in the capacitor?

2.11 A 600pF capacitor is charged by a 200V supply. It is then disconnected from the supply and is connected to another uncharged 600 pF capacitor. How much electrostatic energy is lost in the process?

Reprint 2025-26

Chapter Three

CURRENT ELECTRICITY

3.1 Introduction

In Chapter 1, all charges whether free or bound, were considered to be at rest. Charges in motion constitute an electric current. Such currents occur naturally in many situations. Lightning is one such phenomenon in which charges flow from the clouds to the earth through the atmosphere, sometimes with disastrous results. The flow of charges in lightning is not steady, but in our everyday life we see many devices where charges flow in a steady manner, like water flowing smoothly in a river. A torch and a cell-driven clock are examples of such devices. In the present chapter, we shall study some of the basic laws concerning steady electric currents.

3.2 Electric Current

Imagine a small area held normal to the direction of flow of charges. Both the positive and the negative charges may flow forward and backward across the area. In a given time interval t, let q_+ be the net amount (*i.e.*, forward *minus* backward) of positive charge that flows in the forward direction across the area. Similarly, let q_- be the net amount of negative charge flowing across the area in the forward direction. The net amount of charge flowing across the area in the forward direction in the time interval t, then, is $q = q_+ - q_-$. This is proportional to t for steady current and the quotient

$$I = \frac{q}{t} \tag{3.1}$$

is defined to be the *current* across the area in the forward direction. (If it turn out to be a negative number, it implies a current in the backward direction.)

Currents are not always steady and hence more generally, we define the current as follows. Let ΔQ be the net charge flowing across a cross-section of a conductor during the time interval Δt [i.e., between times t and $(t + \Delta t)$]. Then, the current at time t across the cross-section of the conductor is defined as the value of the ratio of ΔQ to Δt in the limit of Δt tending to zero,

$$I(t) \equiv \lim_{\Delta t \to 0} \frac{\Delta Q}{\Delta t} \tag{3.2}$$

In SI units, the unit of current is ampere. An ampere is defined through magnetic effects of currents that we will study in the following chapter. An ampere is typically the order of magnitude of currents in domestic appliances. An average lightning carries currents of the order of tens of thousands of amperes and at the other extreme, currents in our nerves are in microamperes.

3.3 Electric Currents in Conductors

An electric charge will experience a force if an electric field is applied. If it is free to move, it will thus move contributing to a current. In nature, free charged particles do exist like in upper strata of atmosphere called the *ionosphere*. However, in atoms and molecules, the negatively charged electrons and the positively charged nuclei are bound to each other and are thus not free to move. Bulk matter is made up of many molecules, a gram of water, for example, contains approximately 10^{22} molecules. These molecules are so closely packed that the electrons are no longer attached to individual nuclei. In some materials, the electrons will still be bound, i.e., they will not accelerate even if an electric field is applied. In other materials, notably metals, some of the electrons are practically free to move within the bulk material. These materials, generally called conductors, develop electric currents in them when an electric field is applied.

If we consider solid conductors, then of course the atoms are tightly bound to each other so that the current is carried by the negatively charged electrons. There are, however, other types of conductors like electrolytic solutions where positive and negative charges both can move. In our discussions, we will focus only on solid conductors so that the current is carried by the negatively charged electrons in the background of fixed positive ions.

Consider first the case when no electric field is present. The electrons will be moving due to thermal motion during which they collide with the fixed ions. An electron colliding with an ion emerges with the same speed as before the collision. However, the direction of its velocity after the collision is completely random. At a given time, there is no preferential direction for the velocities of the electrons. Thus on the average, the number of electrons travelling in any direction will be equal to the number of electrons travelling in the opposite direction. So, there will be no net electric current.

Let us now see what happens to such a piece of conductor if an electric field is applied. To focus our thoughts, imagine the conductor in the shape of a cylinder of radius R (Fig. 3.1). Suppose we now take two thin circular discs of a dielectric of the same radius and put positive charge $+Q$ distributed over one disc and similarly $-Q$ at the other disc. We attach the two discs on the two flat surfaces of the cylinder. An electric field will be created and is directed from the positive towards the negative charge. The electrons will be accelerated due to this field towards $+Q$. They will thus move to neutralise the charges. The electrons, as long as they are moving, will constitute an electric current. Hence in the situation considered, there will be a current for a very short while and no current thereafter.

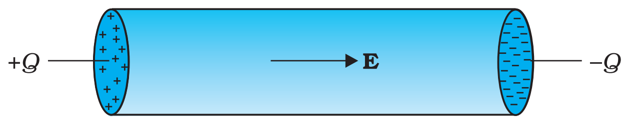


FIGURE 3.1 Charges $+Q$ and $-Q$ put at the ends of a metallic cylinder. The electrons will drift because of the electric field created to neutralise the charges. The current thus will stop after a while unless the charges $+Q$ and $-Q$ are continuously replenished.

We can also imagine a mechanism where the ends of the cylinder are supplied with fresh charges to make up for any charges neutralised by electrons moving inside the conductor. In that case, there will be a steady electric field in the body of the conductor. This will result in a continuous current rather than a current for a short period of time. Mechanisms, which maintain a steady electric field are cells or batteries that we shall study later in this chapter. In the next sections, we shall study the steady current that results from a steady electric field in conductors.

3.4 Ohm's Law

A basic law regarding flow of currents was discovered by G.S. Ohm in 1828, long before the physical mechanism responsible for flow of currents was discovered. Imagine a conductor through which a current I is flowing and let V be the potential difference between the ends of the conductor. Then Ohm's law states that

$$V \propto I$$

$$\text{or, } V = R\,I \qquad (3.3)$$

where the constant of proportionality R is called the *resistance* of the conductor. The SI units of resistance is *ohm*, and is denoted by the symbol Ω. The resistance R not only depends on the material of the conductor but also on the dimensions of the conductor. The dependence of R on the dimensions of the conductor can easily be determined as follows.

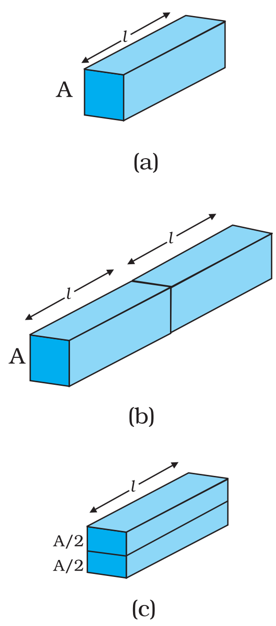


FIGURE 3.2 Illustrating the relation $R = \rho l/A$ for a rectangular slab of length l and area of cross-section A.

Consider a conductor satisfying Eq. (3.3) to be in the form of a slab of length l and cross sectional area A [Fig. 3.2(a)]. Imagine placing two such identical slabs side by side [Fig. 3.2(b)], so that the length of the combination is $2l$. The current flowing through the combination is the same as that flowing through either of the slabs. If V is the potential difference across the ends of the first slab, then V is also the potential difference across the ends of the second slab since the second slab is

GEORG SIMON OHM (1787–1854)

Georg Simon Ohm (1787–1854) German physicist, professor at Munich. Ohm was led to his law by an analogy between the conduction of heat: the electric field is analogous to the temperature gradient, and the electric current is analogous to the heat flow.

identical to the first and the same current I flows through both. The potential difference across the ends of the combination is clearly sum of the potential difference across the two individual slabs and hence equals $2V$. The current through the combination is I and the resistance of the combination R_C is [from Eq. (3.3)],

$$R_C = \frac{2V}{I} = 2R \tag{3.4}$$

since $V/I = R$, the resistance of either of the slabs. Thus, doubling the length of a conductor doubles the resistance. In general, then resistance is proportional to length,

$$R \propto l \tag{3.5}$$

Next, imagine dividing the slab into two by cutting it lengthwise so that the slab can be considered as a combination of two identical slabs of length l, but each having a cross sectional area of $A/2$ [Fig. 3.2(c)].

For a given voltage V across the slab, if I is the current through the entire slab, then clearly the current flowing through each of the two half-slabs is $I/2$. Since the potential difference across the ends of the half-slabs is V, i.e., the same as across the full slab, the resistance of each of the half-slabs R_1 is

$$R_1 = \frac{V}{(I/2)} = 2\frac{V}{I} = 2R. \tag{3.6}$$

Thus, halving the area of the cross-section of a conductor doubles the resistance. In general, then the resistance R is inversely proportional to the cross-sectional area,

$$R \propto \frac{1}{A} \tag{3.7}$$

Combining Eqs. (3.5) and (3.7), we have

$$R \propto \frac{l}{A} \tag{3.8}$$

and hence for a given conductor

$$R = \rho \frac{l}{A} \tag{3.9}$$

where the constant of proportionality ρ depends on the material of the conductor but not on its dimensions. ρ is called *resistivity*.

Using the last equation, Ohm's law reads

$$V = I \times R = \frac{I\rho l}{A} \tag{3.10}$$

Current per unit area (taken normal to the current), I/A, is called *current density* and is denoted by j. The SI units of the current density are A/m^2. Further, if E is the magnitude of uniform electric field in the conductor whose length is l, then the potential difference V across its ends is El. Using these, the last equation reads

$$E\, l = j\, \rho\, l$$

or, $$E = j\, \rho \tag{3.11}$$

The above relation for ***magnitudes*** E and j can indeed be cast in a ***vector*** form. The current density, (which we have defined as the current through unit area ***normal*** to the current) is also directed along **E**, and is also a vector $\mathbf{j}\ (\equiv j\,\mathbf{E}/\mathrm{E})$. Thus, the last equation can be written as,

$$\mathbf{E} = \mathbf{j}\rho \tag{3.12}$$

or, $$\mathbf{j} = \sigma\, \mathbf{E} \tag{3.13}$$

where $\sigma \equiv 1/\rho$ is called the *conductivity*. Ohm's law is often stated in an equivalent form, Eq. (3.13) in addition to Eq.(3.3). In the next section, we will try to understand the origin of the Ohm's law as arising from the characteristics of the drift of electrons.

3.5 Drift of Electrons and the Origin of Resistivity

As remarked before, an electron will suffer collisions with the heavy fixed ions, but after collision, it will emerge with the same speed but in random directions. If we consider all the electrons, their average velocity will be zero since their directions are random. Thus, if there are N electrons and the velocity of the i^{th} electron ($i = 1, 2, 3, \ldots N$) at a given time is $\mathbf{v}_i$, then

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{v}_i = 0 \tag{3.14}$$

Consider now the situation when an electric field is present. Electrons will be accelerated due to this field by

$$\mathbf{a} = \frac{-e\,\mathbf{E}}{m} \tag{3.15}$$

where $-e$ is the charge and m is the mass of an electron. Consider again the i^{th} electron at a given time t. This electron would have had its last collision some time before t, and let t_i be the time elapsed after its last collision. If $\mathbf{v}_i$ was its velocity immediately after the last collision, then its velocity $\mathbf{V}_i$ at time t is

$$\mathbf{V}_i = \mathbf{v}_i + \frac{-e\mathbf{E}}{m}\, t_i \tag{3.16}$$

since starting with its last collision it was accelerated (Fig. 3.3) with an acceleration given by Eq. (3.15) for a time interval t_i. The average velocity of the electrons at time t is the average of all the $\mathbf{V}_i$'s. The average of $\mathbf{v}_i$'s is zero [Eq. (3.14)] since immediately after any collision, the direction of the velocity of an electron is completely random. The collisions of the electrons do not occur at regular intervals but at random times. Let us denote by τ, the average time between successive collisions. Then at a given time, some of the electrons would have spent

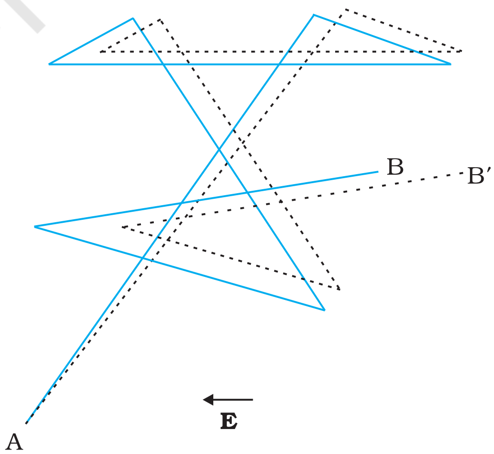


FIGURE 3.3 A schematic picture of an electron moving from a point A to another point B through repeated collisions, and straight line travel between collisions (full lines). If an electric field is applied as shown, the electron ends up at point B′ (dotted lines). A slight drift in a direction opposite the electric field is visible.

time more than τ and some less than τ. In other words, the time t_i in Eq. (3.16) will be less than τ for some and more than τ for others as we go through the values of i = 1, 2 ..... N. The average value of t_i then is τ (known as *relaxation time*). Thus, averaging Eq. (3.16) over the N-electrons at any given time t gives us for the average velocity $\mathbf{v}_d$

$$\mathbf{v}_d \equiv (\mathbf{V}_i)_{average} = (\mathbf{v}_i)_{average} - \frac{e\mathbf{E}}{m}(t_i)_{average}$$

$$= 0 - \frac{e\mathbf{E}}{m}\tau = -\frac{e\mathbf{E}}{m}\tau \tag{3.17}$$

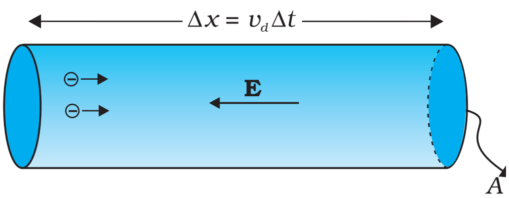


FIGURE 3.4 Current in a metallic conductor. The magnitude of current density in a metal is the magnitude of charge contained in a cylinder of unit area and length v_d.

This last result is surprising. It tells us that the electrons move with an average velocity which is independent of time, although electrons are accelerated. This is the phenomenon of drift and the velocity $\mathbf{v}_d$ in Eq. (3.17) is called the ***drift velocity***.

Because of the drift, there will be net transport of charges across any area perpendicular to **E**. Consider a planar area A, located inside the conductor such that the normal to the area is parallel to **E** (Fig. 3.4). Then because of the drift, in an infinitesimal amount of time Δt, all electrons to the left of the area at distances upto $|\mathbf{v}_d|\Delta t$ would have crossed the area. If n is the number of free electrons per unit volume in the metal, then there are $n\ \Delta t\ |\mathbf{v}_d| A$ such electrons. Since each electron carries a charge $-e$, the total charge transported across this area A to the right in time Δt is $-neA|\mathrm{v}_d|\Delta t$. **E** is directed towards the left and hence the total charge transported along **E** across the area is negative of this. The amount of charge crossing the area A in time Δt is by definition [Eq. (3.2)] $I\,\Delta t$, where I is the magnitude of the current. Hence,

$$I\,\Delta t = +n\,e\,A\,|\mathrm{v}_d|\,\Delta t \tag{3.18}$$

Substituting the value of $|\mathbf{v}_d|$ from Eq. (3.17)

$$I\Delta t = \frac{e^2 A}{m}\tau n\,\Delta t|\mathrm{E}| \tag{3.19}$$

By definition I is related to the magnitude $|\mathrm{j}|$ of the current density by

$$I = |\mathrm{j}|A \tag{3.20}$$

Hence, from Eqs.(3.19) and (3.20),

$$|\mathrm{j}| = \frac{ne^2}{m}\tau|\mathrm{E}| \tag{3.21}$$

The vector **j** is parallel to **E** and hence we can write Eq. (3.21) in the vector form

$$\mathbf{j} = \frac{ne^2}{m}\tau\mathbf{E} \tag{3.22}$$

Comparison with Eq. (3.13) shows that Eq. (3.22) is exactly the Ohm's law, if we identify the conductivity σ as

$$\sigma = \frac{ne^2}{m}\tau \qquad (3.23)$$

We thus see that a very simple picture of electrical conduction reproduces Ohm's law. We have, of course, made assumptions that τ and n are constants, independent of E. We shall, in the next section, discuss the limitations of Ohm's law.

Example 3.1 (a) Estimate the average drift speed of conduction electrons in a copper wire of cross-sectional area 1.0×10^{-7} m^2 carrying a current of 1.5 A. Assume that each copper atom contributes roughly one conduction electron. The density of copper is 9.0×10^3 kg/m^3, and its atomic mass is 63.5 u. (b) Compare the drift speed obtained above with, (i) thermal speeds of copper atoms at ordinary temperatures, (ii) speed of propagation of electric field along the conductor which causes the drift motion.

Solution

(a) The direction of drift velocity of conduction electrons is opposite to the electric field direction, i.e., electrons drift in the direction of increasing potential. The drift speed v_d is given by Eq. (3.18)
$v_d = (I/neA)$
Now, $e = 1.6 \times 10^{-19}$ C, $A = 1.0 \times 10^{-7}$m^2, $I = 1.5$ A. The density of conduction electrons, n is equal to the number of atoms per cubic metre (assuming one conduction electron per Cu atom as is reasonable from its valence electron count of one). A cubic metre of copper has a mass of 9.0×10^3 kg. Since 6.0×10^{23} copper atoms have a mass of 63.5 g,

$$n = \frac{6.0 \times 10^{23}}{63.5} \times 9.0 \times 10^6$$

$$= 8.5 \times 10^{28} \text{ m}^{-3}$$

which gives,

$$v_d = \frac{1.5}{8.5 \times 10^{28} \times 1.6 \times 10^{-19} \times 1.0 \times 10^{-7}}$$

$$= 1.1 \times 10^{-3} \text{ m s}^{-1} = 1.1 \text{ mm s}^{-1}$$

(b) (i) At a temperature T, the thermal speed* of a copper atom of mass M is obtained from [$<(1/2)\, Mv^2> = (3/2)\, k_B T$] and is thus typically of the order of $\sqrt{k_B T/M}$, where k_B is the Boltzmann constant. For copper at 300 K, this is about 2×10^2 m/s. This figure indicates the random vibrational speeds of copper atoms in a conductor. Note that the drift speed of electrons is much smaller, about 10^{-5} times the typical thermal speed at ordinary temperatures.
(ii) An electric field travelling along the conductor has a speed of an electromagnetic wave, namely equal to 3.0×10^8 m s^{-1} (You will learn about this in Chapter 8). The drift speed is, in comparison, extremely small; smaller by a factor of 10^{-11}.

EXAMPLE 3.1

* See Eq. (12.23) of Chapter 12 from Class XI book.

Example 3.2

(a) In Example 3.1, the electron drift speed is estimated to be only a few mm s^{-1} for currents in the range of a few amperes? How then is current established almost the instant a circuit is closed?

(b) The electron drift arises due to the force experienced by electrons in the electric field inside the conductor. But force should cause acceleration. Why then do the electrons acquire a steady average drift speed?

(c) If the electron drift speed is so small, and the electron's charge is small, how can we still obtain large amounts of current in a conductor?

(d) When electrons drift in a metal from lower to higher potential, does it mean that all the 'free' electrons of the metal are moving in the same direction?

(e) Are the paths of electrons straight lines between successive collisions (with the positive ions of the metal) in the (i) absence of electric field, (ii) presence of electric field?

Solution

(a) Electric field is established throughout the circuit, almost instantly (with the speed of light) causing at every point a *local electron drift.* Establishment of a current does not have to wait for electrons from one end of the conductor travelling to the other end. However, it does take a little while for the current to reach its steady value.

(b) Each 'free' electron does accelerate, increasing its drift speed until it collides with a positive ion of the metal. It loses its drift speed after collision but starts to accelerate and increases its drift speed again only to suffer a collision again and so on. On the average, therefore, electrons acquire only a drift speed.

(c) Simple, because the electron number density is enormous, $\sim 10^{29}\ m^{-3}$.

(d) By no means. The drift velocity is superposed over the large random velocities of electrons.

(e) In the absence of electric field, the paths are straight lines; in the presence of electric field, the paths are, in general, curved.

Example 3.2

3.5.1 Mobility

As we have seen, conductivity arises from mobile charge carriers. In metals, these mobile charge carriers are electrons; in an ionised gas, they are electrons and positive charged ions; in an electrolyte, these can be both positive and negative ions.

An important quantity is the *mobility* μ defined as the magnitude of the drift velocity per unit electric field:

$$\mu = \frac{|\mathbf{v}_d|}{E} \tag{3.24}$$

The SI unit of mobility is m^2/Vs and is 10^4 of the mobility in practical units (cm^2/Vs). Mobility is positive. From Eq. (3.17), we have

$$v_d = \frac{e\tau E}{m}$$

Reprint 2025-26

Hence,

$$\mu = \frac{v_d}{E} = \frac{e\tau}{m} \qquad (3.25)$$

where τ is the average collision time for electrons.

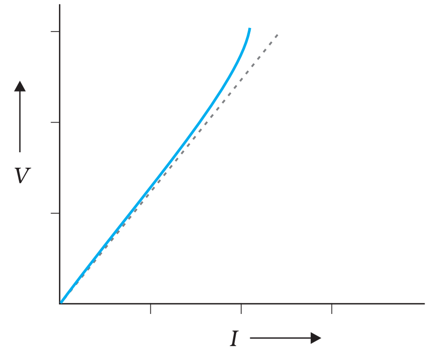


FIGURE 3.5 The dashed line represents the linear Ohm's law. The solid line is the voltage V versus current I for a good conductor.

3.6 Limitations of Ohm's Law

Although Ohm's law has been found valid over a large class of materials, there do exist materials and devices used in electric circuits where the proportionality of V and I does not hold. The deviations broadly are one or more of the following types:

(a) V ceases to be proportional to I (Fig. 3.5).

(b) The relation between V and I depends on the sign of V. In other words, if I is the current for a certain V, then reversing the direction of V keeping its magnitude fixed, does not produce a current of the same magnitude as I in the opposite direction (Fig. 3.6). This happens, for example, in a diode which we will study in Chapter 14.

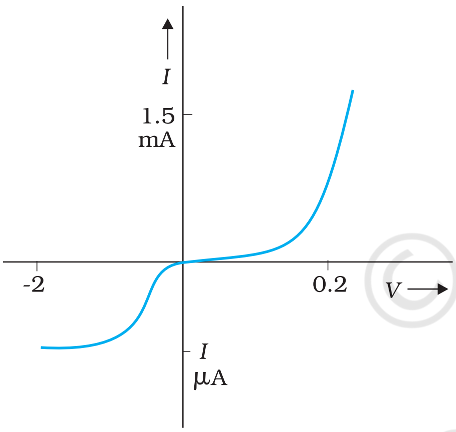


FIGURE 3.6 Characteristic curve of a diode. Note the different scales for negative and positive values of the voltage and current.

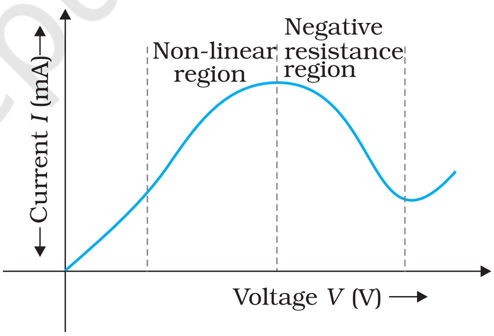


FIGURE 3.7 Variation of current versus voltage for GaAs.

(c) The relation between V and I is not unique, i.e., there is more than one value of V for the same current I (Fig. 3.7). A material exhibiting such behaviour is GaAs.

Materials and devices not obeying Ohm's law in the form of Eq. (3.3) are actually widely used in electronic circuits. In this and a few subsequent chapters, however, we will study the electrical currents in materials that obey Ohm's law.

3.7 Resistivity of Various Materials

The materials are classified as conductors, semiconductors and insulators depending on their resistivities, in an increasing order of their values.

Reprint 2025-26

Metals have low resistivities in the range of 10^{-8} Ωm to 10^{-6} Ωm. At the other end are insulators like ceramic, rubber and plastics having resistivities 10^{18} times greater than metals or more. In between the two are the semiconductors. These, however, have resistivities characteristically decreasing with a rise in temperature. The resistivities of semiconductors can be decreased by adding small amount of suitable impurities. This last feature is exploited in use of semiconductors for electronic devices.

3.8 Temperature Dependence of Resistivity

The resistivity of a material is found to be dependent on the temperature. Different materials do not exhibit the same dependence on temperatures. Over a limited range of temperatures, that is not too large, the resistivity of a metallic conductor is approximately given by,

$$\rho_T = \rho_0 \, [1 + \alpha \, (T - T_0)] \tag{3.26}$$

where ρ_T is the resistivity at a temperature T and ρ_0 is the same at a reference temperature T_0. α is called the *temperature co-efficient of resistivity*, and from Eq. (3.26), the dimension of α is (Temperature)$^{-1}$. For metals, α is positive.

The relation of Eq. (3.26) implies that a graph of ρ_T plotted against T would be a straight line. At temperatures much lower than 0°C, the graph, however, deviates considerably from a straight line (Fig. 3.8).

Equation (3.26) thus, can be used approximately over a limited range of T around any reference temperature T_0, where the graph can be approximated as a straight line.

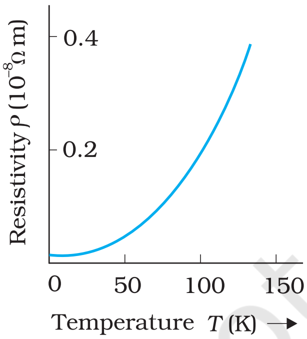


FIGURE 3.8 Resistivity ρ_T of copper as a function of temperature T.

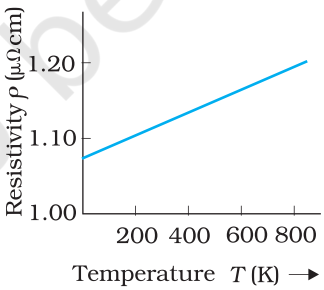


FIGURE 3.9 Resistivity ρ_T of nichrome as a function of absolute temperature T.

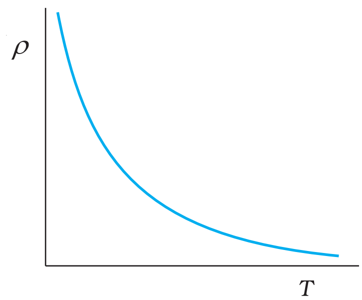


FIGURE 3.10 Temperature dependence of resistivity for a typical semiconductor.

Some materials like Nichrome (which is an alloy of nickel, iron and chromium) exhibit a very weak dependence of resistivity with temperature (Fig. 3.9). Manganin and constantan have similar properties. These materials are thus widely used in wire bound standard resistors since their resistance values would change very little with temperatures.

Reprint 2025-26

Unlike metals, the resistivities of semiconductors decrease with increasing temperatures. A typical dependence is shown in Fig. 3.10.

We can qualitatively understand the temperature dependence of resistivity, in the light of our derivation of Eq. (3.23). From this equation, resistivity of a material is given by

$$\rho = \frac{1}{\sigma} = \frac{m}{n e^2 \tau} \qquad (3.27)$$

ρ thus depends inversely both on the number n of free electrons per unit volume and on the average time τ between collisions. As we increase temperature, average speed of the electrons, which act as the carriers of current, increases resulting in more frequent collisions. The average time of collisions τ, thus decreases with temperature.

In a metal, n is not dependent on temperature to any appreciable extent and thus the decrease in the value of τ with rise in temperature causes ρ to increase as we have observed.

For insulators and semiconductors, however, n increases with temperature. This increase more than compensates any decrease in τ in Eq.(3.23) so that for such materials, ρ decreases with temperature.

Example 3.3 An electric toaster uses nichrome for its heating element. When a negligibly small current passes through it, its resistance at room temperature (27.0 °C) is found to be 75.3 Ω. When the toaster is connected to a 230 V supply, the current settles, after a few seconds, to a steady value of 2.68 A. What is the steady temperature of the nichrome element? The temperature coefficient of resistance of nichrome averaged over the temperature range involved, is 1.70×10^{-4} °C^{-1}.

Solution When the current through the element is very small, heating effects can be ignored and the temperature T_1 of the element is the same as room temperature. When the toaster is connected to the supply, its initial current will be slightly higher than its steady value of 2.68 A. But due to heating effect of the current, the temperature will rise. This will cause an increase in resistance and a slight decrease in current. In a few seconds, a steady state will be reached when temperature will rise no further, and both the resistance of the element and the current drawn will achieve steady values. The resistance R_2 at the steady temperature T_2 is

$$R_2 = \frac{230\ \text{V}}{2.68\ \text{A}} = 85.8\ \Omega$$

Using the relation

$$R_2 = R_1\,[1 + \alpha\,(T_2 - T_1)]$$

with $\alpha = 1.70 \times 10^{-4}$ °C^{-1}, we get

$$T_2 - T_1 = \frac{(85.8 - 75.3)}{(75.3) \times 1.70 \times 10^{-4}} = 820\ °\text{C}$$

that is, $T_2 = (820 + 27.0)$ °C = 847 °C

Thus, the steady temperature of the heating element (when heating effect due to the current equals heat loss to the surroundings) is 847 °C.

EXAMPLE 3.3

Example 3.4 The resistance of the platinum wire of a platinum resistance thermometer at the ice point is 5 Ω and at steam point is 5.23 Ω. When the thermometer is inserted in a hot bath, the resistance of the platinum wire is 5.795 Ω. Calculate the temperature of the bath.

Solution R_0 = 5 Ω, R_{100} = 5.23 Ω and R_t = 5.795 Ω

Now, $t = \frac{R_t - R_0}{R_{100} - R_0} \times 100, \qquad R_t = R_0 (1 + \alpha t)$

$$= \frac{5.795 - 5}{5.23 - 5} \times 100$$

$$= \frac{0.795}{0.23} \times 100 = 345.65\ °C$$

EXAMPLE 3.4

3.9 Electrical Energy, Power

Consider a conductor with end points A and B, in which a current I is flowing from A to B. The electric potential at A and B are denoted by V(A) and V(B) respectively. Since current is flowing from A to B, V(A) > V(B) and the potential difference across AB is $V = V(A) - V(B) > 0$.

In a time interval Δt, an amount of charge $\Delta Q = I\,\Delta t$ travels from A to B. The potential energy of the charge at A, by definition, was $Q\ V$(A) and similarly at B, it is $Q\ V$(B). Thus, change in its potential energy ΔU_{pot} is

$$\Delta U_{pot} = \text{Final potential energy} - \text{Initial potential energy}$$
$$= \Delta Q[(V(B) - V(A)] = -\Delta Q\ V$$
$$= -I\ V\Delta t < 0 \qquad (3.28)$$

If charges moved without collisions through the conductor, their kinetic energy would also change so that the total energy is unchanged. Conservation of total energy would then imply that,

$$\Delta K = -\Delta U_{pot} \qquad (3.29)$$

that is,

$$\Delta K = I\ V\Delta t > 0 \qquad (3.30)$$

Thus, in case charges were moving freely through the conductor under the action of electric field, their kinetic energy would increase as they move. We have, however, seen earlier that on the average, charge carriers do not move with acceleration but with a steady drift velocity. This is because of the collisions with ions and atoms during transit. During collisions, the energy gained by the charges thus is shared with the atoms. The atoms vibrate more vigorously, i.e., the conductor heats up. Thus, in an actual conductor, an amount of energy dissipated as heat in the conductor during the time interval Δt is,

$$\Delta W = I\ V\Delta t \qquad (3.31)$$

The energy dissipated per unit time is the power dissipated $P = \Delta W/\Delta t$ and we have,

$$P = I\ V \qquad (3.32)$$

Using Ohm's law $V = IR$, we get

$$P = I^2 R = V^2/R \tag{3.33}$$

as the power loss ("ohmic loss") in a conductor of resistance R carrying a current I. It is this power which heats up, for example, the coil of an electric bulb to incandescence, radiating out heat and light.

Where does the power come from? As we have reasoned before, we need an external source to keep a steady current through the conductor. It is clearly this source which must supply this power. In the simple circuit shown with a cell (Fig.3.11), it is the chemical energy of the cell which supplies this power for as long as it can.

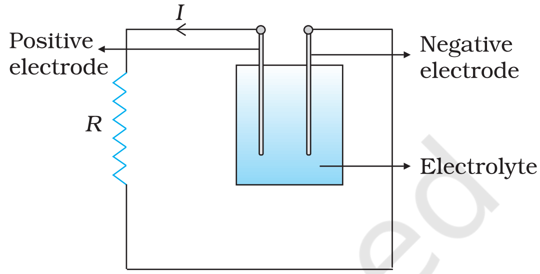


FIGURE 3.11 Heat is produced in the resistor R which is connected across the terminals of a cell. The energy dissipated in the resistor R comes from the chemical energy of the electrolyte.

The expressions for power, Eqs. (3.32) and (3.33), show the dependence of the power dissipated in a resistor R on the current through it and the voltage across it.

Equation (3.33) has an important application to power transmission. Electrical power is transmitted from power stations to homes and factories, which may be hundreds of miles away, via transmission cables. One obviously wants to minimise the power loss in the transmission cables connecting the power stations to homes and factories. We shall see now how this can be achieved. Consider a device R, to which a power P is to be delivered via transmission cables having a resistance R_c to be dissipated by it finally. If V is the voltage across R and I the current through it, then

$$P = V I \tag{3.34}$$

The connecting wires from the power station to the device has a finite resistance R_c. The power dissipated in the connecting wires, which is wasted is P_c with

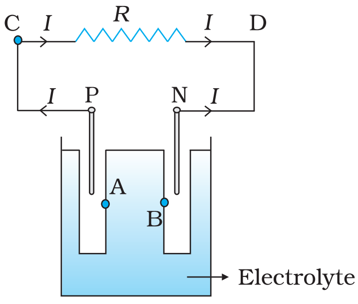
$$P_c = I^2 R_c$$

$$= \frac{P^2 R_c}{V^2} \tag{3.35}$$

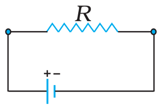
from Eq. (3.32). Thus, to drive a device of power P, the power wasted in the connecting wires is inversely proportional to V^2. The transmission cables from power stations are hundreds of miles long and their resistance R_c is considerable. To reduce P_c, these wires carry current at enormous values of V and this is the reason for the high voltage danger signs on transmission lines — a common sight as we move away from populated areas. Using electricity at such voltages is not safe and hence at the other end, a device called a transformer lowers the voltage to a value suitable for use.

3.10 Cells, emf, Internal Resistance

We have already mentioned that a simple device to maintain a steady current in an electric circuit is the electrolytic cell. Basically a cell has two electrodes, called the positive (P) and the negative (N), as shown in



(a)



Symbol
(b)

FIGURE 3.12 (a) Sketch of an electrolyte cell with positive terminal P and negative terminal N. The gap between the electrodes is exaggerated for clarity. A and B are points in the electrolyte typically close to P and N. (b) the symbol for a cell, + referring to P and – referring to the N electrode. Electrical connections to the cell are made at P and N.

Fig. 3.12. They are immersed in an electrolytic solution. Dipped in the solution, the electrodes exchange charges with the electrolyte. The positive electrode has a potential difference V_+ ($V_+ > 0$) between itself and the electrolyte solution immediately adjacent to it marked A in the figure. Similarly, the negative electrode develops a negative potential $-(V_-)$ ($V_- \geq 0$) relative to the electrolyte adjacent to it, marked as B in the figure. When there is no current, the electrolyte has the same potential throughout, so that the potential difference between P and N is $V_+ - (-V_-) = V_+ + V_-$. This difference is called the *electromotive force* (emf) of the cell and is denoted by ε. Thus

$$\varepsilon = V_+ + V_- > 0 \tag{3.36}$$

Note that ε is, actually, a potential difference and *not a force*. The name emf, however, is used because of historical reasons, and was given at a time when the phenomenon was not understood properly.

To understand the significance of ε, consider a resistor R connected across the cell (Fig. 3.12). A current I flows across R from C to D. As explained before, a steady current is maintained because current flows from N to P through the electrolyte. Clearly, across the electrolyte the same current flows through the electrolyte but from N to P, whereas through R, it flows from P to N.

The electrolyte through which a current flows has a finite resistance r, called the *internal resistance*. Consider first the situation when R is infinite so that $I = V/R = 0$, where V is the potential difference between P and N. Now,

$$\begin{aligned} V &= \text{Potential difference between P and A} \\ &\quad + \text{Potential difference between A and B} \\ &\quad + \text{Potential difference between B and N} \\ &= \varepsilon \end{aligned} \tag{3.37}$$

Thus, emf ε is the potential difference between the positive and negative electrodes in an open circuit, i.e., when no current is flowing through the cell.

If however R is finite, I is not zero. In that case the potential difference between P and N is

$$\begin{aligned} V &= V_+ + V_- - I\,r \\ &= \varepsilon - I\,r \end{aligned} \tag{3.38}$$

Note the negative sign in the expression ($I\,r$) for the potential difference between A and B. This is because the current I flows from B to A in the electrolyte.

In practical calculations, internal resistances of cells in the circuit may be neglected when the current I is such that $\varepsilon >> I\,r$. The actual values of the internal resistances of cells vary from cell to cell. The internal resistance of dry cells, however, is much higher than the common electrolytic cells.

We also observe that since V is the potential difference across R, we have from Ohm's law

$$V = I\,R \tag{3.39}$$

Combining Eqs. (3.38) and (3.39), we get

Reprint 2025-26

$$I\,R \;=\; \varepsilon - I\;r$$

Or, $I = \dfrac{\varepsilon}{R + r}$ (3.40)

The maximum current that can be drawn from a cell is for $R = 0$ and it is $I_{max} = \varepsilon/r$. However, in most cells the maximum allowed current is much lower than this to prevent permanent damage to the cell.

3.11 Cells in Series and in Parallel

Like resistors, cells can be combined together in an electric circuit. And like resistors, one can, for calculating currents and voltages in a circuit, replace a combination of cells by an equivalent cell.

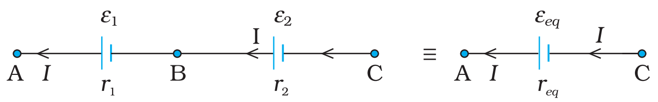


FIGURE 3.13 Two cells of emf's ε_1 and ε_2 in the series. r_1, r_2 are their internal resistances. For connections across A and C, the combination can be considered as one cell of emf ε_{eq} and an internal resistance r_{eq}.

Consider first two cells in series (Fig. 3.13), where one terminal of the two cells is joined together leaving the other terminal in either cell free. ε_1, ε_2 are the emf's of the two cells and r_1, r_2 their internal resistances, respectively.

Let V(A), V(B), V(C) be the potentials at points A, B and C shown in Fig. 3.13. Then V(A) – V(B) is the potential difference between the positive and negative terminals of the first cell. We have already calculated it in Eq. (3.38) and hence,

$$V_{AB} \equiv V(A) - V(B) = \varepsilon_1 - I\,r_1 \tag{3.41}$$

Similarly,

$$V_{BC} \equiv V(B) - V(C) = \varepsilon_2 - I\,r_2 \tag{3.42}$$

Hence, the potential difference between the terminals A and C of the combination is

$$V_{AC} \equiv V(A) - V(C) = \left[V(A) - V(B)\right] + \left[V(B) - V(C)\right]$$

$$= (\varepsilon_1 + \varepsilon_2) - I(r_1 + r_2) \tag{3.43}$$

If we wish to replace the combination by a single cell between A and C of emf ε_{eq} and internal resistance r_{eq}, we would have

$$V_{AC} = \varepsilon_{eq} - I\,r_{eq} \tag{3.44}$$

Comparing the last two equations, we get

$$\varepsilon_{eq} = \varepsilon_1 + \varepsilon_2 \tag{3.45}$$

and $r_{eq} = r_1 + r_2$ (3.46)

In Fig.3.13, we had connected the negative electrode of the first to the positive electrode of the second. If instead we connect the two negatives, Eq. (3.42) would change to $V_{BC} = -\varepsilon_2 - Ir_2$ and we will get

$$\varepsilon_{eq} = \varepsilon_1 - \varepsilon_2 \qquad (\varepsilon_1 > \varepsilon_2) \tag{3.47}$$

The rule for series combination clearly can be extended to any number of cells:

(i) The equivalent emf of a series combination of n cells is just the sum of their individual emf's, and

(ii) The equivalent internal resistance of a series combination of n cells is just the sum of their internal resistances.

This is so, when the current leaves each cell from the positive electrode. If in the combination, the current leaves any cell from the *negative* electrode, the emf of the cell enters the expression for ε_{eq} with a *negative* sign, as in Eq. (3.47).

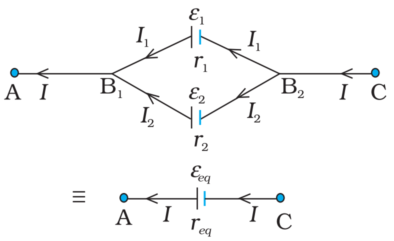


FIGURE 3.14 Two cells in parallel. For connections across A and C, the combination can be replaced by one cell of emf ε_{eq} and internal resistances r_{eq} whose values are given in Eqs. (3.54) and (3.55).

Next, consider a parallel combination of the cells (Fig. 3.14). I_1 and I_2 are the currents leaving the positive electrodes of the cells. At the point B_1, I_1 and I_2 flow in whereas the current I flows out. Since as much charge flows in as out, we have

$$I = I_1 + I_2 \tag{3.48}$$

Let $V(B_1)$ and $V(B_2)$ be the potentials at B_1 and B_2, respectively. Then, considering the first cell, the potential difference across its terminals is $V(B_1) - V(B_2)$. Hence, from Eq. (3.38)

$$V \equiv V(B_1) - V(B_2) = \varepsilon_1 - I_1 r_1 \tag{3.49}$$

Points B_1 and B_2 are connected exactly similarly to the second cell. Hence considering the second cell, we also have

$$V \equiv V(B_1) - V(B_2) = \varepsilon_2 - I_2 r_2 \tag{3.50}$$

Combining the last three equations

$$I = I_1 + I_2$$

$$= \frac{\varepsilon_1 - V}{r_1} + \frac{\varepsilon_2 - V}{r_2} = \left(\frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2}\right) - V\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \tag{3.51}$$

Hence, V is given by,

$$V = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} - I\frac{r_1 r_2}{r_1 + r_2} \tag{3.52}$$

If we want to replace the combination by a single cell, between B_1 and B_2, of emf ε_{eq} and internal resistance r_{eq}, we would have

$$V = \varepsilon_{eq} - I\, r_{eq} \tag{3.53}$$

The last two equations should be the same and hence

$$\varepsilon_{eq} = \frac{\varepsilon_1 r_2 + \varepsilon_2 r_1}{r_1 + r_2} \tag{3.54}$$

$$r_{eq} = \frac{r_1 r_2}{r_1 + r_2} \tag{3.55}$$

We can put these equations in a simpler way,

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + \frac{1}{r_2} \qquad (3.56)$$

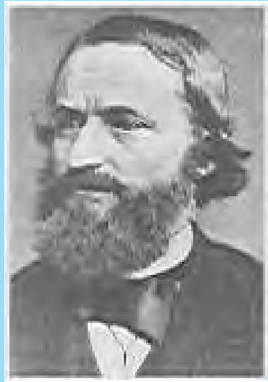
$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + \frac{\varepsilon_2}{r_2} \qquad (3.57)$$

In Fig. (3.14), we had joined the positive terminals together and similarly the two negative ones, so that the currents I_1, I_2 flow out of positive terminals. If the negative terminal of the second is connected to positive terminal of the first, Eqs. (3.56) and (3.57) would still be valid with $\varepsilon_2 \to -\varepsilon_2$

Equations (3.56) and (3.57) can be extended easily. If there are n cells of emf ε_1, . . . ε_n and of internal resistances r_1,... r_n respectively, connected in parallel, the combination is equivalent to a single cell of emf ε_{eq} and internal resistance r_{eq}, such that

$$\frac{1}{r_{eq}} = \frac{1}{r_1} + ... + \frac{1}{r_n} \qquad (3.58)$$

$$\frac{\varepsilon_{eq}}{r_{eq}} = \frac{\varepsilon_1}{r_1} + ... + \frac{\varepsilon_n}{r_n} \qquad (3.59)$$



Gustav Robert Kirchhoff (1824 – 1887) German physicist, professor at Heidelberg and at Berlin. Mainly known for his development of spectroscopy, he also made many important contributions to mathematical physics, among them, his first and second rules for circuits.

GUSTAV ROBERT KIRCHHOFF (1824 – 1887)

3.12 Kirchhoff's Rules

Electric circuits generally consist of a number of resistors and cells interconnected sometimes in a complicated way. The formulae we have derived earlier for series and parallel combinations of resistors are not always sufficient to determine all the currents and potential differences in the circuit. Two rules, called *Kirchhoff's rules*, are very useful for analysis of electric circuits.

Given a circuit, we start by labelling currents in each resistor by a symbol, say I, and a directed arrow to indicate that a current I flows along the resistor in the direction indicated. If ultimately I is determined to be positive, the actual current in the resistor is in the direction of the arrow. If I turns out to be negative, the current actually flows in a direction opposite to the arrow. Similarly, for each source (i.e., cell or some other source of electrical power) the positive and negative electrodes are labelled, as well as, a directed arrow with a symbol for the current flowing through the cell. This will tell us the potential difference, $V = V(\text{P}) - V(\text{N}) = \varepsilon - I\,r$ [Eq. (3.38) between the positive terminal P and the negative terminal N; I here is the current flowing from N to P through the cell]. If, while labelling the current I through the cell one goes from P to N, then of course

$$V = \varepsilon + I\,r \qquad (3.60)$$

Having clarified labelling, we now state the rules and the proof:

(a) *Junction rule: At any junction, the sum of the currents entering the junction is equal to the sum of currents leaving the junction* (Fig. 3.15).

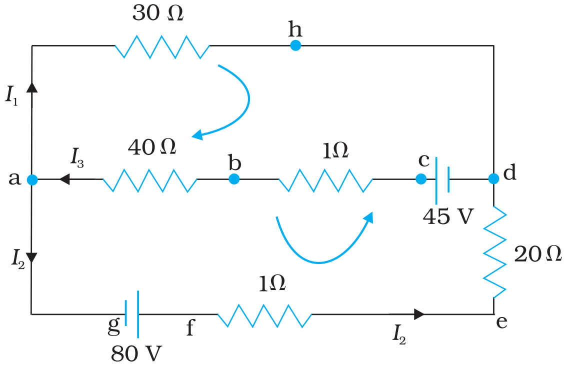


FIGURE 3.15 At junction a the current leaving is $I_1 + I_2$ and current entering is I_3. The junction rule says $I_3 = I_1 + I_2$. At point h current entering is I_1. There is only one current leaving h and by junction rule that will also be I_1. For the loops 'ahdcba' and 'ahdefga', the loop rules give $-30I_1 - 41\ I_3 + 45 = 0$ and $-30I_1 + 21\ I_2 - 80 = 0$.

This applies equally well if instead of a junction of several lines, we consider a point in a line.

The proof of this rule follows from the fact that when currents are steady, there is no accumulation of charges at any junction or at any point in a line. Thus, the total current flowing in, (which is the rate at which charge flows into the junction), must equal the total current flowing out.

(b) *Loop rule: The algebraic sum of changes in potential around any closed loop involving resistors and cells in the loop is zero* (Fig. 3.15).

This rule is also obvious, since electric potential is dependent on the location of the point. Thus starting with any point if we come back to the same point, the total change must be zero. In a closed loop, we do come back to the starting point and hence the rule.

Example 3.5 A battery of 10 V and negligible internal resistance is connected across the diagonally opposite corners of a cubical network consisting of 12 resistors each of resistance 1 Ω (Fig. 3.16). Determine the equivalent resistance of the network and the current along each edge of the cube.

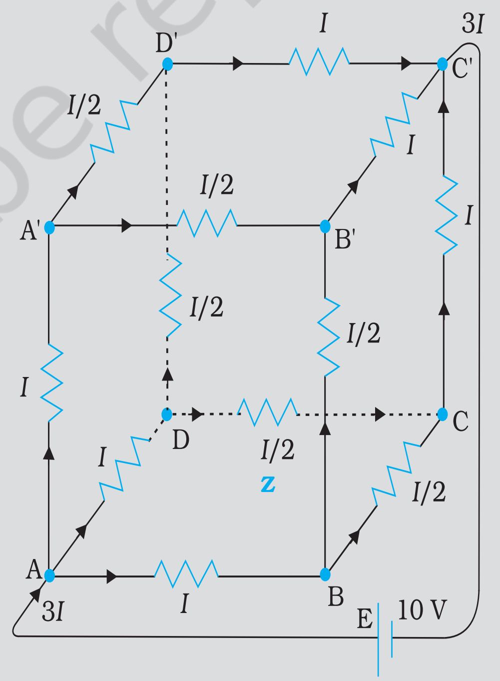


FIGURE 3.16

EXAMPLE 3.5

Solution The network is not reducible to a simple series and parallel combinations of resistors. There is, however, a clear symmetry in the problem which we can exploit to obtain the equivalent resistance of the network.

The paths AA′, AD and AB are obviously symmetrically placed in the network. Thus, the current in each must be the same, say, I. Further, at the corners A′, B and D, the incoming current I must split equally into the two outgoing branches. In this manner, the current in all the 12 edges of the cube are easily written down in terms of I, using Kirchhoff's first rule and the symmetry in the problem.

Next take a closed loop, say, ABCC′EA, and apply Kirchhoff's second rule:

$$-IR - (1/2)IR - IR + \varepsilon = 0$$

where R is the resistance of each edge and ε the emf of battery. Thus,

$$\varepsilon = \frac{5}{2} I R$$

The equivalent resistance R_{eq} of the network is

$$R_{eq} = \frac{\varepsilon}{3I} = \frac{5}{6} R$$

For $R = 1\ \Omega$, $R_{eq} = (5/6)\ \Omega$ and for $\varepsilon = 10$ V, the total current ($= 3I$) in the network is

$$3I = 10\ \text{V}/(5/6)\ \Omega = 12\ \text{A, i.e., } I = 4\ \text{A}$$

The current flowing in each edge can now be read off from the Fig. 3.16.

EXAMPLE 3.5

It should be noted that because of the symmetry of the network, the great power of Kirchhoff's rules has not been very apparent in Example 3.5. In a general network, there will be no such simplification due to symmetry, and only by application of Kirchhoff's rules to junctions and closed loops (as many as necessary to solve the unknowns in the network) can we handle the problem. This will be illustrated in Example 3.6.

Example 3.6 Determine the current in each branch of the network shown in Fig. 3.17.

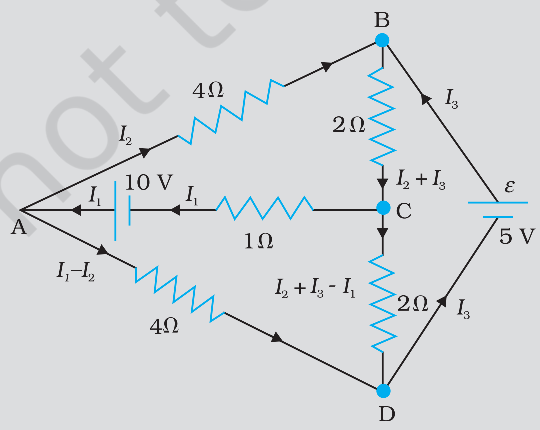


FIGURE 3.17

EXAMPLE 3.6

PHYSICS

Similation for application of Kirchhoff's rules:
http://www.phys.hawaii.edu/~teb/optics/java/kirch3/

Solution Each branch of the network is assigned an unknown current to be determined by the application of Kirchhoff's rules. To reduce the number of unknowns at the outset, the first rule of Kirchhoff is used at every junction to assign the unknown current in each branch. We then have three unknowns I_1, I_2 and I_3 which can be found by applying the second rule of Kirchhoff to three different closed loops. Kirchhoff's second rule for the closed loop ADCA gives,

$$10 - 4(I_1 - I_2) + 2(I_2 + I_3 - I_1) - I_1 = 0 \quad [3.61(a)]$$

that is, $7I_1 - 6I_2 - 2I_3 = 10$

For the closed loop ABCA, we get

$$10 - 4I_2 - 2\,(I_2 + I_3) - I_1 = 0$$

that is, $I_1 + 6I_2 + 2I_3 = 10$ [3.61(b)]

For the closed loop BCDEB, we get

$$5 - 2\,(I_2 + I_3) - 2\,(I_2 + I_3 - I_1) = 0$$

that is, $2I_1 - 4I_2 - 4I_3 = -5$ [3.61(c)]

Equations (3.61 a, b, c) are three simultaneous equations in three unknowns. These can be solved by the usual method to give

$$I_1 = 2.5\text{A}, \quad I_2 = \frac{5}{8}\ \text{A}, \quad I_3 = 1\frac{7}{8}\ \text{A}$$

The currents in the various branches of the network are

$$\text{AB}: \frac{5}{8}\ \text{A}, \quad \text{CA}: 2\frac{1}{2}\ \text{A}, \quad \text{DEB}: 1\frac{7}{8}\ \text{A}$$

$$\text{AD}: 1\frac{7}{8}\ \text{A}, \quad \text{CD}: 0\ \text{A}, \quad \text{BC}: 2\frac{1}{2}\ \text{A}$$

It is easily verified that Kirchhoff's second rule applied to the remaining closed loops does not provide any additional independent equation, that is, the above values of currents satisfy the second rule for every closed loop of the network. For example, the total voltage drop over the closed loop BADEB

$$5\ \text{V} + \left(\frac{5}{8}\times 4\right)\text{V} - \left(\frac{15}{8}\times 4\right)\text{V}$$

equal to zero, as required by Kirchhoff's second rule.

EXAMPLE 3.6

3.13 WHEATSTONE BRIDGE

As an application of Kirchhoff's rules consider the circuit shown in Fig. 3.18, which is called the *Wheatstone bridge*. The bridge has four resistors R_1, R_2, R_3 and R_4. Across one pair of diagonally opposite points (A and C in the figure) a source is connected. This (*i.e.*, AC) is called the battery arm. Between the other two vertices, B and D, a galvanometer G (which is a device to detect currents) is connected. This line, shown as BD in the figure, is called the galvanometer arm.

For simplicity, we assume that the cell has no internal resistance. In general there will be currents flowing across all the resistors as well as a current I_g through G. Of special interest, is the case of a *balanced* bridge where the resistors are such that $I_g = 0$. We can easily get the balance condition, such that there is no current through G. In this case, the Kirchhoff's junction rule applied to junctions D and B (see the figure) immediately gives us the relations $I_1 = I_3$ and $I_2 = I_4$. Next, we apply Kirchhoff's loop rule to closed loops ADBA and CBDC. The first loop gives

$$-I_1 R_1 + 0 + I_2 R_2 = 0 \qquad (I_g = 0) \tag{3.62}$$

and the second loop gives, upon using $I_3 = I_1$, $I_4 = I_2$

$$I_2 R_4 + 0 - I_1 R_3 = 0 \tag{3.63}$$

From Eq. (3.62), we obtain,

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}$$

whereas from Eq. (3.63), we obtain,

$$\frac{I_1}{I_2} = \frac{R_4}{R_3}$$

Hence, we obtain the condition

$$\frac{R_2}{R_1} = \frac{R_4}{R_3} \qquad [3.64(a)]$$

This last equation relating the four resistors is called the *balance condition* for the galvanometer to give zero or null deflection.

The Wheatstone bridge and its balance condition provide a practical method for determination of an unknown resistance. Let us suppose we have an unknown resistance, which we insert in the fourth arm; R_4 is thus not known. Keeping known resistances R_1 and R_2 in the first and second arm of the bridge, we go on varying R_3 till the galvanometer shows a null deflection. The bridge then is balanced, and from the balance condition the value of the unknown resistance R_4 is given by,

$$R_4 = R_3 \frac{R_2}{R_1} \qquad [3.64(b)]$$

A practical device using this principle is called the *meter bridge*.

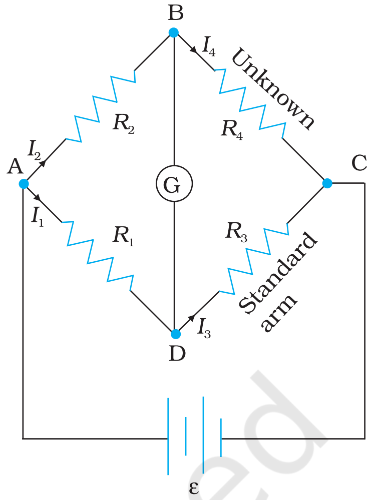


FIGURE 3.18

Example 3.7 The four arms of a Wheatstone bridge (Fig. 3.19) have the following resistances:

AB = 100Ω, BC = 10Ω, CD = 5Ω, and DA = 60Ω.

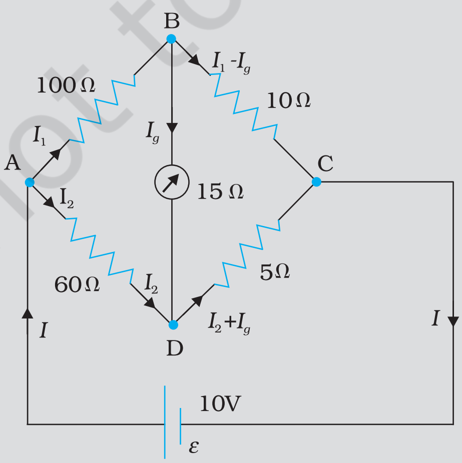


FIGURE 3.19

EXAMPLE 3.7

Reprint 2025-26

A galvanometer of 15Ω resistance is connected across BD. Calculate the current through the galvanometer when a potential difference of 10 V is maintained across AC.

Solution Considering the mesh BADB, we have

$$100I_1 + 15I_g - 60I_2 = 0$$

or $$20I_1 + 3I_g - 12I_2 = 0 \qquad [3.65(a)]$$

Considering the mesh BCDB, we have

$$10\,(I_1 - I_g) - 15I_g - 5\,(I_2 + I_g) = 0$$

$$10I_1 - 30I_g - 5I_2 = 0$$

$$2I_1 - 6I_g - I_2 = 0 \qquad [3.65(b)]$$

Considering the mesh ADCEA,

$$60I_2 + 5\,(I_2 + I_g) = 10$$

$$65I_2 + 5I_g = 10$$

$$13I_2 + I_g = 2 \qquad [3.65(c)]$$

Multiplying Eq. (3.65b) by 10

$$20I_1 - 60I_g - 10I_2 = 0 \qquad [3.65(d)]$$

From Eqs. (3.65d) and (3.65a) we have

$$63I_g - 2I_2 = 0$$

$$I_2 = 31.5I_g \qquad [3.65(e)]$$

Substituting the value of I_2 into Eq. [3.65(c)], we get

$$13\,(31.5I_g) + I_g = 2$$

$$410.5\,I_g = 2$$

$$I_g = 4.87 \text{ mA.}$$

EXAMPLE 3.7

SUMMARY

1. *Current* through a given area of a conductor is the net charge passing per unit time through the area.
2. To maintain a steady current, we must have a closed circuit in which an external agency moves electric charge from lower to higher potential energy. The work done per unit charge by the source in taking the charge from lower to higher potential energy (i.e., from one terminal of the source to the other) is called the electromotive force, or *emf*, of the source. Note that the emf is not a force; it is the voltage difference between the two terminals of a source in open circuit.
3. *Ohm's law*: The electric current I flowing through a substance is proportional to the voltage V across its ends, i.e., $V \propto I$ or $V = RI$, where R is called the *resistance* of the substance. The unit of resistance is ohm: $1\Omega = 1 \text{ V A}^{-1}$.

Reprint 2025-26

4. The *resistance* R of a conductor depends on its length l and cross-sectional area A through the relation,

$$R = \frac{\rho l}{A}$$

where ρ, called *resistivity* is a property of the material and depends on temperature and pressure.

5. *Electrical resistivity* of substances varies over a very wide range. Metals have low resistivity, in the range of 10^{-8} Ω m to 10^{-6} Ω m. Insulators like glass and rubber have 10^{22} to 10^{24} times greater resistivity. Semiconductors like Si and Ge lie roughly in the middle range of resistivity on a logarithmic scale.

6. In most substances, the carriers of current are electrons; in some cases, for example, ionic crystals and electrolytic liquids, positive and negative ions carry the electric current.

7. *Current density* **j** gives the amount of charge flowing per second per unit area normal to the flow,

$$\mathbf{j} = nq\, \mathbf{v}_d$$

where n is the number density (number per unit volume) of charge carriers each of charge q, and $\mathbf{v}_d$ is the *drift velocity* of the charge carriers. For electrons $q = -e$. If **j** is normal to a cross-sectional area **A** and is constant over the area, the magnitude of the current I through the area is $nev_d A$.

8. Using $E = V/l$, $I = nev_d A$, and Ohm's law, one obtains

$$\frac{eE}{m} = \rho \frac{ne^2}{m} v_d$$

The proportionality between the *force* eE on the electrons in a metal due to the external field E and the drift velocity v_d (not acceleration) can be understood, if we assume that the electrons suffer collisions with ions in the metal, which deflect them randomly. If such collisions occur on an average at a time interval τ,

$$v_d = a\tau = eE\tau/m$$

where a is the acceleration of the electron. This gives

$$\rho = \frac{m}{ne^2 \tau}$$

9. In the temperature range in which resistivity increases linearly with temperature, the *temperature coefficient of resistivity* α is defined as the fractional increase in resistivity per unit increase in temperature.

10. Ohm's law is obeyed by many substances, but it is not a fundamental law of nature. It fails if

(a) V depends on I non-linearly.

(b) the relation between V and I depends on the sign of V for the same absolute value of V.

(c) The relation between V and I is non-unique.

An example of (a) is when ρ increases with I (even if temperature is kept fixed). A rectifier combines features (a) and (b). GaAs shows the feature (c).

11. When a source of emf ε is connected to an external resistance R, the voltage V_{ext} across R is given by

$$V_{ext} = IR = \frac{\varepsilon}{R + r} R$$

where r is the *internal resistance* of the source.

12. *Kirchhoff's Rules* –
    (a) *Junction Rule:* At any junction of circuit elements, the sum of currents entering the junction must equal the sum of currents leaving it.
    (b) *Loop Rule:* The algebraic sum of changes in potential around any closed loop must be zero.
13. The *Wheatstone bridge* is an arrangement of four resistances – R_1, R_2, R_3, R_4 as shown in the text. The null-point condition is given by
    $$\frac{R_1}{R_2} = \frac{R_3}{R_4}$$
    using which the value of one resistance can be determined, knowing the other three resistances.

Physical Quantity	Symbol	Dimensions	Unit	Remark
Electric current	I	[A]	A	SI base unit
Charge	Q, q	[T A]	C	
Voltage, Electric potential difference	V	$[M\ L^2\ T^{-3}\ A^{-1}]$	V	Work/charge
Electromotive force	ε	$[M\ L^2\ T^{-3}\ A^{-1}]$	V	Work/charge
Resistance	R	$[M\ L^2\ T^{-3}\ A^{-2}]$	Ω	$R = V/I$
Resistivity	ρ	$[M\ L^3\ T^{-3}\ A^{-2}]$	Ω m	$R = \rho l/A$
Electrical conductivity	σ	$[M^{-1}\ L^{-3}\ T^3\ A^2]$	S	$\sigma = 1/\rho$
Electric field	$\mathbf{E}$	$[M\ L\ T^{-3}\ A^{-1}]$	$V\ m^{-1}$	$\frac{\text{Electric force}}{\text{charge}}$
Drift speed	v_d	$[L\ T^{-1}]$	$m\ s^{-1}$	$v_d = \frac{e\,E\,\tau}{m}$
Relaxation time	τ	[T]	s	
Current density	$\mathbf{j}$	$[L^{-2}\ A]$	$A\ m^{-2}$	current/area
Mobility	μ	$[M\ L^3\ T^{-4}\ A^{-1}]$	$m^2\ V^{-1} s^{-1}$	$v_d\ /E$

POINTS TO PONDER

1. Current is a scalar although we represent current with an arrow. Currents do not obey the law of vector addition. That current is a scalar also follows from it's definition. The current I through an area of cross-section is given by the scalar product of two vectors:
   $I = \mathbf{j} \cdot \Delta\mathbf{S}$
   where $\mathbf{j}$ and $\Delta\mathbf{S}$ are vectors.

Reprint 2025-26

2. Refer to *V-I* curves of a resistor and a diode as drawn in the text. A resistor obeys Ohm's law while a diode does not. The assertion that $V = IR$ is a statement of Ohm's law is not true. This equation defines resistance and it may be applied to all conducting devices whether they obey Ohm's law or not. The Ohm's law asserts that the plot of I versus V is linear i.e., R is independent of V.

   Equation $\mathbf{E} = \rho\, \mathbf{j}$ leads to another statement of *Ohm's* law, i.e., a conducting material obeys Ohm's law when the resistivity of the material does not depend on the magnitude and direction of applied electric field.
3. Homogeneous conductors like silver or semiconductors like pure germanium or germanium containing impurities obey Ohm's law within some range of electric field values. If the field becomes too strong, there are departures from Ohm's law in all cases.
4. Motion of conduction electrons in electric field $\mathbf{E}$ is the sum of (i) motion due to random collisions and (ii) that due to $\mathbf{E}$. The motion due to random collisions averages to zero and does not contribute to v_d (Chapter 10, Textbook of Class XI). v_d, thus is only due to applied electric field on the electron.
5. The relation $\mathbf{j} = \rho\, \mathbf{v}$ should be applied to each type of charge carriers separately. In a conducting wire, the total current and charge density arises from both positive and negative charges:

   $\mathbf{j} = \rho_+ \mathbf{v}_+ + \rho_- \mathbf{v}_-$

   $\boldsymbol{\rho} = \rho_+ + \rho_-$

   Now in a neutral wire carrying electric current,

   $\boldsymbol{\rho}_+ = -\rho_-$

   Further, $v_+ \sim 0$ which gives

   $\boldsymbol{\rho} = 0$

   $\mathbf{j} = \rho_- \mathbf{v}$

   Thus, the relation $\mathbf{j} = \rho\, \mathbf{v}$ does not apply to the total current charge density.
6. Kirchhoff's junction rule is based on conservation of charge and the outgoing currents add up and are equal to incoming current at a junction. Bending or reorienting the wire does not change the validity of Kirchhoff's junction rule.

EXERCISES

3.1 The storage battery of a car has an emf of 12 V. If the internal resistance of the battery is 0.4 Ω, what is the maximum current that can be drawn from the battery?

3.2 A battery of emf 10 V and internal resistance 3 Ω is connected to a resistor. If the current in the circuit is 0.5 A, what is the resistance of the resistor? What is the terminal voltage of the battery when the circuit is closed?

3.3 At room temperature (27.0 °C) the resistance of a heating element is 100 Ω. What is the temperature of the element if the resistance is found to be 117 Ω, given that the temperature coefficient of the material of the resistor is $1.70 \times 10^{-4}\ °C^{-1}$.

Reprint 2025-26

3.4 A negligibly small current is passed through a wire of length 15 m and uniform cross-section 6.0×10^{-7} m^2, and its resistance is measured to be 5.0 Ω. What is the resistivity of the material at the temperature of the experiment?

3.5 A silver wire has a resistance of 2.1 Ω at 27.5 °C, and a resistance of 2.7 Ω at 100 °C. Determine the temperature coefficient of resistivity of silver.

3.6 A heating element using nichrome connected to a 230 V supply draws an initial current of 3.2 A which settles after a few seconds to a steady value of 2.8 A. What is the steady temperature of the heating element if the room temperature is 27.0 °C? Temperature coefficient of resistance of nichrome averaged over the temperature range involved is 1.70×10^{-4} °C^{-1}.

3.7 Determine the current in each branch of the network shown in Fig. 3.20:

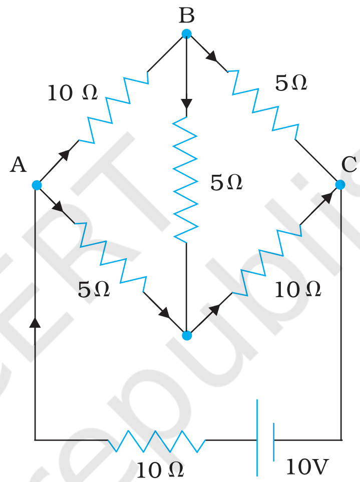


FIGURE 3.20

3.8 A storage battery of emf 8.0 V and internal resistance 0.5 Ω is being charged by a 120 V dc supply using a series resistor of 15.5 Ω. What is the terminal voltage of the battery during charging? What is the purpose of having a series resistor in the charging circuit?

3.9 The number density of free electrons in a copper conductor estimated in Example 3.1 is 8.5×10^{28} m^{-3}. How long does an electron take to drift from one end of a wire 3.0 m long to its other end? The area of cross-section of the wire is 2.0×10^{-6} m^2 and it is carrying a current of 3.0 A.

Chapter Four

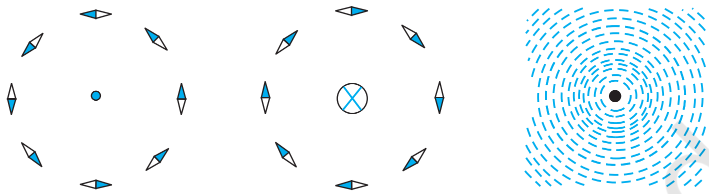
MOVING CHARGES AND MAGNETISM

4.1 Introduction

Both Electricity and Magnetism have been known for more than 2000 years. However, it was only about 200 years ago, in 1820, that it was realised that they were intimately related. During a lecture demonstration in the summer of 1820, Danish physicist Hans Christian Oersted noticed that a current in a straight wire caused a noticeable deflection in a nearby magnetic compass needle. He investigated this phenomenon. He found that the alignment of the needle is tangential to an imaginary circle which has the straight wire as its centre and has its plane perpendicular to the wire. This situation is depicted in Fig.4.1(a). It is noticeable when the current is large and the needle sufficiently close to the wire so that the earth's magnetic field may be ignored. Reversing the direction of the current reverses the orientation of the needle [Fig. 4.1(b)]. The deflection increases on increasing the current or bringing the needle closer to the wire. Iron filings sprinkled around the wire arrange themselves in concentric circles with the wire as the centre [Fig. 4.1(c)]. Oersted concluded that *moving charges or currents produced a magnetic field in the surrounding space.*

Following this, there was intense experimentation. In 1864, the laws obeyed by electricity and magnetism were unified and formulated by

James Maxwell who then realised that light was electromagnetic waves. Radio waves were discovered by Hertz, and produced by J.C.Bose and G. Marconi by the end of the 19th century. A remarkable scientific and technological progress took place in the 20th century. This was due to our increased understanding of electromagnetism and the invention of devices for production, amplification, transmission and detection of electromagnetic waves.



(a) (b) (c)

FIGURE 4.1 The magnetic field due to a straight long current-carrying wire. The wire is perpendicular to the plane of the paper. A ring of compass needles surrounds the wire. The orientation of the needles is shown when (a) the current emerges out of the plane of the paper, (b) the current moves into the plane of the paper. (c) The arrangement of iron filings around the wire. The darkened ends of the needle represent north poles. The effect of the earth's magnetic field is neglected.

HANS CHRISTIAN OERSTED (1777–1851)

Hans Christian Oersted (1777–1851) Danish physicist and chemist, professor at Copenhagen. He observed that a compass needle suffers a deflection when placed near a wire carrying an electric current. This discovery gave the first empirical evidence of a connection between electric and magnetic phenomena.

In this chapter, we will see how magnetic field exerts forces on moving charged particles, like electrons, protons, and current-carrying wires. We shall also learn how currents produce magnetic fields. We shall see how particles can be accelerated to very high energies in a cyclotron. We shall study how currents and voltages are detected by a galvanometer.

In this and subsequent Chapter on magnetism, we adopt the following convention: A current or a field (electric or magnetic) emerging out of the plane of the paper is depicted by a dot (⊙). A current or a field going into the plane of the paper is depicted by a cross (⊗)*. Figures. 4.1(a) and 4.1(b) correspond to these two situations, respectively.

4.2 Magnetic Force

4.2.1 Sources and fields

Before we introduce the concept of a magnetic field **B**, we shall recapitulate what we have learnt in Chapter 1 about the electric field **E.** We have seen that the interaction between two charges can be considered in two stages. The charge Q, the source of the field, produces an electric field **E**, where

* A dot appears like the tip of an arrow pointed at you, a cross is like the feathered tail of an arrow moving away from you.

$$\mathbf{E} = Q\ \hat{\mathbf{r}}\ /\ (4\pi\varepsilon_0)r^2 \qquad (4.1)$$

where $\hat{\mathbf{r}}$ is unit vector along $\mathbf{r}$, and the field $\mathbf{E}$ is a vector field. A charge q interacts with this field and experiences a force $\mathbf{F}$ given by

$$\mathbf{F} = q\ \mathbf{E} = q\,Q\ \hat{\mathbf{r}}\ /\ (4\pi\varepsilon_0)\ r^2 \qquad (4.2)$$

As pointed out in the Chapter 1, the field $\mathbf{E}$ is not just an artefact but has a physical role. It can convey energy and momentum and is not established instantaneously but takes finite time to propagate. The concept of a field was specially stressed by Faraday and was incorporated by Maxwell in his unification of electricity and magnetism. In addition to depending on each point in space, it can also vary with time, i.e., be a function of time. In our discussions in this chapter, we will assume that the fields do not change with time.

The field at a particular point can be due to one or more charges. If there are more charges the fields add vectorially. You have already learnt in Chapter 1 that this is called the principle of superposition. Once the field is known, the force on a test charge is given by Eq. (4.2).

Just as static charges produce an electric field, the currents or moving charges produce (in addition) a magnetic field, denoted by $\mathbf{B}\ (\mathbf{r})$, again a vector field. It has several basic properties identical to the electric field. It is defined at each point in space (and can in addition depend on time). Experimentally, it is found to obey the principle of superposition: *the magnetic field of several sources is the vector addition of magnetic field of each individual source.*

Hendrik Antoon Lorentz (1853 – 1928) Dutch theoretical physicist, professor at Leiden. He investigated the relationship between electricity, magnetism, and mechanics. In order to explain the observed effect of magnetic fields on emitters of light (Zeeman effect), he postulated the existence of electric charges in the atom, for which he was awarded the Nobel Prize in 1902. He derived a set of transformation equations (known after him, as Lorentz transformation equations) by some tangled mathematical arguments, but he was not aware that these equations hinge on a new concept of space and time.

HENDRIK ANTOON LORENTZ (1853 – 1928)

4.2.2 Magnetic Field, Lorentz Force

Let us suppose that there is a point charge q (moving with a velocity $\mathbf{v}$ and, located at $\mathbf{r}$ at a given time t) in presence of both the electric field $\mathbf{E}\ (\mathbf{r})$ and the magnetic field $\mathbf{B}\ (\mathbf{r})$. The force on an electric charge q due to both of them can be written as

$$\mathbf{F} = q\,[\,\mathbf{E}\ (\mathbf{r}) + \mathbf{v} \times \mathbf{B}\ (\mathbf{r})] \equiv \mathbf{F}_{\text{electric}} + \mathbf{F}_{\text{magnetic}} \qquad (4.3)$$

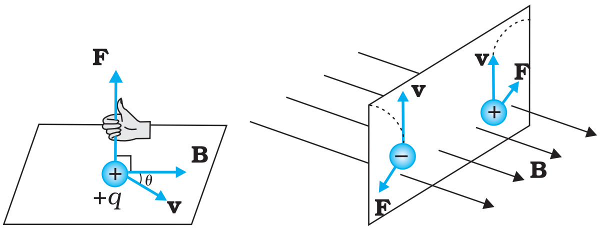
This force was given first by H.A. Lorentz based on the extensive experiments of Ampere and others. It is called the *Lorentz force*. You have already studied in detail the force due to the electric field. If we look at the interaction with the magnetic field, we find the following features.

(i) It depends on q, $\mathbf{v}$ and $\mathbf{B}$ (charge of the particle, the velocity and the magnetic field). *Force on a negative charge is opposite to that on a positive charge.*

(ii) The magnetic force $q\,[\,\mathbf{v} \times \mathbf{B}\,]$ includes a vector product of velocity and magnetic field. The vector product makes the force due to magnetic

Reprint 2025-26

field vanish (become zero) if velocity and magnetic field are parallel or anti-parallel. The force acts in a (sideways) direction perpendicular to both the velocity and the magnetic field. Its direction is given by the screw rule or right hand rule for vector (or cross) product as illustrated in Fig. 4.2.



(a) (b)

FIGURE 4.2 The direction of the magnetic force acting on a charged particle. (a) The force on a positively charged particle with velocity **v** and making an angle θ with the magnetic field **B** is given by the right-hand rule. (b) A moving charged particle q is deflected in an opposite sense to $-q$ in the presence of magnetic field.

(iii) The magnetic force is zero if charge is not moving (as then $|\mathbf{v}| = 0$). Only a moving charge feels the magnetic force.

The expression for the magnetic force helps us to define the unit of the magnetic field, if one takes q, **F** and **v**, all to be unity in the force equation $\mathbf{F} = q\,[\,\mathbf{v} \times \mathbf{B}] = q\,v\,B \sin\theta\ \hat{\mathbf{n}}$, where θ is the angle between **v** and **B** [see Fig. 4.2 (a)]. The magnitude of magnetic field B is 1 SI unit, when the force acting on a unit charge (1 C), moving perpendicular to **B** with a speed 1m/s, is one newton.

Dimensionally, we have $[B] = [F/qv]$ and the unit of **B** are Newton second / (coulomb metre). This unit is called *tesla* (T) named after Nikola Tesla (1856 – 1943). Tesla is a rather large unit. A smaller unit (non-SI) called *gauss* ($=10^{-4}$ tesla) is also often used. The earth's magnetic field is about 3.6×10^{-5} T.

4.2.3 Magnetic force on a current-carrying conductor

We can extend the analysis for force due to magnetic field on a single moving charge to a straight rod carrying current. Consider a rod of a uniform cross-sectional area A and length l. We shall assume one kind of mobile carriers as in a conductor (here electrons). Let the number density of these mobile charge carriers in it be n. Then the total number of mobile charge carriers in it is nlA. For a steady current I in this conducting rod, we may assume that each mobile carrier has an average drift velocity $\mathbf{v}_d$ (see Chapter 3). In the presence of an external magnetic field **B**, the force on these carriers is:

$$\mathbf{F} = (nlA)q\,\mathbf{v}_d \times \mathbf{B}$$

where q is the value of the charge on a carrier. Now $nq\,\mathbf{v}_d$ is the current density **j** and $|(nq\,\mathbf{v}_d)|A$ is the current I (see Chapter 3 for the discussion of current and current density). Thus,

$$\mathbf{F} = [(nq\,\mathbf{v}_d)lA] \times \mathbf{B} = [\,\mathbf{j}Al\,] \times \mathbf{B}$$

$$= I\boldsymbol{l} \times \mathbf{B} \tag{4.4}$$

where $\boldsymbol{l}$ is a vector of magnitude l, the length of the rod, and with a direction identical to the current I. Note that the current I is not a vector. In the last step leading to Eq. (4.4), we have transferred the vector sign from **j** to $\boldsymbol{l}$.

Equation (4.4) holds for a straight rod. In this equation, **B** is the external magnetic field. It is not the field produced by the current-carrying rod. If the wire has an arbitrary shape we can calculate the Lorentz force on it by considering it as a collection of linear strips $\mathrm{d}\boldsymbol{l}_j$ and summing

$$\mathbf{F} = \sum_j I\mathrm{d}\boldsymbol{l}_j \times \mathbf{B}$$

This summation can be converted to an integral in most cases.

Example 4.1 A straight wire of mass 200 g and length 1.5 m carries a current of 2 A. It is suspended in mid-air by a uniform horizontal magnetic field **B** (Fig. 4.3). What is the magnitude of the magnetic field?

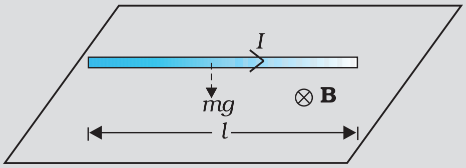


FIGURE 4.3

Solution From Eq. (4.4), we find that there is an upward force **F**, of magnitude IlB,. For mid-air suspension, this must be balanced by the force due to gravity:

$$m\,g = I\,lB$$

$$B = \frac{m\,g}{I\,l}$$

$$= \frac{0.2 \times 9.8}{2 \times 1.5} = 0.65 \text{ T}$$

Note that it would have been sufficient to specify m/l, the mass per unit length of the wire. The earth's magnetic field is approximately 4×10^{-5} T and we have ignored it.

EXAMPLE 4.1

Example 4.2 If the magnetic field is parallel to the positive y-axis and the charged particle is moving along the positive x-axis (Fig. 4.4), which way would the Lorentz force be for (a) an electron (negative charge), (b) a proton (positive charge).

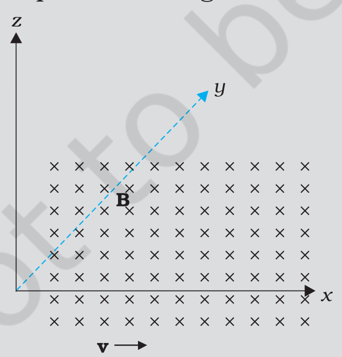


FIGURE 4.4

Solution The velocity **v** of particle is along the x-axis, while **B**, the magnetic field is along the y-axis, so **v × B** is along the z-axis (screw rule or right-hand thumb rule). So, (a) for electron it will be along $-z$ axis. (b) for a positive charge (proton) the force is along $+z$ axis.

EXAMPLE 4.2

PHYSICS

Charged particles moving in a magnetic field.
Interactive demonstration:
http://www.phys.hawaii.edu/~teb/optics/java/partmagn/index.html

4.3 Motion in a Magnetic Field

We will now consider, in greater detail, the motion of a charge moving in a magnetic field. We have learnt in Mechanics (see Class XI book, Chapter 5) that a force on a particle does work if the force has a component along (or opposed to) the direction of motion of the particle. In the case of motion of a charge in a magnetic field, the magnetic force is perpendicular to the velocity of the particle. So no work is done and no change in the magnitude of the velocity is produced (though the direction of momentum may be changed). [Notice that this is unlike the force due to an electric field, $q\mathbf{E}$, which *can* have a component parallel (or antiparallel) to motion and thus can transfer energy in addition to momentum.]

We shall consider motion of a charged particle in a *uniform* magnetic field. First consider the case of **v** perpendicular to **B**. The perpendicular force, $q\,\mathbf{v} \times \mathbf{B}$, acts as a centripetal force and produces a circular motion perpendicular to the magnetic field. *The particle will describe a circle if* **v** *and* **B** *are perpendicular to each other* (Fig. 4.5).

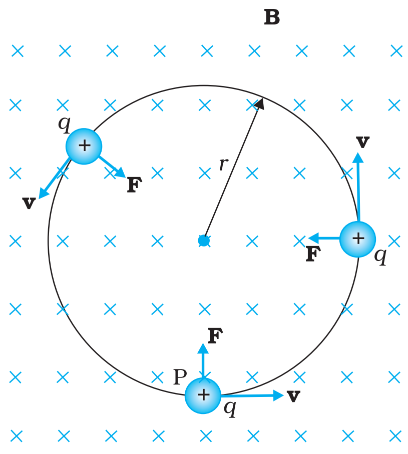


FIGURE 4.5 Circular motion

If velocity has a component along **B**, this component remains unchanged as the motion along the magnetic field will not be affected by the magnetic field. The motion in a plane perpendicular to **B** is as before a circular one, thereby producing a *helical motion* (Fig. 4.6).

You have already learnt in earlier classes (See Class XI, Chapter 3) that if r is the radius of the circular path of a particle, then a force of $m\,v^2 / r$, acts perpendicular to the path towards the centre of the circle, and is called the centripetal force. If the velocity **v** is perpendicular to the magnetic field **B**, the magnetic force is perpendicular to both **v** and **B** and acts like a centripetal force. It has a magnitude $q\ v\ B$. Equating the two expressions for centripetal force,

$$m\,v^2/r = q\,v\,B, \text{ which gives}$$

$$r = m\,v\ /\ qB \tag{4.5}$$

for the radius of the circle described by the charged particle. The larger the momentum, the larger is the radius and bigger the circle described. If ω is the angular frequency, then $v = \omega\ r$. So,

$$\omega = 2\pi\ \nu = q\,B/\ m \tag{4.6(a)}$$

which is independent of the velocity or energy . Here ν is the frequency of rotation. The independence of ν from energy has important application in the design of a cyclotron.

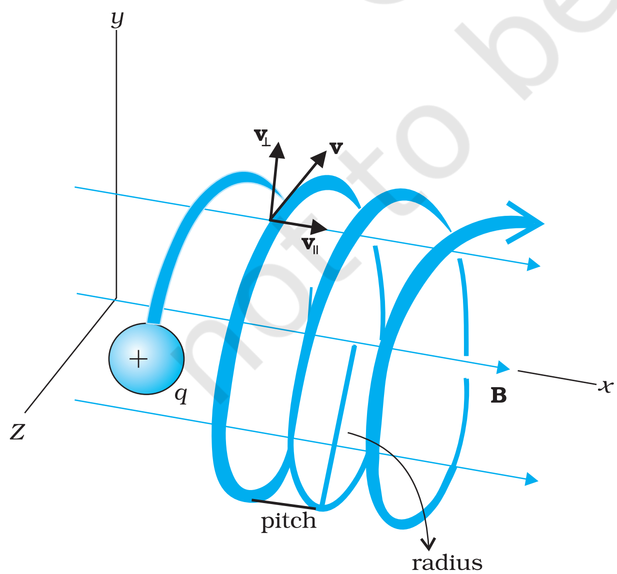


FIGURE 4.6 Helical motion

The time taken for one revolution is $T = 2\pi/\omega \equiv 1/\nu$. If there is a component of the velocity parallel to the magnetic field (denoted by $v_{||}$), it will make the particle move along the field and the path of the particle would be a helical one (Fig. 4.6). The distance moved along the magnetic field in one rotation is called pitch p. Using Eq. [4.6 (a)], we have

$$p = v_{||}T = 2\pi m\, v_{||} \,/\, q\,B \qquad [4.6(b)]$$

The radius of the circular component of motion is called the *radius* of the *helix*.

Example 4.3 What is the radius of the path of an electron (mass 9×10^{-31} kg and charge 1.6×10^{-19} C) moving at a speed of 3×10^7 m/s in a magnetic field of 6×10^{-4} T perpendicular to it? What is its frequency? Calculate its energy in keV. (1 eV = 1.6×10^{-19} J).

Solution Using Eq. (4.5) we find

$r = m\,v \,/\, (qB) = 9 \times 10^{-31}\ \text{kg} \times 3 \times 10^{7}\ \text{m s}^{-1} \,/\, (1.6 \times 10^{-19}\ \text{C} \times 6 \times 10^{-4}\ \text{T})$
$= 28 \times 10^{-2}$ m = 28 cm

$\nu = v \,/\, (2\,\pi r) = 17\times10^{6}\ \text{s}^{-1} = 17\times10^{6}$ Hz $=17$ MHz.

$E = (½)mv^2 = (½)\ 9 \times 10^{-31}\ \text{kg} \times 9 \times 10^{14}\ \text{m}^2/\text{s}^2 = 40.5 \times10^{-17}$ J
$\approx 4\times10^{-16}$ J = 2.5 keV.

EXAMPLE 4.3

4.4 Magnetic Field due to a Current Element, Biot-Savart Law

All magnetic fields that we know are due to currents (or moving charges) and due to intrinsic magnetic moments of particles. Here, we shall study the relation between current and the magnetic field it produces. It is given by the Biot-Savart's law. Fig. 4.7 shows a finite conductor XY carrying current I. Consider an infinitesimal element d$\boldsymbol{l}$ of the conductor. The magnetic field dB due to this element is to be determined at a point P which is at a distance r from it. Let θ be the angle between d$\boldsymbol{l}$ and the displacement vector **r**. According to Biot-Savart's law, the magnitude of the magnetic field dB is proportional to the current I, the element length |d$\boldsymbol{l}$|, and inversely proportional to the square of the distance r. Its direction* is perpendicular to the plane containing d$\boldsymbol{l}$ and **r**. Thus, in vector notation,

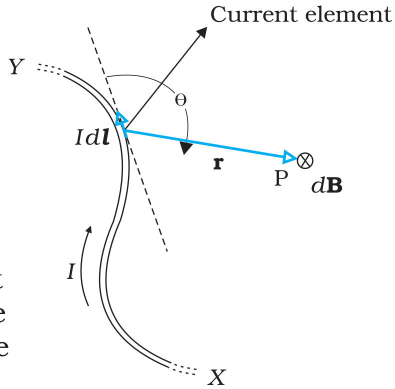


FIGURE 4.7 Illustration of the Biot-Savart law. The current element I d$\boldsymbol{l}$ produces a field dB at a distance r. The ⊗ sign indicates that the field is perpendicular to the plane of this page and directed into it.

$$d\mathbf{B} \propto \frac{I\, d\boldsymbol{l} \times \mathbf{r}}{r^3}$$

$$= \frac{\mu_0}{4\pi}\frac{I\, d\boldsymbol{l} \times \mathbf{r}}{r^3} \qquad [4.7(a)]$$

where $\mu_0/4\pi$ is a constant of proportionality. The above expression holds when the medium is vacuum.

* The sense of d$\boldsymbol{l}$ × **r** is also given by the *Right Hand Screw rule* : Look at the plane containing vectors d$\boldsymbol{l}$ and **r**. Imagine moving from the first vector towards second vector. If the movement is anticlockwise, the resultant is towards you. If it is clockwise, the resultant is away from you.

The magnitude of this field is,

$$|d\mathbf{B}| = \frac{\mu_0}{4\pi} \frac{I\,dl \sin\theta}{r^2} \quad [4.7(b)]$$

where we have used the property of cross-product. Equation [4.7 (a)] constitutes our basic equation for the magnetic field. The proportionality constant in SI units has the exact value,

$$\frac{\mu_0}{4\pi} = 10^{-7}\ \mathrm{Tm/A} \quad [4.7(c)]$$

We call μ_0 the *permeability* of free space (or vacuum).

The Biot-Savart law for the magnetic field has certain similarities, as well as, differences with the Coulomb's law for the electrostatic field. Some of these are:

(i) Both are long range, since both depend inversely on the square of distance from the source to the point of interest. The principle of superposition applies to both fields. [In this connection, note that the magnetic field is *linear* in the *source* $I\,d\boldsymbol{l}$ just as the electrostatic field is linear in its source: the electric charge.]

(ii) The electrostatic field is produced by a scalar source, namely, the electric charge. The magnetic field is produced by a vector source $I\,d\boldsymbol{l}$.

(iii) The electrostatic field is along the displacement vector joining the source and the field point. The magnetic field is perpendicular to the plane containing the displacement vector $\mathbf{r}$ and the current element $I\,d\boldsymbol{l}$.

(iv) There is an angle dependence in the Biot-Savart law which is not present in the electrostatic case. In Fig. 4.7, the magnetic field at any point in the direction of $d\boldsymbol{l}$ (the dashed line) is zero. Along this line, $\theta = 0$, $\sin\theta = 0$ and from Eq. [4.7(a)], $|d\mathbf{B}| = 0$.

There is an interesting relation between ε_0, the permittivity of free space; μ_0, the permeability of free space; and c, the speed of light in vacuum:

$$\varepsilon_0\mu_0 = (4\pi\varepsilon_0)\frac{\mu_0}{4\pi} = \frac{1}{9\times10^9}\left(10^{-7}\right) = \frac{1}{(3\times10^8)^2} = \frac{1}{c^2}$$

We will discuss this connection further in Chapter 8 on the electromagnetic waves. Since the speed of light in vacuum is constant, the product $\mu_0\varepsilon_0$ is fixed in magnitude. Choosing the value of either ε_0 or μ_0, fixes the value of the other. In SI units, μ_0 is fixed to be equal to $4\pi \times 10^{-7}$ in magnitude.

Example 4.4 An element $\Delta\boldsymbol{l} = \Delta x\,\hat{\mathbf{i}}$ is placed at the origin and carries a large current $I = 10$ A (Fig. 4.8). What is the magnetic field on the y-axis at a distance of 0.5 m. $\Delta x = 1$ cm.

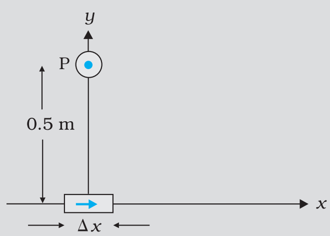


FIGURE 4.8

EXAMPLE 4.4

Solution

$$|\,d\mathbf{B}\,| = \frac{\mu_0}{4\pi}\frac{I\,dl\sin\theta}{r^2} \quad \text{[using Eq. (4.7)]}$$

$dl = \Delta x = 10^{-2}\,\text{m}\,,\; I = 10\text{ A},\;\; r = 0.5\text{ m} = y,\;\; \mu_0/4\pi = 10^{-7}\,\dfrac{\text{T m}}{\text{A}}$

$\theta = 90°\;;\; \sin\,\theta = 1$

$$|d\mathbf{B}| = \frac{10^{-7}\times 10\times 10^{-2}}{25\times 10^{-2}} = 4\times 10^{-8}\,\text{T}$$

The direction of the field is in the +z-direction. This is so since,

$$d\boldsymbol{l}\times\mathbf{r} = \Delta x\,\hat{\mathbf{i}}\times y\,\hat{\mathbf{j}} = y\,\Delta x\left(\hat{\mathbf{i}}\times\hat{\mathbf{j}}\right) = y\,\Delta x\,\hat{\mathbf{k}}$$

We remind you of the following cyclic property of cross-products,

$\hat{\mathbf{i}}\times\hat{\mathbf{j}} = \hat{\mathbf{k}}\,;\; \hat{\mathbf{j}}\times\hat{\mathbf{k}} = \hat{\mathbf{i}}\,;\; \hat{\mathbf{k}}\times\hat{\mathbf{i}} = \hat{\mathbf{j}}$

Note that the field is small in magnitude.

EXAMPLE 4.4

In the next section, we shall use the Biot-Savart law to calculate the magnetic field due to a circular loop.

4.5 MAGNETIC FIELD ON THE AXIS OF A CIRCULAR CURRENT LOOP

In this section, we shall evaluate the magnetic field due to a circular coil along its axis. The evaluation entails summing up the effect of infinitesimal current elements ($I\,d\boldsymbol{l}$) mentioned in the previous section. We assume that the current I is steady and that the evaluation is carried out in free space (i.e., vacuum).

Fig. 4.9 depicts a circular loop carrying a steady current I. The loop is placed in the y-z plane with its centre at the origin O and has a radius R. The x-axis is the axis of the loop. We wish to calculate the magnetic field at the point P on this axis. Let x be the distance of P from the centre O of the loop.

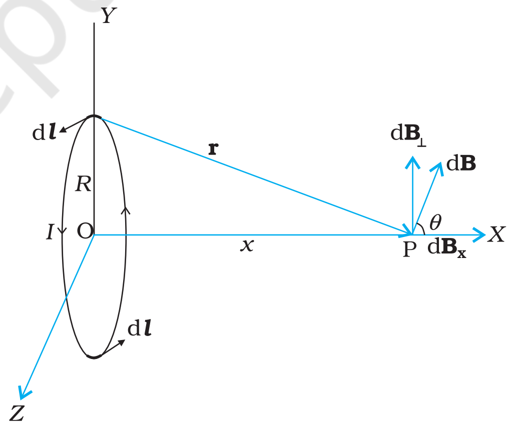


FIGURE 4.9 Magnetic field on the axis of a current carrying circular loop of radius R. Shown are the magnetic field $d\mathbf{B}$ (due to a line element $d\boldsymbol{l}$) and its components along and perpendicular to the axis.

Consider a conducting element $d\boldsymbol{l}$ of the loop. This is shown in Fig. 4.9. The magnitude dB of the magnetic field due to $d\boldsymbol{l}$ is given by the Biot-Savart law [Eq. 4.7(a)],

$$dB = \frac{\mu_0}{4\pi}\frac{I\left|d\boldsymbol{l}\times\mathbf{r}\right|}{r^3} \tag{4.8}$$

Now $r^2 = x^2 + R^2$. Further, any element of the loop will be perpendicular to the displacement vector from the element to the axial point. For example, the element $d\boldsymbol{l}$ in Fig. 4.9 is in the y-z plane, whereas, the displacement vector $\mathbf{r}$ from $d\boldsymbol{l}$ to the axial point P is in the x–y plane. Hence $|\,d\boldsymbol{l}\times\mathbf{r}\,| = r\,dl$. Thus,

$$dB = \frac{\mu_0}{4\pi}\frac{I dl}{\left(x^2 + R^2\right)} \tag{4.9}$$

Reprint 2025-26

The direction of dB is shown in Fig. 4.9. It is perpendicular to the plane formed by d$\boldsymbol{l}$ and **r**. It has an x-component d$\mathbf{B}_x$ and a component perpendicular to x-axis, d$\mathbf{B}_\perp$. When the components perpendicular to the x-axis are summed over, they cancel out and we obtain a null result. For example, the d$\mathbf{B}_\perp$ component due to d$\boldsymbol{l}$ is cancelled by the contribution due to the diametrically opposite d$\boldsymbol{l}$ element, shown in Fig. 4.9. Thus, only the x-component survives. The net contribution along x-direction can be obtained by integrating $\mathrm{d}B_x = \mathrm{d}B\cos\theta$ over the loop. For Fig. 4.9,

$$\cos\theta = \frac{R}{(x^2+R^2)^{1/2}} \tag{4.10}$$

From Eqs. (4.9) and (4.10),

$$\mathrm{d}B_x = \frac{\mu_0 I\mathrm{d}l}{4\pi}\frac{R}{\left(x^2+R^2\right)^{3/2}}$$

The summation of elements dl over the loop yields $2\pi R$, the circumference of the loop. Thus, the magnetic field at P due to entire circular loop is

$$\mathbf{B} = B_x\hat{\mathbf{i}} = \frac{\mu_0 I R^2}{2\left(x^2+R^2\right)^{3/2}}\hat{\mathbf{i}} \tag{4.11}$$

As a special case of the above result, we may obtain the field at the centre of the loop. Here $x = 0$, and we obtain,

$$\mathbf{B}_0 = \frac{\mu_0 I}{2R}\hat{\mathbf{i}} \tag{4.12}$$

The magnetic field lines due to a circular wire form closed loops and are shown in Fig. 4.10. The direction of the magnetic field is given by (another) *right-hand thumb rule* stated below:

Curl the palm of your right hand around the circular wire with the fingers pointing in the direction of the current. The right-hand thumb gives the direction of the magnetic field.

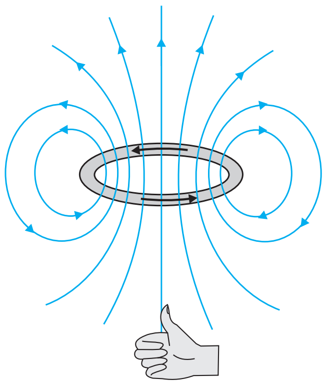


FIGURE 4.10 The magnetic field lines for a current loop. The direction of the field is given by the right-hand thumb rule described in the text. The upper side of the loop may be thought of as the north pole and the lower side as the south pole of a magnet.

Example 4.5 A straight wire carrying a current of 12 A is bent into a semi-circular arc of radius 2.0 cm as shown in Fig. 4.11(a). Consider the magnetic field **B** at the centre of the arc. (a) What is the magnetic field due to the straight segments? (b) In what way the contribution to **B** from the semicircle differs from that of a circular loop and in what way does it resemble? (c) Would your answer be different if the wire were bent into a semi-circular arc of the same radius but in the opposite way as shown in Fig. 4.11(b)?

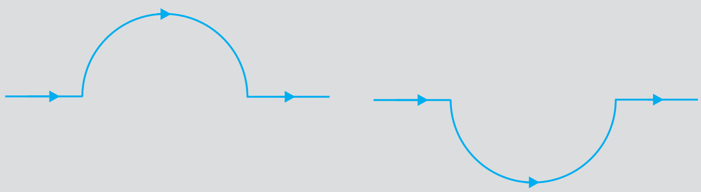


FIGURE 4.11

Solution

(a) d**l** and **r** for each element of the straight segments are parallel. Therefore, d$\boldsymbol{l}$ × **r** = 0. Straight segments do not contribute to |**B**|.

(b) For all segments of the semicircular arc, d$\boldsymbol{l}$ × **r** are all parallel to each other (into the plane of the paper). All such contributions add up in magnitude. Hence direction of **B** for a semicircular arc is given by the right-hand rule and magnitude is half that of a circular loop. Thus **B** is 1.9×10^{-4} T normal to the plane of the paper going into it.

(c) Same magnitude of **B** but opposite in direction to that in (b).

EXAMPLE 4.5

Example 4.6 Consider a tightly wound 100 turn coil of radius 10 cm, carrying a current of 1 A. What is the magnitude of the magnetic field at the centre of the coil?

Solution Since the coil is tightly wound, we may take each circular element to have the same radius R = 10 cm = 0.1 m. The number of turns N = 100. The magnitude of the magnetic field is,

$$B = \frac{\mu_0 NI}{2R} = \frac{4\pi \times 10^{-7} \times 10^2 \times 1}{2 \times 10^{-1}} = 2\pi \times 10^{-4} = 6.28 \times 10^{-4} \text{ T}$$

EXAMPLE 4.6

4.6 Ampere's Circuital Law

There is an alternative and appealing way in which the Biot-Savart law may be expressed. Ampere's circuital law considers an open surface with a boundary (Fig. 4.12). The surface has current passing through it. We consider the boundary to be made up of a number of small line elements. Consider one such element of length *dl*. We take the value of the tangential component of the magnetic field, B_t, at this element and multiply it by the

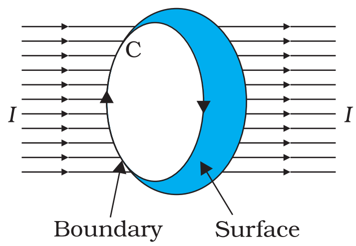


FIGURE 4.12

Reprint 2025-26

Andre Ampere (1775 – 1836) Andre Marie Ampere was a French physicist, mathematician and chemist who founded the science of electrodynamics. Ampere was a child prodigy who mastered advanced mathematics by the age of 12. Ampere grasped the significance of Oersted's discovery. He carried out a large series of experiments to explore the relationship between current electricity and magnetism. These investigations culminated in 1827 with the publication of 'Mathematical Theory of Electrodynamic Phenomena Deduced Solely from Experiments'. He hypothesised that *all* magnetic phenomena are due to circulating electric currents. Ampere was humble and absent-minded. He once forgot an invitation to dine with the Emperor Napoleon. He died of pneumonia at the age of 61. His gravestone bears the epitaph: *Tandem Felix* (Happy at last).

ANDRE AMPERE (1775 –1836)

length of that element dl [Note: $B_t dl = \mathbf{B}\cdot d\boldsymbol{l}$]. All such products are added together. We consider the limit as the lengths of elements get smaller and their number gets larger. The sum then tends to an integral. Ampere's law states that this integral is equal to μ_0 times the total current passing through the surface, i.e.,

$$\oint \mathbf{B}\cdot d\boldsymbol{l} = \mu_0 I \qquad [4.13(a)]$$

where I is the total current through the surface. The integral is taken over the closed loop coinciding with the boundary C of the surface. The relation above involves a sign-convention, given by the right-hand rule. Let the fingers of the right-hand be curled in the sense the boundary is traversed in the loop integral $\oint \mathbf{B}\cdot d\boldsymbol{l}$. Then the direction of the thumb gives the sense in which the current I is regarded as positive.

For several applications, a much simplified version of Eq. [4.13(a)] proves sufficient. We shall assume that, in such cases, it is possible to choose the loop (called an *amperian loop*) such that at each point of the loop, *either*

(i) **B** is tangential to the loop and is a non-zero *constant* B, *or*

(ii) **B** is normal to the loop, *or*

(iii) **B** vanishes.

Now, let L be the length (part) of the loop for which **B** is tangential. Let I_e be the current enclosed by the loop. Then, Eq. (4.13) reduces to,

$$BL = \mu_0 I_e \qquad [4.13(b)]$$

When there is a system with a symmetry such as for a *straight infinite current-carrying wire* in Fig. 4.13, the Ampere's law enables an easy evaluation of the magnetic field, much the same way Gauss' law helps in determination of the electric field. This is exhibited in the Example 4.8 below. The boundary of the loop chosen is a circle and magnetic field is tangential to the circumference of the circle. The law gives, for the left hand side of Eq. [4.13 (b)], B. $2\pi r$. We find that the magnetic field at a distance r outside the wire is *tangential* and given by

$$B \times 2\pi r = \mu_0 I,$$

$$B = \mu_0 I/(2\pi r) \qquad (4.14)$$

The above result for the infinite wire is interesting from several points of view.

(i) It implies that the field at every point on a circle of radius r, (with the wire along the axis), is same in magnitude. In other words, the magnetic field possesses what is called a *cylindrical symmetry*. The field that normally can depend on three coordinates depends only on one: r. Whenever there is symmetry, the solutions simplify.

(ii) The field direction at any point on this circle is tangential to it. Thus, the lines of constant magnitude of magnetic field form concentric circles. Notice now, in Fig. 4.1(c), the iron filings form concentric circles. These lines called *magnetic field lines* form closed loops. This is unlike the electrostatic field lines which originate from positive charges and end at negative charges. The expression for the magnetic field of a straight wire provides a theoretical justification to Oersted's experiments.

(iii) Another interesting point to note is that even though the wire is infinite, the field due to it at a non-zero distance is *not* infinite. It tends to blow up only when we come very close to the wire. The field is directly proportional to the current and inversely proportional to the distance from the (infinitely long) current source.

(iv) There exists a simple rule to determine the direction of the magnetic field due to a long wire. This rule, called the *right-hand rule**, is:

Grasp the wire in your right hand with your extended thumb pointing in the direction of the current. Your fingers will curl around in the direction of the magnetic field.

Ampere's circuital law is not new in content from Biot-Savart law. Both relate the magnetic field and the current, and both express the same physical consequences of a steady electrical current. Ampere's law is to Biot-Savart law, what Gauss's law is to Coulomb's law. Both, Ampere's and Gauss's law relate a physical quantity on the periphery or boundary (magnetic or electric field) to another physical quantity, namely, the source, in the interior (current or charge). We also note that Ampere's circuital law holds for steady currents which do not fluctuate with time. The following example will help us understand what is meant by the term *enclosed* current.

Example 4.7 Figure 4.13 shows a long straight wire of a circular cross-section (radius a) carrying steady current I. The current I is uniformly distributed across this cross-section. Calculate the magnetic field in the region $r < a$ and $r > a$.

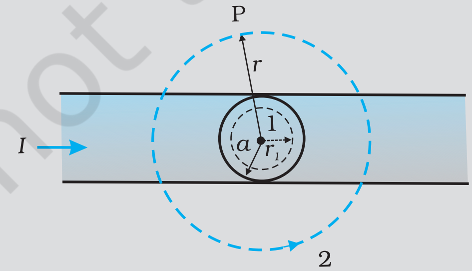


FIGURE 4.13

EXAMPLE 4.7

* Note that there are *two distinct* right-hand rules: One which gives the direction of **B** on the axis of current-loop and the other which gives direction of **B** for a straight conducting wire. Fingers and thumb play different roles in the two.

Solution (a) Consider the case $r > a$. The Amperian loop, labelled 2, is a circle concentric with the cross-section. For this loop,

$L = 2\pi r$

I_e = Current enclosed by the loop = I

The result is the familiar expression for a long straight wire

$B(2\pi r) = \mu_0 I$

$$B = \frac{\mu_0 I}{2\pi r} \qquad [4.15(a)]$$

$$B \propto \frac{1}{r} \qquad (r > a)$$

Now the current enclosed I_e is not I, but is less than this value. Since the current distribution is uniform, the current enclosed is,

$$I_e = I\left(\frac{\pi r^2}{\pi a^2}\right) = \frac{Ir^2}{a^2}$$

Using Ampere's law, $B(2\pi r) = \mu_0 \dfrac{I r^2}{a^2}$

$$B = \left(\frac{\mu_0 I}{2\pi a^2}\right) r \qquad [4.15(b)]$$

$$B \propto r \qquad (r < a)$$

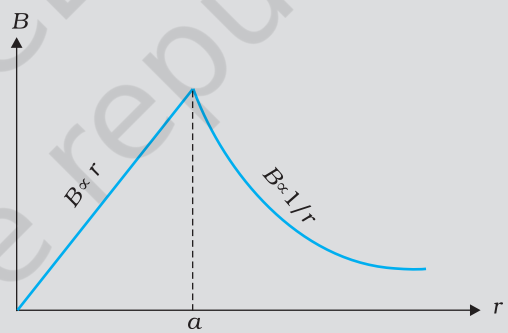


FIGURE 4.14

Figure (4.14) shows a plot of the magnitude of **B** with distance r from the centre of the wire. The direction of the field is tangential to the respective circular loop (1 or 2) and given by the right-hand rule described earlier in this section.

This example possesses the required symmetry so that Ampere's law can be applied readily.

EXAMPLE 4.7

It should be noted that while Ampere's circuital law holds for any loop, it may not always facilitate an evaluation of the magnetic field in every case. For example, for the case of the circular loop discussed in Section 4.5, it cannot be applied to extract the simple expression $B = \mu_0 I/2R$ [Eq. (4.12)] for the field at the centre of the loop. However, there exists a large number of situations of high symmetry where the law can be conveniently applied. We shall use it in the next section to calculate the magnetic field produced by a commonly used and very useful magnetic system: the *solenoid*.

4.7 The Solenoid

We shall discuss a long solenoid. By long solenoid we mean that the solenoid's length is large compared to its radius. It consists of a long wire wound in the form of a helix where the neighbouring turns are closely spaced. So each turn can be regarded as a circular loop. The net magnetic field is the vector sum of the fields due to all the turns. Enamelled wires are used for winding so that turns are insulated from each other.

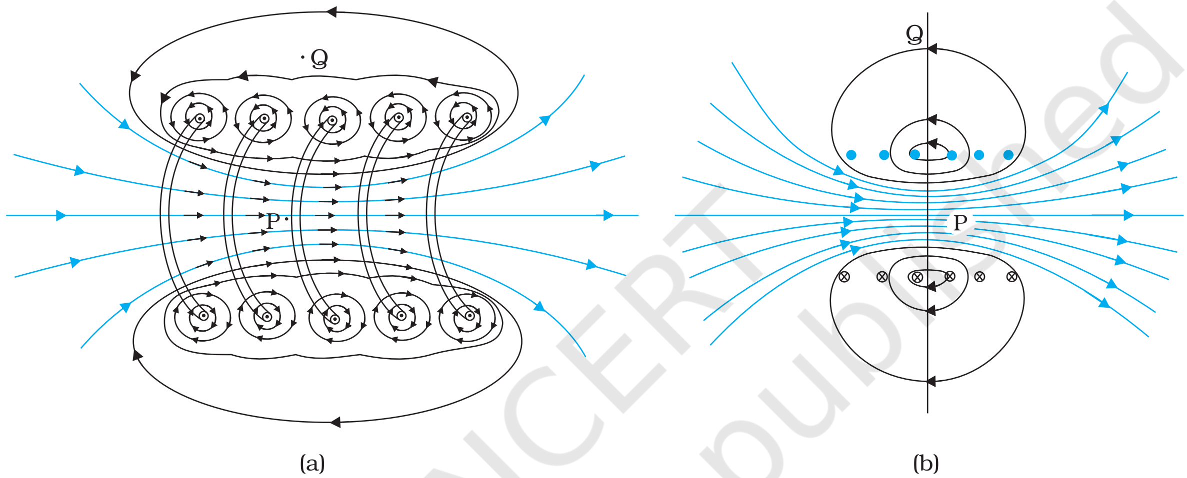


FIGURE 4.15 (a) The magnetic field due to a section of the solenoid which has been stretched out for clarity. Only the exterior semi-circular part is shown. Notice how the circular loops between neighbouring turns tend to cancel. (b) The magnetic field of a finite solenoid.

Figure 4.15 displays the magnetic field lines for a finite solenoid. We show a section of this solenoid in an enlarged manner in Fig. 4.15(a). Figure 4.15(b) shows the entire finite solenoid with its magnetic field. In Fig. 4.15(a), it is clear from the circular loops that the field between two neighbouring turns vanishes. In Fig. 4.15(b), we see that the field at the interior mid-point P is uniform, strong and along the axis of the solenoid. The field at the exterior mid-point Q is weak and moreover is along the axis of the solenoid with no perpendicular or normal component. As the

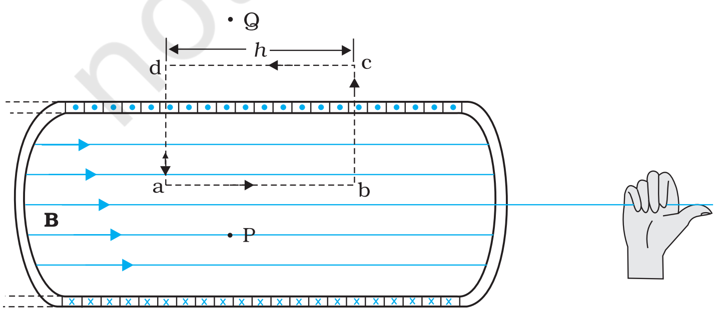


FIGURE 4.16 The magnetic field of a very long solenoid. We consider a rectangular Amperian loop abcd to determine the field.

Reprint 2025-26

solenoid is made longer it appears like a long cylindrical metal sheet. Figure 4.16 represents this idealised picture. The field outside the solenoid approaches zero. We shall assume that the field outside is zero. The field inside becomes everywhere parallel to the axis.

Consider a rectangular Amperian loop abcd. Along cd the field is zero as argued above. Along transverse sections bc and ad, the field component is zero. Thus, these two sections make no contribution. Let the field along ab be B. Thus, the relevant length of the Amperian loop is, $L = h$.

Let n be the number of turns per unit length, then the total number of turns is nh. The enclosed current is, $I_e = I\,(n\,h)$, where I is the current in the solenoid. From Ampere's circuital law [Eq. 4.13 (b)]

$$BL = \mu_0 I_e, \quad B\,h = \mu_0 I\,(n\,h)$$

$$B = \mu_0\, n\, I \tag{4.16}$$

The direction of the field is given by the right-hand rule. The solenoid is commonly used to obtain a uniform magnetic field. We shall see in the next chapter that a large field is possible by inserting a soft iron core inside the solenoid.

Example 4.8 A solenoid of length 0.5 m has a radius of 1 cm and is made up of 500 turns. It carries a current of 5 A. What is the magnitude of the magnetic field inside the solenoid?

Solution The number of turns per unit length is,

$$n = \frac{500}{0.5} = 1000 \text{ turns/m}$$

The length l = 0.5 m and radius r = 0.01 m. Thus, l/a = 50 i.e., $l >> a$. Hence, we can use the *long* solenoid formula, namely, Eq. (4.20)

$$B = \mu_0 n\, I$$
$$= 4\pi \times 10^{-7} \times 10^{3} \times 5$$
$$= 6.28 \times 10^{-3} \text{ T}$$

EXAMPLE 4.8

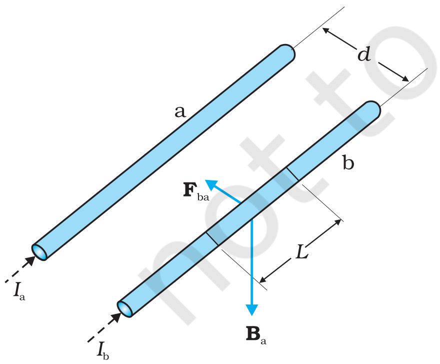


FIGURE 4.17 Two long straight parallel conductors carrying steady currents I_a and I_b and separated by a distance d. $\mathbf{B}_a$ is the magnetic field set up by conductor 'a' at conductor 'b'.

4.8 Force between Two Parallel Currents, the Ampere

We have learnt that there exists a magnetic field due to a conductor carrying a current which obeys the Biot-Savart law. Further, we have learnt that an external magnetic field will exert a force on a current-carrying conductor. This follows from the Lorentz force formula. Thus, it is logical to expect that two current-carrying conductors placed near each other will exert (magnetic) forces on each other. In the period 1820-25, Ampere studied the nature of this magnetic force and its dependence on the magnitude of the current, on the shape and size of the conductors, as well as, the distances between the conductors. In this section, we shall take the simple example of two parallel current- carrying conductors, which will perhaps help us to appreciate Ampere's painstaking work.

Figure 4.17 shows two long parallel conductors a and b separated by a distance d and carrying (parallel) currents I_a and I_b, respectively. The conductor 'a' produces, the same magnetic field $\mathbf{B}_a$ at all points along the conductor 'b'. The right-hand rule tells us that the direction of this field is downwards (when the conductors are placed horizontally). Its magnitude is given by Eq. [4.15(a)] or from Ampere's circuital law,

$$B_a = \frac{\mu_0 I_a}{2\pi d}$$

The conductor 'b' carrying a current I_b will experience a sideways force due to the field $\mathbf{B}_a$. The direction of this force is towards the conductor 'a' (Verify this). We label this force as $\mathbf{F}_{ba}$, the force on a segment L of 'b' due to 'a'. The magnitude of this force is given by Eq. (4.4),

$$F_{ba} = I_b L B_a$$

$$= \frac{\mu_0 I_a I_b}{2\pi d} L \qquad (4.17)$$

It is of course possible to compute the force on 'a' due to 'b'. From considerations similar to above we can find the force $\mathbf{F}_{ab}$, on a segment of length L of 'a' due to the current in 'b'. It is equal in magnitude to F_{ba}, and directed towards 'b'. Thus,

$$\mathbf{F}_{ba} = -\mathbf{F}_{ab} \qquad (4.18)$$

Note that this is consistent with Newton's third Law. Thus, at least for parallel conductors and steady currents, we have shown that the Biot-Savart law and the Lorentz force yield results in accordance with Newton's third Law*.

We have seen from above that currents flowing in the same direction attract each other. One can show that oppositely directed currents repel each other. Thus,

Parallel currents attract, and antiparallel currents repel.

This rule is the opposite of what we find in electrostatics. Like (same sign) charges repel each other, but like (parallel) currents attract each other.

Let f_{ba} represent the magnitude of the force $\mathbf{F}_{ba}$ per unit length. Then, from Eq. (4.17),

$$f_{ba} = \frac{\mu_0 I_a I_b}{2\pi d} \qquad (4.19)$$

The above expression is used to define the ampere (A), which is one of the seven SI base units.

* It turns out that when we have time-dependent currents and/or charges in motion, Newton's third law may not hold for forces between charges and/or conductors. An essential consequence of the Newton's third law in mechanics is conservation of momentum of an isolated system. This, however, holds even for the case of time-dependent situations with electromagnetic fields, provided the momentum carried by fields is also taken into account.

The *ampere* is the value of that steady current which, when maintained in each of the two very long, straight, parallel conductors of negligible cross-section, and placed one metre apart in vacuum, would produce on each of these conductors a force equal to 2×10^{-7} newtons per metre of length.

This definition of the ampere was adopted in 1946. It is a theoretical definition. In practice, one must eliminate the effect of the earth's magnetic field and substitute very long wires by multiturn coils of appropriate geometries. An instrument called the current balance is used to measure this mechanical force.

The SI unit of charge, namely, the coulomb, can now be defined in terms of the ampere.

When a steady current of 1A is set up in a conductor, the quantity of charge that flows through its cross-section in 1s is one coulomb (1C).

Example 4.9 The horizontal component of the earth's magnetic field at a certain place is 3.0×10^{-5} T and the direction of the field is from the geographic south to the geographic north. A very long straight conductor is carrying a steady current of 1A. What is the force per unit length on it when it is placed on a horizontal table and the direction of the current is (a) east to west; (b) south to north?

Solution $\mathbf{F} = I\mathbf{l} \times \mathbf{B}$

$F = IlB \sin\theta$

The force per unit length is

$f = F/l = I\,B \sin\theta$

(a) When the current is flowing from east to west,

$\theta = 90°$

Hence,

$f = I\,B$

$= 1 \times 3 \times 10^{-5} = 3 \times 10^{-5}\ \mathrm{N\,m^{-1}}$

This is larger than the value $2 \times 10^{-7}\ \mathrm{Nm^{-1}}$ quoted in the definition of the ampere. Hence it is important to eliminate the effect of the earth's magnetic field and other stray fields while standardising the ampere.

The direction of the force is downwards. This direction may be obtained by the directional property of cross product of vectors.

(b) When the current is flowing from south to north,

$\theta = 0°$

$f = 0$

Hence there is no force on the conductor.

EXAMPLE 4.9

4.9 Torque on Current Loop, Magnetic Dipole

4.9.1 Torque on a rectangular current loop in a uniform magnetic field

We now show that a rectangular loop carrying a steady current I and placed in a uniform magnetic field experiences a torque. It does not experience a net force. This behaviour is analogous to that of electric dipole in a uniform electric field (Section 1.11).

Reprint 2025-26

We first consider the simple case when the rectangular loop is placed such that the uniform magnetic field **B** is in the plane of the loop. This is illustrated in Fig. 4.18(a).

The field exerts no force on the two arms AD and BC of the loop. It is perpendicular to the arm AB of the loop and exerts a force $\mathbf{F}_1$ on it which is directed into the plane of the loop. Its magnitude is,

$F_1 = I\,b\,B$

Similarly, it exerts a force $\mathbf{F}_2$ on the arm CD and $\mathbf{F}_2$ is directed out of the plane of the paper.

$F_2 = I\,b\,B = F_1$

Thus, the *net force* on the loop *is zero*. There is a torque on the loop due to the pair of forces $\mathbf{F}_1$ and $\mathbf{F}_2$. Figure 4.18(b) shows a view of the loop from the AD end. It shows that the torque on the loop tends to rotate it anticlockwise. This torque is (in magnitude),

$$\tau = F_1\frac{a}{2} + F_2\frac{a}{2}$$

$$= IbB\frac{a}{2} + IbB\frac{a}{2} = I(ab)B$$

$$= I\,A\,B \qquad (4.20)$$

where $A = ab$ is the area of the rectangle.

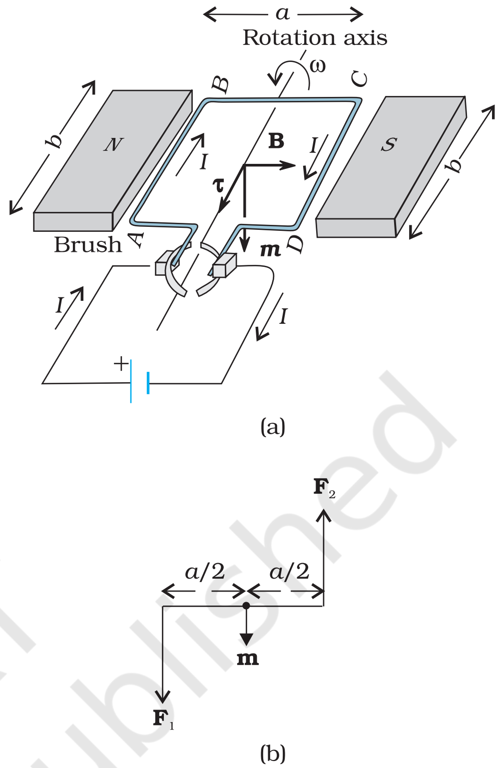


FIGURE 4.18 (a) A rectangular current-carrying coil in uniform magnetic field. The magnetic moment **m** points downwards. The torque **τ** is along the axis and tends to rotate the coil anticlockwise. (b) The couple acting on the coil.

We next consider the case when the plane of the loop, is not along the magnetic field, but makes an angle with it. We take the angle between the field and the normal to the coil to be angle θ (The previous case corresponds to $\theta = \pi/2$). Figure 4.19 illustrates this general case.

The forces on the arms BC and DA are equal, opposite, and act along the axis of the coil, which connects the centres of mass of BC and DA. Being collinear along the axis they cancel each other, resulting in no net force or torque. The forces on arms AB and CD are $\mathbf{F}_1$ and $\mathbf{F}_2$. They too are equal and opposite, with magnitude,

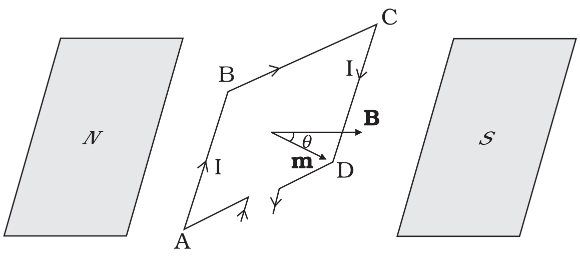
$F_1 = F_2 = I\,b\,B$

But they are not collinear! This results in a couple as before. The torque is, however, less than the earlier case when plane of loop was along the magnetic field. This is because the perpendicular distance between the forces of the couple has decreased. Figure 4.19(b) is a view of the arrangement from the AD end and it illustrates these two forces constituting a couple. The magnitude of the torque on the loop is,

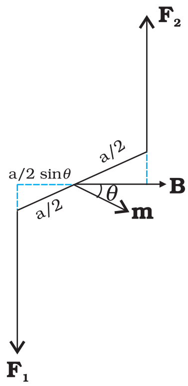
$$\tau = F_1\frac{a}{2}\sin\theta + F_2\frac{a}{2}\sin\theta$$

$$= I\,ab\,B\sin\theta$$

$$= I\,A\,B\sin\theta \qquad (4.21)$$



(a)



(b)

FIGURE 4.19 (a) The area vector of the loop ABCD makes an arbitrary angle θ with the magnetic field. (b) Top view of the loop. The forces $\mathbf{F}_1$ and $\mathbf{F}_2$ acting on the arms AB and CD are indicated.

As $\theta \to 0$, the perpendicular distance between the forces of the couple also approaches zero. This makes the forces collinear and the net force and torque zero. The torques in Eqs. (4.20) and (4.21) can be expressed as vector product of the magnetic moment of the coil and the magnetic field. We define the *magnetic moment* of the current loop as,

$$\mathbf{m} = I\,\mathbf{A} \tag{4.22}$$

where the direction of the area vector **A** is given by the right-hand thumb rule and is directed into the plane of the paper in Fig. 4.18. Then as the angle between **m** and **B** is θ, Eqs. (4.20) and (4.21) can be expressed by one expression

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B} \tag{4.23}$$

This is analogous to the electrostatic case (Electric dipole of dipole moment $\mathbf{p}_e$ in an electric field **E**).

$$\tau = \mathbf{p}_e \times \mathbf{E}$$

As is clear from Eq. (4.22), the dimensions of the magnetic moment are [A][L^2] and its unit is Am^2.

From Eq. (4.23), we see that the torque $\boldsymbol{\tau}$ vanishes when **m** is either parallel or antiparallel to the magnetic field **B**. This indicates a state of equilibrium as there is no torque on the coil (this also applies to any object with a magnetic moment **m**). When **m** and **B** are parallel the equilibrium is a stable one. Any small rotation of the coil produces a torque which brings it back to its original position. When they are antiparallel, the equilibrium is unstable as any rotation produces a torque which increases with the amount of rotation. The presence of this torque is also the reason why a small magnet or any magnetic dipole aligns itself with the external magnetic field.

If the loop has N closely wound turns, the expression for torque, Eq. (4.23), still holds, with

$$\mathbf{m} = N\,I\,\mathbf{A} \tag{4.24}$$

EXAMPLE 4.10

Example 4.10 A 100 turn closely wound circular coil of radius 10 cm carries a current of 3.2 A. (a) What is the field at the centre of the coil? (b) What is the magnetic moment of this coil?

The coil is placed in a vertical plane and is free to rotate about a horizontal axis which coincides with its diameter. A uniform magnetic field of 2T in the horizontal direction exists such that initially the axis of the coil is in the direction of the field. The coil rotates through an angle of 90° under the influence of the magnetic field. (c) What are the magnitudes of the torques on the coil in the initial and final position? (d) What is the angular speed acquired by the coil when it has rotated by 90°? The moment of inertia of the coil is 0.1 kg m^2.

Solution

(a) From Eq. (4.12)

$$B = \frac{\mu_0 NI}{2R}$$

Here, $N = 100$; $I = 3.2$ A, and $R = 0.1$ m. Hence,

$$B = \frac{4\pi \times 10^{-7} \times 3.2}{2 \times 10^{-1}} = \frac{4 \times 10^{-5} \times 10}{2 \times 10^{-1}} \quad \text{(using } \pi \times 3.2 = 10\text{)}$$

$= 2 \times 10^{-3}$ T

The direction is given by the right-hand thumb rule.

(b) The magnetic moment is given by Eq. (4.24),

$m = N I A = N I \pi r^2 = 100 \times 3.2 \times 3.14 \times 10^{-2} = 10$ A m^2

The direction is once again given by the right-hand thumb rule.

(c) $\tau = |\mathbf{m} \times \mathbf{B}|$ [from Eq. (4.23)]

$= m B \sin\theta$

Initially, $\theta = 0$. Thus, initial torque $\tau_i = 0$. Finally, $\theta = \pi/2$ (or 90°).
Thus, final torque $\tau_f = m B = 10 \times 2 = 20$ N m.

(d) From Newton's second law,

$$\mathscr{I} \frac{d\omega}{dt} = mB \sin\theta$$

where $\mathscr{I}$ is the moment of inertia of the coil. From chain rule,

$$\frac{d\omega}{dt} = \frac{d\omega}{d\theta}\frac{d\theta}{dt} = \frac{d\omega}{d\theta}\omega$$

Using this,

$$\mathscr{I} \omega d\omega = m B \sin\theta d\theta$$

Integrating from $\theta = 0$ to $\theta = \pi/2$,

$$\mathscr{I} \int_0^{\omega_f} \omega \, d\omega = mB \int_0^{\pi/2} \sin\theta \, d\theta$$

$$\mathscr{I} \frac{\omega_f^2}{2} = -m B \cos\theta \Big|_0^{\pi/2} = m B$$

$$\omega_f = \left(\frac{2mB}{\mathscr{I}}\right)^{1/2} = \left(\frac{2 \times 20}{10^{-1}}\right)^{1/2} = 20 \text{s}^{-1}$$

EXAMPLE 4.10

Example 4.11

(a) A current-carrying circular loop lies on a smooth horizontal plane. Can a uniform magnetic field be set up in such a manner that the loop turns around itself (i.e., turns about the vertical axis).

(b) A current-carrying circular loop is located in a uniform external magnetic field. If the loop is free to turn, what is its orientation of stable equilibrium? Show that in this orientation, the flux of

EXAMPLE 4.11

the total field (external field + field produced by the loop) is maximum.

(c) A loop of irregular shape carrying current is located in an external magnetic field. If the wire is flexible, why does it change to a circular shape?

Solution

(a) No, because that would require $\boldsymbol{\tau}$ to be in the vertical direction. But $\boldsymbol{\tau} = I\,\mathbf{A} \times \mathbf{B}$, and since $\mathbf{A}$ of the horizontal loop is in the vertical direction, τ would be in the plane of the loop for any $\mathbf{B}$.

(b) Orientation of stable equilibrium is one where the area vector $\mathbf{A}$ of the loop is in the direction of external magnetic field. In this orientation, the magnetic field produced by the loop is in the same direction as external field, both normal to the plane of the loop, thus giving rise to maximum flux of the total field.

(c) It assumes circular shape with its plane normal to the field to maximise flux, since for a given perimeter, a circle encloses greater area than any other shape.

EXAMPLE 4.11

4.9.2 Circular current loop as a magnetic dipole

In this section, we shall consider the elementary magnetic element: the current loop. We shall show that the magnetic field (at large distances) due to current in a circular current loop is very similar in behaviour to the electric field of an electric dipole. In Section 4.5, we have evaluated the magnetic field on the axis of a circular loop, of a radius R, carrying a steady current I. The magnitude of this field is [(Eq. (4.11)],

$$B = \frac{\mu_0 I R^2}{2\left(x^2 + R^2\right)^{3/2}}$$

and its direction is along the axis and given by the right-hand thumb rule (Fig. 4.10). Here, x is the distance along the axis from the centre of the loop. For $x >> R$, we may drop the R^2 term in the denominator. Thus,

$$B = \frac{\mu_0 I R^2}{2x^3}$$

Note that the area of the loop $A = \pi R^2$. Thus,

$$B = \frac{\mu_0 I A}{2\pi x^3}$$

As earlier, we define the magnetic moment $\mathbf{m}$ to have a magnitude IA, $\boldsymbol{m} = I\mathbf{A}$. Hence,

$$\mathrm{B} \simeq \frac{\mu_0 m}{2\pi x^3}$$

$$= \frac{\mu_0}{4\pi}\frac{2\mathbf{m}}{x^3} \qquad [4.25(a)]$$

The expression of Eq. [4.25(a)] is very similar to an expression obtained earlier for the electric field of a dipole. The similarity may be seen if we substitute,

$$\mu_0 \to 1/\varepsilon_0$$

$\mathbf{m} \to \mathbf{p}_e$ (electrostatic dipole)

$\mathbf{B} \to \mathbf{E}$ (electrostatic field)

We then obtain,

$$\mathbf{E} = \frac{2\mathbf{p}_e}{4\pi\varepsilon_0 x^3}$$

which is precisely the field for an electric dipole at a point on its axis. considered in Chapter 1, Section 1.9 [Eq. (1.20)].

It can be shown that the above analogy can be carried further. We had found in Chapter 1 that the electric field on the perpendicular bisector of the dipole is given by [See Eq.(1.21)],

$$E \simeq \frac{\mathbf{p}_e}{4\pi\varepsilon_0 x^3}$$

where x is the distance from the dipole. If we replace $\mathbf{p} \to \mathbf{m}$ and $\mu_0 \to 1/\varepsilon_0$ in the above expression, we obtain the result for **B** for a point *in the plane of the loop* at a distance x from the centre. For $x >> R$,

$$\mathbf{B} \simeq \frac{\mu_0}{4\pi}\frac{\mathbf{m}}{x^3}; \qquad x >> R \qquad [4.25(b)]$$

The results given by Eqs. [4.25(a)] and [4.25(b)] become exact for a *point* magnetic dipole.

The results obtained above can be shown to apply to any planar loop: a planar current loop is equivalent to a magnetic dipole of dipole moment $\mathbf{m} = I\mathbf{A}$, which is the analogue of electric dipole moment **p**. Note, however, a fundamental difference: an electric dipole is built up of two elementary units — the charges (or electric monopoles). In magnetism, a magnetic dipole (or a current loop) is the most elementary element. The equivalent of electric charges, i.e., magnetic monopoles, are not known to exist.

We have shown that a current loop (i) produces a magnetic field (see Fig. 4.10) and behaves like a magnetic dipole at large distances, and (ii) is subject to torque like a magnetic needle. This led Ampere to suggest that all magnetism is due to circulating currents. This seems to be partly true and no magnetic monopoles have been seen so far. However, elementary particles such as an electron or a proton also carry an *intrinsic* magnetic moment, not accounted by circulating currents.

4.10 The Moving Coil Galvanometer

Currents and voltages in circuits have been discussed extensively in Chapters 3. But how do we measure them? How do we claim that current in a circuit is 1.5 A or the voltage drop across a resistor is 1.2 V? Figure 4.20 exhibits a very useful instrument for this purpose: the *moving coil galvanometer* (MCG). It is a device whose principle can be understood on the basis of our discussion in Section 4.9.

The galvanometer consists of a coil, with many turns, free to rotate about a fixed axis (Fig. 4.20), in a uniform radial magnetic field. There is a cylindrical soft iron core which not only makes the field radial but also increases the strength of the magnetic field. When a current flows through the coil, a torque acts on it. This torque is given by Eq. (4.20) to be

$$\tau = NI\,AB$$

Reprint 2025-26

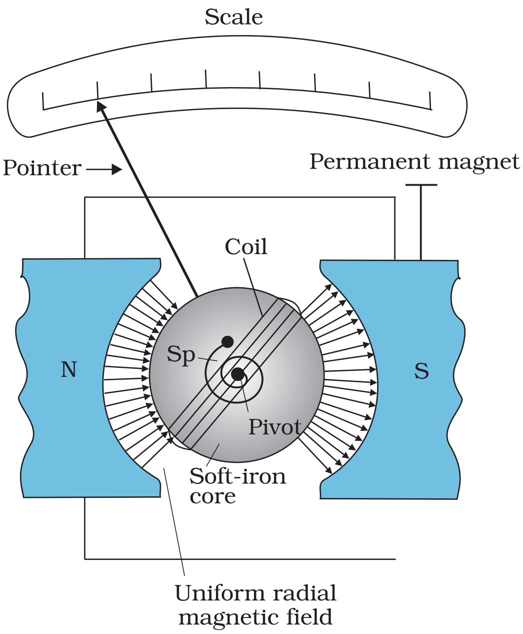


FIGURE 4.20 The moving coil galvanometer. Its elements are described in the text. Depending on the requirement, this device can be used as a current detector or for measuring the value of the current (ammeter) or voltage (voltmeter).

where the symbols have their usual meaning. Since the field is radial by design, we have taken sin θ = 1 in the above expression for the torque. The magnetic torque *NIAB* tends to rotate the coil. A spring S_p provides a counter torque $k\phi$ that balances the magnetic torque *NIAB;* resulting in a steady angular deflection ϕ. In equilibrium

$$k\phi = NI\,AB$$

where *k* is the torsional constant of the spring; i.e. the restoring torque per unit twist. The deflection ϕ is indicated on the scale by a pointer attached to the spring. We have

$$\phi = \left(\frac{NAB}{k}\right)I \tag{4.26}$$

The quantity in brackets is a constant for a given galvanometer.

The galvanometer can be used in a number of ways. It can be used as a detector to check if a current is flowing in the circuit. We have come across this usage in the Wheatstone's bridge arrangement. In this usage the neutral position of the pointer (when no current is flowing through the galvanometer) is in the middle of the scale and not at the left end as shown in Fig.4.20. Depending on the direction of the current, the pointer's deflection is either to the right or the left.

The galvanometer cannot as such be used as an ammeter to measure the value of the current in a given circuit. This is for two reasons: (i) Galvanometer is a very sensitive device, it gives a full-scale deflection for a current of the order of μA. (ii) For measuring currents, the galvanometer has to be connected in series, and as it has a large resistance, this will change the value of the current in the circuit. To overcome these difficulties, one attaches a small resistance r_s, called *shunt resistance*, in parallel with the galvanometer coil; so that most of the current passes through the shunt. The resistance of this arrangement is,

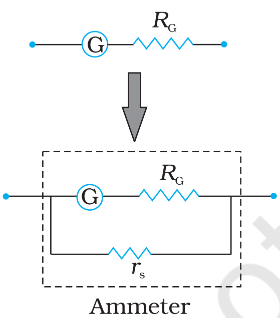


FIGURE 4.21 Conversion of a galvanometer (G) to an ammeter by the introduction of a shunt resistance r_s of very small value in parallel.

$$R_G\, r_s / (R_G + r_s) \simeq r_s \qquad \text{if} \quad R_G >> r_s$$

If r_s has small value, in relation to the resistance of the rest of the circuit R_c, the effect of introducing the measuring instrument is also small and negligible. This arrangement is schematically shown in Fig. 4.21. The scale of this ammeter is calibrated and then graduated to read off the current value with ease. We define the *current sensitivity of the galvanometer as the deflection per unit current*. From Eq. (4.26) this current sensitivity is,

$$\frac{\phi}{I} = \frac{NAB}{k} \tag{4.27}$$

A convenient way for the manufacturer to increase the sensitivity is to increase the number of turns *N*. We choose galvanometers having sensitivities of value, required by our experiment.

Reprint 2025-26

The galvanometer can also be used as a voltmeter to measure the voltage across a given section of the circuit. For this it must be connected *in parallel* with that section of the circuit. Further, it must draw a very small current, otherwise the voltage measurement will disturb the original set up by an amount which is very large. Usually we like to keep the disturbance due to the measuring device below one per cent. To ensure this, a large resistance R is connected *in series* with the galvanometer. This arrangement is schematically depicted in Fig.4.22. Note that the resistance of the voltmeter is now,

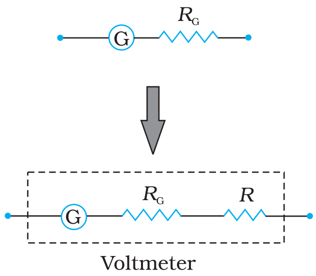


FIGURE 4.22 Conversion of a galvanometer (G) to a voltmeter by the introduction of a resistance R of large value in series.

$R_G + R \simeq R$: large

The scale of the voltmeter is calibrated to read off the voltage value with ease. We define the *voltage sensitivity as the deflection per unit voltage*. From Eq. (4.26),

$$\frac{\phi}{V} = \left(\frac{NAB}{k}\right)\frac{I}{V} = \left(\frac{NAB}{k}\right)\frac{1}{R} \qquad (4.28)$$

An interesting point to note is that increasing the current sensitivity may not necessarily increase the voltage sensitivity. Let us take Eq. (4.27) which provides a measure of current sensitivity. If $N \rightarrow 2N$, i.e., we double the number of turns, then

$$\frac{\phi}{I} \rightarrow 2\frac{\phi}{I}$$

Thus, the current sensitivity doubles. However, the resistance of the galvanometer is also likely to double, since it is proportional to the length of the wire. In Eq. (4.28), $N \rightarrow 2N$, and $R \rightarrow 2R$, thus the voltage sensitivity,

$$\frac{\phi}{V} \rightarrow \frac{\phi}{V}$$

remains unchanged. So in general, the modification needed for conversion of a galvanometer to an ammeter will be different from what is needed for converting it into a voltmeter.

Example 4.12 In the circuit (Fig. 4.23) the current is to be measured. What is the value of the current if the ammeter shown (a) is a galvanometer with a resistance R_G = 60.00 Ω; (b) is a galvanometer described in (a) but converted to an ammeter by a shunt resistance r_s = 0.02 Ω; (c) is an ideal ammeter with zero resistance?

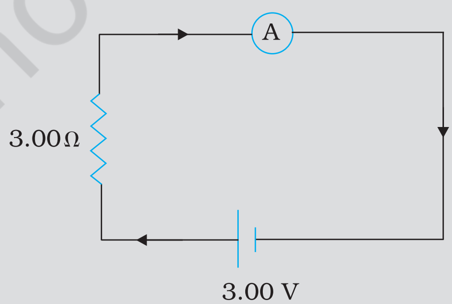


FIGURE 4.23

EXAMPLE 4.12

EXAMPLE 4.12

Solution

(a) Total resistance in the circuit is,

$R_G + 3 = 63\,\Omega$. Hence, I = 3/63 = 0.048 A.

(b) Resistance of the galvanometer converted to an ammeter is,

$$\frac{R_G\, r_s}{R_G + r_s} = \frac{60\ \ \Omega \times 0.02\Omega}{(60+0.02)\Omega} \simeq 0.02\Omega$$

Total resistance in the circuit is,

$0.02\,\Omega + 3\,\Omega = 3.02\,\Omega$. Hence, I = 3/3.02 = 0.99 A.

(c) For the ideal ammeter with zero resistance,

I = 3/3 = 1.00 A

SUMMARY

1. The total force on a charge q moving with velocity **v** in the presence of magnetic and electric fields **B** and **E,** respectively is called the *Lorentz force*. It is given by the expression:
   F = q (**v** × **B** + **E**)
   The magnetic force q (**v** × **B**) is normal to **v** and work done by it is zero.
2. A straight conductor of length l and carrying a steady current I experiences a force **F** in a uniform external magnetic field **B**,
   F = I **l** × **B**
   where |**l**| = l and the direction of **l** is given by the direction of the current.
3. In a uniform magnetic field **B**, a charge q executes a circular orbit in a plane normal to **B**. Its frequency of uniform circular motion is called the *cyclotron frequency* and is given by:
   $$\nu_c = \frac{qB}{2\pi m}$$
   This frequency is independent of the particle's speed and radius. This fact is exploited in a machine, the cyclotron, which is used to accelerate charged particles.
4. The *Biot-Savart* law asserts that the magnetic field d**B** due to an element d**l** carrying a steady current I at a point P at a distance r from the current element is:
   $$\mathrm{d}\mathbf{B} = \frac{\mu_0}{4\pi} I \frac{\mathrm{d}\mathbf{l} \times \mathbf{r}}{r^3}$$
   To obtain the total field at P, we must integrate this vector expression over the entire length of the conductor.
5. The magnitude of the magnetic field due to a circular coil of radius R carrying a current I at an axial distance x from the centre is

$$B = \frac{\mu_0 I R^2}{2(x^2 + R^2)^{3/2}}$$

At the centre this reduces to

$$B = \frac{\mu_0 I}{2R}$$

6. *Ampere's Circuital Law:* Let an open surface S be bounded by a loop C. Then the Ampere's law states that $\oint_C \mathbf{B}.d\mathbf{l} = \mu_0 I$ where I refers to the current passing through S. The sign of I is determined from the right-hand rule. We have discussed a simplified form of this law. If **B** is directed along the tangent to every point on the perimeter L of a closed curve and is constant in magnitude along perimeter then,

   $BL = \mu_0 I_e$

   where I_e is the net current enclosed by the closed circuit.
7. The magnitude of the magnetic field at a distance R from a long, straight wire carrying a current I is given by:

$$B = \frac{\mu_0 I}{2\pi R}$$

   The field lines are circles concentric with the wire.
8. The magnitude of the field B inside a *long solenoid* carrying a current I is

   $B = \mu_0 n I$

   where n is the number of turns per unit length.
9. Parallel currents attract and anti-parallel currents repel.
10. A planar loop carrying a current I, having N closely wound turns, and an area A possesses a magnetic moment **m** where,

    $\mathbf{m} = N\, I\, \mathbf{A}$

    and the direction of **m** is given by the right-hand thumb rule : curl the palm of your right hand along the loop with the fingers pointing in the direction of the current. The thumb sticking out gives the direction of **m** (and **A**)

    When this loop is placed in a uniform magnetic field **B**, the force **F** on it is: $F = 0$

    And the torque on it is,

    $\tau = \mathbf{m} \times \mathbf{B}$

    In a moving coil galvanometer, this torque is balanced by a counter-torque due to a spring, yielding

    $k\phi = NI\,AB$

    where ϕ is the equilibrium deflection and k the torsion constant of the spring.
11. A moving coil galvanometer can be converted into a ammeter by introducing a shunt resistance r_s, of small value in parallel. It can be converted into a voltmeter by introducing a resistance of a large value in series.

Physical Quantity	Symbol	Nature	Dimensions	Units	Remarks
Permeability of free space	μ_0	Scalar	$[MLT^{-2}A^{-2}]$	$T\ m\ A^{-1}$	$4\pi \times 10^{-7}$ T m A^{-1}
Magnetic Field	**B**	Vector	$[M\ T^{-2}A^{-1}]$	T (telsa)	
Magnetic Moment	**m**	Vector	$[L^2A]$	A m^2 or J/T	
Torsion Constant	k	Scalar	$[M\ L^2T^{-2}]$	N m rad^{-1}	Appears in MCG

POINTS TO PONDER

1. Electrostatic field lines originate at a positive charge and terminate at a negative charge or fade at infinity. Magnetic field lines always form closed loops.
2. The discussion in this Chapter holds only for steady currents which do not vary with time.
   When currents vary with time Newton's third law is valid only if momentum carried by the electromagnetic field is taken into account.
3. Recall the expression for the Lorentz force,

   $\mathbf{F} = q\,(\mathbf{v} \times \mathbf{B} + \mathbf{E})$

   This velocity dependent force has occupied the attention of some of the greatest scientific thinkers. If one switches to a frame with instantaneous velocity **v**, the magnetic part of the force vanishes. The motion of the charged particle is then explained by arguing that there exists an appropriate electric field in the new frame. We shall not discuss the details of this mechanism. However, we stress that the resolution of this paradox implies that electricity and magnetism are linked phenomena (*electromagnetism*) and that the Lorentz force expression *does not* imply a universal preferred frame of reference in nature.
4. Ampere's Circuital law is not independent of the Biot-Savart law. It can be derived from the Biot-Savart law. Its relationship to the Biot-Savart law is similar to the relationship between Gauss's law and Coulomb's law.

EXERCISES

4.1 A circular coil of wire consisting of 100 turns, each of radius 8.0 cm carries a current of 0.40 A. What is the magnitude of the magnetic field **B** at the centre of the coil?

4.2 A long straight wire carries a current of 35 A. What is the magnitude of the field **B** at a point 20 cm from the wire?

4.3 A long straight wire in the horizontal plane carries a current of 50 A in north to south direction. Give the magnitude and direction of **B** at a point 2.5 m east of the wire.

Reprint 2025-26

4.4 A horizontal overhead power line carries a current of 90 A in east to west direction. What is the magnitude and direction of the magnetic field due to the current 1.5 m below the line?

4.5 What is the magnitude of magnetic force per unit length on a wire carrying a current of 8 A and making an angle of 30º with the direction of a uniform magnetic field of 0.15 T?

4.6 A 3.0 cm wire carrying a current of 10 A is placed inside a solenoid perpendicular to its axis. The magnetic field inside the solenoid is given to be 0.27 T. What is the magnetic force on the wire?

4.7 Two long and parallel straight wires A and B carrying currents of 8.0 A and 5.0 A in the same direction are separated by a distance of 4.0 cm. Estimate the force on a 10 cm section of wire A.

4.8 A closely wound solenoid 80 cm long has 5 layers of windings of 400 turns each. The diameter of the solenoid is 1.8 cm. If the current carried is 8.0 A, estimate the magnitude of **B** inside the solenoid near its centre.

4.9 A square coil of side 10 cm consists of 20 turns and carries a current of 12 A. The coil is suspended vertically and the normal to the plane of the coil makes an angle of 30º with the direction of a uniform horizontal magnetic field of magnitude 0.80 T. What is the magnitude of torque experienced by the coil?

4.10 Two moving coil meters, M_1 and M_2 have the following particulars:

$R_1 = 10\ \Omega$, $N_1 = 30$,

$A_1 = 3.6 \times 10^{-3}\ \text{m}^2$, $B_1 = 0.25$ T

$R_2 = 14\ \Omega$, $N_2 = 42$,

$A_2 = 1.8 \times 10^{-3}\ \text{m}^2$, $B_2 = 0.50$ T

(The spring constants are identical for the two meters).

Determine the ratio of (a) current sensitivity and (b) voltage sensitivity of M_2 and M_1.

4.11 In a chamber, a uniform magnetic field of 6.5 G (1 G = 10^{-4} T) is maintained. An electron is shot into the field with a speed of $4.8 \times 10^6\ \text{m s}^{-1}$ normal to the field. Explain why the path of the electron is a circle. Determine the radius of the circular orbit. ($e = 1.5 \times 10^{-19}$ C, $m_e = 9.1 \times 10^{-31}$ kg)

4.12 In Exercise 4.11 obtain the frequency of revolution of the electron in its circular orbit. Does the answer depend on the speed of the electron? Explain.

4.13 (a) A circular coil of 30 turns and radius 8.0 cm carrying a current of 6.0 A is suspended vertically in a uniform horizontal magnetic field of magnitude 1.0 T. The field lines make an angle of 60° with the normal of the coil. Calculate the magnitude of the counter torque that must be applied to prevent the coil from turning.

(b) Would your answer change, if the circular coil in (a) were replaced by a planar coil of some irregular shape that encloses the same area? (All other particulars are also unaltered.)

Chapter Five

MAGNETISM AND MATTER

5.1 INTRODUCTION

Magnetic phenomena are universal in nature. Vast, distant galaxies, the tiny invisible atoms, humans and beasts all are permeated through and through with a host of magnetic fields from a variety of sources. The earth's magnetism predates human evolution. The word magnet is derived from the name of an island in Greece called *magnesia* where magnetic ore deposits were found, as early as 600 BC.

In the previous chapter we have learned that moving charges or electric currents produce magnetic fields. This discovery, which was made in the early part of the nineteenth century is credited to Oersted, Ampere, Biot and Savart, among others.

In the present chapter, we take a look at magnetism as a subject in its own right.

Some of the commonly known ideas regarding magnetism are:

(i) The earth behaves as a magnet with the magnetic field pointing approximately from the geographic south to the north.

(ii) When a bar magnet is freely suspended, it points in the north-south direction. The tip which points to the geographic north is called the *north pole* and the tip which points to the geographic south is called the *south pole* of the magnet.

(iii) There is a repulsive force when north poles (or south poles) of two magnets are brought close together. Conversely, there is an attractive force between the north pole of one magnet and the south pole of the other.

(iv) We cannot isolate the north, or south pole of a magnet. If a bar magnet is broken into two halves, we get two similar bar magnets with somewhat weaker properties. Unlike electric charges, isolated magnetic north and south poles known as *magnetic monopoles* do not exist.

(v) It is possible to make magnets out of iron and its alloys.

We begin with a description of a bar magnet and its behaviour in an external magnetic field. We describe Gauss's law of magnetism. We next describe how materials can be classified on the basis of their magnetic properties. We describe para-, dia-, and ferromagnetism.

5.2 THE BAR MAGNET

We begin our study by examining iron filings sprinkled on a sheet of glass placed over a short bar magnet. The arrangement of iron filings is shown in Fig. 5.1.

The pattern of iron filings suggests that the magnet has two poles similar to the positive and negative charge of an electric dipole. As mentioned in the introductory section, one pole is designated the *North pole* and the other, the *South pole*. When suspended freely, these poles point approximately towards the geographic north and south poles, respectively. A similar pattern of iron filings is observed around a current carrying solenoid.

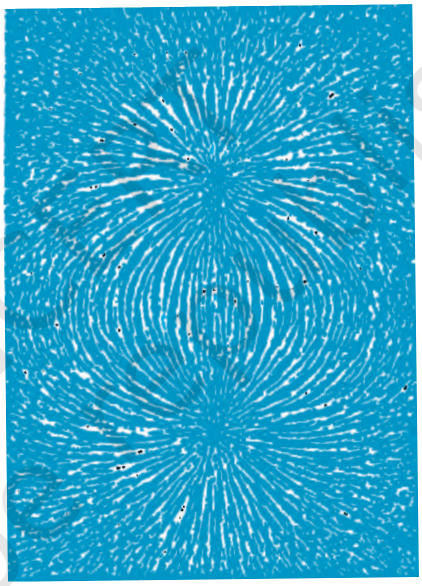


FIGURE 5.1 The arrangement of iron filings surrounding a bar magnet. The pattern mimics magnetic field lines. The pattern suggests that the bar magnet is a magnetic dipole.

5.2.1 The magnetic field lines

The pattern of iron filings permits us to plot the magnetic field lines*. This is shown both for the bar-magnet and the current-carrying solenoid in Fig. 5.2. For comparison refer to the Chapter 1, Figure 1.14(d). Electric field lines of an electric dipole are also displayed in Fig. 5.2(c). The magnetic field lines are a visual and intuitive realisation of the magnetic field. Their properties are:

(i) The magnetic field lines of a magnet (or a solenoid) form continuous closed loops. This is unlike the electric dipole where these field lines begin from a positive charge and end on the negative charge or escape to infinity.

* In some textbooks the magnetic field lines are called *magnetic lines of force*. This nomenclature is avoided since it can be confusing. Unlike electrostatics the field lines in magnetism do not indicate the direction of the force on a (moving) charge.

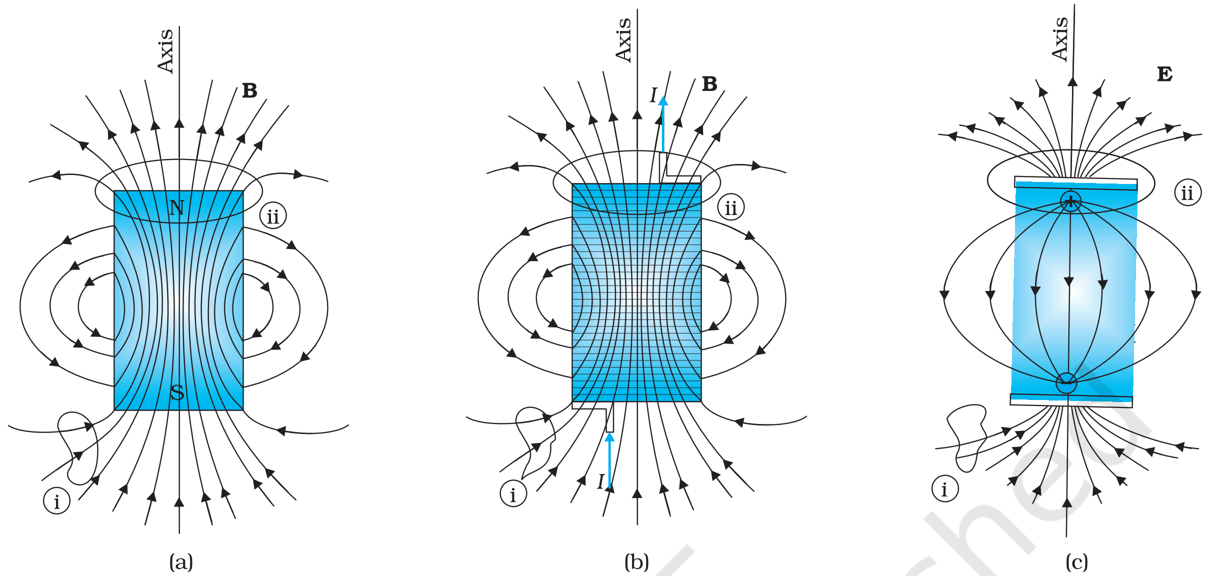


FIGURE 5.2 The field lines of (a) a bar magnet, (b) a current-carrying finite solenoid and (c) electric dipole. At large distances, the field lines are very similar. The curves labelled (i) and (ii) are closed Gaussian surfaces.

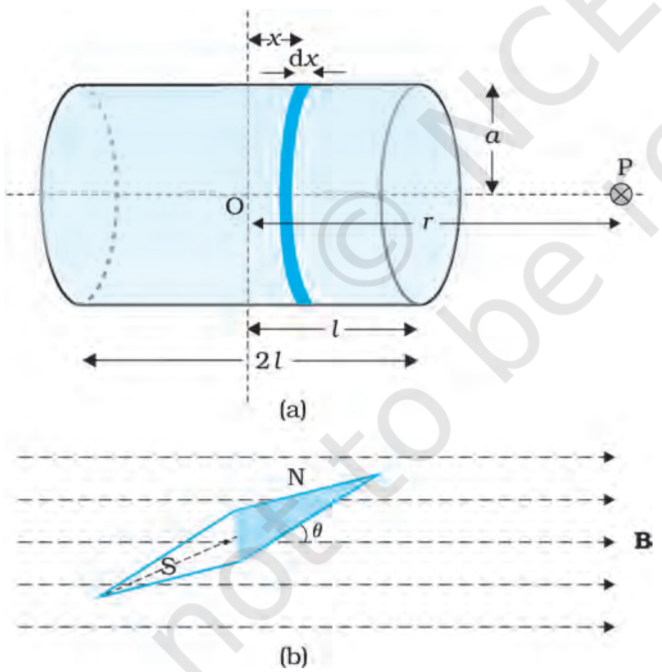


FIGURE 5.3 Calculation of (a) The axial field of a finite solenoid in order to demonstrate its similarity to that of a bar magnet. (b) A magnetic needle in a uniform magnetic field **B**. The arrangement may be used to determine either **B** or the magnetic moment **m** of the needle.

(ii) The tangent to the field line at a given point represents the direction of the net magnetic field **B** at that point.

(iii) The larger the number of field lines crossing per unit area, the stronger is the magnitude of the magnetic field **B**. In Fig. 5.2(a), **B** is larger around region ii than in region (i).

(iv) The magnetic field lines do not intersect, for if they did, the direction of the magnetic field would not be unique at the point of intersection.

One can plot the magnetic field lines in a variety of ways. One way is to place a small magnetic compass needle at various positions and note its orientation. This gives us an idea of the magnetic field direction at various points in space.

5.2.2 Bar magnet as an equivalent solenoid

In the previous chapter, we have explained how a current loop acts as a magnetic dipole (Section 4.9). We mentioned Ampere's hypothesis that all magnetic phenomena can be explained in terms of circulating currents.

Reprint 2025-26

The resemblance of magnetic field lines for a bar magnet and a solenoid suggest that a bar magnet may be thought of as a large number of circulating currents in analogy with a solenoid. Cutting a bar magnet in half is like cutting a solenoid. We get two smaller solenoids with weaker magnetic properties. The field lines remain continuous, emerging from one face of the solenoid and entering into the other face. One can test this analogy by moving a small compass needle in the neighbourhood of a bar magnet and a current-carrying finite solenoid and noting that the deflections of the needle are similar in both cases.

To make this analogy more firm we may calculate the axial field of a finite solenoid depicted in Fig. 5.3 (a). We can demonstrate that at large distances this axial field resembles that of a bar magnet.

The magnitude of the field at point P due to the solenoid is

$$B = \frac{\mu_0}{4\pi}\frac{2m}{r^3} \tag{5.1}$$

This is also the far axial magnetic field of a bar magnet which one may obtain experimentally. Thus, a bar magnet and a solenoid produce similar magnetic fields. The magnetic moment of a bar magnet is thus equal to the magnetic moment of an equivalent solenoid that produces the same magnetic field.

5.2.3 The dipole in a uniform magnetic field

Let's place a small compass needle of known magnetic moment **m** allowing it to oscillate in the magnetic field. This arrangement is shown in Fig. 5.3(b).

The torque on the needle is [see Eq. (4.23)],

$$\boldsymbol{\tau} = \mathbf{m} \times \mathbf{B} \tag{5.2}$$

In magnitude $\tau = mB\sin\theta$

Here $\boldsymbol{\tau}$ is restoring torque and θ is the angle between **m** and **B**.

An expression for magnetic potential energy can be obtained on lines similar to electrostatic potential energy.

The magnetic potential energy U_m is given by

$$U_m = \int \tau(\theta) d\theta$$

$$= \int mB\sin\theta\, d\theta = -mB\cos\theta$$

$$= -\mathbf{m}.\mathbf{B} \tag{5.3}$$

We have emphasised in Chapter 2 that the zero of potential energy can be fixed at one's convenience. Taking the constant of integration to be zero means fixing the zero of potential energy at $\theta = 90°$, i.e., when the needle is perpendicular to the field. Equation (5.3) shows that potential energy is minimum (= $-m$B) at $\theta = 0°$ (most stable position) and maximum (= $+m$B) at $\theta = 180°$ (most unstable position).

Example 5.1

(a) What happens if a bar magnet is cut into two pieces: (i) transverse to its length, (ii) along its length?

(b) A magnetised needle in a uniform magnetic field experiences a torque but no net force. An iron nail near a bar magnet, however, experiences a force of attraction in addition to a torque. Why?

EXAMPLE 5.1

(c) Must every magnetic configuration have a north pole and a south pole? What about the field due to a toroid?
(d) Two identical looking iron bars A and B are given, one of which is definitely known to be magnetised. (We do not know which one.) How would one ascertain whether or not both are magnetised? If only one is magnetised, how does one ascertain which one? [Use nothing else but the bars A and B.]

Solution
(a) In either case, one gets two magnets, each with a north and south pole.
(b) No force if the field is uniform. The iron nail experiences a non-uniform field due to the bar magnet. There is induced magnetic moment in the nail, therefore, it experiences both force and torque. The net force is attractive because the induced south pole (say) in the nail is closer to the north pole of magnet than induced north pole.
(c) Not necessarily. True only if the source of the field has a net non-zero magnetic moment. This is not so for a toroid or even for a straight infinite conductor.
(d) Try to bring different ends of the bars closer. A repulsive force in some situation establishes that both are magnetised. If it is always attractive, then one of them is not magnetised. In a bar magnet the intensity of the magnetic field is the strongest at the two ends (poles) and weakest at the central region. This fact may be used to determine whether A or B is the magnet. In this case, to see which one of the two bars is a magnet, pick up one, (say, A) and lower one of its ends; first on one of the ends of the other (say, B), and then on the middle of B. If you notice that in the middle of B, A experiences no force, then B is magnetised. If you do not notice any change from the end to the middle of B, then A is magnetised.

EXAMPLE 5.1

5.2.4 The electrostatic analog

Comparison of Eqs. (5.1), (5.2) and (5.3) with the corresponding equations for electric dipole (Chapter 1), suggests that magnetic field at large distances due to a bar magnet of magnetic moment **m** can be obtained from the equation for electric field due to an electric dipole of dipole moment **p**, by making the following replacements:

$$\mathbf{E} \rightarrow \mathbf{B}\ ,\ \mathbf{p} \rightarrow \mathbf{m}\ ,\ \frac{1}{4\pi\varepsilon_0} \rightarrow \frac{\mu_0}{4\pi}$$

In particular, we can write down the equatorial field ($\mathbf{B}_E$) of a bar magnet at a distance r, for $r >> l$, where l is the size of the magnet:

$$\mathbf{B}_E = -\frac{\mu_0 \mathbf{m}}{4\pi r^3} \tag{5.4}$$

Likewise, the axial field ($\mathbf{B}_A$) of a bar magnet for $r >> l$ is:

$$\mathbf{B}_A = \frac{\mu_0}{4\pi}\frac{2\mathbf{m}}{r^3} \tag{5.5}$$

Equation (5.5) is just Eq. (5.1) in the vector form. Table 5.1 summarises the analogy between electric and magnetic dipoles.

TABLE 5.1 THE DIPOLE ANALOGY

	Electrostatics	Magnetism
	$1/\varepsilon_0$	μ_0
Dipole moment	$\mathbf{p}$	$\mathbf{m}$
Equatorial Field for a short dipole	$-\mathbf{p}/4\pi\varepsilon_0 r^3$	$-\mu_0\,\mathbf{m}\,/\,4\pi\,r^3$
Axial Field for a short dipole	$2\mathbf{p}/4\pi\varepsilon_0 r^3$	$\mu_0\,2\mathbf{m}\,/\,4\pi\,r^3$
External Field: torque	$\mathbf{p}\times\mathbf{E}$	$\mathbf{m}\times\mathbf{B}$
External Field: Energy	$-\mathbf{p}\cdot\mathbf{E}$	$-\mathbf{m}\cdot\mathbf{B}$

Example 5.2 Figure 5.4 shows a small magnetised needle P placed at a point O. The arrow shows the direction of its magnetic moment. The other arrows show different positions (and orientations of the magnetic moment) of another identical magnetised needle Q.

(a) In which configuration the system is not in equilibrium?

(b) In which configuration is the system in (i) stable, and (ii) unstable equilibrium?

(c) Which configuration corresponds to the lowest potential energy among all the configurations shown?

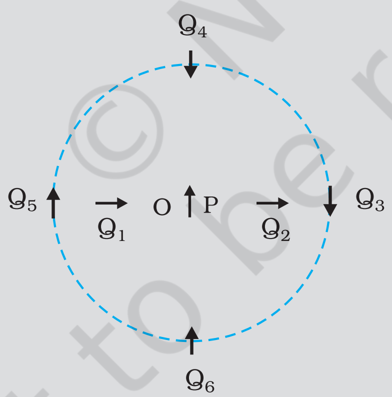


FIGURE 5.4

Solution

Potential energy of the configuration arises due to the potential energy of one dipole (say, Q) in the magnetic field due to other (P). Use the result that the field due to P is given by the expression [Eqs. (5.4) and (5.5)]:

$$\mathbf{B}_P = -\frac{\mu_0}{4\pi}\frac{\mathbf{m}_P}{r^3} \quad \text{(on the normal bisector)}$$

$$\mathbf{B}_P = \frac{\mu_0 2}{4\pi}\frac{\mathbf{m}_P}{r^3} \quad \text{(on the axis)}$$

where $\mathbf{m}_P$ is the magnetic moment of the dipole P.

Equilibrium is stable when $\mathbf{m}_Q$ is parallel to $\mathbf{B}_P$, and unstable when it is anti-parallel to $\mathbf{B}_P$.

EXAMPLE 5.2

EXAMPLE 5.2

For instance for the configuration Q_3 for which Q is along the perpendicular bisector of the dipole P, the magnetic moment of Q is parallel to the magnetic field at the position 3. Hence Q_3 is stable. Thus,

(a) PQ_1 and PQ_2

(b) (i) PQ_3, PQ_6 (stable); (ii) PQ_5, PQ_4 (unstable)

(c) PQ_6

KARL FRIEDRICH GAUSS (1777 – 1855)

Karl Friedrich Gauss (1777 – 1855) He was a child prodigy and was gifted in mathematics, physics, engineering, astronomy and even land surveying. The properties of numbers fascinated him, and in his work he anticipated major mathematical development of later times. Along with Wilhelm Welser, he built the first electric telegraph in 1833. His mathematical theory of curved surface laid the foundation for the later work of Riemann.

5.3 MAGNETISM AND GAUSS'S LAW

In Chapter 1, we studied Gauss's law for electrostatics. In Fig 5.2(c), we see that for a closed surface represented by (i), the number of lines leaving the surface is equal to the number of lines entering it. This is consistent with the fact that no net charge is enclosed by the surface. However, in the same figure, for the closed surface (ii), there is a net outward flux, since it does include a net (positive) charge.

The situation is radically different for magnetic fields which are continuous and form closed loops. Examine the Gaussian surfaces represented by (i) or (ii) in Fig 5.2(a) or Fig. 5.2(b). Both cases visually demonstrate that the number of magnetic field lines leaving the surface is balanced by the number of lines entering it. The *net magnetic flux is zero for both the surfaces*. This is true for any closed surface.

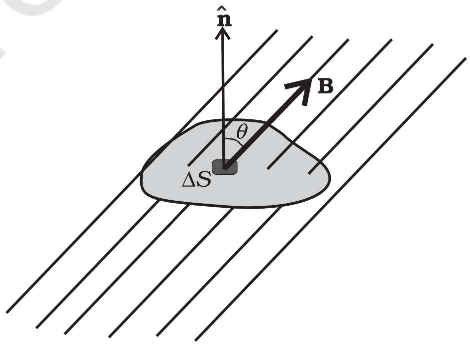


FIGURE 5.5

Consider a small vector area element ΔS of a closed surface S as in Fig. 5.5. The magnetic flux through ÄS is defined as $\Delta\phi_B = \mathbf{B}\cdot\Delta\mathbf{S}$, where **B** is the field at $\Delta\mathbf{S}$. We divide *S* into many small area elements and calculate the individual flux through each. Then, the net flux ϕ_B is,

$$\phi_B = \sum_{'all'} \Delta\phi_B = \sum_{'all'} \mathbf{B}.\Delta\mathbf{S} = 0 \qquad (5.6)$$

where 'all' stands for 'all area elements ΔS′. Compare this with the Gauss's law of electrostatics. The flux through a closed surface in that case is given by

$$\sum \mathbf{E}.\Delta\mathbf{S} = \frac{q}{\varepsilon_0}$$

where q is the electric charge enclosed by the surface.

The difference between the Gauss's law of magnetism and that for electrostatics is a reflection of the fact that isolated magnetic poles (also called monopoles) are not known to exist. There are no sources or sinks of **B**; the simplest magnetic element is a dipole or a current loop. All magnetic phenomena can be explained in terms of an arrangement of dipoles and/or current loops.

Thus, Gauss's law for magnetism is:

The net magnetic flux through any closed surface is zero.

Example 5.3 Many of the diagrams given in Fig. 5.6 show magnetic field lines (thick lines in the figure) *wrongly*. Point out what is wrong with them. Some of them may describe electrostatic field lines correctly. Point out which ones.

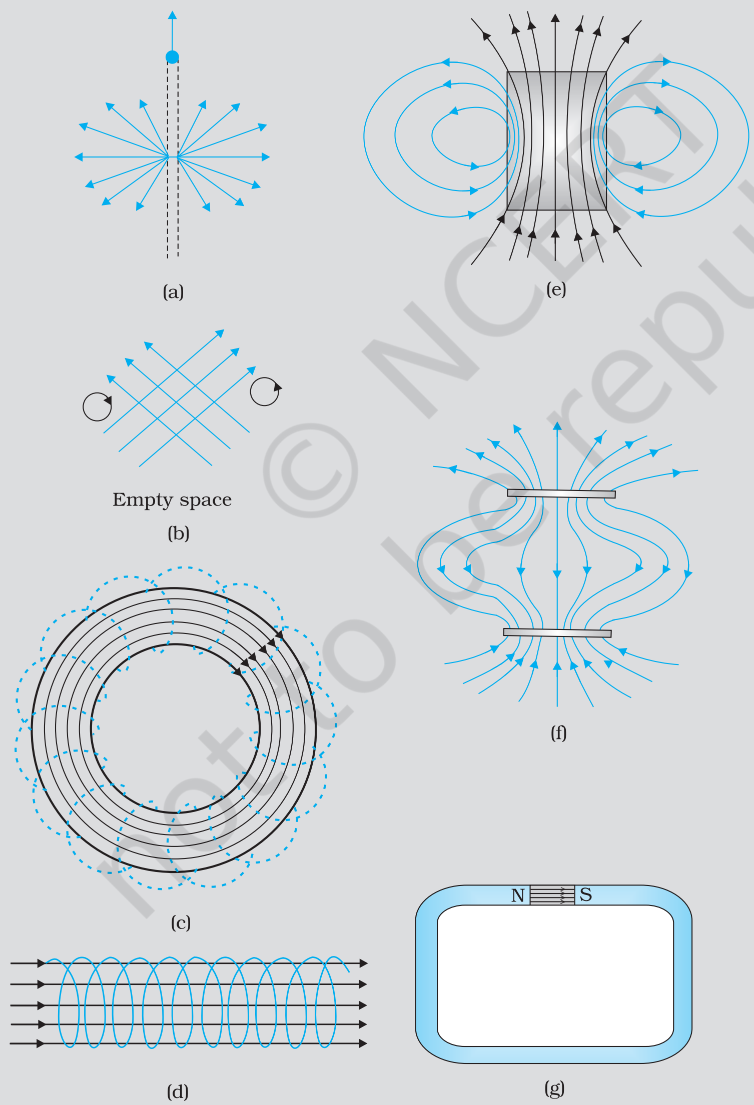


FIGURE 5.6

Solution

(a) *Wrong*. Magnetic field lines can never emanate from a point, as shown in figure. Over any closed surface, the net flux of **B** must always be zero, i.e., pictorially as many field lines should seem to enter the surface as the number of lines leaving it. The field lines shown, in fact, represent electric field of a long positively charged wire. The correct magnetic field lines are circling the straight conductor, as described in Chapter 4.

(b) *Wrong*. Magnetic field lines (like electric field lines) can never cross each other, because otherwise the direction of field at the point of intersection is ambiguous. There is further error in the figure. Magnetostatic field lines can never form closed loops around empty space. A closed loop of static magnetic field line must enclose a region across which a current is passing. By contrast, electrostatic field lines can never form closed loops, neither in empty space, nor when the loop encloses charges.

(c) *Right*. Magnetic lines are completely confined within a toroid. Nothing wrong here in field lines forming closed loops, since each loop encloses a region across which a current passes. Note, for clarity of figure, only a few field lines within the toroid have been shown. Actually, the entire region enclosed by the windings contains magnetic field.

(d) *Wrong*. Field lines due to a solenoid at its ends and outside cannot be so completely straight and confined; such a thing violates Ampere's law. The lines should curve out at both ends, and meet eventually to form closed loops.

(e) *Right*. These are field lines outside and inside a bar magnet. Note carefully the direction of field lines inside. Not all field lines emanate out of a north pole (or converge into a south pole). Around both the N-pole, and the S-pole, the net flux of the field is zero.

(f) *Wrong*. These field lines cannot possibly represent a magnetic field. Look at the upper region. All the field lines seem to emanate out of the shaded plate. The net flux through a surface surrounding the shaded plate is not zero. This is impossible for a magnetic field. The given field lines, in fact, show the electrostatic field lines around a positively charged upper plate and a negatively charged lower plate. The difference between Fig. [5.6(e) and (f)] should be carefully grasped.

(g) *Wrong*. Magnetic field lines between two pole pieces cannot be precisely straight at the ends. Some fringing of lines is inevitable. Otherwise, Ampere's law is violated. This is also true for electric field lines.

EXAMPLE 5.3

Example 5.4

(a) Magnetic field lines show the direction (at every point) along which a small magnetised needle aligns (at the point). Do the magnetic field lines also represent the *lines of force* on a moving charged particle at every point?

(b) If magnetic monopoles existed, how would the Gauss's law of magnetism be modified?

(c) Does a bar magnet exert a torque on itself due to its own field? Does one element of a current-carrying wire exert a force on another element of the *same wire*?

EXAMPLE 5.4

Reprint 2025-26

(d) Magnetic field arises due to charges in motion. Can a system have magnetic moments even though its net charge is zero?

Solution

(a) *No*. The magnetic force is always normal to **B** (remember magnetic force = $q\mathbf{v} \times \mathbf{B}$). It is misleading to call *magnetic field lines* as *lines of force*.

(b) Gauss's law of magnetism states that the flux of **B** through any closed surface is always zero $\int_S \mathbf{B}.\Delta\mathbf{s} = 0$.

If monopoles existed, the right hand side would be equal to the monopole (magnetic charge) q_m enclosed by S. [Analogous to Gauss's law of electrostatics, $\int_S \mathbf{B}.\Delta\mathbf{s} = \mu_0 q_m$ where q_m is the (monopole) magnetic charge enclosed by S.]

(c) *No*. There is no force or torque on an element due to the field produced by that element itself. But there is a force (or torque) on an element of the same wire. (For the special case of a straight wire, this force is zero.)

(d) *Yes*. The average of the charge in the system may be zero. Yet, the mean of the magnetic moments due to various current loops may not be zero. We will come across such examples in connection with paramagnetic material where atoms have net dipole moment through their net charge is zero.

EXAMPLE 5.4

5.4 MAGNETISATION AND MAGNETIC INTENSITY

The earth abounds with a bewildering variety of elements and compounds. In addition, we have been synthesising new alloys, compounds and even elements. One would like to classify the magnetic properties of these substances. In the present section, we define and explain certain terms which will help us to carry out this exercise.

We have seen that a circulating electron in an atom has a magnetic moment. In a bulk material, these moments add up vectorially and they can give a net magnetic moment which is non-zero. We define *magnetisation* **M** of a sample to be equal to its net magnetic moment per unit volume:

$$\mathbf{M} = \frac{\mathbf{m}_{net}}{V} \tag{5.7}$$

M is a vector with dimensions L^{-1} A and is measured in a units of A m^{-1}.

Consider a long solenoid of n turns per unit length and carrying a current I. The magnetic field in the interior of the solenoid was shown to be given by

$$\mathbf{B}_0 = \mu_0 \, nI \tag{5.8}$$

If the interior of the solenoid is filled with a material with non-zero magnetisation, the field inside the solenoid will be greater than $\mathbf{B}_0$. The net **B** field in the interior of the solenoid may be expressed as

$$\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_m \tag{5.9}$$

Reprint 2025-26

where $\mathbf{B}_m$ is the field contributed by the material core. It turns out that this additional field $\mathbf{B}_m$ is proportional to the magnetisation $\mathbf{M}$ of the material and is expressed as

$$\mathbf{B}_m = \mu_0 \mathbf{M} \tag{5.10}$$

where μ_0 is the same constant (permittivity of vacuum) that appears in Biot-Savart's law.

It is convenient to introduce another vector field $\mathbf{H}$, called the *magnetic intensity*, which is defined by

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{5.11}$$

where $\mathbf{H}$ has the same dimensions as $\mathbf{M}$ and is measured in units of A m^{-1}. Thus, the total magnetic field $\mathbf{B}$ is written as

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \tag{5.12}$$

We repeat our defining procedure. We have partitioned the contribution to the total magnetic field inside the sample into two parts: *one*, due to external factors such as the current in the solenoid. This is represented by $\mathbf{H}$. The *other* is due to the specific nature of the magnetic material, namely $\mathbf{M}$. The latter quantity can be influenced by external factors. This influence is mathematically expressed as

$$\mathbf{M} = \chi \mathbf{H} \tag{5.13}$$

where χ, a dimensionless quantity, is appropriately called the *magnetic susceptibility*. It is a measure of how a magnetic material responds to an external field. χ is small and positive for materials, which are called *paramagnetic*. It is small and negative for materials, which are termed *diamagnetic*. In the latter case $\mathbf{M}$ and $\mathbf{H}$ are opposite in direction. From Eqs. (5.12) and (5.13) we obtain,

$$\mathbf{B} = \mu_0 (1 + \chi)\mathbf{H} \tag{5.14}$$

$$= \mu_0 \mu_r \mathbf{H}$$

$$= \mu \mathbf{H} \tag{5.15}$$

where $\mu_r = 1 + \chi$, is a dimensionless quantity called the *relative magnetic permeability* of the substance. It is the analog of the dielectric constant in electrostatics. The *magnetic permeability* of the substance is μ and it has the same dimensions and units as μ_0;

$$\mu = \mu_0 \mu_r = \mu_0 (1+\chi).$$

The three quantities χ, μ_r and μ are interrelated and only one of them is independent. Given one, the other two may be easily determined.

EXAMPLE 5.5

Example 5.5 A solenoid has a core of a material with relative permeability 400. The windings of the solenoid are insulated from the core and carry a current of 2A. If the number of turns is 1000 per metre, calculate (a) H, (b) M, (c) B and (d) the magnetising current I_m.

Solution

(a) The field H is dependent of the material of the core, and is $H = nI = 1000 \times 2.0 = 2 \times 10^3$ A/m.

(b) The magnetic field B is given by

$$B = \mu_r \mu_0 H$$
$$= 400 \times 4\pi \times 10^{-7} \text{ (N/A}^2) \times 2 \times 10^3 \text{ (A/m)}$$
$$= 1.0 \text{ T}$$

(c) Magnetisation is given by

$$M = (B - \mu_0 H)/\mu_0$$
$$= (\mu_r \mu_0 H - \mu_0 H)/\mu_0 = (\mu_r - 1)H = 399 \times H$$
$$\cong 8 \times 10^5 \text{ A/m}$$

(d) The magnetising current I_M is the additional current that needs to be passed through the windings of the solenoid in the absence of the core which would give a B value as in the presence of the core. Thus $B = \mu_r n (I + I_M)$. Using $I = 2$A, $B = 1$ T, we get $I_M = 794$ A.

EXAMPLE 5.5

5.5 Magnetic Properties of Materials

The discussion in the previous section helps us to classify materials as diamagnetic, paramagnetic or ferromagnetic. In terms of the susceptibility χ, a material is diamagnetic if χ is negative, para- if χ is positive and small, and ferro- if χ is large and positive.

A glance at Table 5.2 gives one a better feeling for these materials. Here ε is a small positive number introduced to quantify paramagnetic materials. Next, we describe these materials in some detail.

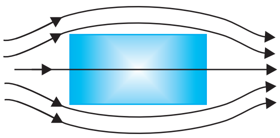
TABLE 5.2

Diamagnetic	Paramagnetic	Ferromagnetic
$-1 \le \chi < 0$	$0 < \chi < \varepsilon$	$\chi >> 1$
$0 \le \mu_r < 1$	$1 < \mu_r < 1 + \varepsilon$	$\mu_r >> 1$
$\mu < \mu_0$	$\mu > \mu_0$	$\mu >> \mu_0$

5.5.1 Diamagnetism

Diamagnetic substances are those which have tendency to move from stronger to the weaker part of the external magnetic field. In other words, unlike the way a magnet attracts metals like iron, it would repel a diamagnetic substance.

Figure 5.7(a) shows a bar of diamagnetic material placed in an external magnetic field. The field lines are repelled or expelled and the field inside the material is reduced. In most cases, this reduction is slight, being one part in 10^5. When placed in a non-uniform magnetic field, the bar will tend to move from high to low field.


(a)

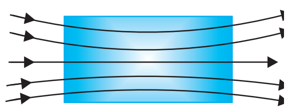

(b)

FIGURE 5.7 Behaviour of magnetic field lines near a (a) diamagnetic, (b) paramagnetic substance.

Reprint 2025-26

The simplest explanation for diamagnetism is as follows. Electrons in an atom orbiting around nucleus possess orbital angular momentum. These orbiting electrons are equivalent to current-carrying loop and thus possess orbital magnetic moment. Diamagnetic substances are the ones in which resultant magnetic moment in an atom is zero. When magnetic field is applied, those electrons having orbital magnetic moment in the same direction slow down and those in the opposite direction speed up. This happens due to induced current in accordance with Lenz's law which you will study in Chapter 6. Thus, the substance develops a net magnetic moment in direction opposite to that of the applied field and hence repulsion.

Some diamagnetic materials are bismuth, copper, lead, silicon, nitrogen (at STP), water and sodium chloride. Diamagnetism is present in all the substances. However, the effect is so weak in most cases that it gets shifted by other effects like paramagnetism, ferromagnetism, etc.

The most exotic diamagnetic materials are *superconductors*. These are metals, cooled to very low temperatures which exhibits both *perfect conductivity* and *perfect diamagnetism*. Here the field lines are completely expelled! $\chi = -1$ and $\mu_r = 0$. A superconductor repels a magnet and (by Newton's third law) is repelled by the magnet. The phenomenon of perfect diamagnetism in superconductors is called the *Meissner effect*, after the name of its discoverer. Superconducting magnets can be gainfully exploited in variety of situations, for example, for running magnetically levitated superfast trains.

5.5.2 Paramagnetism

Paramagnetic substances are those which get weakly magnetised when placed in an external magnetic field. They have tendency to move from a region of weak magnetic field to strong magnetic field, i.e., they get weakly attracted to a magnet.

The individual atoms (or ions or molecules) of a paramagnetic material possess a permanent magnetic dipole moment of their own. On account of the ceaseless random thermal motion of the atoms, no net magnetisation is seen. In the presence of an external field $\mathbf{B}_0$, which is strong enough, and at low temperatures, the individual atomic dipole moment can be made to align and point in the same direction as $\mathbf{B}_0$. Figure 5.7(b) shows a bar of paramagnetic material placed in an external field. The field lines gets concentrated inside the material, and the field inside is enhanced. In most cases, this enhancement is slight, being one part in 10^5. When placed in a non-uniform magnetic field, the bar will tend to move from weak field to strong.

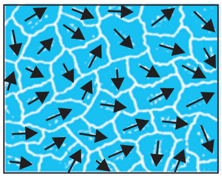
Some paramagnetic materials are aluminium, sodium, calcium, oxygen (at STP) and copper chloride. For a paramagnetic material both χ and μ_r depend not only on the material, but also (in a simple fashion) on the sample temperature. As the field is increased or the temperature is lowered, the magnetisation increases until it reaches the saturation value at which point all the dipoles are perfectly aligned with the field.

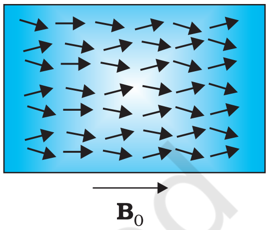
Reprint 2025-26

5.5.3 Ferromagnetism

Ferromagnetic substances are those which gets strongly magnetised when placed in an external magnetic field. They have strong tendency to move from a region of weak magnetic field to strong magnetic field, i.e., they get strongly attracted to a magnet.

The individual atoms (or ions or molecules) in a ferromagnetic material possess a dipole moment as in a paramagnetic material. However, they interact with one another in such a way that they spontaneously align themselves in a common direction over a macroscopic volume called *domain*. The explanation of this cooperative effect requires quantum mechanics and is beyond the scope of this textbook. Each domain has a net magnetisation. Typical domain size is 1mm and the domain contains about 10^{11} atoms. In the first instant, the magnetisation varies randomly from domain to domain and there is no bulk magnetisation. This is shown in Fig. 5.8(a). When we apply an external magnetic field $\mathbf{B}_0$, the domains orient themselves in the direction of $\mathbf{B}_0$ and simultaneously the domain oriented in the direction of $\mathbf{B}_0$ grow in size. This existence of domains and their motion in $\mathbf{B}_0$ are not speculations. One may observe this under a microscope after sprinkling a liquid suspension of powdered ferromagnetic substance of samples. This motion of suspension can be observed. Fig. 5.8(b) shows the situation when the domains have aligned and amalgamated to form a single 'giant' domain.


(a)



(b)

FIGURE 5.8 (a) Randomly oriented domains, (b) Aligned domains.

Thus, in a ferromagnetic material the field lines are highly concentrated. In non-uniform magnetic field, the sample tends to move towards the region of high field. We may wonder as to what happens when the external field is removed. In some ferromagnetic materials the magnetisation persists. Such materials are called *hard* magnetic materials or *hard ferromagnets*. Alnico, an alloy of iron, aluminium, nickel, cobalt and copper, is one such material. The naturally occurring lodestone is another. Such materials form permanent magnets to be used among other things as a compass needle. On the other hand, there is a class of ferromagnetic materials in which the magnetisation disappears on removal of the external field. Soft iron is one such material. Appropriately enough, such materials are called *soft ferromagnetic materials*. There are a number of elements, which are ferromagnetic: iron, cobalt, nickel, gadolinium, etc. The relative magnetic permeability is >1000!

The ferromagnetic property depends on temperature. At high enough temperature, a ferromagnet becomes a paramagnet. The domain structure disintegrates with temperature. This disappearance of magnetisation with temperature is gradual.

SUMMARY

1. The science of magnetism is old. It has been known since ancient times that magnetic materials tend to point in the north-south direction; like

magnetic poles repel and unlike ones attract; and cutting a bar magnet in two leads to two smaller magnets. Magnetic poles cannot be isolated.

2. When a bar magnet of dipole moment **m** is placed in a uniform magnetic field **B**,
   (a) the force on it is zero,
   (b) the torque on it is **m** × **B**,
   (c) its potential energy is –**m**·**B**, where we choose the zero of energy at the orientation when **m** is perpendicular to **B**.

3. Consider a bar magnet of size *l* and magnetic moment **m**, at a distance *r* from its mid-point, where *r* >>*l*, the magnetic field **B** due to this bar is,

$$\mathbf{B} = \frac{\mu_0 \mathbf{m}}{2\pi r^3} \quad \text{(along axis)}$$

$$= -\frac{\mu_0 \mathbf{m}}{4\pi r^3} \quad \text{(along equator)}$$

4. Gauss's law for magnetism states that the net magnetic flux through any closed surface is zero

$$\phi_B = \sum_{\substack{\text{all area} \\ \text{elements } \Delta\mathbf{S}}} \mathbf{B}\cdot\Delta\mathbf{S} = 0$$

5. Consider a material placed in an external magnetic field $\mathbf{B}_0$. The magnetic intensity is defined as,

$$\mathbf{H} = \frac{\mathbf{B}_0}{\mu_0}$$

The magnetisation **M** of the material is its dipole moment per unit volume. The magnetic field **B** in the material is,

$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M})$

6. For a linear material $\mathbf{M} = \chi\, \mathbf{H}$. So that $\mathbf{B} = \mu\, \mathbf{H}$ and χ is called the magnetic susceptibility of the material. The three quantities, χ, the relative magnetic permeability μ_r, and the magnetic permeability μ are related as follows:

$\mu = \mu_0\, \mu_r$

$\mu_r = 1 + \chi$

7. Magnetic materials are broadly classified as: diamagnetic, paramagnetic, and ferromagnetic. For diamagnetic materials χ is negative and small and for paramagnetic materials it is positive and small. Ferromagnetic materials have large χ.

8. Substances, which at room temperature, retain their ferromagnetic property for a long period of time are called permanent magnets.

Reprint 2025-26

Physical quantity	Symbol	Nature	Dimensions	Units	Remarks
Permeability of free space	μ_0	Scalar	$[MLT^{-2}\ A^{-2}]$	$T\ m\ A^{-1}$	$\mu_0/4\pi = 10^{-7}$
Magnetic field, Magnetic induction, Magnetic flux density	**B**	Vector	$[MT^{-2}\ A^{-1}]$	T (tesla)	10^4 G (gauss) = 1 T
Magnetic moment	**m**	Vector	$[L^{-2}\ A]$	$A\ m^2$	
Magnetic flux	ϕ_B	Scalar	$[ML^2T^{-2}\ A^{-1}]$	W (weber)	$W = T\ m^2$
Magnetisation	**M**	Vector	$[L^{-1}\ A]$	$A\ m^{-1}$	$\dfrac{\text{Magnetic moment}}{\text{Volume}}$
Magnetic intensity Magnetic field strength	**H**	Vector	$[L^{-1}\ A]$	$A\ m^{-1}$	$\mathbf{B} = \mu_0\ (\mathbf{H} + \mathbf{M})$
Magnetic susceptibility	χ	Scalar	-	-	$\mathbf{M} = \chi \mathbf{H}$
Relative magnetic permeability	μ_r	Scalar	-	-	$\mathbf{B} = \mu_0 \mu_r \mathbf{H}$
Magnetic permeability	μ	Scalar	$[MLT^{-2}\ A^{-2}]$	$T\ m\ A^{-1}$ $N\ A^{-2}$	$\mu = \mu_0 \mu_r$ $\mathbf{B} = \mu\,\mathbf{H}$

POINTS TO PONDER

1. A satisfactory understanding of magnetic phenomenon in terms of moving charges/currents was arrived at after 1800 AD. But technological exploitation of the directional properties of magnets predates this scientific understanding by two thousand years. Thus, scientific understanding is not a necessary condition for engineering applications. Ideally, science and engineering go hand-in-hand, one leading and assisting the other in tandem.
2. Magnetic monopoles do not exist. If you slice a magnet in half, you get two smaller magnets. On the other hand, isolated positive and negative charges exist. There exists a smallest unit of charge, for example, the electronic charge with value $|e| = 1.6 \times 10^{-19}$ C. All other charges are integral multiples of this smallest unit charge. In other words, charge is quantised. We do not know why magnetic monopoles do not exist or why electric charge is quantised.
3. A consequence of the fact that magnetic monopoles do not exist is that the magnetic field lines are continuous and form closed loops. In contrast, the electrostatic lines of force begin on a positive charge and terminate on the negative charge (or fade out at infinity).
4. A miniscule difference in the value of χ, the magnetic susceptibility, yields radically different behaviour: diamagnetic versus paramagnetic. For diamagnetic materials $\chi = -10^{-5}$ whereas $\chi = +10^{-5}$ for paramagnetic materials.

Reprint 2025-26

5. There exists *a perfect diamagnet*, namely, a superconductor. This is a metal at very low temperatures. In this case $\chi = -1$, $\mu_r = 0$, $\mu = 0$. The external magnetic field is totally expelled. Interestingly, this material is also a perfect conductor. However, there exists no classical theory which ties these two properties together. A quantum-mechanical theory by Bardeen, Cooper, and Schrieffer (BCS theory) explains these effects. The BCS theory was proposed in1957 and was eventually recognised by a Nobel Prize in physics in 1970.
6. Diamagnetism is universal. It is present in all materials. But it is weak and hard to detect if the substance is para- or ferromagnetic.
7. We have classified materials as diamagnetic, paramagnetic, and ferromagnetic. However, there exist additional types of magnetic material such as ferrimagnetic, anti-ferromagnetic, spin glass, etc. with properties which are exotic and mysterious.

EXERCISES

5.1 A short bar magnet placed with its axis at 30° with a uniform external magnetic field of 0.25 T experiences a torque of magnitude equal to 4.5×10^{-2} J. What is the magnitude of magnetic moment of the magnet?

5.2 A short bar magnet of magnetic moment m = 0.32 JT^{-1} is placed in a uniform magnetic field of 0.15 T. If the bar is free to rotate in the plane of the field, which orientation would correspond to its (a) stable, and (b) unstable equilibrium? What is the potential energy of the magnet in each case?

5.3 A closely wound solenoid of 800 turns and area of cross section 2.5×10^{-4} m^2 carries a current of 3.0 A. Explain the sense in which the solenoid acts like a bar magnet. What is its associated magnetic moment?

5.4 If the solenoid in Exercise 5.5 is free to turn about the vertical direction and a uniform horizontal magnetic field of 0.25 T is applied, what is the magnitude of torque on the solenoid when its axis makes an angle of 30° with the direction of applied field?

5.5 A bar magnet of magnetic moment 1.5 J T^{-1} lies aligned with the direction of a uniform magnetic field of 0.22 T.

(a) What is the amount of work required by an external torque to turn the magnet so as to align its magnetic moment: (i) normal to the field direction, (ii) opposite to the field direction?

(b) What is the torque on the magnet in cases (i) and (ii)?

5.6 A closely wound solenoid of 2000 turns and area of cross-section 1.6×10^{-4} m^2, carrying a current of 4.0 A, is suspended through its centre allowing it to turn in a horizontal plane.

(a) What is the magnetic moment associated with the solenoid?
(b) What is the force and torque on the solenoid if a uniform horizontal magnetic field of 7.5×10^{-2} T is set up at an angle of 30° with the axis of the solenoid?

5.7 A short bar magnet has a magnetic moment of 0.48 J T^{-1}. Give the direction and magnitude of the magnetic field produced by the magnet at a distance of 10 cm from the centre of the magnet on (a) the axis, (b) the equatorial lines (normal bisector) of the magnet.

Reprint 2025-26

Chapter Six

ELECTROMAGNETIC INDUCTION

6.1 Introduction

Electricity and magnetism were considered separate and unrelated phenomena for a long time. In the early decades of the nineteenth century, experiments on electric current by Oersted, Ampere and a few others established the fact that electricity and magnetism are inter-related. They found that moving electric charges produce magnetic fields. For example, an electric current deflects a magnetic compass needle placed in its vicinity. This naturally raises the questions like: Is the converse effect possible? Can moving magnets produce electric currents? Does the nature permit such a relation between electricity and magnetism? The answer is resounding yes! The experiments of Michael Faraday in England and Joseph Henry in USA, conducted around 1830, demonstrated conclusively that electric currents were induced in closed coils when subjected to changing magnetic fields. In this chapter, we will study the phenomena associated with changing magnetic fields and understand the underlying principles. The phenomenon in which electric current is generated by varying magnetic fields is appropriately called *electromagnetic induction.*

When Faraday first made public his discovery that relative motion between a bar magnet and a wire loop produced a small current in the latter, he was asked, "What is the use of it?" His reply was: "What is the use of a new born baby?" The phenomenon of electromagnetic induction is not merely of theoretical or academic interest but also of practical utility. Imagine a world where there is no electricity – no electric lights, no trains, no telephones and no personal computers. The pioneering experiments of Faraday and Henry have led directly to the development of modern day generators and transformers. Today's civilisation owes its progress to a great extent to the discovery of electromagnetic induction.

Josheph Henry [1797 – 1878] American experimental physicist, professor at Princeton University and first director of the Smithsonian Institution. He made important improvements in electro-magnets by winding coils of insulated wire around iron pole pieces and invented an electromagnetic motor and a new, efficient telegraph. He discovered self-induction and investigated how currents in one circuit induce currents in another.

JOSEPH HENRY (1797 – 1878)

6.2 The Experiments of Faraday and Henry

The discovery and understanding of electromagnetic induction are based on a long series of experiments carried out by Faraday and Henry. We shall now describe some of these experiments.

Experiment 6.1

Figure 6.1 shows a coil C_1* connected to a galvanometer G. When the North-pole of a bar magnet is pushed towards the coil, the pointer in the galvanometer deflects, indicating the presence of electric current in the coil. The deflection lasts as long as the bar magnet is in motion. The galvanometer does not show any deflection when the magnet is held stationary. When the magnet is pulled away from the coil, the galvanometer shows deflection in the opposite direction, which indicates reversal of the current's direction. Moreover, when the South-pole of the bar magnet is moved towards or away from the coil, the deflections in the galvanometer are opposite to that observed with the North-pole for similar movements. Further, the deflection (and hence current) is found to be larger when the magnet is pushed towards or pulled away from the coil faster. Instead, when the bar magnet is held fixed and the coil C_1 is moved towards or away from the magnet, the same effects are observed. It shows that *it is the relative motion between the magnet and the coil that is responsible for generation (induction) of electric current in the coil.*

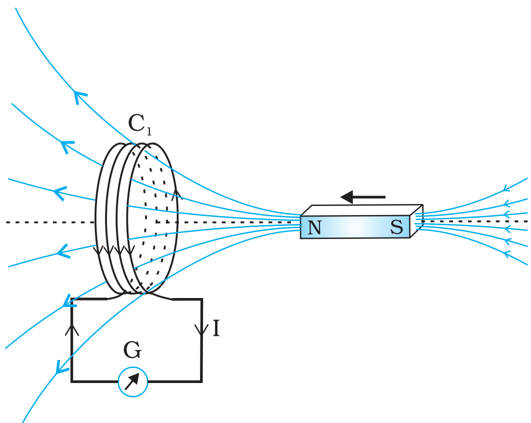


FIGURE 6.1 When the bar magnet is pushed towards the coil, the pointer in the galvanometer G deflects.

Experiment 6.2

In Fig. 6.2 the bar magnet is replaced by a second coil C_2 connected to a battery. The steady current in the coil C_2 produces a steady magnetic field. As coil C_2 is

* Wherever the term 'coil' or 'loop' is used, it is assumed that they are made up of conducting material and are prepared using wires which are coated with insulating material.

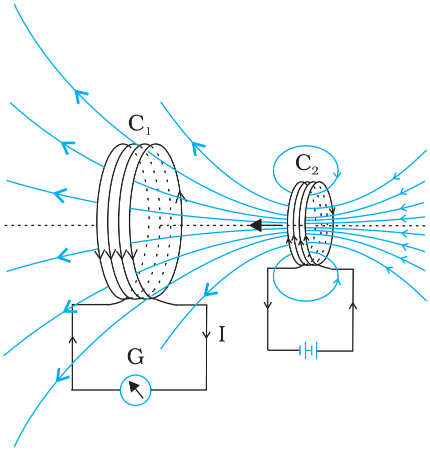


FIGURE 6.2 Current is induced in coil C_1 due to motion of the current carrying coil C_2.

moved towards the coil C_1, the galvanometer shows a deflection. This indicates that electric current is induced in coil C_1. When C_2 is moved away, the galvanometer shows a deflection again, but this time in the opposite direction. The deflection lasts as long as coil C_2 is in motion. When the coil C_2 is held fixed and C_1 is moved, the same effects are observed. Again, *it is the relative motion between the coils that induces the electric current.*

Experiment 6.3

The above two experiments involved relative motion between a magnet and a coil and between two coils, respectively. Through another experiment, Faraday showed that this relative motion is not an absolute requirement. Figure 6.3 shows two coils C_1 and C_2 held stationary. Coil C_1 is connected to galvanometer G while the second coil C_2 is connected to a battery through a tapping key K.

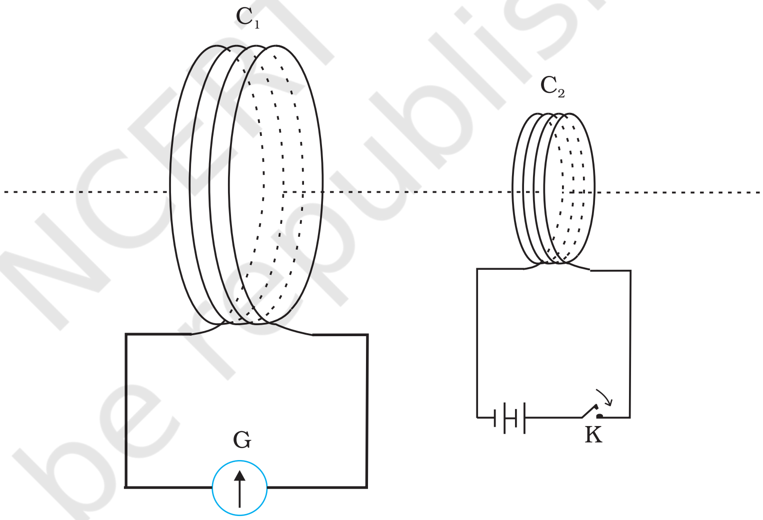


FIGURE 6.3 Experimental set-up for Experiment 6.3.

It is observed that the galvanometer shows a momentary deflection when the tapping key K is pressed. The pointer in the galvanometer returns to zero immediately. If the key is held pressed continuously, there is no deflection in the galvanometer. When the key is released, a momentory deflection is observed again, but in the opposite direction. It is also observed that the deflection increases dramatically when an iron rod is inserted into the coils along their axis.

6.3 Magnetic Flux

Faraday's great insight lay in discovering a simple mathematical relation to explain the series of experiments he carried out on electromagnetic induction. However, before we state and appreciate his laws, we must get familiar with the notion of magnetic flux, Φ_B. Magnetic flux is defined in the same way as electric flux is defined in Chapter 1. Magnetic flux through a plane of area A placed in a uniform magnetic field **B** (Fig. 6.4) can be written as

$$\Phi_B = \mathbf{B} \cdot \mathbf{A} = BA \cos \theta \qquad (6.1)$$

where θ is angle between **B** and **A**. The notion of the area as a vector has been discussed earlier in Chapter 1. Equation (6.1) can be extended to curved surfaces and nonuniform fields.

If the magnetic field has different magnitudes and directions at various parts of a surface as shown in Fig. 6.5, then the magnetic flux through the surface is given by

$$\Phi_B = \mathbf{B}_1 \cdot d\mathbf{A}_1 + \mathbf{B}_2 \cdot d\mathbf{A}_2 + \cdots = \sum_{\text{all}} \mathbf{B}_i \cdot d\mathbf{A}_i \qquad (6.2)$$

where 'all' stands for summation over all the area elements $d\mathbf{A}_i$ comprising the surface and $\mathbf{B}_i$ is the magnetic field at the area element $d\mathbf{A}_i$. The SI unit of magnetic flux is weber (Wb) or tesla meter squared (T m^2). Magnetic flux is a scalar quantity.

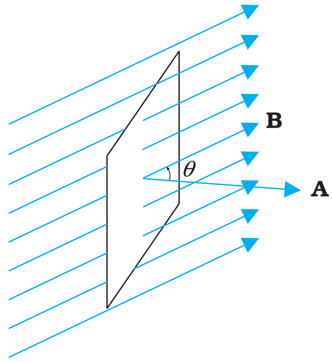


FIGURE 6.4 A plane of surface area **A** placed in a uniform magnetic field **B**.

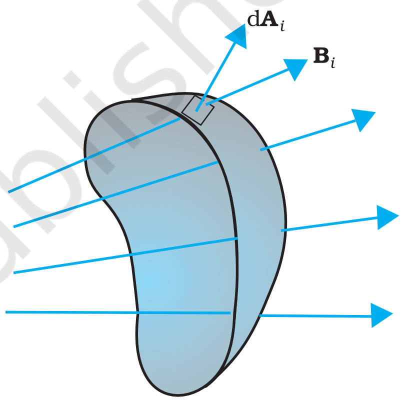


FIGURE 6.5 Magnetic field $\mathbf{B}_i$ at the i^{th} area element. $d\mathbf{A}_i$ represents area vector of the i^{th} area element.

6.4 Faraday's Law of Induction

From the experimental observations, Faraday arrived at a conclusion that an emf is induced in a coil when magnetic flux through the coil changes with time. Experimental observations discussed in Section 6.2 can be explained using this concept.

The motion of a magnet towards or away from coil C_1 in Experiment 6.1 and moving a current-carrying coil C_2 towards or away from coil C_1 in Experiment 6.2, change the magnetic flux associated with coil C_1. The change in magnetic flux induces emf in coil C_1. It was this induced emf which caused electric current to flow in coil C_1 and through the galvanometer. A plausible explanation for the observations of Experiment 6.3 is as follows: When the tapping key K is pressed, the current in coil C_2 (and the resulting magnetic field) rises from zero to a maximum value in a short time. Consequently, the magnetic flux through the neighbouring coil C_1 also increases. It is the change in magnetic flux through coil C_1 that produces an induced emf in coil C_1. When the key is held pressed, current in coil C_2 is constant. Therefore, there is no change in the magnetic flux through coil C_1 and the current in coil C_1 drops to zero. When the key is released, the current in C_2 and the resulting magnetic field decreases from the maximum value to zero in a short time. This results in a decrease in magnetic flux through coil C_1 and hence again induces an electric current in coil C_1*. The common point in all these observations is that the time rate of change of magnetic flux through a circuit induces emf in it. Faraday stated experimental observations in the form of a law called *Faraday's law of electromagnetic induction*. The law is stated below.

* Note that sensitive electrical instruments in the vicinity of an electromagnet can be damaged due to the induced emfs (and the resulting currents) when the electromagnet is turned on or off.

MICHAEL FARADAY (1791–1867)

Michael Faraday [1791–1867] Faraday made numerous contributions to science, viz., the discovery of electromagnetic induction, the laws of electrolysis, benzene, and the fact that the plane of polarisation is rotated in an electric field. He is also credited with the invention of the electric motor, the electric generator and the transformer. He is widely regarded as the greatest experimental scientist of the nineteenth century.

The magnitude of the induced emf *in a circuit is equal to the time rate of change of magnetic flux through the circuit.*

Mathematically, the induced emf is given by

$$\varepsilon = -\frac{\mathrm{d}\Phi_B}{\mathrm{d}t} \tag{6.3}$$

The negative sign indicates the direction of ε and hence the direction of current in a closed loop. This will be discussed in detail in the next section.

In the case of a closely wound coil of N turns, change of flux associated with each turn, is the same. Therefore, the expression for the total induced emf is given by

$$\varepsilon = -N\frac{\mathrm{d}\Phi_B}{\mathrm{d}t} \tag{6.4}$$

The induced emf can be increased by increasing the number of turns N of a closed coil.

From Eqs. (6.1) and (6.2), we see that the flux can be varied by changing any one or more of the terms **B**, **A** and θ. In Experiments 6.1 and 6.2 in Section 6.2, the flux is changed by varying **B**. The flux can also be altered by changing the shape of a coil (that is, by shrinking it or stretching it) in a magnetic field, or rotating a coil in a magnetic field such that the angle θ between **B** and **A** changes. In these cases too, an emf is induced in the respective coils.

EXAMPLE 6.1

Example 6.1 Consider Experiment 6.2. (a) What would you do to obtain a large deflection of the galvanometer? (b) How would you demonstrate the presence of an induced current in the absence of a galvanometer?

Solution

(a) To obtain a large deflection, one or more of the following steps can be taken: (i) Use a rod made of soft iron inside the coil C_2, (ii) Connect the coil to a powerful battery, and (iii) Move the arrangement rapidly towards the test coil C_1.

(b) Replace the galvanometer by a small bulb, the kind one finds in a small torch light. The relative motion between the two coils will cause the bulb to glow and thus demonstrate the presence of an induced current.

In experimental physics one must learn to innovate. Michael Faraday who is ranked as one of the best experimentalists ever, was legendary for his innovative skills.

EXAMPLE 6.2

Example 6.2 A square loop of side 10 cm and resistance 0.5 Ω is placed vertically in the east-west plane. A uniform magnetic field of 0.10 T is set up across the plane in the north-east direction. The magnetic field is decreased to zero in 0.70 s at a steady rate. Determine the magnitudes of induced emf and current during this time-interval.

Solution The angle θ made by the area vector of the coil with the magnetic field is 45°. From Eq. (6.1), the initial magnetic flux is

$$\Phi = BA\cos\theta$$

$$= \frac{0.1 \times 10^{-2}}{\sqrt{2}}\,\text{Wb}$$

Final flux, $\Phi_{min} = 0$

The change in flux is brought about in 0.70 s. From Eq. (6.3), the magnitude of the induced emf is given by

$$\varepsilon = \frac{|\Delta\Phi_B|}{\Delta t} = \frac{|(\Phi - 0)|}{\Delta t} = \frac{10^{-3}}{\sqrt{2} \times 0.7} = 1.0\,\text{mV}$$

And the magnitude of the current is

$$I = \frac{\varepsilon}{R} = \frac{10^{-3}\,\text{V}}{0.5\Omega} = 2\,\text{mA}$$

Note that the earth's magnetic field also produces a flux through the loop. But it is a steady field (which does not change within the time span of the experiment) and hence does not induce any emf.

EXAMPLE 6.2

Example 6.3

A circular coil of radius 10 cm, 500 turns and resistance 2 Ω is placed with its plane perpendicular to the horizontal component of the earth's magnetic field. It is rotated about its vertical diameter through 180° in 0.25 s. Estimate the magnitudes of the emf and current induced in the coil. Horizontal component of the earth's magnetic field at the place is 3.0×10^{-5} T.

Solution

Initial flux through the coil,

$$\Phi_{B\,(initial)} = BA\cos\theta$$
$$= 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 0°$$
$$= 3\pi \times 10^{-7}\,\text{Wb}$$

Final flux after the rotation,

$$\Phi_{B\,(final)} = 3.0 \times 10^{-5} \times (\pi \times 10^{-2}) \times \cos 180°$$
$$= -3\pi \times 10^{-7}\,\text{Wb}$$

Therefore, estimated value of the induced emf is,

$$\varepsilon = N\frac{\Delta\Phi}{\Delta t}$$

$$= 500 \times (6\pi \times 10^{-7})/0.25$$
$$= 3.8 \times 10^{-3}\,\text{V}$$

$I = \varepsilon/R = 1.9 \times 10^{-3}$ A

Note that the magnitudes of ε and I are the estimated values. Their instantaneous values are different and depend upon the speed of rotation at the particular instant.

EXAMPLE 6.3

6.5 Lenz's Law and Conservation of Energy

In 1834, German physicist Heinrich Friedrich Lenz (1804-1865) deduced a rule, known as *Lenz's law* which gives the polarity of the induced emf in a clear and concise fashion. The statement of the law is:

The polarity of induced emf is such that it tends to produce a current which opposes the change in magnetic flux that produced it.

The negative sign shown in Eq. (6.3) represents this effect. We can understand Lenz's law by examining Experiment 6.1 in Section 6.2.1. In Fig. 6.1, we see that the North-pole of a bar magnet is being pushed towards the closed coil. As the North-pole of the bar magnet moves towards the coil, the magnetic flux through the coil increases. Hence current is induced in the coil in such a direction that it opposes the increase in flux. This is possible only if the current in the coil is in a counter-clockwise direction with respect to an observer situated on the side of the magnet. Note that magnetic moment associated with this current has North polarity towards the North-pole of the approaching magnet. Similarly, if the North-pole of the magnet is being withdrawn from the coil, the magnetic flux through the coil will decrease. To counter this decrease in magnetic flux, the induced current in the coil flows in clockwise direction and its South-pole faces the receding North-pole of the bar magnet. This would result in an attractive force which opposes the motion of the magnet and the corresponding decrease in flux.

What will happen if an open circuit is used in place of the closed loop in the above example? In this case too, an emf is induced across the open ends of the circuit. The direction of the induced emf can be found using Lenz's law. Consider Figs. 6.6 (a) and (b). They provide an easier way to understand the direction of induced currents. Note that the direction shown by $\circlearrowleft$ and $\circlearrowright$ indicate the directions of the induced currents.

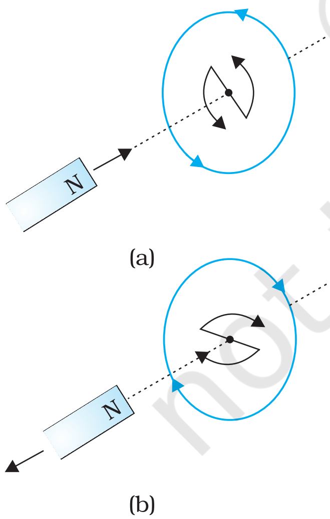


FIGURE 6.6 Illustration of Lenz's law.

A little reflection on this matter should convince us on the correctness of Lenz's law. Suppose that the induced current was in the direction opposite to the one depicted in Fig. 6.6(a). In that case, the South-pole due to the induced current will face the approaching North-pole of the magnet. The bar magnet will then be attracted towards the coil at an ever increasing acceleration. A gentle push on the magnet will initiate the process and its velocity and kinetic energy will continuously increase without expending any energy. If this can happen, one could construct a perpetual-motion machine by a suitable arrangement. This violates the law of conservation of energy and hence can not happen.

Now consider the correct case shown in Fig. 6.6(a). In this situation, the bar magnet experiences a repulsive force due to the induced current. Therefore, a person has to do work in moving the magnet. Where does the energy spent by the person go? This energy is dissipated by Joule heating produced by the induced current.

Reprint 2025-26

Example 6.4
Figure 6.7 shows planar loops of different shapes moving out of or into a region of a magnetic field which is directed normal to the plane of the loop away from the reader. Determine the direction of induced current in each loop using Lenz's law.

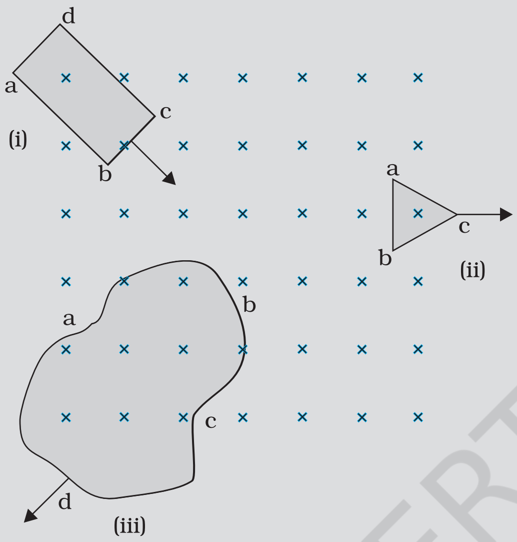


FIGURE 6.7

Solution
(i) The magnetic flux through the rectangular loop abcd increases, due to the motion of the loop into the region of magnetic field, The induced current must flow along the path bcdab so that it opposes the increasing flux.
(ii) Due to the outward motion, magnetic flux through the triangular loop abc decreases due to which the induced current flows along bacb, so as to oppose the change in flux.
(iii) As the magnetic flux decreases due to motion of the irregular shaped loop abcd out of the region of magnetic field, the induced current flows along cdabc, so as to oppose change in flux.
Note that there are no induced current as long as the loops are completely inside or outside the region of the magnetic field.

EXAMPLE 6.4

Example 6.5
(a) A closed loop is held stationary in the magnetic field between the north and south poles of two permanent magnets held fixed. Can we hope to generate current in the loop by using very strong magnets?
(b) A closed loop moves normal to the constant electric field between the plates of a large capacitor. Is a current induced in the loop (i) when it is wholly inside the region between the capacitor plates (ii) when it is partially outside the plates of the capacitor? The electric field is normal to the plane of the loop.
(c) A rectangular loop and a circular loop are moving out of a uniform magnetic field region (Fig. 6.8) to a field-free region with a *constant velocity* **v**. In which loop do you expect the induced emf to be constant *during* the passage out of the field region? The field is normal to the loops.

EXAMPLE 6.5

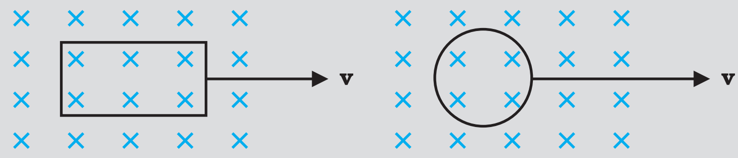


FIGURE 6.8

(d) Predict the polarity of the capacitor in the situation described by Fig. 6.9.

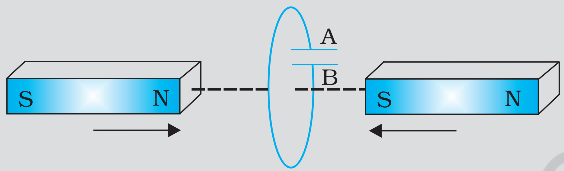


FIGURE 6.9

Solution

(a) No. However strong the magnet may be, current can be induced only by changing the magnetic flux through the loop.

(b) No current is induced in *either* case. Current can not be induced by changing the electric flux.

(c) The induced emf is expected to be constant only in the case of the rectangular loop. In the case of circular loop, the rate of change of area of the loop during its passage out of the field region is not constant, hence induced *emf* will vary accordingly.

(d) The polarity of plate 'A' will be positive with respect to plate 'B' in the capacitor.

EXAMPLE 6.5

6.6 Motional Electromotive Force

Let us consider a straight conductor moving in a uniform and time-independent magnetic field. Figure 6.10 shows a rectangular conductor PQRS in which the conductor PQ is free to move. The rod PQ is moved towards the left with a constant velocity **v** as shown in the figure. Assume that there is no loss of energy due to friction. PQRS forms a closed circuit enclosing an area that changes as PQ moves. It is placed in a uniform magnetic field **B** which is perpendicular to the plane of this system. If the length RQ = x and RS = l, the magnetic flux Φ_B enclosed by the loop PQRS will be

$$\Phi_B = Blx$$

Since x is changing with time, the rate of change of flux Φ_B will induce an emf given by:

$$\varepsilon = \frac{-\mathrm{d}\Phi_B}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}(Blx)$$

$$= -Bl\frac{\mathrm{d}x}{\mathrm{d}t} = Blv \qquad (6.5)$$

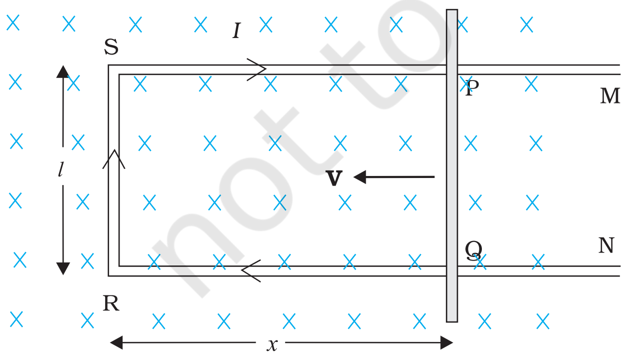


FIGURE 6.10 The arm PQ is moved to the left side, thus decreasing the area of the rectangular loop. This movement induces a current I as shown.

where we have used $dx/dt = -v$ which is the speed of the conductor PQ. The induced emf *Blv* is called *motional emf*. Thus, we are able to produce induced emf by moving a conductor instead of varying the magnetic field, that is, by changing the magnetic flux enclosed by the circuit.

It is also possible to explain the motional emf expression in Eq. (6.5) by invoking the Lorentz force acting on the free charge carriers of conductor PQ. Consider any arbitrary charge q in the conductor PQ. When the rod moves with speed v, the charge will also be moving with speed v in the magnetic field **B**. The Lorentz force on this charge is qvB in magnitude, and its direction is towards Q. All charges experience the same force, in magnitude and direction, irrespective of their position in the rod PQ. The work done in moving the charge from P to Q is,

$$W = qvBl$$

Since emf is the work done per unit charge,

$$\varepsilon = \frac{W}{q}$$

$$= Blv$$

This equation gives emf induced across the rod PQ and is identical to Eq. (6.5). We stress that our presentation is not wholly rigorous. But it does help us to understand the basis of Faraday's law when the conductor is moving in a uniform and time-independent magnetic field.

On the other hand, it is not obvious how an emf is induced when a conductor is stationary and the magnetic field is changing – a fact which Faraday verified by numerous experiments. In the case of a stationary conductor, the force on its charges is given by

$$\mathbf{F} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = q\mathbf{E} \tag{6.6}$$

since $\mathbf{v} = 0$. Thus, any force on the charge must arise from the electric field term **E** alone. Therefore, to explain the existence of induced emf or induced current, we must assume that a time-varying magnetic field generates an electric field. However, we hasten to add that electric fields produced by static electric charges have properties different from those produced by time-varying magnetic fields. In Chapter 4, we learnt that charges in motion (current) can exert force/torque on a stationary magnet. Conversely, a bar magnet in motion (or more generally, a changing magnetic field) can exert a force on the stationary charge. This is the fundamental significance of the Faraday's discovery. Electricity and magnetism are related.

Example 6.6 A metallic rod of 1 m length is rotated with a frequency of 50 rev/s, with one end hinged at the centre and the other end at the circumference of a circular metallic ring of radius 1 m, about an axis passing through the centre and perpendicular to the plane of the ring (Fig. 6.11). A constant and uniform magnetic field of 1 T parallel to the axis is present everywhere. What is the emf between the centre and the metallic ring?

EXAMPLE 6.6

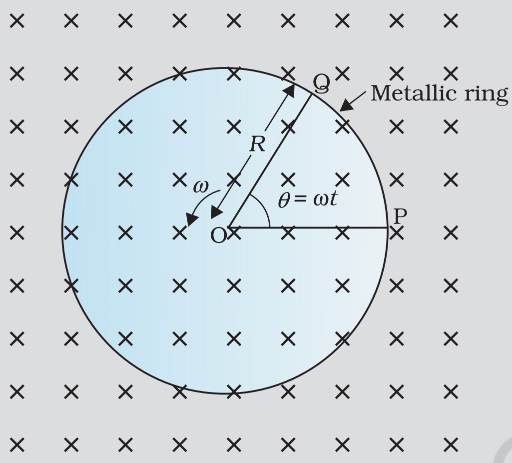


FIGURE 6.11

Solution

Method I

As the rod is rotated, free electrons in the rod move towards the outer end due to Lorentz force and get distributed over the ring. Thus, the resulting separation of charges produces an emf across the ends of the rod. At a certain value of emf, there is no more flow of electrons and a steady state is reached. Using Eq. (6.5), the magnitude of the emf generated across a length d*r* of the rod as it moves at right angles to the magnetic field is given by

$d\varepsilon = Bv\,dr$. Hence,

$$\varepsilon = \int d\varepsilon = \int_0^R Bv\,dr = \int_0^R B\omega r\,dr = \frac{B\omega R^2}{2}$$

Note that we have used $v = \omega\, r$. This gives

$$\varepsilon = \frac{1}{2} \times 1.0 \times 2\pi \times 50 \times (1^2)$$

= 157 V

Method II

To calculate the emf, we can imagine a closed loop OPQ in which point O and P are connected with a resistor *R* and OQ is the rotating rod. The potential difference across the resistor is then equal to the induced emf and equals *B* × (rate of change of area of loop). If θ is the angle between the rod and the radius of the circle at P at time *t*, the area of the sector OPQ is given by

$$\pi R^2 \times \frac{\theta}{2\pi} = \frac{1}{2} R^2 \theta$$

where *R* is the radius of the circle. Hence, the induced emf is

$$\varepsilon = B \times \frac{d}{dt}\left[\frac{1}{2} R^2 \theta\right] = \frac{1}{2} BR^2 \frac{d\theta}{dt} = \frac{B\omega R^2}{2}$$

[Note: $\frac{d\theta}{dt} = \omega = 2\pi\nu$]

This expression is identical to the expression obtained by Method I and we get the same value of ε.

EXAMPLE 6.6

Example 6.7
A wheel with 10 metallic spokes each 0.5 m long is rotated with a speed of 120 rev/min in a plane normal to the horizontal component of earth's magnetic field H_E at a place. If H_E = 0.4 G at the place, what is the induced emf between the axle and the rim of the wheel? Note that 1 G = 10^{-4} T.

Solution

Induced emf = $(1/2)\ \omega B R^2$

$$= (1/2) \times 4\pi \times 0.4 \times 10^{-4} \times (0.5)^2$$

$$= 6.28 \times 10^{-5}\ \text{V}$$

The number of spokes is immaterial because the emf's across the spokes are *in parallel*.

EXAMPLE 6.7

6.7 INDUCTANCE

An electric current can be induced in a coil by flux change produced by another coil in its vicinity or flux change produced by the same coil. These two situations are described separately in the next two sub-sections. However, in both the cases, the flux through a coil is proportional to the current. That is, $\Phi_B \propto I$.

Further, if the geometry of the coil does not vary with time then,

$$\frac{d\Phi_B}{dt} \propto \frac{dI}{dt}$$

For a closely wound coil of N turns, the same magnetic flux is linked with all the turns. When the flux Φ_B through the coil changes, each turn contributes to the induced emf. Therefore, a term called *flux linkage* is used which is equal to $N\Phi_B$ for a closely wound coil and in such a case

$$N\Phi_B \propto I$$

The constant of proportionality, in this relation, is called *inductance*. We shall see that inductance depends only on the geometry of the coil and intrinsic material properties. This aspect is akin to capacitance which for a parallel plate capacitor depends on the plate area and plate separation (geometry) and the dielectric constant K of the intervening medium (intrinsic material property).

Inductance is a scalar quantity. It has the dimensions of $[\text{M L}^2\text{T}^{-2}\text{A}^{-2}]$ given by the dimensions of flux divided by the dimensions of current. The SI unit of inductance is *henry* and is denoted by H. It is named in honour of Joseph Henry who discovered electromagnetic induction in USA, independently of Faraday in England.

6.7.1 Mutual inductance

Consider Fig. 6.12 which shows two long co-axial solenoids each of length l. We denote the radius of the inner solenoid S_1 by r_1 and the number of turns per unit length by n_1. The corresponding quantities for the outer solenoid S_2 are r_2 and n_2, respectively. Let N_1 and N_2 be the total number of turns of coils S_1 and S_2, respectively.

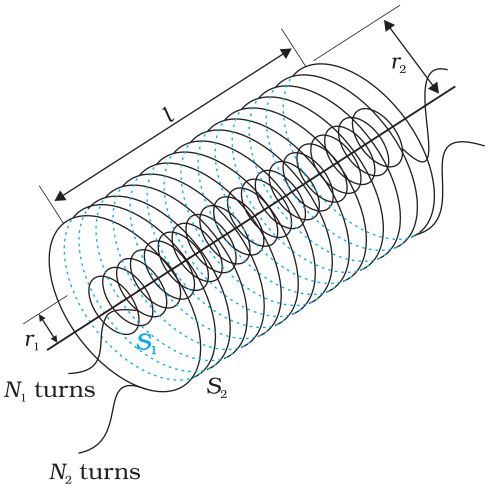


FIGURE 6.12 Two long co-axial solenoids of same length l.

When a current I_2 is set up through S_2, it in turn sets up a magnetic flux through S_1. Let us denote it by Φ_1. The corresponding flux linkage with solenoid S_1 is

$$N_1 \Phi_1 = M_{12} I_2 \tag{6.7}$$

M_{12} is called the *mutual inductance* of solenoid S_1 with respect to solenoid S_2. It is also referred to as the *coefficient of mutual induction*.

For these simple co-axial solenoids it is possible to calculate M_{12}. The magnetic field due to the current I_2 in S_2 is $\mu_0 n_2 I_2$. The resulting flux linkage with coil S_1 is,

$$N_1 \Phi_1 = (n_1 l)(\pi r_1^2)(\mu_0 n_2 I_2)$$

$$= \mu_0 n_1 n_2 \pi r_1^2 l I_2 \tag{6.8}$$

where $n_1 l$ is the total number of turns in solenoid S_1. Thus, from Eq. (6.7) and Eq. (6.8),

$$M_{12} = \mu_0 n_1 n_2 \pi r_1^2 l \tag{6.9}$$

Note that we neglected the edge effects and considered the magnetic field $\mu_0 n_2 I_2$ to be uniform throughout the length and width of the solenoid S_2. This is a good approximation keeping in mind that the solenoid is long, implying $l >> r_2$.

We now consider the reverse case. A current I_1 is passed through the solenoid S_1 and the flux linkage with coil S_2 is,

$$N_2 \Phi_2 = M_{21} I_1 \tag{6.10}$$

M_{21} is called the *mutual inductance* of solenoid S_2 with respect to solenoid S_1.

The flux due to the current I_1 in S_1 can be assumed to be confined solely inside S_1 since the solenoids are very long. Thus, flux linkage with solenoid S_2 is

$$N_2 \Phi_2 = (n_2 l)(\pi r_1^2)(\mu_0 n_1 I_1)$$

where $n_2 l$ is the total number of turns of S_2. From Eq. (6.10),

$$M_{21} = \mu_0 n_1 n_2 \pi r_1^2 l \tag{6.11}$$

Using Eq. (6.9) and Eq. (6.10), we get

$$M_{12} = M_{21} = M \text{ (say)} \tag{6.12}$$

We have demonstrated this equality for long co-axial solenoids. However, the relation is far more general. Note that if the inner solenoid was much shorter than (and placed well inside) the outer solenoid, then we could still have calculated the flux linkage $N_1 \Phi_1$ because the inner solenoid is effectively immersed in a uniform magnetic field due to the outer solenoid. In this case, the calculation of M_{12} would be easy. However, it would be extremely difficult to calculate the flux linkage with the outer solenoid as the magnetic field due to the inner solenoid would vary across the length as well as cross section of the outer solenoid. Therefore, the calculation of M_{21} would also be extremely difficult in this case. The equality M_{12}=M_{21} is very useful in such situations.

We explained the above example with air as the medium within the solenoids. Instead, if a medium of relative permeability μ_r had been present, the mutual inductance would be

$$M = \mu_r \mu_0 \, n_1 n_2 \pi \, r_1^2 \, l$$

It is also important to know that the mutual inductance of a pair of coils, solenoids, etc., depends on their separation as well as their relative orientation.

Example 6.8 Two concentric circular coils, one of small radius r_1 and the other of large radius r_2, such that $r_1 << r_2$, are placed co-axially with centres coinciding. Obtain the mutual inductance of the arrangement.

Solution Let a current I_2 flow through the outer circular coil. The field at the centre of the coil is $B_2 = \mu_0 I_2 / 2r_2$. Since the other co-axially placed coil has a very small radius, B_2 may be considered constant over its cross-sectional area. Hence,

$$\Phi_1 = \pi r_1^2 B_2$$

$$= \frac{\mu_0 \pi r_1^2}{2r_2} I_2$$

$$= M_{12} I_2$$

Thus,

$$M_{12} = \frac{\mu_0 \pi r_1^2}{2r_2}$$

From Eq. (6.12)

$$M_{12} = M_{21} = \frac{\mu_0 \pi r_1^2}{2r_2}$$

Note that we calculated M_{12} from an approximate value of Φ_1, assuming the magnetic field B_2 to be uniform over the area πr_1^2. However, we can accept this value because $r_1 << r_2$.

EXAMPLE 6.8

Now, let us recollect Experiment 6.3 in Section 6.2. In that experiment, emf is induced in coil C_1 wherever there was any change in current through coil C_2. Let Φ_1 be the flux through coil C_1 (say of N_1 turns) when current in coil C_2 is I_2.

Then, from Eq. (6.7), we have

$$N_1 \Phi_1 = M I_2$$

For currents varrying with time,

$$\frac{d(N_1 \Phi_1)}{dt} = \frac{d(M I_2)}{dt}$$

Since induced emf in coil C_1 is given by

$$\varepsilon_1 = -\frac{d(N_1 \Phi_1)}{dt}$$

We get,

$$\varepsilon_1 = -M \frac{dI_2}{dt}$$

It shows that varying current in a coil can induce emf in a neighbouring coil. The magnitude of the induced emf depends upon the rate of change of current and mutual inductance of the two coils.

6.7.2 Self-inductance

In the previous sub-section, we considered the flux in one solenoid due to the current in the other. It is also possible that emf is induced in a single isolated coil due to change of flux through the coil by means of varying the current through the same coil. This phenomenon is called *self-induction*. In this case, flux linkage through a coil of N turns is proportional to the current through the coil and is expressed as

$$N\Phi_B \propto I$$

$$N\Phi_B = L\,I \tag{6.13}$$

where constant of proportionality L is called *self-inductance* of the coil. It is also called the *coefficient of self-induction* of the coil. When the current is varied, the flux linked with the coil also changes and an emf is induced in the coil. Using Eq. (6.13), the induced emf is given by

$$\varepsilon = -\frac{\mathrm{d}(N\Phi_B)}{\mathrm{d}t}$$

$$\varepsilon = -L\frac{\mathrm{d}I}{\mathrm{d}t} \tag{6.14}$$

Thus, the self-induced emf always opposes any change (increase or decrease) of current in the coil.

It is possible to calculate the self-inductance for circuits with simple geometries. Let us calculate the self-inductance of a long solenoid of cross-sectional area A and length l, having n turns per unit length. The magnetic field due to a current I flowing in the solenoid is $B = \mu_0\, n\, I$ (neglecting edge effects, as before). The total flux linked with the solenoid is

$$N\Phi_B = (nl)(\mu_0 n\, I)(A)$$

$$= \mu_0 n^2 Al\, I$$

where nl is the total number of turns. Thus, the self-inductance is,

$$L = \frac{N\Phi_B}{I}$$

$$= \mu_0 n^2 Al \tag{6.15}$$

If we fill the inside of the solenoid with a material of relative permeability μ_r (for example soft iron, which has a high value of relative permeability), then,

$$L = \mu_r\, \mu_0\, n^2 Al \tag{6.16}$$

The self-inductance of the coil depends on its geometry and on the permeability of the medium.

The self-induced emf is also called the *back emf* as it opposes any change in the current in a circuit. Physically, the *self-inductance plays* *the role of inertia*. It is the electromagnetic analogue of mass in mechanics. So, work needs to be done against the back emf (ε) in establishing the current. This work done is stored as magnetic potential energy. For the current I at an instant in a circuit, the rate of work done is

$$\frac{dW}{dt} = |\varepsilon| I$$

If we ignore the resistive losses and consider only inductive effect, then using Eq. (6.14),

$$\frac{dW}{dt} = L\, I \frac{dI}{dt}$$

Total amount of work done in establishing the current I is

$$W = \int dW = \int_0^I L\, I\, dI$$

Thus, the energy required to build up the current I is,

$$W = \frac{1}{2} L I^2 \qquad (6.17)$$

This expression reminds us of $mv^2/2$ for the (mechanical) kinetic energy of a particle of mass m, and shows that L is analogous to m (i.e., L is electrical inertia and opposes growth and decay of current in the circuit).

Consider the general case of currents flowing simultaneously in two nearby coils. The flux linked with one coil will be the sum of two fluxes which exist independently. Equation (6.7) would be modified into

$$N_1 \Phi_1 = M_{11} I_1 + M_{12} I_2$$

where M_{11} represents inductance due to the same coil.

Therefore, using Faraday's law,

$$\varepsilon_1 = -M_{11} \frac{dI_1}{dt} - M_{12} \frac{dI_2}{dt}$$

M_{11} is the *self-inductance* and is written as L_1. Therefore,

$$\varepsilon_1 = -L_1 \frac{dI_1}{dt} - M_{12} \frac{dI_2}{dt}$$

Example 6.9 (a) Obtain the expression for the magnetic energy stored in a solenoid in terms of magnetic field B, area A and length l of the solenoid. (b) How does this magnetic energy compare with the electrostatic energy stored in a capacitor?

Solution

(a) From Eq. (6.17), the magnetic energy is

$$U_B = \frac{1}{2} L I^2$$

$$= \frac{1}{2} L \left(\frac{B}{\mu_0 n}\right)^2 \qquad (\text{since } B = \mu_0 n I, \text{for a solenoid})$$

EXAMPLE 6.9

Interactive animation on ac generator:
http://micro.magnet.fsu.edu/electromag/java/generator/ac.html

PHYSICS

EXAMPLE 6.9

$$= \frac{1}{2}(\mu_0 n^2 Al)\left(\frac{B}{\mu_0 n}\right)^2 \qquad \text{[from Eq. (6.15)]}$$

$$= \frac{1}{2\mu_0} B^2 Al$$

(b) The magnetic energy per unit volume is,

$$u_B = \frac{U_B}{V} \qquad \text{(where } V \text{ is volume that contains flux)}$$

$$= \frac{U_B}{Al}$$

$$= \frac{B^2}{2\mu_0} \tag{6.18}$$

We have already obtained the relation for the electrostatic energy stored per unit volume in a parallel plate capacitor (refer to Chapter 2, Eq. 2.73),

$$u_E = \frac{1}{2}\varepsilon_0 E^2 \tag{2.73}$$

In both the cases energy is proportional to the square of the field strength. Equations (6.18) and (2.73) have been derived for special cases: a solenoid and a parallel plate capacitor, respectively. But they are general and valid for any region of space in which a magnetic field or/and an electric field exist.

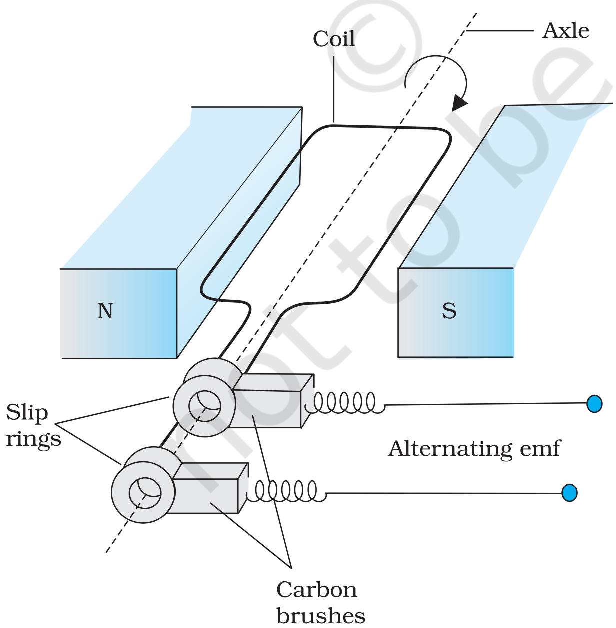


FIGURE 6.13 AC Generator

6.8 AC GENERATOR

The phenomenon of electromagnetic induction has been technologically exploited in many ways. An exceptionally important application is the generation of alternating currents (ac). The modern ac generator with a typical output capacity of 100 MW is a highly evolved machine. In this section, we shall describe the basic principles behind this machine. The Yugoslav inventor Nicola Tesla is credited with the development of the machine. As was pointed out in Section 6.3, one method to induce an emf or current in a loop is through a change in the loop's orientation or a change in its effective area. As the coil rotates in a magnetic field **B**, the effective area of the loop (the face perpendicular to the field) is $A \cos \theta$, where θ is the angle between **A** and **B**. This method of producing a flux change is the principle of operation of a simple ac generator. An ac generator converts mechanical energy into electrical energy.

The basic elements of an ac generator are shown in Fig. 6.13. It consists of a coil mounted on a rotor shaft. The axis of rotation of the coil is perpendicular to the direction of the magnetic field. The coil (called armature) is mechanically rotated in the uniform magnetic field by some external means. The rotation of the coil causes the magnetic flux through it to change, so an emf is induced in the coil. The ends of the coil are connected to an external circuit by means of slip rings and brushes.

When the coil is rotated with a constant angular speed ω, the angle θ between the magnetic field vector **B** and the area vector **A** of the coil at any instant t is $\theta = \omega t$ (assuming $\theta = 0°$ at $t = 0$). As a result, the effective area of the coil exposed to the magnetic field lines changes with time, and from Eq. (6.1), the flux at any time t is

$$\Phi_B = BA \cos\theta = BA \cos\omega t$$

From Faraday's law, the induced emf for the rotating coil of N turns is then,

$$\varepsilon = -N\frac{d\Phi_B}{dt} = -NBA\frac{d}{dt}(\cos\omega t)$$

Thus, the instantaneous value of the emf is

$$\varepsilon = NBA\,\omega \sin\omega t \qquad (6.19)$$

where $NBA\omega$ is the maximum value of the emf, which occurs when $\sin\omega t = \pm 1$. If we denote $NBA\omega$ as ε_0, then

$$\varepsilon = \varepsilon_0 \sin\omega t \qquad (6.20)$$

Since the value of the sine fuction varies between +1 and –1, the sign, or polarity of the emf changes with time. Note from Fig. 6.14 that the emf has its extremum value when $\theta = 90°$ or $\theta = 270°$, as the change of flux is greatest at these points.

The direction of the current changes periodically and therefore the current is called *alternating current* (ac). Since $\omega = 2\pi\nu$, Eq (6.20) can be written as

$$\varepsilon = \varepsilon_0 \sin 2\pi\,\nu\,t \qquad (6.21)$$

where ν is the frequency of revolution of the generator's coil.

Note that Eq. (6.20) and (6.21) give the instantaneous value of the emf and ε varies between $+\varepsilon_0$ and $-\varepsilon_0$ periodically. We shall learn how to determine the time-averaged value for the alternating voltage and current in the next chapter.

In commercial generators, the mechanical energy required for rotation of the armature is provided by water falling from a height, for example, from dams. These are called *hydro-electric generators*. Alternatively, water is heated to produce steam using coal or other sources. The steam at high pressure produces the rotation of the armature. These are called *thermal generators*. Instead of coal, if a nuclear fuel is used, we get *nuclear power generators*. Modern day generators produce electric power as high as 500 MW, i.e., one can light

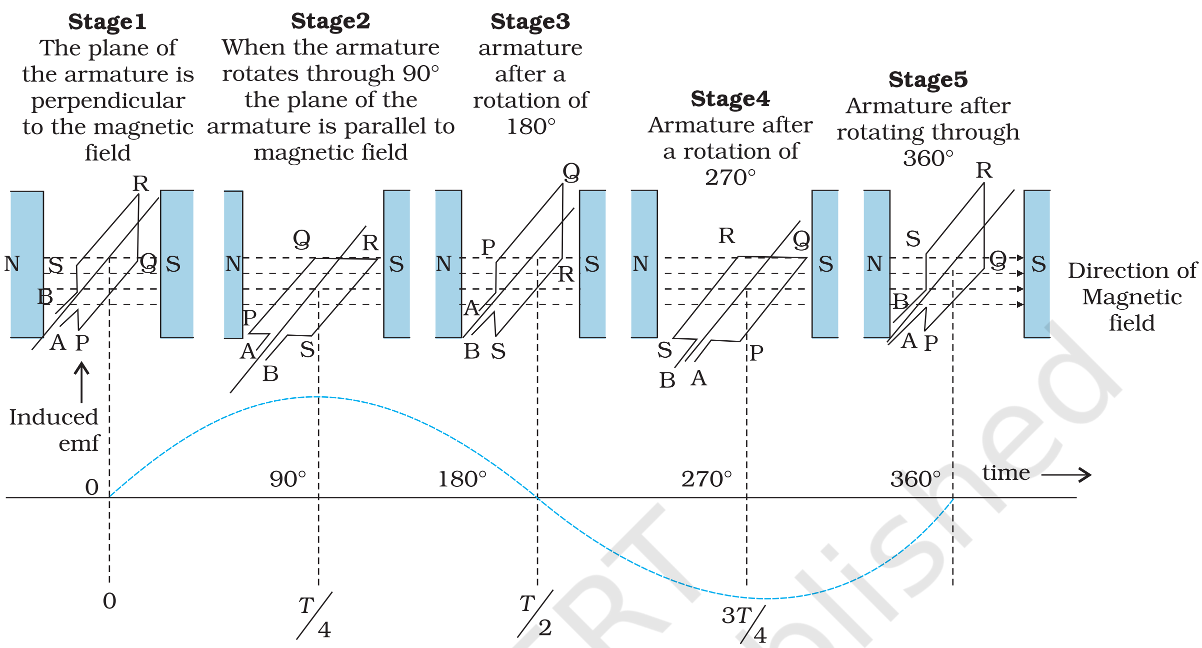


FIGURE 6.14 An alternating emf is generated by a loop of wire rotating in a magnetic field.

up 5 million 100 W bulbs! In most generators, the coils are held stationary and it is the electromagnets which are rotated. The frequency of rotation is 50 Hz in India. In certain countries such as USA, it is 60 Hz.

EXAMPLE 6.10

Example 6.10 Kamla peddles a stationary bicycle. The pedals of the bicycle are attached to a 100 turn coil of area 0.10 m^2. The coil rotates at half a revolution per second and it is placed in a uniform magnetic field of 0.01 T perpendicular to the axis of rotation of the coil. What is the maximum voltage generated in the coil?

Solution Here $\nu = 0.5$ Hz; $N = 100$, $A = 0.1$ m^2 and $B = 0.01$ T. Employing Eq. (6.19)

$$\varepsilon_0 = NBA\,(2\,\pi\,\nu)$$

$$= 100 \times 0.01 \times 0.1 \times 2 \times 3.14 \times 0.5$$

$$= 0.314 \text{ V}$$

The maximum voltage is 0.314 V.

We urge you to explore such alternative possibilities for power generation.

SUMMARY

1. The magnetic flux through a surface of area **A** placed in a uniform magnetic field **B** is defined as,

   $\Phi_B = \mathbf{B.A} = BA \cos\theta$

   where θ is the angle between **B** and **A**.
2. Faraday's laws of induction imply that the emf induced in a coil of N turns is directly related to the rate of change of flux through it,

   $$\varepsilon = -N\frac{d\Phi_B}{dt}$$

   Here Φ_B is the flux linked with one turn of the coil. If the circuit is closed, a current $I = \varepsilon/R$ is set up in it, where R is the resistance of the circuit.
3. Lenz's law states that the polarity of the induced emf is such that it tends to produce a current which opposes the change in magnetic flux that produces it. The negative sign in the expression for Faraday's law indicates this fact.
4. When a metal rod of length l is placed normal to a uniform magnetic field B and moved with a velocity v perpendicular to the field, the induced emf (called motional emf) across its ends is

   $\varepsilon = Blv$
5. Inductance is the ratio of the flux-linkage to current. It is equal to $N\Phi/I$.
6. A changing current in a coil (coil 2) can induce an emf in a nearby coil (coil 1). This relation is given by,

   $$\varepsilon_1 = -M_{12}\frac{dI_2}{dt}$$

   The quantity M_{12} is called mutual inductance of coil 1 with respect to coil 2. One can similarly define M_{21}. There exists a general equality,

   $M_{12} = M_{21}$
7. When a current in a coil changes, it induces a back emf in the same coil. The self-induced emf is given by,

   $$\varepsilon = -L\frac{dI}{dt}$$

   L is the self-inductance of the coil. It is a measure of the inertia of the coil against the change of current through it.
8. The self-inductance of a long solenoid, the core of which consists of a magnetic material of relative permeability μ_r, is given by

   $L = \mu_r\, \mu_0\, n^2 A\, l$

   where A is the area of cross-section of the solenoid, l its length and n the number of turns per unit length.
9. In an ac generator, mechanical energy is converted to electrical energy by virtue of electromagnetic induction. If coil of N turn and area A is rotated at ν revolutions per second in a uniform magnetic field B, then the motional emf produced is

   $\varepsilon = NBA\,(2\pi\nu)\sin(2\pi\nu t)$

   where we have assumed that at time $t = 0$ s, the coil is perpendicular to the field.

Quantity	Symbol	Units	Dimensions	Equations
Magnetic Flux	Φ_B	Wb (weber)	$[M L^2 T^{-2} A^{-1}]$	$\Phi_B = \mathbf{B} \cdot \mathbf{A}$
EMF	ε	V (volt)	$[M L^2 T^{-3} A^{-1}]$	$\varepsilon = -d(N\Phi_B)/dt$
Mutual Inductance	M	H (henry)	$[M L^2 T^{-2} A^{-2}]$	$\varepsilon_1 = -M_{12}(dI_2/dt)$
Self Inductance	L	H (henry)	$[M L^2 T^{-2} A^{-2}]$	$\varepsilon = -L(dI/dt)$

POINTS TO PONDER

1. Electricity and magnetism are intimately related. In the early part of the nineteenth century, the experiments of Oersted, Ampere and others established that moving charges (currents) produce a magnetic field. Somewhat later, around 1830, the experiments of Faraday and Henry demonstrated that a moving magnet can induce electric current.
2. In a closed circuit, electric currents are induced so as to oppose the changing magnetic flux. It is as per the law of conservation of energy. However, in case of an open circuit, an emf is induced across its ends. How is it related to the flux change?
3. The motional emf discussed in Section 6.5 can be argued independently from Faraday's law using the Lorentz force on moving charges. However, even if the charges are stationary [and the q ($\mathbf{v} \times \mathbf{B}$) term of the Lorentz force is not operative], an emf is nevertheless induced in the presence of a time-varying magnetic field. Thus, moving charges in static field and static charges in a time-varying field seem to be symmetric situation for Faraday's law. This gives a tantalising hint on the relevance of the principle of relativity for Faraday's law.

EXERCISES

6.1 Predict the direction of induced current in the situations described by the following Figs. 6.15(a) to (f).

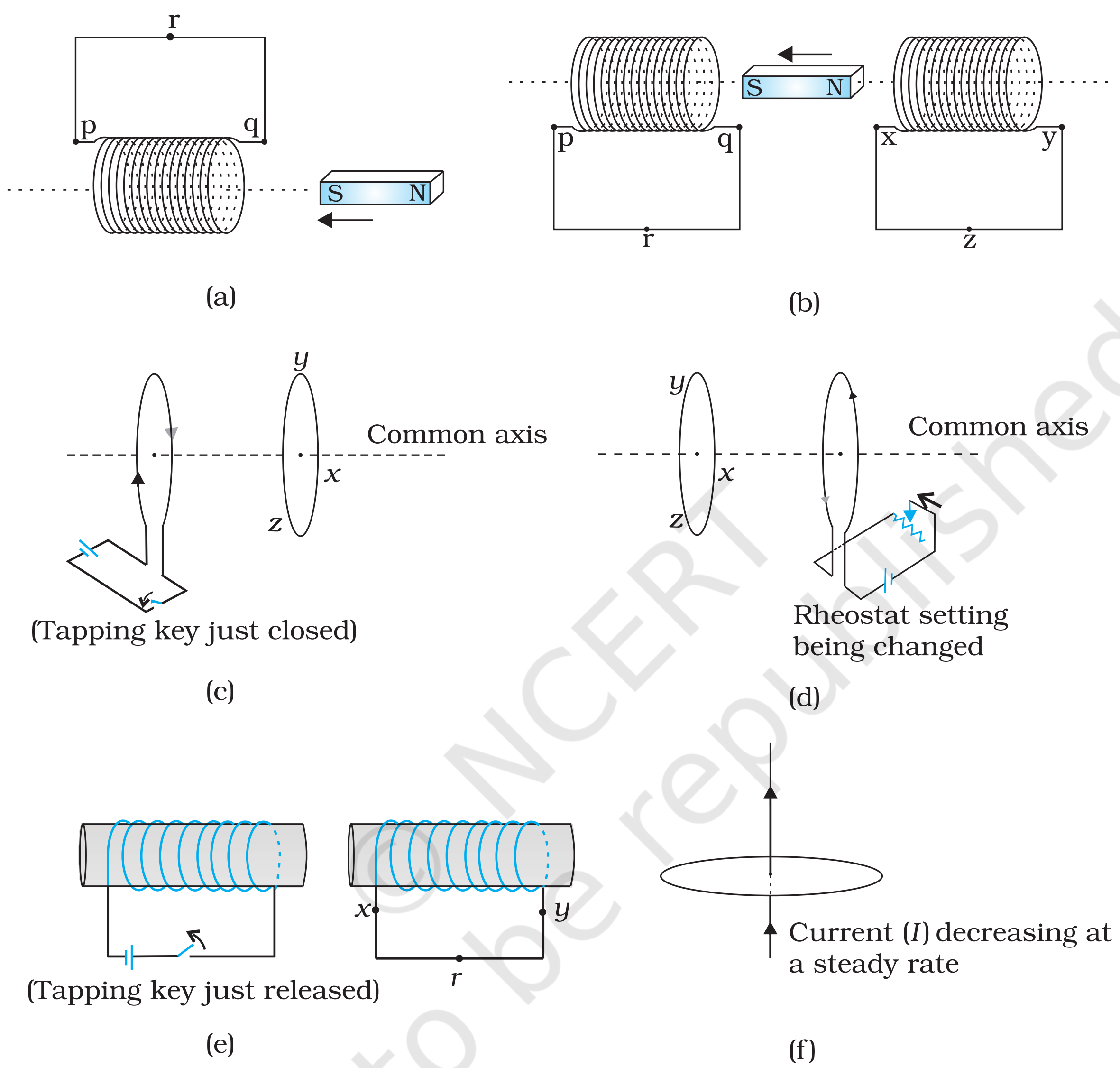


FIGURE 6.15

6.2 Use Lenz's law to determine the direction of induced current in the situations described by Fig. 6.16:

(a) A wire of irregular shape turning into a circular shape;

(b) A circular loop being deformed into a narrow straight wire.

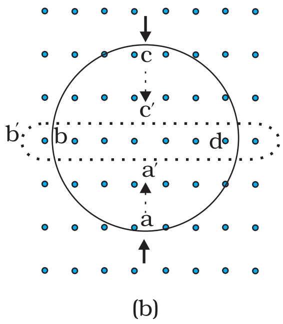


FIGURE 6.16

6.3 A long solenoid with 15 turns per cm has a small loop of area 2.0 cm^2 placed inside the solenoid normal to its axis. If the current carried by the solenoid changes steadily from 2.0 A to 4.0 A in 0.1 s, what is the induced emf in the loop while the current is changing?

6.4 A rectangular wire loop of sides 8 cm and 2 cm with a small cut is moving out of a region of uniform magnetic field of magnitude 0.3 T directed normal to the loop. What is the emf developed across the cut if the velocity of the loop is 1 cm s^{-1} in a direction normal to the (a) longer side, (b) shorter side of the loop? For how long does the induced voltage last in each case?

6.5 A 1.0 m long metallic rod is rotated with an angular frequency of 400 rad s^{-1} about an axis normal to the rod passing through its one end. The other end of the rod is in contact with a circular metallic ring. A constant and uniform magnetic field of 0.5 T parallel to the axis exists everywhere. Calculate the emf developed between the centre and the ring.

6.6 A horizontal straight wire 10 m long extending from east to west is falling with a speed of 5.0 m s^{-1}, at right angles to the horizontal component of the earth's magnetic field, 0.30×10^{-4} Wb m^{-2}.
(a) What is the instantaneous value of the emf induced in the wire?
(b) What is the direction of the emf?
(c) Which end of the wire is at the higher electrical potential?

6.7 Current in a circuit falls from 5.0 A to 0.0 A in 0.1 s. If an average emf of 200 V induced, give an estimate of the self-inductance of the circuit.

6.8 A pair of adjacent coils has a mutual inductance of 1.5 H. If the current in one coil changes from 0 to 20 A in 0.5 s, what is the change of flux linkage with the other coil?

Chapter Seven

ALTERNATING CURRENT

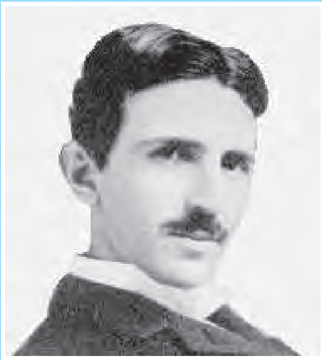
7.1 Introduction

We have so far considered direct current (dc) sources and circuits with dc sources. These currents do not change direction with time. But voltages and currents that vary with time are very common. The electric mains supply in our homes and offices is a voltage that varies like a sine function with time. Such a voltage is called *alternating voltage* (ac voltage) and the current driven by it in a circuit is called the *alternating current* (ac current)*. Today, most of the electrical devices we use require ac voltage. This is mainly because most of the electrical energy sold by power companies is transmitted and distributed as alternating current. The main reason for preferring use of ac voltage over dc voltage is that ac voltages can be easily and efficiently converted from one voltage to the other by means of transformers. Further, electrical energy can also be transmitted economically over long distances. AC circuits exhibit characteristics which are exploited in many devices of daily use. For example, whenever we tune our radio to a favourite station, we are taking advantage of a special property of ac circuits – one of many that you will study in this chapter.

---

* The phrases *ac voltage* and *ac current* are contradictory and redundant, respectively, since they mean, literally, *alternating current voltage* and *alternating current current*. Still, the abbreviation *ac* to designate an electrical quantity displaying simple harmonic time dependance has become so universally accepted that we follow others in its use. Further, *voltage* – another phrase commonly used means potential difference between two points.

NICOLA TESLA (1856 – 1943)



Nicola Tesla (1856 – 1943) Serbian-American scientist, inventor and genius. He conceived the idea of the rotating magnetic field, which is the basis of practically all alternating current machinery, and which helped usher in the age of electric power. He also invented among other things the induction motor, the polyphase system of ac power, and the high frequency induction coil (the Tesla coil) used in radio and television sets and other electronic equipment. The SI unit of magnetic field is named in his honour.

7.2 AC Voltage Applied to a Resistor

Figure 7.1 shows a resistor connected to a source ε of ac voltage. The symbol for an ac source in a circuit diagram is ⊝. We consider a source which produces sinusoidally varying potential difference across its terminals. Let this potential difference, also called ac voltage, be given by

$$v = v_m \sin \omega t \tag{7.1}$$

where v_m is the amplitude of the oscillating potential difference and ω is its angular frequency.

FIGURE 7.1 AC voltage applied to a resistor.

To find the value of current through the resistor, we apply Kirchhoff's loop rule $\sum \varepsilon(t) = 0$ (refer to Section 3.12), to the circuit shown in Fig. 7.1 to get

$$v_m \sin \omega t = i\,R$$

or $$i = \frac{v_m}{R} \sin \omega t$$

Since R is a constant, we can write this equation as

$$i = i_m \sin \omega t \tag{7.2}$$

where the current amplitude i_m is given by

$$i_m = \frac{v_m}{R} \tag{7.3}$$

Equation (7.3) is Ohm's law, which for resistors, works equally well for both ac and dc voltages. The voltage across a pure resistor and the current through it, given by Eqs. (7.1) and (7.2) are plotted as a function of time in Fig. 7.2. Note, in particular that both v and i reach zero, minimum and maximum values at the same time. Clearly, *the voltage* and *current are in phase with each other.*

We see that, like the applied voltage, the current varies sinusoidally and has corresponding positive and negative values during each cycle. Thus, the sum of the instantaneous current values over one complete cycle is zero, and the average current is zero. The fact that the average current is zero, however, does

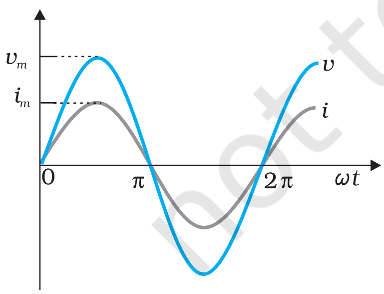


FIGURE 7.2 In a pure resistor, the voltage and current are in phase. The minima, zero and maxima occur at the same respective times.

Reprint 2025-26

not mean that the average power consumed is zero and that there is no dissipation of electrical energy. As you know, Joule heating is given by i^2R and depends on i^2 (which is always positive whether i is positive or negative) and not on i. Thus, there is Joule heating and dissipation of electrical energy when an ac current passes through a resistor.

The instantaneous power dissipated in the resistor is

$$p = i^2 R = i_m^2 R \sin^2 \omega t \tag{7.4}$$

The average value of p over a cycle is*

$$\bar{p} = \langle i^2 R \rangle = \langle i_m^2 R \sin^2 \omega t \rangle \qquad [7.5(a)]$$

where the bar over a letter (here, p) denotes its average value and <......> denotes taking average of the quantity inside the bracket. Since, i_m^2 and R are constants,

$$\bar{p} = i_m^2 R \langle \sin^2 \omega t \rangle \qquad [7.5(b)]$$

Using the trigonometric identity, $\sin^2 \omega t = 1/2\,(1 - \cos 2\omega t)$, we have $\langle \sin^2 \omega t \rangle = (1/2)\,(1 - \langle \cos 2\omega t \rangle)$ and since $\langle \cos 2\omega t \rangle = 0$**, we have,

$$\langle \sin^2 \omega t \rangle = \frac{1}{2}$$

Thus,

$$\bar{p} = \frac{1}{2} i_m^2 R \qquad [7.5(c)]$$

To express ac power in the same form as dc power ($P = I^2R$), a special value of current is defined and used. It is called, *root mean square* (rms) or *effective current* (Fig. 7.3) and is denoted by I_{rms} *or I*.

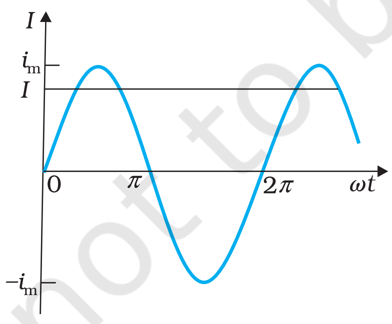
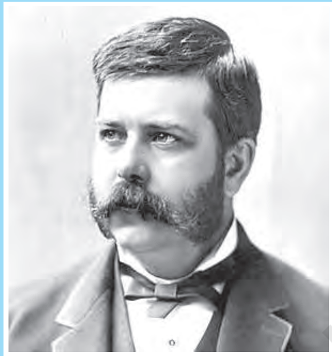


FIGURE 7.3 The rms current I is related to the peak current i_m by $I = i_m/\sqrt{2} = 0.707\, i_m$.



George Westinghouse (1846 – 1914) A leading proponent of the use of alternating current over direct current. Thus, he came into conflict with Thomas Alva Edison, an advocate of direct current. Westinghouse was convinced that the technology of alternating current was the key to the electrical future. He founded the famous Company named after him and enlisted the services of Nicola Tesla and other inventors in the development of alternating current motors and apparatus for the transmission of high tension current, pioneering in large scale lighting.

GEORGE WESTINGHOUSE (1846 – 1914)

* The average value of a function $F(t)$ over a period T is given by $\langle F(t) \rangle = \frac{1}{T}\int_0^T F(t)\,dt$

** $\langle \cos 2\omega t \rangle = \frac{1}{T}\int_0^T \cos 2\omega t\, dt = \frac{1}{T}\left[\frac{\sin 2\omega t}{2\omega}\right]_0^T = \frac{1}{2\omega T}[\sin 2\omega T - 0] = 0$

It is defined by

$$I=\sqrt{\overline{i^2}} = \sqrt{\frac{1}{2}i_m^2} = \frac{i_m}{\sqrt{2}}$$

$$= 0.707\ i_m \tag{7.6}$$

In terms of I, the average power, denoted by P is

$$P = \overline{p} = \frac{1}{2}i_m^2 R = I^2 R \tag{7.7}$$

Similarly, we define the *rms voltage* or *effective voltage* by

$$V = \frac{v_m}{\sqrt{2}} = 0.707\ v_m \tag{7.8}$$

From Eq. (7.3), we have

$$v_m = i_m R$$

or, $\dfrac{v_m}{\sqrt{2}} = \dfrac{i_m}{\sqrt{2}} R$

or, $V = IR$ (7.9)

Equation (7.9) gives the relation between ac current and ac voltage and is similar to that in the dc case. This shows the advantage of introducing the concept of rms values. In terms of rms values, the equation for power [Eq. (7.7)] and relation between current and voltage in ac circuits are essentially the same as those for the dc case.

It is customary to measure and specify rms values for ac quantities. For example, the household line voltage of 220 V is an rms value with a peak voltage of

$$v_m = \sqrt{2}\ V = (1.414)(220\ \text{V}) = 311\ \text{V}$$

In fact, the I or rms current is the equivalent dc current that would produce the same average power loss as the alternating current. Equation (7.7) can also be written as

$$P = V^2 / R = I\,V \quad (\text{since } V = I\,R)$$

Example 7.1 A light bulb is rated at 100W for a 220 V supply. Find (a) the resistance of the bulb; (b) the peak voltage of the source; and (c) the rms current through the bulb.

Solution

(a) We are given P = 100 W and V = 220 V. The resistance of the bulb is

$$R = \frac{V^2}{P} = \frac{(220\ \text{V})^2}{100\ \text{W}} = 484\,\Omega$$

(b) The peak voltage of the source is

$$v_m = \sqrt{2}V = 311\ \text{V}$$

(c) Since, $P = I\,V$

$$I = \frac{P}{V} = \frac{100\ \text{W}}{220\ \text{V}} = 0.454\,\text{A}$$

EXAMPLE 7.1

7.3 Representation of AC Current and Voltage by Rotating Vectors — Phasors

In the previous section, we learnt that the current through a resistor is in phase with the ac voltage. But this is not so in the case of an inductor, a capacitor or a combination of these circuit elements. In order to show phase relationship between voltage and current in an ac circuit, we use the notion of *phasors*. The analysis of an ac circuit is facilitated by the use of a phasor diagram. A phasor* is a vector which rotates about the origin with angular speed ω, as shown in Fig. 7.4. The vertical components of phasors **V** and **I** represent the sinusoidally varying quantities v and i. The magnitudes of phasors **V** and **I** represent the amplitudes or the peak values v_m and i_m of these oscillating quantities. Figure 7.4(a) shows the voltage and current phasors and their relationship at time t_1 for the case of an ac source connected to a resistor i.e., corresponding to the circuit shown in Fig. 7.1. The projection of voltage and current phasors on vertical axis, i.e., $v_m \sin\omega t$ and $i_m \sin\omega t$, respectively represent the value of voltage and current at that instant. As they rotate with frequency ω, curves in Fig. 7.4(b) are generated.

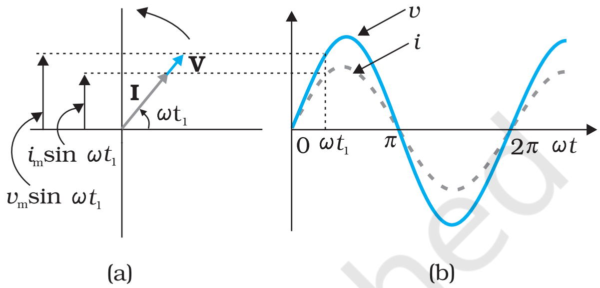


FIGURE 7.4 (a) A phasor diagram for the circuit in Fig 7.1. (b) Graph of v and i versus ωt.

From Fig. 7.4(a) we see that phasors **V** and **I** for the case of a resistor are in the same direction. This is so for all times. This means that the phase angle between the voltage and the current is zero.

7.4 AC Voltage Applied to an Inductor

Figure 7.5 shows an ac source connected to an inductor. Usually, inductors have appreciable resistance in their windings, but we shall assume that this inductor has negligible resistance. Thus, the circuit is a purely inductive ac circuit. Let the voltage across the source be $v = v_m \sin\omega t$. Using the Kirchhoff's loop rule, $\sum \varepsilon(t) = 0$, and since there is no resistor in the circuit,

$$v - L\frac{di}{dt} = 0 \tag{7.10}$$

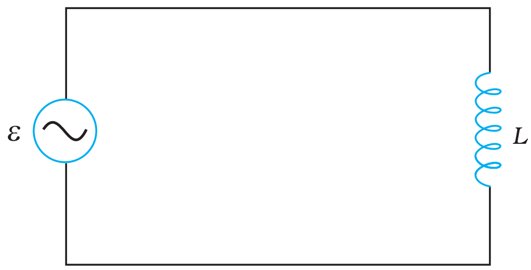


FIGURE 7.5 An ac source connected to an inductor.

where the second term is the self-induced Faraday emf in the inductor; and L is the self-inductance of

* Though voltage and current in ac circuit are represented by phasors – rotating vectors, they are not vectors themselves. They are scalar quantities. It so happens that the amplitudes and phases of harmonically varying scalars combine mathematically in the same way as do the projections of rotating vectors of corresponding magnitudes and directions. The *rotating vectors* that *represent* harmonically varying scalar quantities are introduced only to provide us with a simple way of adding these quantities using a rule that we already know.

the inductor. The negative sign follows from Lenz's law (Chapter 6). Combining Eqs. (7.1) and (7.10), we have

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \frac{v}{L} = \frac{v_m}{L}\sin\omega t \tag{7.11}$$

Equation (7.11) implies that the equation for $i(t)$, the current as a function of time, must be such that its slope $\mathrm{d}i/\mathrm{d}t$ is a sinusoidally varying quantity, with the same phase as the source voltage and an amplitude given by v_m/L. To obtain the current, we integrate $\mathrm{d}i/\mathrm{d}t$ with respect to time:

$$\int \frac{\mathrm{d}i}{\mathrm{d}t}\mathrm{d}t = \frac{v_m}{L}\int \sin(\omega t)\mathrm{d}t$$

and get,

$$i = -\frac{v_m}{\omega L}\cos(\omega t) + \text{constant}$$

The integration constant has the dimension of current and is time-independent. Since the source has an emf which oscillates symmetrically about zero, the current it sustains also oscillates symmetrically about zero, so that no constant or time-independent component of the current exists. Therefore, the integration constant is zero.

Using

$$-\cos(\omega t) = \sin\left(\omega t - \frac{\pi}{2}\right)$$, we have

$$i = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \tag{7.12}$$

where $i_m = \dfrac{v_m}{\omega L}$ is the amplitude of the current. The quantity ωL is analogous to the resistance and is called *inductive reactance*, denoted by X_L:

$$X_L = \omega L \tag{7.13}$$

The amplitude of the current is, then

$$i_m = \frac{v_m}{X_L} \tag{7.14}$$

The dimension of inductive reactance is the same as that of resistance and its SI unit is ohm (Ω). The inductive reactance limits the current in a purely inductive circuit in the same way as the resistance limits the current in a purely resistive circuit. The inductive reactance is directly proportional to the inductance and to the frequency of the current.

A comparison of Eqs. (7.1) and (7.12) for the source voltage and the current in an inductor shows that the current lags the voltage by $\pi/2$ or one-quarter (1/4) cycle. Figure 7.6 (a) shows the voltage and the current phasors in the present case at instant t_1. The current phasor **I** is $\pi/2$ behind the voltage phasor **V**. When rotated with frequency ω counter-clockwise, they generate the voltage and current given by Eqs. (7.1) and (7.12), respectively and as shown in Fig. 7.6(b).

PHYSICS

Interactive animation on Phasor diagrams of ac circuits containing, R, L, C and RLC series circuits:
http://www.animations.physics.unsw.edu.au//jw/AC.html

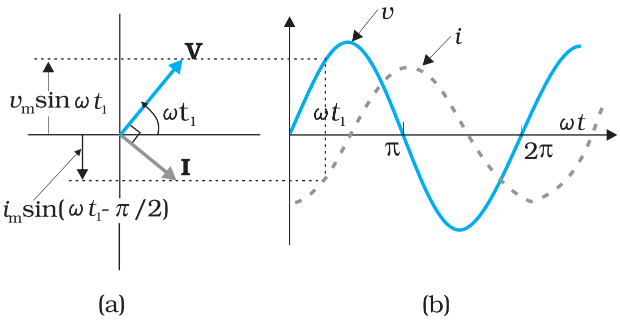


FIGURE 7.6 (a) A Phasor diagram for the circuit in Fig. 7.5. (b) Graph of v and i versus ωt.

We see that the current reaches its maximum value later than the voltage by one-fourth of a period $\left[\frac{T}{4} = \frac{\pi/2}{\omega}\right]$. You have seen that an inductor has reactance that limits current similar to resistance in a dc circuit. Does it also consume power like a resistance? Let us try to find out.

The instantaneous power supplied to the inductor is

$$p_L = i\,v = i_m \sin\left(\omega t - \frac{\pi}{2}\right) \times v_m \sin(\omega t)$$

$$= -i_m v_m \cos(\omega t) \sin(\omega t)$$

$$= -\frac{i_m v_m}{2} \sin(2\omega t)$$

So, the average power over a complete cycle is

$$P_L = \left\langle -\frac{i_m v_m}{2} \sin(2\omega t) \right\rangle$$

$$= -\frac{i_m v_m}{2} \langle \sin(2\omega t) \rangle = 0,$$

since the average of sin $(2\omega t)$ over a complete cycle is zero.

Thus, the *average power supplied to an inductor over one complete cycle is zero.*

Example 7.2 A pure inductor of 25.0 mH is connected to a source of 220 V. Find the inductive reactance and rms current in the circuit if the frequency of the source is 50 Hz.

Solution The inductive reactance,

$$X_L = 2\pi\nu L = 2 \times 3.14 \times 50 \times 25 \times 10^{-3}\ \Omega$$

$$= 7.85\Omega$$

The rms current in the circuit is

$$I = \frac{V}{X_L} = \frac{220\text{ V}}{7.85\ \Omega} = 28\text{A}$$

EXAMPLE 7.2

Reprint 2025-26

7.5 AC Voltage Applied to a Capacitor

Figure 7.7 shows an ac source ε generating ac voltage $v = v_m \sin \omega t$ connected to a capacitor only, a purely capacitive ac circuit.

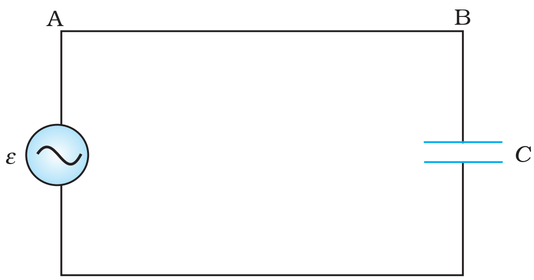


FIGURE 7.7 An ac source connected to a capacitor.

When a capacitor is connected to a voltage source in a dc circuit, current will flow for the short time required to charge the capacitor. As charge accumulates on the capacitor plates, the voltage across them increases, opposing the current. That is, a capacitor in a dc circuit will limit or oppose the current as it charges. When the capacitor is fully charged, the current in the circuit falls to zero.

When the capacitor is connected to an ac source, as in Fig. 7.7, it limits or regulates the current, but does not completely prevent the flow of charge. The capacitor is alternately charged and discharged as the current reverses each half cycle. Let q be the charge on the capacitor at any time t. The instantaneous voltage v across the capacitor is

$$v = \frac{q}{C} \tag{7.15}$$

From the Kirchhoff's loop rule, the voltage across the source and the capacitor are equal,

$$v_m \sin \omega t = \frac{q}{C}$$

To find the current, we use the relation $i = \frac{\mathrm{d}q}{\mathrm{d}t}$

$$i = \frac{\mathrm{d}}{\mathrm{d}t}\left(v_m C \sin \omega t\right) = \omega C v_m \cos(\omega t)$$

Using the relation, $\cos(\omega t) = \sin\left(\omega t + \frac{\pi}{2}\right)$, we have

$$i = i_m \sin\left(\omega t + \frac{\pi}{2}\right) \tag{7.16}$$

where the amplitude of the oscillating current is $i_m = \omega C v_m$. We can rewrite it as

$$i_m = \frac{v_m}{(1/\omega C)}$$

Comparing it to $i_m = v_m/R$ for a purely resistive circuit, we find that $(1/\omega C)$ plays the role of resistance. It is called *capacitive reactance* and is denoted by X_c,

$$X_c = 1/\omega C \tag{7.17}$$

so that the amplitude of the current is

$$i_m = \frac{v_m}{X_C} \tag{7.18}$$

The dimension of capacitive reactance is the same as that of resistance and its SI unit is ohm (Ω). The capacitive reactance limits the amplitude of the current in a purely capacitive circuit in the same way as the resistance limits the current in a purely resistive circuit. But it is inversely proportional to the frequency and the capacitance.

A comparison of Eq. (7.16) with the equation of source voltage, Eq. (7.1) shows that the current is $\pi/2$ ahead of voltage.

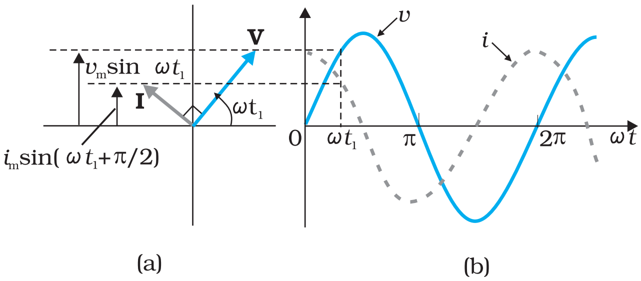


FIGURE 7.8 (a) A Phasor diagram for the circuit in Fig. 7.7. (b) Graph of v and i versus ωt.

Figure 7.8(a) shows the phasor diagram at an instant t_1. Here the current phasor **I** is $\pi/2$ ahead of the voltage phasor **V** as they rotate counterclockwise. Figure 7.8(b) shows the variation of voltage and current with time. We see that the current reaches its maximum value earlier than the voltage by one-fourth of a period.

The instantaneous power supplied to the capacitor is

$$p_c = i\,v = i_m \cos(\omega t)\, v_m \sin(\omega t)$$
$$= i_m v_m \cos(\omega t)\sin(\omega t)$$
$$= \frac{i_m v_m}{2}\sin(2\omega t) \qquad (7.19)$$

So, as in the case of an inductor, the average power

$$P_C = \left\langle \frac{i_m v_m}{2}\sin(2\omega t)\right\rangle = \frac{i_m v_m}{2}\langle \sin(2\omega t)\rangle = 0$$

since $\langle \sin(2\omega t)\rangle = 0$ over a complete cycle.

Thus, we see that in the case of an inductor, the current lags the voltage by $\pi/2$ and in the case of a capacitor, the current leads the voltage by $\pi/2$.

Example 7.3 A lamp is connected in series with a capacitor. Predict your observations for dc and ac connections. What happens in each case if the capacitance of the capacitor is reduced?

Solution When a dc source is connected to a capacitor, the capacitor gets charged and after charging no current flows in the circuit and the lamp will not glow. There will be no change even if C is reduced. With ac source, the capacitor offers capacitative reactance $(1/\omega C)$ and the current flows in the circuit. Consequently, the lamp will shine. Reducing C will increase reactance and the lamp will shine less brightly than before.

EXAMPLE 7.3

Example 7.4 A 15.0 μF capacitor is connected to a 220 V, 50 Hz source. Find the capacitive reactance and the current (rms and peak) in the circuit. If the frequency is doubled, what happens to the capacitive reactance and the current?

Solution The capacitive reactance is

$$X_C = \frac{1}{2\pi \nu C} = \frac{1}{2\pi(50\text{Hz})(15.0\times10^{-6}\text{F})} = 212\,\Omega$$

The rms current is

EXAMPLE 7.4

$$I = \frac{V}{X_C} = \frac{220\,\text{V}}{212\,\Omega} = 1.04\,\text{A}$$

The peak current is

$$i_m = \sqrt{2}I = (1.41)(1.04\,A) = 1.47\,A$$

This current oscillates between +1.47A and –1.47 A, and is ahead of the voltage by $\pi/2$.

If the frequency is doubled, the capacitive reactance is halved and consequently, the current is doubled.

EXAMPLE 7.4

Example 7.5 A light bulb and an open coil inductor are connected to an ac source through a key as shown in Fig. 7.9.

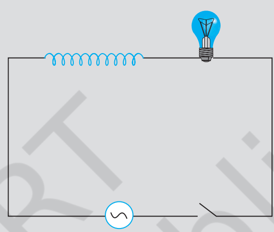


FIGURE 7.9

The switch is closed and after sometime, an iron rod is inserted into the interior of the inductor. The glow of the light bulb (a) increases; (b) decreases; (c) is unchanged, as the iron rod is inserted. Give your answer with reasons.

Solution As the iron rod is inserted, the magnetic field inside the coil magnetizes the iron increasing the magnetic field inside it. Hence, the inductance of the coil increases. Consequently, the inductive reactance of the coil increases. As a result, a larger fraction of the applied ac voltage appears across the inductor, leaving less voltage across the bulb. Therefore, the glow of the light bulb decreases.

EXAMPLE 7.5

7.6 AC Voltage Applied to a Series LCR Circuit

Figure 7.10 shows a series LCR circuit connected to an ac source ε. As usual, we take the voltage of the source to be $v = v_m \sin \omega t$.

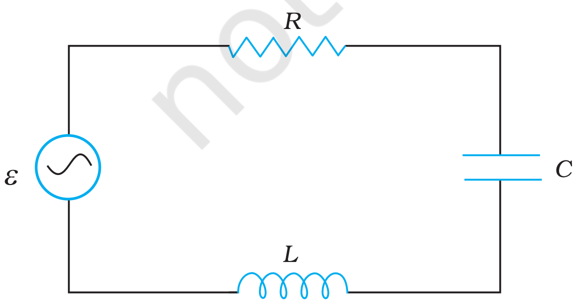


FIGURE 7.10 A series *LCR* circuit connected to an ac source.

If q is the charge on the capacitor and i the current, at time t, we have, from Kirchhoff's loop rule:

$$L\frac{di}{dt} + iR + \frac{q}{C} = v \tag{7.20}$$

We want to determine the instantaneous current i and its phase relationship to the applied alternating voltage v. We shall solve this problem by two methods. First, we use the technique of phasors and in the second method, we solve Eq. (7.20) analytically to obtain the time–dependence of i.

7.6.1 Phasor-diagram solution

From the circuit shown in Fig. 7.10, we see that the resistor, inductor and capacitor are in series. Therefore, the ac current in each element is the same at any time, having the same amplitude and phase. Let it be

$$i = i_m \sin(\omega t+\phi) \tag{7.21}$$

where ϕ is the phase difference between the voltage across the source and the current in the circuit. On the basis of what we have learnt in the previous sections, we shall construct a phasor diagram for the present case.

Let **I** be the phasor representing the current in the circuit as given by Eq. (7.21). Further, let $\mathbf{V_L}$, $\mathbf{V_R}$, $\mathbf{V_C}$, and **V** represent the voltage across the inductor, resistor, capacitor and the source, respectively. From previous section, we know that $\mathbf{V_R}$ is parallel to **I**, $\mathbf{V_C}$ is $\pi/2$ behind **I** and $\mathbf{V_L}$ is $\pi/2$ ahead of **I**. $\mathbf{V_L}$, $\mathbf{V_R}$, $\mathbf{V_C}$ and **I** are shown in Fig. 7.11(a) with apppropriate phase-relations.

The length of these phasors or the amplitude of $\mathbf{V_R}$, $\mathbf{V_C}$ and $\mathbf{V_L}$ are:

$$v_{Rm} = i_m R,\ v_{Cm} = i_m X_C,\ v_{Lm} = i_m X_L \tag{7.22}$$

The voltage Equation (7.20) for the circuit can be written as

$$v_L + v_R + v_C = v \tag{7.23}$$

The phasor relation whose vertical component gives the above equation is

$$\mathbf{V_L} + \mathbf{V_R} + \mathbf{V_C} = \mathbf{V} \tag{7.24}$$

This relation is represented in Fig. 7.11(b). Since $\mathbf{V_C}$ and $\mathbf{V_L}$ are always along the same line and in opposite directions, they can be combined into a single phasor $(\mathbf{V_C} + \mathbf{V_L})$ which has a magnitude $|v_{Cm} - v_{Lm}|$. Since **V** is represented as the hypotenuse of a right-triangle whose sides are $\mathbf{V_R}$ and $(\mathbf{V_C} + \mathbf{V_L})$, the pythagorean theorem gives:

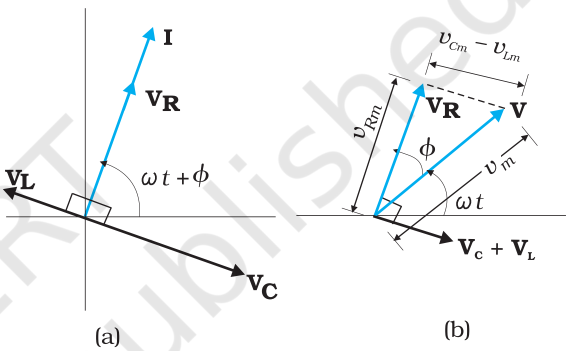


FIGURE 7.11 (a) Relation between the phasors $\mathbf{V_L}$, $\mathbf{V_R}$, $\mathbf{V_C}$, and *I*, (b) Relation between the phasors $\mathbf{V_L}$, $\mathbf{V_R}$, and $(\mathbf{V_L} + \mathbf{V_C})$ for the circuit in Fig. 7.10.

$$v_m^2 = v_{Rm}^2 + \left(v_{Cm} - v_{Lm}\right)^2$$

Substituting the values of v_{Rm}, v_{Cm}, and v_{Lm} from Eq. (7.22) into the above equation, we have

$$v_m^2 = (i_m R)^2 + (i_m X_C - i_m X_L)^2$$

$$= i_m^2 \left[R^2 + (X_C - X_L)^2\right]$$

or, $$i_m = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}} \tag{7.25(a)}$$

By analogy to the resistance in a circuit, we introduce the *impedance Z* in an ac circuit:

$$i_m = \frac{v_m}{Z} \tag{7.25(b)}$$

where $$Z = \sqrt{R^2 + (X_C - X_L)^2} \tag{7.26}$$

Reprint 2025-26

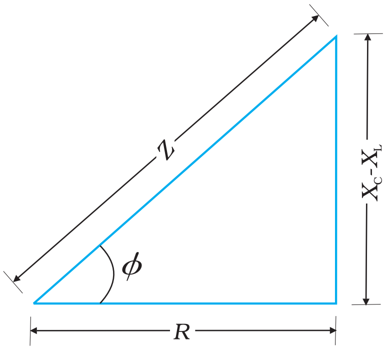


FIGURE 7.12 Impedance diagram.

Since phasor **I** is always parallel to phasor $\mathbf{V_R}$, the phase angle ϕ is the angle between $\mathbf{V_R}$ and **V** and can be determined from Fig. 7.12:

$$\tan\phi = \frac{v_{Cm} - v_{Lm}}{v_{Rm}}$$

Using Eq. (7.22), we have

$$\tan\phi = \frac{X_C - X_L}{R} \qquad (7.27)$$

Equations (7.26) and (7.27) are graphically shown in Fig. (7.12). This is called *Impedance diagram* which is a right-triangle with *Z* as its hypotenuse.

Equation 7.25(a) gives the amplitude of the current and Eq. (7.27) gives the phase angle. With these, Eq. (7.21) is completely specified.

If $X_C > X_L$, ϕ is positive and the circuit is predominantly capacitive. Consequently, the current in the circuit leads the source voltage. If $X_C < X_L$, ϕ is negative and the circuit is predominantly inductive. Consequently, the current in the circuit lags the source voltage.

Figure 7.13 shows the phasor diagram and variation of v and i with ωt for the case $X_C > X_L$.

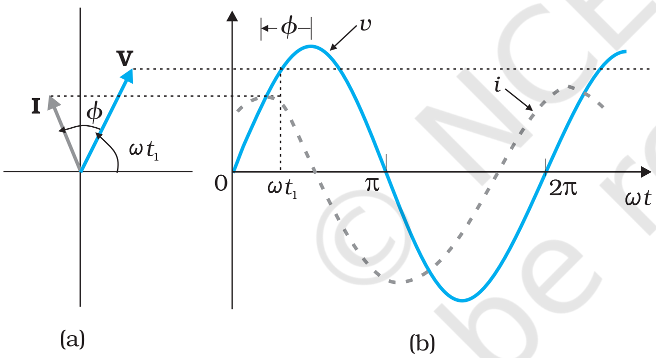


FIGURE 7.13 (a) Phasor diagram of **V** and **I**. (b) Graphs of v and i versus ωt for a series *LCR* circuit where $X_C > X_L$.

Thus, we have obtained the amplitude and phase of current for an *LCR* series circuit using the technique of phasors. But this method of analysing ac circuits suffers from certain disadvantages. First, the phasor diagram say nothing about the initial condition. One can take any arbitrary value of *t* (say, t_1, as done throughout this chapter) and draw different phasors which show the relative angle between different phasors. The solution so obtained is called the *steady-state solution*. This is not a general solution. Additionally, we do have a *transient solution* which exists even for $v = 0$. The general solution is the sum of the transient solution and the steady-state solution. After a sufficiently long time, the effects of the transient solution die out and the behaviour of the circuit is described by the steady-state solution.

7.6.2 Resonance

An interesting characteristic of the series *RLC* circuit is the phenomenon of resonance. The phenomenon of resonance is common among systems that have a tendency to oscillate at a particular frequency. This frequency is called the system's *natural frequency*. If such a system is driven by an energy source at a frequency that is near the natural frequency, the amplitude of oscillation is found to be large. A familiar example of this phenomenon is a child on a swing. The swing has a natural frequency for swinging back and forth like a pendulum. If the child pulls on the rope at regular intervals and the frequency of the pulls is almost the same as the frequency of swinging, the amplitude of the swinging will be large.

For an *RLC* circuit driven with voltage of amplitude v_m and frequency ω, we found that the current amplitude is given by

$$i_m = \frac{v_m}{Z} = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}}$$

with $X_c = 1/\omega C$ and $X_L = \omega L$. So if ω is varied, then at a particular frequency ω_0, $X_c = X_L$, and the impedance is minimum $\left(Z = \sqrt{R^2 + 0^2} = R\right)$. This frequency is called the *resonant frequency*:

$$X_c = X_L \text{ or } \frac{1}{\omega_0 C} = \omega_0 L$$

or $$\omega_0 = \frac{1}{\sqrt{LC}} \qquad (7.28)$$

At resonant frequency, the current amplitude is maximum; $i_m = v_m/R$.

Figure 7.16 shows the variation of i_m with ω in a *RLC* series circuit with L = 1.00 mH, C = 1.00 nF for two values of R: (i) R = 100 Ω and (ii) R = 200 Ω. For the source applied v_m = 100 V. ω_0 for this case is $\frac{1}{\sqrt{LC}}$ = 1.00×10^6 rad/s.

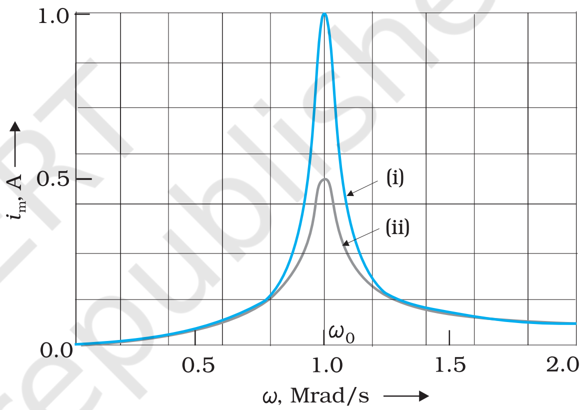


FIGURE 7.14 Variation of i_m with ω for two cases: (i) R = 100 Ω, (ii) R = 200 Ω, L = 1.00 mH.

We see that the current amplitude is maximum at the resonant frequency. Since $i_m = v_m / R$ at resonance, the current amplitude for case (i) is twice to that for case (ii).

Resonant circuits have a variety of applications, for example, in the tuning mechanism of a radio or a TV set. The antenna of a radio accepts signals from many broadcasting stations. The signals picked up in the antenna acts as a source in the tuning circuit of the radio, so the circuit can be driven at many frequencies. But to hear one particular radio station, we tune the radio. In tuning, we vary the capacitance of a capacitor in the tuning circuit such that the resonant frequency of the circuit becomes nearly equal to the frequency of the radio signal received. When this happens, the amplitude of the current with the frequency of the signal of the particular radio station in the circuit is maximum.

It is important to note that resonance phenomenon is exhibited by a circuit only if both L and C are present in the circuit. Only then do the voltages across L and C cancel each other (both being out of phase) and the current amplitude is v_m/R, the total source voltage appearing across R. This means that we cannot have resonance in a RL or RC circuit.

Reprint 2025-26

Example 7.6 A resistor of 200 Ω and a capacitor of 15.0 μF are connected in series to a 220 V, 50 Hz ac source. (a) Calculate the current in the circuit; (b) Calculate the voltage (rms) across the resistor and the capacitor. Is the algebraic sum of these voltages more than the source voltage? If yes, resolve the paradox.

Solution

Given

$R = 200\,\Omega,\ C = 15.0\,\mu\text{F} = 15.0 \times 10^{-6}\,\text{F}$

$V = 220\,\text{V},\ \nu = 50\,\text{Hz}$

(a) In order to calculate the current, we need the impedance of the circuit. It is

$$Z = \sqrt{R^2 + X_C^2} = \sqrt{R^2 + (2\pi\nu C)^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (2 \times 3.14 \times 50 \times 15.0 \times 10^{-6}\,\text{F})^{-2}}$$

$$= \sqrt{(200\,\Omega)^2 + (212.3\,\Omega)^2}$$

$$= 291.67\,\Omega$$

Therefore, the current in the circuit is

$$I = \frac{V}{Z} = \frac{220\,\text{V}}{291.5\,\Omega} = 0.755\,\text{A}$$

(b) Since the current is the same throughout the circuit, we have

$$V_R = I\,R = (0.755\,\text{A})(200\,\Omega) = 151\,\text{V}$$

$$V_C = I\,X_C = (0.755\,\text{A})(212.3\,\Omega) = 160.3\,\text{V}$$

The algebraic sum of the two voltages, V_R and V_C is 311.3 V which is more than the source voltage of 220 V. How to resolve this paradox? As you have learnt in the text, the two voltages are not in the same phase. Therefore, *they cannot be added like ordinary numbers*. The two voltages are out of phase by ninety degrees. Therefore, the total of these voltages must be obtained using the Pythagorean theorem:

$$V_{R+C} = \sqrt{V_R^2 + V_C^2}$$

$$= 220\text{ V}$$

Thus, if the phase difference between two voltages is properly taken into account, the total voltage across the resistor and the capacitor is equal to the voltage of the source.

Example 7.6

7.7 Power in AC Circuit: The Power Factor

We have seen that a voltage $v = v_m \sin\omega t$ applied to a series *RLC* circuit drives a current in the circuit given by $i = i_m \sin(\omega t + \phi)$ where

$$i_m = \frac{v_m}{Z} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{X_C - X_L}{R}\right)$$

Therefore, the instantaneous power *p* supplied by the source is

$$p = v\,i = (v_m \sin \omega t) \times [i_m \sin(\omega t + \phi)]$$

$$= \frac{v_m i_m}{2}[\cos\phi - \cos(2\omega t + \phi)] \qquad (7.29)$$

The average power over a cycle is given by the average of the two terms in R.H.S. of Eq. (7.29). It is only the second term which is time-dependent. Its average is zero (the positive half of the cosine cancels the negative half). Therefore,

$$P = \frac{v_m i_m}{2}\cos\phi = \frac{v_m}{\sqrt{2}}\frac{i_m}{\sqrt{2}}\cos\phi$$

$$= V\,I\cos\phi \qquad [7.30(a)]$$

This can also be written as,

$$P = I^2\,Z\cos\phi \qquad [7.30(b)]$$

So, the average power dissipated depends not only on the voltage and current but also on the cosine of the phase angle ϕ between them. The quantity $\cos\phi$ is called the *power factor*. Let us discuss the following cases:

Case (i) *Resistive circuit*: If the circuit contains only pure R, it is called *resistive*. In that case $\phi = 0$, $\cos\phi = 1$. There is maximum power dissipation.

Case (ii) *Purely inductive or capacitive circuit*: If the circuit contains only an inductor or capacitor, we know that the phase difference between voltage and current is $\pi/2$. Therefore, $\cos\phi = 0$, and no power is dissipated even though a current is flowing in the circuit. This current is sometimes referred to as *wattless current*.

Case (iii) *LCR series circuit*: In an *LCR* series circuit, power dissipated is given by Eq. (7.30) where $\phi = \tan^{-1}(X_c - X_L)/R$. So, ϕ may be non-zero in a *RL* or *RC* or *RCL* circuit. Even in such cases, power is dissipated only in the resistor.

Case (iv) *Power dissipated at resonance in LCR circuit*: At resonance $X_c - X_L = 0$, and $\phi = 0$. Therefore, $\cos\phi = 1$ and $P = I^2Z = I^2 R$. That is, maximum power is dissipated in a circuit (through R) at resonance.

Example 7.7 (a) For circuits used for transporting electric power, a low power factor implies large power loss in transmission. Explain.

(b) Power factor can often be improved by the use of a capacitor of appropriate capacitance in the circuit. Explain.

Solution (a) We know that $P = I\,V\cos\phi$ where $\cos\phi$ is the power factor. To supply a given power at a given voltage, if $\cos\phi$ is small, we have to increase current accordingly. But this will lead to large power loss (I^2R) in transmission.

(b)Suppose in a circuit, current I lags the voltage by an angle ϕ. Then power factor $\cos\phi = R/Z$.

We can improve the power factor (tending to 1) by making Z tend to R. Let us understand, with the help of a phasor diagram (Fig. 7.15)

EXAMPLE 7.7

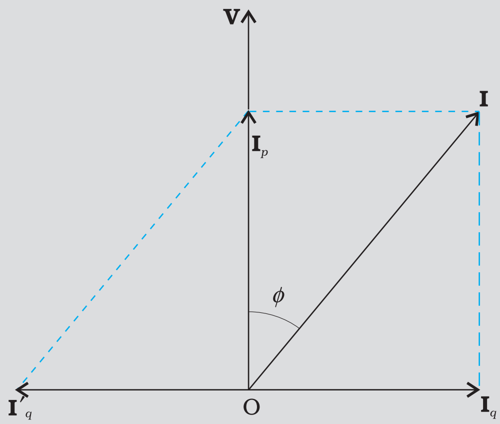


FIGURE 7.15

how this can be achieved. Let us resolve **I** into two components. $\mathbf{I}_p$ along the applied voltage **V** and $\mathbf{I}_q$ perpendicular to the applied voltage. $\mathbf{I}_q$ as you have learnt in Section 7.7, is called the wattless component since corresponding to this component of current, there is no power loss. $\mathbf{I}_P$ is known as the power component because it is in phase with the voltage and corresponds to power loss in the circuit.

It's clear from this analysis that if we want to improve power factor, we must completely neutralize the lagging wattless current $\mathbf{I}_q$ by an equal leading wattless current $\mathbf{I}'_q$. This can be done by connecting a capacitor of appropriate value in parallel so that $\mathbf{I}_q$ and $\mathbf{I}'_q$ cancel each other and P is effectively $I_p V$.

EXAMPLE 7.7

Example 7.8 A sinusoidal voltage of peak value 283 V and frequency 50 Hz is applied to a series *LCR* circuit in which R = 3 Ω, L = 25.48 mH, and C = 796 μF. Find (a) the impedance of the circuit; (b) the phase difference between the voltage across the source and the current; (c) the power dissipated in the circuit; and (d) the power factor.

Solution

(a) To find the impedance of the circuit, we first calculate X_L and X_C.

$$X_L = 2\,\pi\nu L$$
$$= 2 \times 3.14 \times 50 \times 25.48 \times 10^{-3}\ \Omega = 8\ \Omega$$

$$X_C = \frac{1}{2\pi\nu C}$$

$$= \frac{1}{2 \times 3.14 \times 50 \times 796 \times 10^{-6}} = 4\Omega$$

Therefore,

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{3^2 + (8-4)^2}$$
$$= 5\ \Omega$$

(b) Phase difference, $\phi = \tan^{-1}\dfrac{X_C - X_L}{R}$

$$= \tan^{-1}\left(\frac{4-8}{3}\right) = -53.1°$$

EXAMPLE 7.8

Since ϕ is negative, the current in the circuit lags the voltage across the source.

(c) The power dissipated in the circuit is

$$P = I^2 R$$

Now, $I = \frac{i_m}{\sqrt{2}} = \frac{1}{\sqrt{2}}\left(\frac{283}{5}\right) = 40\text{A}$

Therefore, $P = (40\text{A})^2 \times 3\,\Omega = 4800\text{ W}$

(d) Power factor $= \cos\phi = \cos(-53.1°) = 0.6$

EXAMPLE 7.8

Example 7.9 Suppose the frequency of the source in the previous example can be varied. (a) What is the frequency of the source at which resonance occurs? (b) Calculate the impedance, the current, and the power dissipated at the resonant condition.

Solution

(a) The frequency at which the resonance occurs is

$$\omega_0 = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{25.48 \times 10^{-3} \times 796 \times 10^{-6}}}$$

$$= 222.1\,\text{rad/s}$$

$$\nu_r = \frac{\omega_0}{2\pi} = \frac{221.1}{2 \times 3.14}\text{Hz} = 35.4\text{Hz}$$

(b) The impedance Z at resonant condition is equal to the resistance:

$$Z = R = 3\,\Omega$$

The rms current at resonance is

$$= \frac{V}{Z} = \frac{V}{R} = \left(\frac{283}{\sqrt{2}}\right)\frac{1}{3} = 66.7\text{A}$$

The power dissipated at resonance is

$$P = I^2 \times R = (66.7)^2 \times 3 = 13.35\text{ kW}$$

You can see that in the present case, power dissipated at resonance is more than the power dissipated in Example 7.8.

EXAMPLE 7.9

Example 7.10 At an airport, a person is made to walk through the doorway of a metal detector, for security reasons. If she/he is carrying anything made of metal, the metal detector emits a sound. On what principle does this detector work?

Solution The metal detector works on the principle of resonance in ac circuits. When you walk through a metal detector, you are, in fact, walking through a coil of many turns. The coil is connected to a capacitor tuned so that the circuit is in resonance. When you walk through with metal in your pocket, the impedance of the circuit changes – resulting in significant change in current in the circuit. This change in current is detected and the electronic circuitry causes a sound to be emitted as an alarm.

EXAMPLE 7.10

7.8 Transformers

For many purposes, it is necessary to change (or transform) an alternating voltage from one to another of greater or smaller value. This is done with a device called *transformer* using the principle of mutual induction.

A transformer consists of two sets of coils, insulated from each other. They are wound on a soft-iron core, either one on top of the other as in Fig. 7.16(a) or on separate limbs of the core as in Fig. 7.16(b). One of the coils called the *primary coil* has N_p turns. The other coil is called the *secondary coil*; it has N_s turns. Often the primary coil is the input coil and the secondary coil is the output coil of the transformer.

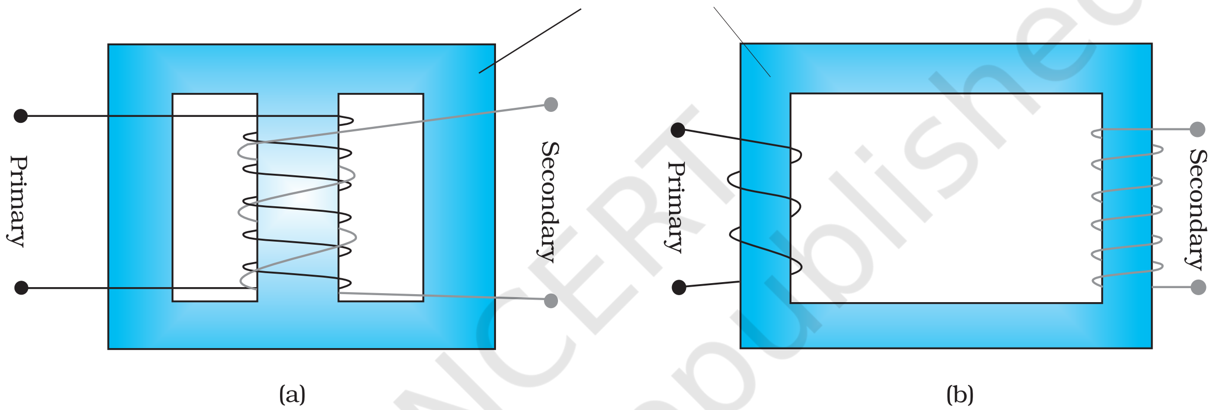


FIGURE 7.16 Two arrangements for winding of primary and secondary coil in a transformer: (a) two coils on top of each other, (b) two coils on separate limbs of the core.

When an alternating voltage is applied to the primary, the resulting current produces an alternating magnetic flux which links the secondary and induces an emf in it. The value of this emf depends on the number of turns in the secondary. We consider an ideal transformer in which the primary has negligible resistance and all the flux in the core links both primary and secondary windings. Let ϕ be the flux in each turn in the core at time t due to current in the primary when a voltage v_p is applied to it.

Then the induced emf or voltage ε_s, in the secondary with N_s turns is

$$\varepsilon_s = -N_s \frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{7.31}$$

The alternating flux ϕ also induces an emf, called back emf in the primary. This is

$$\varepsilon_p = -N_p \frac{\mathrm{d}\phi}{\mathrm{d}t} \tag{7.32}$$

But $\varepsilon_p = v_p$. If this were not so, the primary current would be infinite since the primary has zero resistance (as assumed). If the secondary is an open circuit or the current taken from it is small, then to a good approximation

$$\varepsilon_s = v_s$$

where v_s is the voltage across the secondary. Therefore, Eqs. (7.31) and (7.32) can be written as

$$v_s = -N_s \frac{d\phi}{dt} \qquad [7.31(a)]$$

$$v_p = -N_p \frac{d\phi}{dt} \qquad [7.32(a)]$$

From Eqs. [7.31 (a)] and [7.32 (a)], we have

$$\frac{v_s}{v_p} = \frac{N_s}{N_p} \qquad (7.33)$$

Note that the above relation has been obtained using three assumptions: (i) the primary resistance and current are small; (ii) the same flux links both the primary and the secondary as very little flux escapes from the core, and (iii) the secondary current is small.

If the transformer is assumed to be 100% efficient (no energy losses), the power input is equal to the power output, and since $p = i\,v$,

$$i_p v_p = i_s v_s \qquad (7.34)$$

Although some energy is always lost, this is a good approximation, since a well designed transformer may have an efficiency of more than 95%. Combining Eqs. (7.33) and (7.34), we have

$$\frac{i_p}{i_s} = \frac{v_s}{v_p} = \frac{N_s}{N_p} \qquad (7.35)$$

Since i and v both oscillate with the same frequency as the ac source, Eq. (7.35) also gives the ratio of the amplitudes or rms values of corresponding quantities.

Now, we can see how a transformer affects the voltage and current. We have:

$$V_s = \left(\frac{N_s}{N_p}\right) V_p \quad \text{and} \quad I_s = \left(\frac{N_p}{N_s}\right) I_p \qquad (7.36)$$

That is, if the secondary coil has a greater number of turns than the primary ($N_s > N_p$), the voltage is stepped up ($V_s > V_p$). This type of arrangement is called a *step-up transformer*. However, in this arrangement, there is less current in the secondary than in the primary ($N_p/N_s < 1$ and $I_s < I_p$). For example, if the primary coil of a transformer has 100 turns and the secondary has 200 turns, $N_s/N_p = 2$ and $N_p/N_s = 1/2$. Thus, a 220V input at 10A will step-up to 440 V output at 5.0 A.

If the secondary coil has less turns than the primary ($N_s < N_p$), we have a *step-down transformer*. In this case, $V_s < V_p$ and $I_s > I_p$. That is, the voltage is stepped down, or reduced, and the current is increased.

The equations obtained above apply to ideal transformers (without any energy losses). But in actual transformers, small energy losses do occur due to the following reasons:

(i) *Flux Leakage*: There is always some flux leakage; that is, not all of the flux due to primary passes through the secondary due to poor design of the core or the air gaps in the core. It can be reduced by winding the primary and secondary coils one over the other.

(ii) *Resistance of the windings*: The wire used for the windings has some resistance and so, energy is lost due to heat produced in the wire (I^2R). In high current, low voltage windings, these are minimised by using thick wire.

(iii) *Eddy currents*: The alternating magnetic flux induces eddy currents in the iron core and causes heating. The effect is reduced by using a laminated core.

(iv) *Hysteresis*: The magnetisation of the core is repeatedly reversed by the alternating magnetic field. The resulting expenditure of energy in the core appears as heat and is kept to a minimum by using a magnetic material which has a low hysteresis loss.

The large scale transmission and distribution of electrical energy over long distances is done with the use of transformers. The voltage output of the generator is stepped-up (so that current is reduced and consequently, the I^2R loss is cut down). It is then transmitted over long distances to an area sub-station near the consumers. There the voltage is stepped down. It is further stepped down at distributing sub-stations and utility poles before a power supply of 240 V reaches our homes.

SUMMARY

1. An alternating voltage $v = v_m \sin \omega t$ applied to a resistor R drives a current $i = i_m \sin\omega t$ in the resistor, $i_m = \dfrac{v_m}{R}$. The current is in phase with the applied voltage.
2. For an alternating current $i = i_m \sin \omega t$ passing through a resistor R, the average power loss P (averaged over a cycle) due to joule heating is $(1/2)i_m^2R$. To express it in the same form as the dc power ($P = I^2R$), a special value of current is used. It is called *root mean square* (rms) *current* and is donoted by I:
$$I = \frac{i_m}{\sqrt{2}} = 0.707\, i_m$$
Similarly, the *rms voltage* is defined by
$$V = \frac{v_m}{\sqrt{2}} = 0.707\, v_m$$
We have $P = IV = I^2R$
3. An ac voltage $v = v_m \sin \omega t$ applied to a pure inductor L, drives a current in the inductor $i = i_m \sin(\omega t - \pi/2)$, where $i_m = v_m/X_L$. $X_L = \omega L$ is called *inductive reactance*. The current in the inductor lags the voltage by $\pi/2$. The average power supplied to an inductor over one complete cycle is zero.

4. An ac voltage $v = v_m \sin\omega t$ applied to a capacitor drives a current in the capacitor: $i = i_m \sin(\omega t + \pi/2)$. Here,

$$i_m = \frac{v_m}{X_C},\ X_C = \frac{1}{\omega C}$$ is called *capacitive reactance*.

The current through the capacitor is $\pi/2$ ahead of the applied voltage. As in the case of inductor, the average power supplied to a capacitor over one complete cycle is zero.

5. For a series *RLC* circuit driven by voltage $v = v_m \sin \omega t$, the current is given by $i = i_m \sin(\omega t + \phi)$

where $$i_m = \frac{v_m}{\sqrt{R^2 + (X_C - X_L)^2}}$$

and $$\phi = \tan^{-1}\frac{X_C - X_L}{R}$$

$$Z = \sqrt{R^2 + (X_C - X_L)^2}$$ is called the *impedance* of the circuit.

The average power loss over a complete cycle is given by

$$P = V\,I\cos\phi$$

The term $\cos\phi$ is called the *power factor*.

6. In a purely inductive or capacitive circuit, $\cos\phi = 0$ and no power is dissipated even though a current is flowing in the circuit. In such cases, current is referred to as a *wattless current*.

7. The phase relationship between current and voltage in an ac circuit can be shown conveniently by representing voltage and current by rotating vectors called *phasors*. A phasor is a vector which rotates about the origin with angular speed ω. The magnitude of a phasor represents the amplitude or peak value of the quantity (voltage or current) represented by the phasor.

The analysis of an ac circuit is facilitated by the use of a phasor diagram.

8. A transformer consists of an iron core on which are bound a primary coil of N_p turns and a secondary coil of N_s turns. If the primary coil is connected to an ac source, the primary and secondary voltages are related by

$$V_s = \left(\frac{N_s}{N_p}\right)V_p$$

and the currents are related by

$$I_s = \left(\frac{N_p}{N_s}\right)I_p$$

If the secondary coil has a greater number of turns than the primary, the voltage is stepped-up ($V_s > V_p$). This type of arrangement is called a *step-up transformer*. If the secondary coil has turns less than the primary, we have a *step-down transformer*.

Physical quantity	Symbol	Dimensions	Unit	Remarks
rms voltage	V	[M $L^2T^{-3}A^{-1}$]	V	$V = \frac{v_m}{\sqrt{2}}$, v_m is the amplitude of the ac voltage.
rms current	I	[A]	A	$I = \frac{i_m}{\sqrt{2}}$, i_m is the amplitude of the ac current.
Reactance: Inductive	X_L	[M $L^2T^{-3}A^{-2}$]		$X_L = \omega L$
Capacitive	X_C	[M $L^2T^{-3}A^{-2}$]		$X_C = 1/\omega C$
Impedance	Z	[M $L^2T^{-3}A^{-2}$]		Depends on elements present in the circuit.
Resonant frequency	ω_r or ω_0	[T^{-1}]	Hz	$\omega_0 = \frac{1}{\sqrt{LC}}$ for a series *RLC* circuit
Quality factor	Q	Dimensionless		$Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR}$ for a series *RLC* circuit.
Power factor		Dimensionless		$= \cos\phi$, ϕ is the phase difference between voltage applied and current in the circuit.

POINTS TO PONDER

1. When a value is given for ac voltage or current, it is ordinarily the rms value. The voltage across the terminals of an outlet in your room is normally 240 V. This refers to the *rms* value of the voltage. The amplitude of this voltage is

$$v_m = \sqrt{2}V = \sqrt{2}(240) = 340\text{ V}$$

2. The power rating of an element used in ac circuits refers to its average power rating.
3. The power consumed in an ac circuit is never negative.
4. Both alternating current and direct current are measured in amperes. But how is the ampere defined for an alternating current? It cannot be derived from the mutual attraction of two parallel wires carrying ac currents, as the dc ampere is derived. An ac current changes direction

with the source frequency and the attractive force would average to zero. Thus, the ac ampere must be defined in terms of some property that is independent of the direction of the current. Joule heating is such a property, and there is one ampere of *rms* value of alternating current in a circuit if the current produces the same average heating effect as one ampere of dc current would produce under the same conditions.

5. In an ac circuit, while adding voltages across different elements, one should take care of their phases properly. For example, if V_R and V_C are voltages across R and C, respectively in an RC circuit, then the total voltage across RC combination is $V_{RC} = \sqrt{V_R^2 + V_C^2}$ and not $V_R + V_C$ since V_C is $\pi/2$ out of phase of V_R.
6. Though in a phasor diagram, voltage and current are represented by vectors, these quantities are not really vectors themselves. They are scalar quantities. It so happens that the amplitudes and phases of harmonically varying scalars combine mathematically in the same way as do the projections of rotating vectors of corresponding magnitudes and directions. The 'rotating vectors' that represent harmonically varying scalar quantities are introduced only to provide us with a simple way of adding these quantities using a rule that we already know as the law of vector addition.
7. There are no power losses associated with pure capacitances and pure inductances in an ac circuit. The only element that dissipates energy in an ac circuit is the resistive element.
8. In a RLC circuit, resonance phenomenon occur when $X_L = X_C$ or $\omega_0 = \frac{1}{\sqrt{LC}}$. For resonance to occur, the presence of both L and C elements in the circuit is a must. With only one of these (L or C) elements, there is no possibility of voltage cancellation and hence, no resonance is possible.
9. The power factor in a RLC circuit is a measure of how close the circuit is to expending the maximum power.
10. In generators and motors, the roles of input and output are reversed. In a motor, electric energy is the input and mechanical energy is the output. In a generator, mechanical energy is the input and electric energy is the output. Both devices simply transform energy from one form to another.
11. A transformer (step-up) changes a low-voltage into a high-voltage. This does not violate the law of conservation of energy. The current is reduced by the same proportion.

EXERCISES

7.1 A 100 Ω resistor is connected to a 220 V, 50 Hz ac supply.

(a) What is the rms value of current in the circuit?

(b) What is the net power consumed over a full cycle?

7.2 (a) The peak voltage of an ac supply is 300 V. What is the rms voltage?

(b) The rms value of current in an ac circuit is 10 A. What is the peak current?

7.3 A 44 mH inductor is connected to 220 V, 50 Hz ac supply. Determine the rms value of the current in the circuit.

7.4 A 60 μF capacitor is connected to a 110 V, 60 Hz ac supply. Determine the rms value of the current in the circuit.

7.5 In Exercises 7.3 and 7.4, what is the net power absorbed by each circuit over a complete cycle. Explain your answer.

7.6 A charged 30 μF capacitor is connected to a 27 mH inductor. What is the angular frequency of free oscillations of the circuit?

7.7 A series *LCR* circuit with $R = 20\ \Omega$, $L = 1.5$ H and $C = 35\ \mu$F is connected to a variable-frequency 200 V ac supply. When the frequency of the supply equals the natural frequency of the circuit, what is the average power transferred to the circuit in one complete cycle?

7.8 Figure 7.17 shows a series *LCR* circuit connected to a variable frequency 230 V source. $L = 5.0$ H, $C = 80\mu$F, $R = 40\ \Omega$.

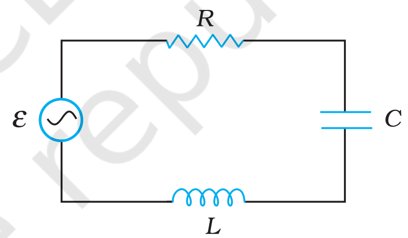


FIGURE 7.17

(a) Determine the source frequency which drives the circuit in resonance.

(b) Obtain the impedance of the circuit and the amplitude of current at the resonating frequency.

(c) Determine the rms potential drops across the three elements of the circuit. Show that the potential drop across the *LC* combination is zero at the resonating frequency.

Reprint 2025-26

Chapter Eight

ELECTROMAGNETIC WAVES

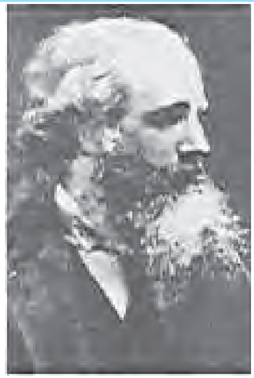
8.1 Introduction

In Chapter 4, we learnt that an electric current produces magnetic field and that two current-carrying wires exert a magnetic force on each other. Further, in Chapter 6, we have seen that a magnetic field changing with time gives rise to an electric field. Is the converse also true? Does an electric field changing with time give rise to a magnetic field? James Clerk Maxwell (1831-1879), argued that this was indeed the case – not only an electric current but also a time-varying electric field generates magnetic field. While applying the Ampere's circuital law to find magnetic field at a point outside a capacitor connected to a time-varying current, Maxwell noticed an inconsistency in the Ampere's circuital law. He suggested the existence of an additional current, called by him, the displacement current to remove this inconsistency.

Maxwell formulated a set of equations involving electric and magnetic fields, and their sources, the charge and current densities. These equations are known as Maxwell's equations. Together with the Lorentz force formula (Chapter 4), they mathematically express all the basic laws of electromagnetism.

The most important prediction to emerge from Maxwell's equations is the existence of electromagnetic waves, which are (coupled) time-varying electric and magnetic fields that propagate in space. The speed of the waves, according to these equations, turned out to be very close to

JAMES CLERK MAXWELL (1831–1879)



James Clerk Maxwell (1831 – 1879) Born in Edinburgh, Scotland, was among the greatest physicists of the nineteenth century. He derived the thermal velocity distribution of molecules in a gas and was among the first to obtain reliable estimates of molecular parameters from measurable quantities like viscosity, etc. Maxwell's greatest acheivement was the unification of the laws of electricity and magnetism (discovered by Coulomb, Oersted, Ampere and Faraday) into a consistent set of equations now called Maxwell's equations. From these he arrived at the most important conclusion that light is an electromagnetic wave. Interestingly, Maxwell did not agree with the idea (strongly suggested by the Faraday's laws of electrolysis) that electricity was particulate in nature.

the speed of light(3×10^8 m/s), obtained from optical measurements. This led to the remarkable conclusion that light is an electromagnetic wave. Maxwell's work thus unified the domain of electricity, magnetism and light. Hertz, in 1885, experimentally demonstrated the existence of electromagnetic waves. Its technological use by Marconi and others led in due course to the revolution in communication that we are witnessing today.

In this chapter, we first discuss the need for displacement current and its consequences. Then we present a descriptive account of electromagnetic waves. The broad spectrum of electromagnetic waves, stretching from γ rays (wavelength ~10^{-12} m) to long radio waves (wavelength ~10^6 m) is described.

8.2 Displacement Current

We have seen in Chapter 4 that an electrical current produces a magnetic field around it. Maxwell showed that for logical consistency, a changing electric field *must also* produce a magnetic field. This effect is of great importance because it explains the existence of radio waves, gamma rays and visible light, as well as all other forms of electromagnetic waves.

To see how a changing electric field gives rise to a magnetic field, let us consider the process of charging of a capacitor and apply Ampere's circuital law given by (Chapter 4)

$$\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0 \, i\,(t) \tag{8.1}$$

to find magnetic field at a point outside the capacitor. Figure 8.1(a) shows a parallel plate capacitor C which is a part of circuit through which a time-dependent current $i\,(t)$ flows . Let us find the magnetic field at a point such as P, in a region outside the parallel plate capacitor. For this, we consider a plane circular loop of radius r whose plane is perpendicular to the direction of the current-carrying wire, and which is centred symmetrically with respect to the wire [Fig. 8.1(a)]. From symmetry, the magnetic field is directed along the circumference of the circular loop and is the same in magnitude at all points on the loop so that if B is the magnitude of the field, the left side of Eq. (8.1) is $B\,(2\pi r)$. So we have

$$B\,(2\pi r) = \mu_0 i\,(t) \tag{8.2}$$

Now, consider a different surface, which has the same boundary. This is a pot like surface [Fig. 8.1(b)] which nowhere touches the current, but has its bottom between the capacitor plates; its mouth is the circular loop mentioned above. Another such surface is shaped like a tiffin box (without the lid) [Fig. 8.1(c)]. On applying Ampere's circuital law to such surfaces with the *same* perimeter, we find that the left hand side of Eq. (8.1) has not changed but the right hand side is *zero* and *not* $\mu_0 i$, since *no* current passes through the surface of Fig. 8.1(b) and (c). So we have a *contradiction*; calculated one way, there is a magnetic field at a point P; calculated another way, the magnetic field at P is zero.
Since the contradiction arises from our use of Ampere's circuital law, this law must be missing something. The missing term must be such that one gets the same magnetic field at point P, no matter what surface is used.

We can actually guess the missing term by looking carefully at Fig. 8.1(c). Is there anything passing through the surface S *between* the plates of the capacitor? Yes, of course, the electric field! If the plates of the capacitor have an area A, and a total charge Q, the magnitude of the electric field $\mathbf{E}$ between the plates is $(Q/A)/\varepsilon_0$ (see Eq. 2.41). The field is perpendicular to the surface S of Fig. 8.1(c). It has the same magnitude over the area A of the capacitor plates, and vanishes outside it. So what is the *electric* flux Φ_E through the surface S ? Using Gauss's law, it is

$$\Phi_E = |\mathbf{E}|A = \frac{1}{\varepsilon_0}\frac{Q}{A}A = \frac{Q}{\varepsilon_0} \tag{8.3}$$

Now if the charge Q on the capacitor plates changes with time, there is a current $i = (\mathrm{d}Q/\mathrm{d}t)$, so that using Eq. (8.3), we have

$$\frac{\mathrm{d}\Phi_E}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Q}{\varepsilon_0}\right) = \frac{1}{\varepsilon_0}\frac{\mathrm{d}Q}{\mathrm{d}t}$$

This implies that for consistency,

$$\varepsilon_0\left(\frac{\mathrm{d}\Phi_E}{\mathrm{d}t}\right) = i \tag{8.4}$$

This is the missing term in Ampere's circuital law. If we generalise this law by adding to the total current carried by conductors through the surface, another term which is ε_0 times the rate of change of electric flux through the same surface, the *total* has the same value of current i for all surfaces. If this is done, there is no contradiction in the value of B obtained anywhere using the generalised Ampere's law. B at the point P is non-zero no matter which surface is used for calculating it. B at a point P outside the plates [Fig. 8.1(a)] is the same as at a point M just inside, as it should be. The current carried by conductors due to flow of charges is called *conduction current*. The current, given by Eq. (8.4), is a new term, and is due to changing electric field (or electric *displacement*, an old term still used sometimes). It is, therefore, called *displacement current* or Maxwell's displacement current. Figure 8.2 shows the electric and magnetic fields inside the parallel plate capacitor discussed above.

The generalisation made by Maxwell then is the following. The source of a magnetic field is not *just* the conduction electric current due to flowing

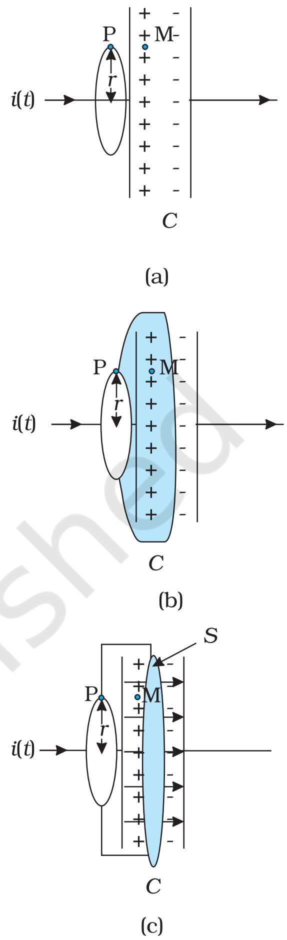
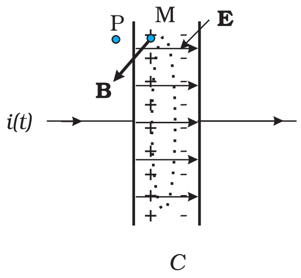
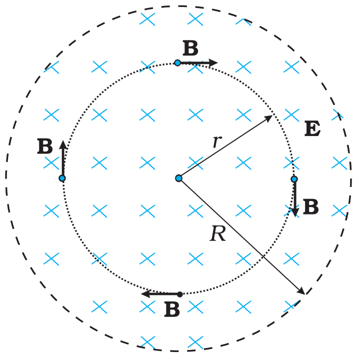


FIGURE 8.1 A parallel plate capacitor C, as part of a circuit through which a time dependent current i (t) flows, (a) a loop of radius r, to determine magnetic field at a point P on the loop; (b) a pot-shaped surface passing through the interior between the capacitor plates with the loop shown in (a) as its rim; (c) a tiffin-shaped surface with the circular loop as its rim and a flat circular bottom S between the capacitor plates. The arrows show uniform electric field between the capacitor plates.

Reprint 2025-26



(a)



(b)

FIGURE 8.2 (a) The electric and magnetic fields **E** and **B** between the capacitor plates, at the point M. (b) A cross sectional view of Fig. (a).

charges, but also the time rate of change of electric field. More precisely, the total current *i* is the sum of the conduction current denoted by i_c, and the displacement current denoted by i_d (= ε_0 ($d\Phi_E/dt$)). So we have

$$i = i_c + i_d = i_c + \varepsilon_0 \frac{d\Phi_E}{dt} \qquad (8.5)$$

In explicit terms, this means that outside the capacitor plates, we have only conduction current $i_c = i$, and no displacement current, i.e., $i_d = 0$. On the other hand, inside the capacitor, there is no conduction current, i.e., $i_c = 0$, and there is only displacement current, so that $i_d = i$.

The generalised (and correct) Ampere's circuital law has the same form as Eq. (8.1), with one difference: "the *total current* passing through any surface of which the closed loop is the perimeter" is the sum of the conduction current and the displacement current. The generalised law is

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 i_c + \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt} \qquad (8.6)$$

and is known as Ampere-Maxwell law.

In all respects, the displacement current has the same physical effects as the conduction current. In some cases, for example, steady electric fields in a conducting wire, the displacement current may be zero since the electric field **E** does not change with time. In other cases, for example, the charging capacitor above, both conduction and displacement currents may be present in different regions of space. In most of the cases, they both may be present in the same region of space, as there exist no perfectly conducting or perfectly insulating medium. Most interestingly, there may be large regions of space where there is *no* conduction current, but there is only a displacement current due to time-varying electric fields. In such a region, we expect a magnetic field, though there is no (conduction) current source nearby! The prediction of such a displacement current can be verified experimentally. For example, a *magnetic* field (say at point M) between the plates of the capacitor in Fig. 8.2(a) can be measured and is seen to be the same as that just outside (at P).

The displacement current has (literally) far reaching consequences. One thing we immediately notice is that the laws of electricity and magnetism are now more symmetrical*. Faraday's law of induction states that there is an induced emf *equal to the rate of* change of magnetic flux. Now, since the emf between two points 1 and 2 is the work done per unit charge in taking it from 1 to 2, the existence of an emf implies the existence of an electric field. So, we can rephrase Faraday's law of electromagnetic induction by saying that a *magnetic field,* changing with time, gives rise to an *electric field.* Then, the fact that an *electric field* changing with time gives rise to a *magnetic field,* is the symmetrical counterpart, and is

* They are still not perfectly symmetrical; there are no known sources of magnetic field (magnetic monopoles) analogous to electric charges which are sources of electric field.

Reprint 2025-26

a consequence of the displacement current being a source of a magnetic field. Thus, time- dependent electric and magnetic fields give rise to each other! Faraday's law of electromagnetic induction and Ampere-Maxwell law give a quantitative expression of this statement, with the current being the total current, as in Eq. (8.5). One very important consequence of this symmetry is the existence of electromagnetic waves, which we discuss qualitatively in the next section.

Maxwell's equations in Vacuum

1. $\oint \mathbf{E} \cdot d\mathbf{A} = Q/\varepsilon_0$ (Gauss's Law for electricity)
2. $\oint \mathbf{B} \cdot d\mathbf{A} = 0$ (Gauss's Law for magnetism)
3. $\oint \mathbf{E} \cdot d\mathbf{l} = \dfrac{-d\Phi_B}{dt}$ (Faraday's Law)
4. $\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 i_c + \mu_0 \varepsilon_0 \dfrac{d\Phi_E}{dt}$ (Ampere – Maxwell Law)

8.3 Electromagnetic Waves

8.3.1 Sources of electromagnetic waves

How are electromagnetic waves produced? Neither stationary charges nor charges in uniform motion (steady currents) can be sources of electromagnetic waves. The former produces only electrostatic fields, while the latter produces magnetic fields that, however, do not vary with time. It is an important result of Maxwell's theory that accelerated charges radiate electromagnetic waves. The proof of this basic result is beyond the scope of this book, but we can accept it on the basis of rough, qualitative reasoning. Consider a charge oscillating with some frequency. (An oscillating charge is an example of accelerating charge.) This produces an oscillating electric field in space, which produces an oscillating magnetic field, which in turn, is a source of oscillating electric field, and so on. The oscillating electric and magnetic fields thus regenerate each other, so to speak, as the wave propagates through the space. The frequency of the electromagnetic wave naturally equals the frequency of oscillation of the charge. The energy associated with the propagating wave comes at the expense of the energy of the source – the accelerated charge.

From the preceding discussion, it might appear easy to test the prediction that light is an electromagnetic wave. We might think that all we needed to do was to set up an ac circuit in which the current oscillate at the frequency of visible light, say, yellow light. But, alas, that is not possible. The frequency of yellow light is about 6×10^{14} Hz, while the frequency that we get even with modern electronic circuits is hardly about 10^{11} Hz. This is why the experimental demonstration of electromagnetic

HEINRICH RUDOLF HERTZ (1857–1894)

Heinrich Rudolf Hertz (1857 – 1894) German physicist who was the first to broadcast and receive radio waves. He produced electromagnetic waves, sent them through space, and measured their wavelength and speed. He showed that the nature of their vibration, reflection and refraction was the same as that of light and heat waves, establishing their identity for the first time. He also pioneered research on discharge of electricity through gases, and discovered the photoelectric effect.

wave had to come in the low frequency region (the radio wave region), as in the Hertz's experiment (1887).

Hertz's successful experimental test of Maxwell's theory created a sensation and sparked off other important works in this field. Two important achievements in this connection deserve mention. Seven years after Hertz, Jagdish Chandra Bose, working at Calcutta (now Kolkata), succeeded in producing and observing electromagnetic waves of much shorter wavelength (25 mm to 5 mm). His experiment, like that of Hertz's, was confined to the laboratory.

At around the same time, Guglielmo Marconi in Italy followed Hertz's work and succeeded in transmitting electromagnetic waves over distances of many kilometres. Marconi's experiment marks the beginning of the field of communication using electromagnetic waves.

8.3.2 Nature of electromagnetic waves

It can be shown from Maxwell's equations that electric and magnetic fields in an electromagnetic wave are perpendicular to each other, *and* to the direction of propagation. It appears reasonable, say from our discussion of the displacement current. Consider Fig. 8.2. The electric field inside the plates of the capacitor is directed perpendicular to the plates. The magnetic field this gives rise to via the displacement current is along the perimeter of a circle parallel to the capacitor plates. So **B** and **E** are perpendicular in this case. This is a general feature.

In Fig. 8.3, we show a typical example of a plane electromagnetic wave propagating along the *z* direction (the fields are shown as a function of the *z* coordinate, at a given time *t*). The electric field E_x is along the *x*-axis, and varies sinusoidally with *z*, at a given time. The magnetic field B_y is along the *y*-axis, and again varies sinusoidally with *z*. The electric and magnetic fields E_x and B_y are perpendicular to each other, and to the direction *z* of propagation. We can write E_x and B_y as follows:

$$E_x = E_0 \sin (kz - \omega t) \qquad [8.7(a)]$$

$$B_y = B_0 \sin (kz - \omega t) \qquad [8.7(b)]$$

Here *k* is related to the wave length λ of the wave by the usual equation

$$k = \frac{2\pi}{\lambda} \qquad (8.8)$$

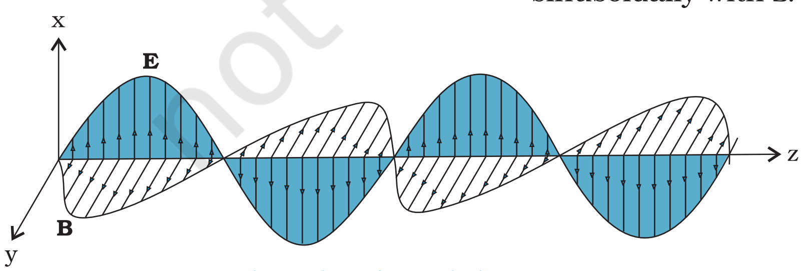


FIGURE 8.3 A linearly polarised electromagnetic wave, propagating in the *z*-direction with the oscillating electric field **E** along the *x*-direction and the oscillating magnetic field **B** along the *y*-direction.

and ω is the angular frequency. k is the magnitude of the wave vector (or propagation vector) **k** and its direction describes the direction of propagation of the wave. The speed of propagation of the wave is (ω/k). Using Eqs. [8.7(a) and (b)] for E_x and B_y and Maxwell's equations, one finds that

$$\omega = ck, \text{ where, } c = 1/\sqrt{\mu_0 \varepsilon_0} \qquad [8.9(a)]$$

The relation $\omega = ck$ is the standard one for waves (see for example, Section 14.4 of class XI Physics textbook). This relation is often written in terms of frequency, ν $(=\omega/2\pi)$ and wavelength, λ $(=2\pi/k)$ as

$$2\pi\nu = c\left(\frac{2\pi}{\lambda}\right) \quad \text{or}$$

$$\nu\lambda = c \qquad [8.9(b)]$$

It is also seen from Maxwell's equations that the magnitude of the electric and the magnetic fields in an electromagnetic wave are related as

$$B_0 = (E_0/c) \qquad (8.10)$$

We here make remarks on some features of electromagnetic waves. They are self-sustaining oscillations of electric and magnetic fields in free space, or vacuum. They differ from all the other waves we have studied so far, in respect that *no material medium* is involved in the vibrations of the electric and magnetic fields.

But what if a material medium is actually there? We know that light, an electromagnetic wave, does propagate through glass, for example. We have seen earlier that the total electric and magnetic fields inside a medium are described in terms of a permittivity ε and a magnetic permeability μ (these describe the factors by which the total fields differ from the external fields). These replace ε_0 and μ_0 in the description to electric and magnetic fields in Maxwell's equations with the result that in a material medium of permittivity ε and magnetic permeability μ, the velocity of light becomes,

$$v = \frac{1}{\sqrt{\mu\varepsilon}} \qquad (8.11)$$

Thus, the velocity of light depends on electric and magnetic properties of the medium. We shall see in the next chapter that the *refractive index* of one medium with respect to the other is equal to the ratio of velocities of light in the two media.

The velocity of electromagnetic waves in free space or vacuum is an important fundamental constant. It has been shown by experiments on electromagnetic waves of different wavelengths that this velocity is the same (independent of wavelength) to within a few metres per second, out of a value of 3×10^8 m/s. The constancy of the velocity of em waves in vacuum is so strongly supported by experiments and the actual value is so well known now that this is used to define a standard of *length*.

The great technological importance of electromagnetic waves stems from their capability to carry energy from one place to another. The radio and TV signals from broadcasting stations carry energy. Light carries energy from the sun to the earth, thus making life possible on the earth.

Reprint 2025-26

Example 8.1 A plane electromagnetic wave of frequency 25 MHz travels in free space along the x-direction. At a particular point in space and time, $\mathbf{E} = 6.3\ \hat{\mathbf{j}}$ V/m. What is **B** at this point?

Solution Using Eq. (8.10), the magnitude of **B** is

$$B = \frac{E}{c}$$

$$= \frac{6.3\ \text{V/m}}{3 \times 10^8\ \text{m/s}} = 2.1 \times 10^{-8}\ \text{T}$$

To find the direction, we note that **E** is along y-direction and the wave propagates along x-axis. Therefore, **B** should be in a direction perpendicular to both x- and y-axes. Using vector algebra, **E × B** should be along x-direction. Since, $(+\hat{\mathbf{j}}) \times (+\hat{\mathbf{k}}) = \hat{\mathbf{i}}$, **B** is along the z-direction.

Thus, $\mathbf{B} = 2.1 \times 10^{-8}\ \hat{\mathbf{k}}$ T

EXAMPLE 8.1

Example 8.2 The magnetic field in a plane electromagnetic wave is given by $B_y = (2 \times 10^{-7})$ T sin $(0.5\times10^3 x+1.5\times10^{11} t)$.

(a) What is the wavelength and frequency of the wave?

(b) Write an expression for the electric field.

Solution

(a) Comparing the given equation with

$$B_y = B_0 \text{Sin}\left[2p\left(\frac{x}{\lambda} + \frac{t}{T}\right)\right]$$

We get, $\lambda = \dfrac{2\pi}{0.5 \times 10^3}$ m = 1.26 cm,

and $\dfrac{1}{T} = \nu = \left(1.5 \times 10^{11}\right)/2\pi = 23.9$ GHz

(b) $E_0 = B_0 c = 2\times10^{-7}$ T × 3 × 10^8 m/s = 6 × 10^1 V/m

The electric field component is perpendicular to the direction of propagation and the direction of magnetic field. Therefore, the electric field component along the z-axis is obtained as

$E_z = 60 \sin (0.5 \times 10^3 x + 1.5 \times 10^{11}\ t)$ V/m

EXAMPLE 8.2

Electromagnetic spectrum
http://www.fnal.gov/pub/inquiring/more/light
http://imagine.gsfc.nasa.gov/docs/science/

PHYSICS

8.4 Electromagnetic Spectrum

At the time Maxwell predicted the existence of electromagnetic waves, the only familiar electromagnetic waves were the visible light waves. The existence of ultraviolet and infrared waves was barely established. By the end of the nineteenth century, X-rays and gamma rays had also been discovered. We now know that, electromagnetic waves include visible light waves, X-rays, gamma rays, radio waves, microwaves, ultraviolet and infrared waves. The classification of em waves according to frequency is the electromagnetic spectrum (Fig. 8.4). *There is no sharp division between one kind of wave and the next*. The classification is based roughly on how the waves are produced and/or detected.

We briefly describe these different types of electromagnetic waves, in order of decreasing wavelengths.

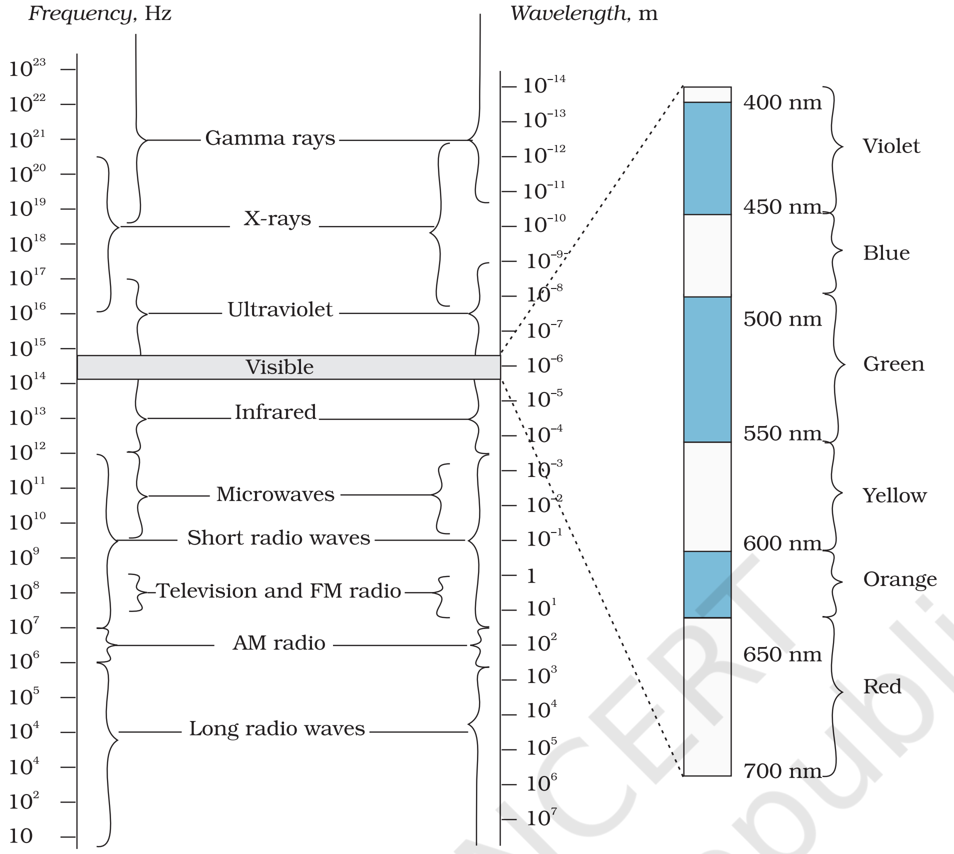


FIGURE 8.4 The electromagnetic spectrum, with common names for various part of it. The various regions do not have sharply defined boundaries.

8.4.1 Radio waves

Radio waves are produced by the accelerated motion of charges in conducting wires. They are used in radio and television communication systems. They are generally in the frequency range from 500 kHz to about 1000 MHz. The AM (amplitude modulated) band is from 530 kHz to 1710 kHz. Higher frequencies upto 54 MHz are used for *short wave* bands. TV waves range from 54 MHz to 890 MHz. The FM (frequency modulated) radio band extends from 88 MHz to 108 MHz. Cellular phones use radio waves to transmit voice communication in the ultrahigh frequency (UHF) band. How these waves are transmitted and received is described in Chapter 15.

8.4.2 Microwaves

Microwaves (short-wavelength radio waves), with frequencies in the gigahertz (GHz) range, are produced by special vacuum tubes (called klystrons, magnetrons and Gunn diodes). Due to their short wavelengths, they are suitable for the radar systems used in aircraft navigation. Radar also provides the basis for the speed guns used to time fast balls, tennis-serves, and automobiles. Microwave ovens are an interesting domestic application of these waves. In such ovens, the frequency of the microwaves is selected to match the resonant frequency of water molecules so that energy from the waves is transferred efficiently to the kinetic energy of the molecules. This raises the temperature of any food containing water.

8.4.3 Infrared waves

Infrared waves are produced by hot bodies and molecules. This band lies adjacent to the low-frequency or long-wave length end of the visible spectrum. Infrared waves are sometimes referred to as *heat waves*. This is because water molecules present in most materials readily absorb infrared waves (many other molecules, for example, CO_2, NH_3, also absorb infrared waves). After absorption, their thermal motion increases, that is, they heat up and heat their surroundings. Infrared lamps are used in physical therapy. Infrared radiation also plays an important role in maintaining the earth's warmth or average temperature through the greenhouse effect. Incoming visible light (which passes relatively easily through the atmosphere) is absorbed by the earth's surface and re-radiated as infrared (longer wavelength) radiations. This radiation is trapped by greenhouse gases such as carbon dioxide and water vapour. Infrared detectors are used in Earth satellites, both for military purposes and to observe growth of crops. Electronic devices (for example semiconductor light emitting diodes) also emit infrared and are widely used in the remote switches of household electronic systems such as TV sets, video recorders and hi-fi systems.

8.4.4 Visible rays

It is the most familiar form of electromagnetic waves. It is the part of the spectrum that is detected by the human eye. It runs from about 4×10^{14} Hz to about 7×10^{14} Hz or a wavelength range of about 700 – 400 nm. Visible light emitted or reflected from objects around us provides us information about the world. Our eyes are sensitive to this range of wavelengths. Different animals are sensitive to different range of wavelengths. For example, snakes can detect infrared waves, and the 'visible' range of many insects extends well into the utraviolet.

8.4.5 Ultraviolet rays

It covers wavelengths ranging from about 4×10^{-7} m (400 nm) down to 6×10^{-10}m (0.6 nm). Ultraviolet (UV) radiation is produced by special lamps and very hot bodies. The sun is an important source of ultraviolet light. But fortunately, most of it is absorbed in the ozone layer in the atmosphere at an altitude of about 40 – 50 km. UV light in large quantities has harmful effects on humans. Exposure to UV radiation induces the production of more melanin, causing tanning of the skin. UV radiation is absorbed by ordinary glass. Hence, one cannot get tanned or sunburn through glass windows.

Welders wear special glass goggles or face masks with glass windows to protect their eyes from large amount of UV produced by welding arcs. Due to its shorter wavelengths, UV radiations can be focussed into very narrow beams for high precision applications such as LASIK (*Laser-assisted in situ keratomileusis*) eye surgery. UV lamps are used to kill germs in water purifiers.

Ozone layer in the atmosphere plays a protective role, and hence its depletion by chlorofluorocarbons (CFCs) gas (such as freon) is a matter of international concern.

Reprint 2025-26

8.4.6 X-rays

Beyond the UV region of the electromagnetic spectrum lies the X-ray region. We are familiar with X-rays because of its medical applications. It covers wavelengths from about 10^{-8} m (10 nm) down to 10^{-13} m (10^{-4} nm). One common way to generate X-rays is to bombard a metal target by high energy electrons. X-rays are used as a diagnostic tool in medicine and as a treatment for certain forms of cancer. Because X-rays damage or destroy living tissues and organisms, care must be taken to avoid unnecessary or over exposure.

8.4.7 Gamma rays

They lie in the upper frequency range of the electromagnetic spectrum and have wavelengths of from about 10^{-10}m to less than 10^{-14}m. This high frequency radiation is produced in nuclear reactions and also emitted by radioactive nuclei. They are used in medicine to destroy cancer cells.

Table 8.1 summarises different types of electromagnetic waves, their production and detections. As mentioned earlier, the demarcation between different regions is not sharp and there are overlaps.

TABLE 8.1 DIFFERENT TYPES OF ELECTROMAGNETIC WAVES

Type	Wavelength range	Production	Detection
Radio	> 0.1 m	Rapid acceleration and decelerations of electrons in aerials	Receiver's aerials
Microwave	0.1m to 1 mm	Klystron valve or magnetron valve	Point contact diodes
Infra-red	1mm to 700 nm	Vibration of atoms and molecules	Thermopiles Bolometer, Infrared photographic film
Light	700 nm to 400 nm	Electrons in atoms emit light when they move from one energy level to a lower energy level	The eye Photocells Photographic film
Ultraviolet	400 nm to 1nm	Inner shell electrons in atoms moving from one energy level to a lower level	Photocells Photographic film
X-rays	1nm to 10^{-3} nm	X-ray tubes or inner shell electrons	Photographic film Geiger tubes Ionisation chamber
Gamma rays	$<10^{-3}$ nm	Radioactive decay of the nucleus	-do-

Reprint 2025-26

SUMMARY

1. Maxwell found an inconsistency in the Ampere's law and suggested the existence of an additional current, called displacement current, to remove this inconsistency. This displacement current is due to time-varying electric field and is given by

$$i_d = \varepsilon_0 \frac{d\Phi_E}{dt}$$

and acts as a source of magnetic field in exactly the same way as conduction current.

2. An accelerating charge produces electromagnetic waves. An electric charge oscillating harmonically with frequency v, produces electromagnetic waves of the same frequency v. An electric dipole is a basic source of electromagnetic waves.
3. Electromagnetic waves with wavelength of the order of a few metres were first produced and detected in the laboratory by Hertz in 1887. He thus verified a basic prediction of Maxwell's equations.
4. Electric and magnetic fields oscillate sinusoidally in space and time in an electromagnetic wave. The oscillating electric and magnetic fields, **E** and **B** are perpendicular to each other, and to the direction of propagation of the electromagnetic wave. For a wave of frequency v, wavelength λ, propagating along z-direction, we have

$$E = E_x(t) = E_0 \sin (kz - \omega t)$$

$$= E_0 \sin \left[2\pi\left(\frac{z}{\lambda} - vt\right)\right] = E_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \frac{t}{T}\right)\right]$$

$$B = B_y(t) = B_0 \sin (kz - \omega t)$$

$$= B_0 \sin\left[2\pi\left(\frac{z}{\lambda} - vt\right)\right] = B_0 \sin\left[2\pi\left(\frac{z}{\lambda} - \frac{t}{T}\right)\right]$$

They are related by $E_0/B_0 = c$.

5. The speed c of electromagnetic wave in vacuum is related to μ_0 and ε_0 (the free space permeability and permittivity constants) as follows: $c = 1/\sqrt{\mu_0 \varepsilon_0}$. The value of c equals the speed of light obtained from optical measurements.

Light is an electromagnetic wave; c is, therefore, also the speed of light. Electromagnetic waves other than light also have the same velocity c in free space.

The speed of light, or of electromagnetic waves in a material medium is given by $v = 1/\sqrt{\mu \varepsilon}$

where μ is the permeability of the medium and ε its permittivity.

6. The spectrum of electromagnetic waves stretches, in principle, over an infinite range of wavelengths. Different regions are known by different names; γ-rays, X-rays, ultraviolet rays, visible rays, infrared rays, microwaves and radio waves in order of increasing wavelength from 10^{-2} Å or 10^{-12} m to 10^{6} m.

They interact with matter via their electric and magnetic fields which set in oscillation charges present in all matter. The detailed interaction and so the mechanism of absorption, scattering, etc., depend on the wavelength of the electromagnetic wave, and the nature of the atoms and molecules in the medium.

POINTS TO PONDER

1. The basic difference between various types of electromagnetic waves lies in their wavelengths or frequencies since all of them travel through vacuum with the same speed. Consequently, the waves differ considerably in their mode of interaction with matter.
2. Accelerated charged particles radiate electromagnetic waves. The wavelength of the electromagnetic wave is often correlated with the characteristic size of the system that radiates. Thus, gamma radiation, having wavelength of 10^{-14} m to 10^{-15} m, typically originate from an atomic nucleus. X-rays are emitted from heavy atoms. Radio waves are produced by accelerating electrons in a circuit. A transmitting antenna can most efficiently radiate waves having a wavelength of about the same size as the antenna. Visible radiation emitted by atoms is, however, much longer in wavelength than atomic size.
3. Infrared waves, with frequencies lower than those of visible light, vibrate not only the electrons, but entire atoms or molecules of a substance. This vibration increases the internal energy and consequently, the temperature of the substance. This is why infrared waves are often called *heat waves*.
4. The centre of sensitivity of our eyes coincides with the centre of the wavelength distribution of the sun. It is because humans have evolved with visions most sensitive to the strongest wavelengths from the sun.

EXERCISES

8.1 Figure 8.5 shows a capacitor made of two circular plates each of radius 12 cm, and separated by 5.0 cm. The capacitor is being charged by an external source (not shown in the figure). The charging current is constant and equal to 0.15A.

(a) Calculate the capacitance and the rate of change of potential difference between the plates.

(b) Obtain the displacement current across the plates.

(c) Is Kirchhoff's first rule (junction rule) valid at each plate of the capacitor? Explain.

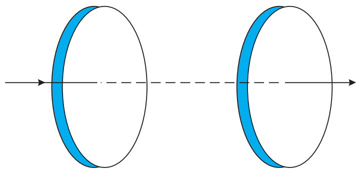


FIGURE 8.5

8.2 A parallel plate capacitor (Fig. 8.6) made of circular plates each of radius R = 6.0 cm has a capacitance C = 100 pF. The capacitor is connected to a 230 V ac supply with a (angular) frequency of 300 rad s^{-1}.

Reprint 2025-26

(a) What is the rms value of the conduction current?
(b) Is the conduction current equal to the displacement current?
(c) Determine the amplitude of **B** at a point 3.0 cm from the axis between the plates.

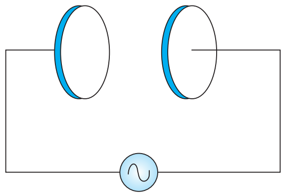


FIGURE 8.6

8.3 What physical quantity is the same for X-rays of wavelength 10^{-10}m, red light of wavelength 6800 Å and radiowaves of wavelength 500m?

8.4 A plane electromagnetic wave travels in vacuum along *z*-direction. What can you say about the directions of its electric and magnetic field vectors? If the frequency of the wave is 30 MHz, what is its wavelength?

8.5 A radio can tune in to any station in the 7.5 MHz to 12 MHz band. What is the corresponding wavelength band?

8.6 A charged particle oscillates about its mean equilibrium position with a frequency of 10^9 Hz. What is the frequency of the electromagnetic waves produced by the oscillator?

8.7 The amplitude of the magnetic field part of a harmonic electromagnetic wave in vacuum is B_0 = 510 nT. What is the amplitude of the electric field part of the wave?

8.8 Suppose that the electric field amplitude of an electromagnetic wave is E_0 = 120 N/C and that its frequency is ν = 50.0 MHz. (a) Determine, B_0,ω, *k*, and λ. (b) Find expressions for **E** and **B**.

8.9 The terminology of different parts of the electromagnetic spectrum is given in the text. Use the formula $E = h\nu$ (for energy of a quantum of radiation: photon) and obtain the photon energy in units of eV for different parts of the electromagnetic spectrum. In what way are the different scales of photon energies that you obtain related to the sources of electromagnetic radiation?

8.10 In a plane electromagnetic wave, the electric field oscillates sinusoidally at a frequency of 2.0×10^{10} Hz and amplitude 48 V m^{-1}.
(a) What is the wavelength of the wave?
(b) What is the amplitude of the oscillating magnetic field?
(c) Show that the average energy density of the **E** field equals the average energy density of the **B** field. [$c = 3 \times 10^8$ m s^{-1}.]

Reprint 2025-26

12090CH09

Chapter Nine

RAY OPTICS AND OPTICAL INSTRUMENTS

9.1 Introduction

Nature has endowed the human eye (retina) with the sensitivity to detect electromagnetic waves within a small range of the electromagnetic spectrum. Electromagnetic radiation belonging to this region of the spectrum (wavelength of about 400 nm to 750 nm) is called light. It is mainly through light and the sense of vision that we know and interpret the world around us.

There are two things that we can intuitively mention about light from common experience. First, that it travels with enormous speed and second, that it travels in a straight line. It took some time for people to realise that the speed of light is finite and measurable. Its presently accepted value in vacuum is $c = 2.99792458 \times 10^8$ m s^{-1}. For many purposes, it suffices to take $c = 3 \times 10^8$ m s^{-1}. The speed of light in vacuum is the highest speed attainable in nature.

The intuitive notion that light travels in a straight line seems to contradict what we have learnt in Chapter 8, that light is an electromagnetic wave of wavelength belonging to the visible part of the spectrum. How to reconcile the two facts? The answer is that the wavelength of light is very small compared to the size of ordinary objects that we encounter commonly (generally of the order of a few cm or larger). In this situation, as you will learn in Chapter 10, a light wave can be considered to travel from one point to another, along a straight line joining them. The path is called a *ray* of light, and a bundle of such rays constitutes a *beam* of light.

In this chapter, we consider the phenomena of reflection, refraction and dispersion of light, using the ray picture of light. Using the basic laws of reflection and refraction, we shall study the image formation by plane and spherical reflecting and refracting surfaces. We then go on to describe the construction and working of some important optical instruments, including the human eye.

9.2 Reflection of Light by Spherical Mirrors

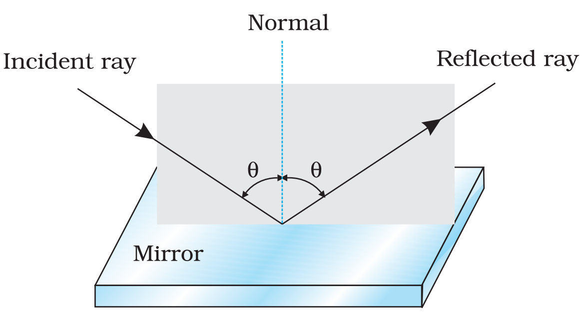


FIGURE 9.1 The incident ray, reflected ray and the normal to the reflecting surface lie in the same plane.

We are familiar with the laws of reflection. The angle of reflection (i.e., the angle between reflected ray and the normal to the reflecting surface or the mirror) equals the angle of incidence (angle between incident ray and the normal). Also that the incident ray, reflected ray and the normal to the reflecting surface at the point of incidence lie in the same plane (Fig. 9.1). These laws are valid at each point on any reflecting surface whether plane or curved. However, we shall restrict our discussion to the special case of curved surfaces, that is, spherical surfaces. The normal in this case is to be taken as normal to the tangent to surface at the point of incidence. That is, the normal is along the radius, the line joining the centre of curvature of the mirror to the point of incidence.

We have already studied that the geometric centre of a spherical mirror is called its pole while that of a spherical lens is called its optical centre. The line joining the pole and the centre of curvature of the spherical mirror is known as the *principal axis*. In the case of spherical lenses, the principal axis is the line joining the optical centre with its principal focus as you will see later.

9.2.1 Sign convention

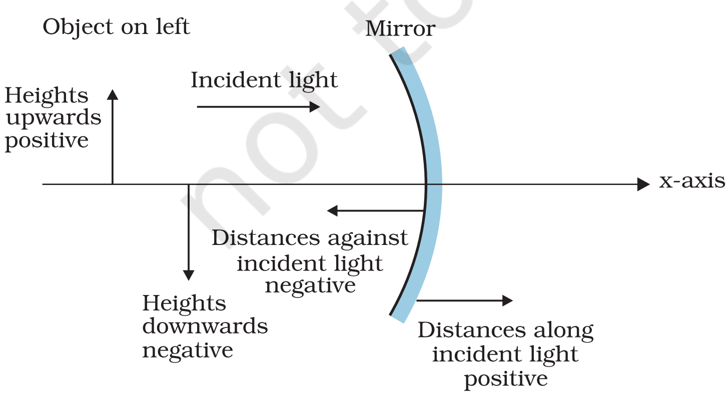


FIGURE 9.2 The Cartesian Sign Convention.

To derive the relevant formulae for reflection by spherical mirrors and refraction by spherical lenses, we must first adopt a sign convention for measuring distances. In this book, we shall follow the *Cartesian sign convention*. According to this convention, all distances are measured from the pole of the mirror or the optical centre of the lens. The distances measured in the same direction as the incident light are taken as positive and those measured in the direction opposite to the direction of incident light are taken as negative (Fig. 9.2). The heights measured upwards with respect to x-axis and normal to the

Reprint 2025-26

principal axis (x-axis) of the mirror/lens are taken as positive (Fig. 9.2). The heights measured downwards are taken as negative.

With a common accepted convention, it turns out that a single formula for spherical mirrors and a single formula for spherical lenses can handle all different cases.

9.2.2 Focal length of spherical mirrors

Figure 9.3 shows what happens when a parallel beam of light is incident on (a) a concave mirror, and (b) a convex mirror. We assume that the rays are *paraxial*, i.e., they are incident at points close to the pole P of the mirror and make small angles with the principal axis. The reflected rays converge at a point F on the principal axis of a concave mirror [Fig. 9.3(a)]. For a convex mirror, the reflected rays appear to diverge from a point F on its principal axis [Fig. 9.3(b)]. The point F is called the *principal focus* of the mirror. If the parallel paraxial beam of light were incident, making some angle with the principal axis, the reflected rays would converge (or appear to diverge) from a point in a plane through F normal to the principal axis. This is called the *focal plane* of the mirror [Fig. 9.3(c)].

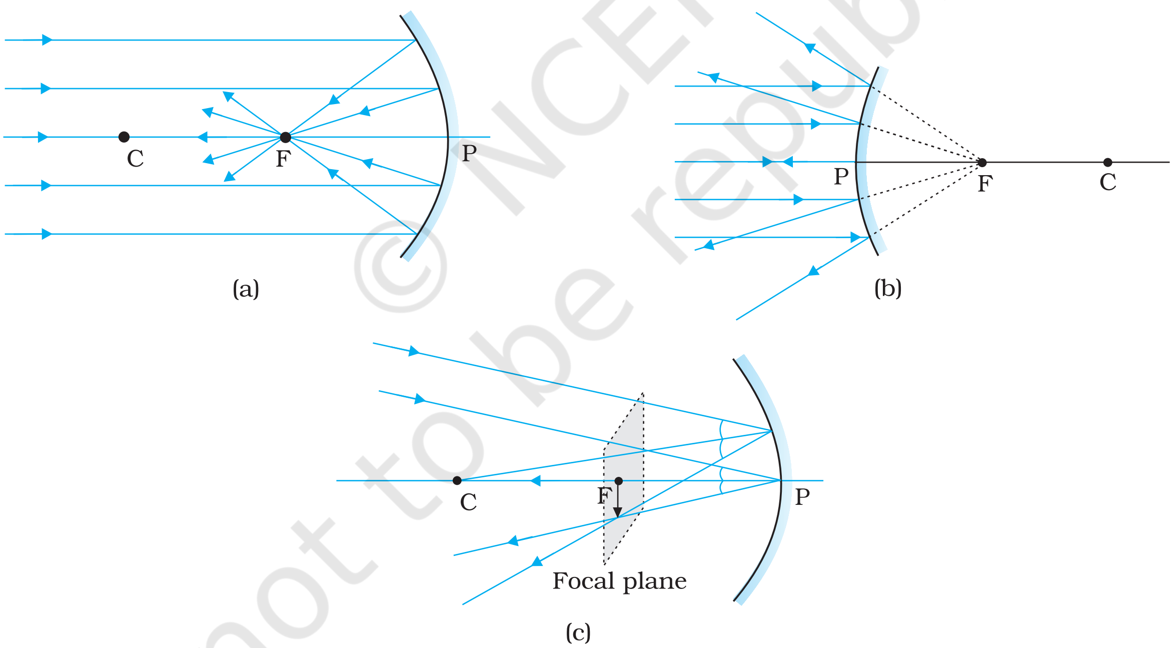
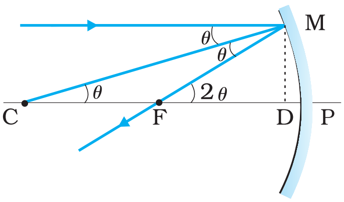


FIGURE 9.3 Focus of a concave and convex mirror.

The distance between the focus F and the pole P of the mirror is called the *focal length* of the mirror, denoted by f. We now show that $f = R/2$, where R is the radius of curvature of the mirror. The geometry of reflection of an incident ray is shown in Fig. 9.4.

Let C be the centre of curvature of the mirror. Consider a ray parallel to the principal axis striking the mirror at M. Then CM will be perpendicular to the mirror at M. Let θ be the angle of incidence, and MD



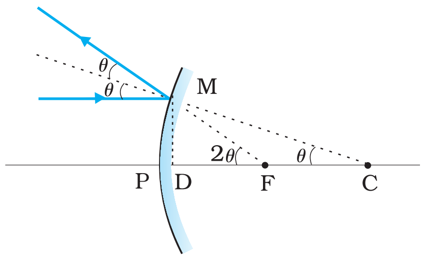


FIGURE 9.4 Geometry of reflection of an incident ray on (a) concave spherical mirror, and (b) convex spherical mirror.

be the perpendicular from M on the principal axis. Then,

$\angle MCP = \theta$ and $\angle MFP = 2\theta$

Now,

$$\tan\theta = \frac{MD}{CD} \text{ and } \tan 2\theta = \frac{MD}{FD} \qquad (9.1)$$

For small θ, which is true for paraxial rays, $\tan\theta \approx \theta$, $\tan 2\theta \approx 2\theta$. Therefore, Eq. (9.1) gives

$$\frac{MD}{FD} = 2\,\frac{MD}{CD}$$

$$\text{or, } FD = \frac{CD}{2} \qquad (9.2)$$

Now, for small θ, the point D is very close to the point P. Therefore, FD = f and CD = R. Equation (9.2) then gives

$$f = R/2 \qquad (9.3)$$

9.2.3 The mirror equation

If rays emanating from a point actually meet at another point after reflection and/or refraction, that point is called the *image* of the first point. The image is *real* if the rays actually converge to the point; it is *virtual* if the rays do not actually meet but appear to diverge from the point when produced backwards. An image is thus a point-to-point correspondence with the object established through reflection and/or refraction.

In principle, we can take any two rays emanating from a point on an object, trace their paths, find their point of intersection and thus, obtain the image of the point due to reflection at a spherical mirror. In practice, however, it is convenient to choose any two of the following rays:

(i) The ray from the point which is parallel to the principal axis. The reflected ray goes through the focus of the mirror.

(ii) The ray passing through the centre of curvature of a concave mirror or appearing to pass through it for a convex mirror. The reflected ray simply retraces the path.

(iii) The ray passing through (or directed towards) the focus of the concave mirror or appearing to pass through (or directed towards) the focus of a convex mirror. The reflected ray is parallel to the principal axis.

(iv) The ray incident at any angle at the pole. The reflected ray follows laws of reflection.

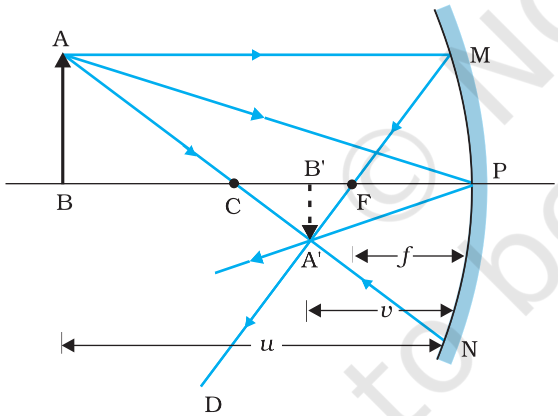


FIGURE 9.5 Ray diagram for image formation by a concave mirror.

Figure 9.5 shows the ray diagram considering three rays. It shows the image A′B′ (in this case, real) of an object AB formed by a concave mirror. It does not mean that only three rays emanate from the point A. An infinite number of rays emanate from any source, in all directions. Thus, point A′ is image point of A if every ray originating at point A and falling on the concave mirror after reflection passes through the point A′.

We now derive the mirror equation or the relation between the object distance (u), image distance (v) and the focal length (f).

From Fig. 9.5, the two right-angled triangles A′B′F and MPF are similar. (For paraxial rays, MP can be considered to be a straight line perpendicular to CP.) Therefore,

$$\frac{B'A'}{PM} = \frac{B'F}{FP}$$

or $\frac{B'A'}{BA} = \frac{B'F}{FP}$ ($\because$ PM = AB) (9.4)

Since $\angle$ APB = $\angle$ A′PB′, the right angled triangles A′B′P and ABP are also similar. Therefore,

$$\frac{B'A'}{BA} = \frac{B'P}{BP} \qquad (9.5)$$

Comparing Eqs. (9.4) and (9.5), we get

$$\frac{B'F}{FP} = \frac{B'P - FP}{FP} = \frac{B'P}{BP} \qquad (9.6)$$

Equation (9.6) is a relation involving magnitude of distances. We now apply the sign convention. We note that light travels from the object to the mirror MPN. Hence this is taken as the positive direction. To reach the object AB, image A′B′ as well as the focus F from the pole P, we have to travel opposite to the direction of incident light. Hence, all the three will have negative signs. Thus,

B′ P = $-v$, FP = $-f$, BP = $-u$

Using these in Eq. (9.6), we get

$$\frac{-v+f}{-f} = \frac{-v}{-u}$$

or $\frac{v-f}{f} = \frac{v}{u}$

$$\frac{v}{f} = 1 + \frac{v}{u}$$

Dividing it by v, we get

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f} \qquad (9.7)$$

This relation is known as the *mirror equation*.

The size of the image relative to the size of the object is another important quantity to consider. We define linear *magnification* (m) as the ratio of the height of the image (h') to the height of the object (h):

$$m = \frac{h'}{h} \qquad (9.8)$$

h and h' will be taken positive or negative in accordance with the accepted sign convention. In triangles A′B′P and ABP, we have,

$$\frac{B'A'}{BA} = \frac{B'P}{BP}$$

With the sign convention, this becomes

$$\frac{-h'}{h} = \frac{-v}{-u}$$

so that

$$m = \frac{h'}{h} = -\frac{v}{u} \qquad (9.9)$$

We have derived here the mirror equation, Eq. (9.7), and the magnification formula, Eq. (9.9), for the case of real, inverted image formed by a concave mirror. With the proper use of sign convention, these are, in fact, valid for all the cases of reflection by a spherical mirror (concave or convex) whether the image formed is real or virtual. Figure 9.6 shows the ray diagrams for virtual image formed by a concave and convex mirror. You should verify that Eqs. (9.7) and (9.9) are valid for these cases as well.

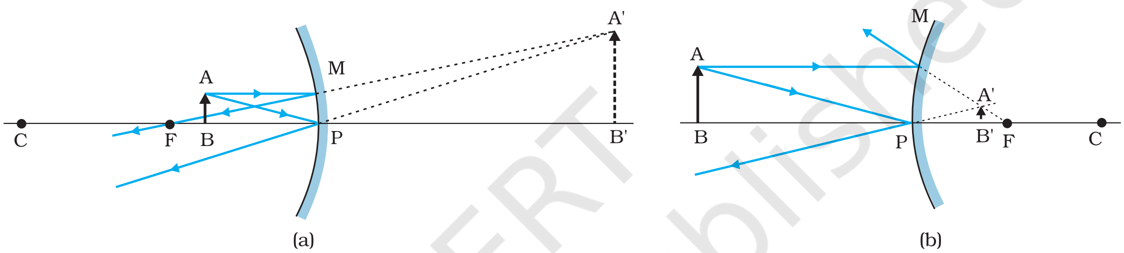


FIGURE 9.6 Image formation by (a) a concave mirror with object between P and F, and (b) a convex mirror.

EXAMPLE 9.1

Example 9.1 Suppose that the lower half of the concave mirror's reflecting surface in Fig. 9.6 is covered with an opaque (non-reflective) material. What effect will this have on the image of an object placed in front of the mirror?

Solution You may think that the image will now show only half of the object, but taking the laws of reflection to be true for all points of the remaining part of the mirror, the image will be that of the whole object. However, as the area of the reflecting surface has been reduced, the intensity of the image will be low (in this case, half).

EXAMPLE 9.2

Example 9.2 A mobile phone lies along the principal axis of a concave mirror, as shown in Fig. 9.7. Show by suitable diagram, the formation of its image. Explain why the magnification is not uniform. Will the distortion of image depend on the location of the phone with respect to the mirror?

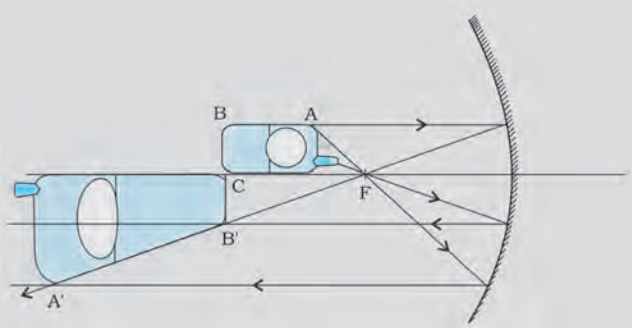


FIGURE 9.7

Solution
The ray diagram for the formation of the image of the phone is shown in Fig. 9.7. The image of the part which is on the plane perpendicular to principal axis will be on the same plane. It will be of the same size, i.e., B′C = BC. You can yourself realise why the image is distorted.

EXAMPLE 9.2

Example 9.3 An object is placed at (i) 10 cm, (ii) 5 cm in front of a concave mirror of radius of curvature 15 cm. Find the position, nature, and magnification of the image in each case.

Solution
The focal length $f = -15/2$ cm $= -7.5$ cm

(i) The object distance $u = -10$ cm. Then Eq. (9.7) gives

$$\frac{1}{v} + \frac{1}{-10} = \frac{1}{-7.5}$$

$$\text{or} \quad v = \frac{10 \times 7.5}{-2.5} = -\,30 \text{ cm}$$

The image is 30 cm from the mirror on the same side as the object.

Also, magnification $m = -\dfrac{v}{u} = -\dfrac{(-30)}{(-10)} = -3$

The image is magnified, real and inverted.

(ii) The object distance $u = -5$ cm. Then from Eq. (9.7),

$$\frac{1}{v} + \frac{1}{-5} = \frac{1}{-7.5}$$

$$\text{or} \quad v = \frac{5 \times 7.5}{(7.5 - 5)} = 15 \text{ cm}$$

This image is formed at 15 cm behind the mirror. It is a virtual image.

Magnification $m = -\dfrac{v}{u} = -\dfrac{15}{(-5)} = 3$

The image is magnified, virtual and erect.

EXAMPLE 9.3

Example 9.4 Suppose while sitting in a parked car, you notice a jogger approaching towards you in the side view mirror of $R = 2$ m. If the jogger is running at a speed of 5 m s^{-1}, how fast the image of the jogger appear to move when the jogger is (a) 39 m, (b) 29 m, (c) 19 m, and (d) 9 m away.

Solution
From the mirror equation, Eq. (9.7), we get

$$v = \frac{fu}{u - f}$$

For convex mirror, since $R = 2$ m, $f = 1$ m. Then

for $u = -39$ m, $v = \dfrac{(-39) \times 1}{-39 - 1} = \dfrac{39}{40}$ m

Since the jogger moves at a constant speed of 5 m s^{-1}, after 1 s the position of the image v (for $u = -39 + 5 = -34$) is (34/35)m.

EXAMPLE 9.4

EXAMPLE 9.4

The shift in the position of image in 1 s is

$$\frac{39}{40}-\frac{34}{35}=\frac{1365-1360}{1400}=\frac{5}{1400}=\frac{1}{280}\,\text{m}$$

Therefore, the average speed of the image when the jogger is between 39 m and 34 m from the mirror, is (1/280) m s^{-1}

Similarly, it can be seen that for u = –29 m, –19 m and –9 m, the speed with which the image appears to move is

$$\frac{1}{150}\,\text{m s}^{-1},\ \frac{1}{60}\,\text{m s}^{-1}\ \text{and}\ \frac{1}{10}\,\text{m s}^{-1},\ \text{respectively.}$$

Although the jogger has been moving with a constant speed, the speed of his/her image appears to increase substantially as he/she moves closer to the mirror. This phenomenon can be noticed by any person sitting in a stationary car or a bus. In case of moving vehicles, a similar phenomenon could be observed if the vehicle in the rear is moving closer with a constant speed.

9.3 REFRACTION

When a beam of light encounters another transparent medium, a part of light gets reflected back into the first medium while the rest enters the other. A ray of light represents a beam. The direction of propagation of an obliquely incident (0°< i < 90°) ray of light that enters the other medium, changes at the interface of the two media. This phenomenon is called *refraction of light*. Snell experimentally obtained the following laws of refraction:

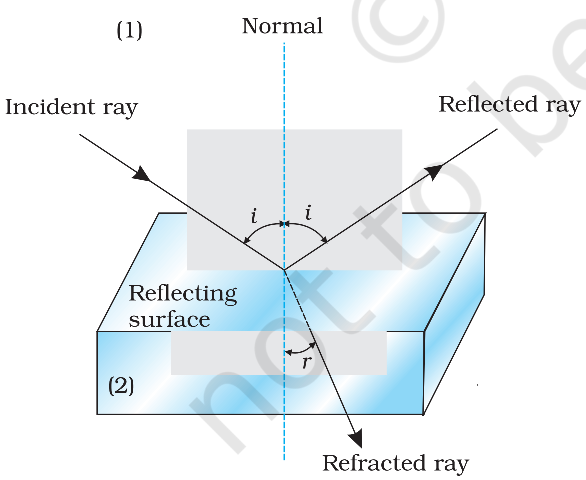


FIGURE 9.8 Refraction and reflection of light.

(i) The incident ray, the refracted ray and the normal to the interface at the point of incidence, all lie in the same plane.

(ii) The ratio of the sine of the angle of incidence to the sine of angle of refraction is constant. Remember that the angles of incidence (i) and refraction (r) are the angles that the incident and its refracted ray make with the normal, respectively. We have

$$\frac{\sin i}{\sin r}=n_{21} \qquad (9.10)$$

where n_{21} is a constant, called the *refractive index* of the second medium with respect to the first medium. Equation (9.10) is the well-known Snell's law of refraction. We note that n_{21} is a characteristic of the pair of media (and also depends on the wavelength of light), but is independent of the angle of incidence.

From Eq. (9.10), if $n_{21} > 1$, $r < i$, i.e., the refracted ray bends towards the normal. In such a case medium 2 is said to be *optically denser* (or *denser*, in short) than medium 1. On the other hand, if $n_{21} <1$, $r > i$, the refracted ray bends away from the normal. This is the case when incident ray in a denser medium refracts into a rarer medium.

Note: *Optical density should not be confused with mass density, which is mass per unit volume. It is possible that mass density of an optically denser medium may be less than that of an optically rarer medium (optical density is the ratio of the speed of light in two media). For example, turpentine and water. Mass density of turpentine is less than that of water but its optical density is higher.*

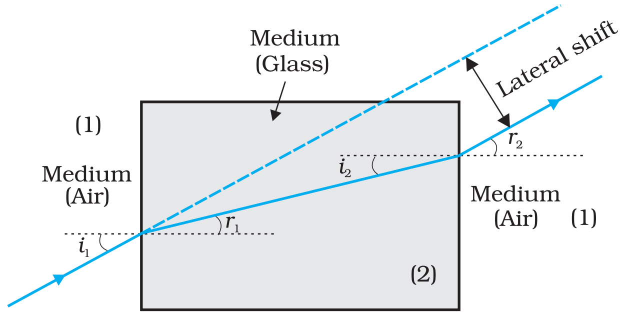


FIGURE 9.9 Lateral shift of a ray refracted through a parallel-sided slab.

If n_{21} is the refractive index of medium 2 with respect to medium 1 and n_{12} the refractive index of medium 1 with respect to medium 2, then it should be clear that

$$n_{12} = \frac{1}{n_{21}} \qquad (9.11)$$

It also follows that if n_{32} is the refractive index of medium 3 with respect to medium 2 then $n_{32} = n_{31} \times n_{12}$, where n_{31} is the refractive index of medium 3 with respect to medium 1.

Some elementary results based on the laws of refraction follow immediately. For a rectangular slab, refraction takes place at two interfaces (air-glass and glass-air). It is easily seen from Fig. 9.9 that $r_2 = i_1$, i.e., the emergent ray is parallel to the incident ray—there is no deviation, but it does suffer lateral displacement/shift with respect to the incident ray. Another familiar observation is that the bottom of a tank filled with water appears to be raised (Fig. 9.10). For viewing near the normal direction, it can be shown that the apparent depth (h_1) is real depth (h_2) divided by the refractive index of the medium (water).

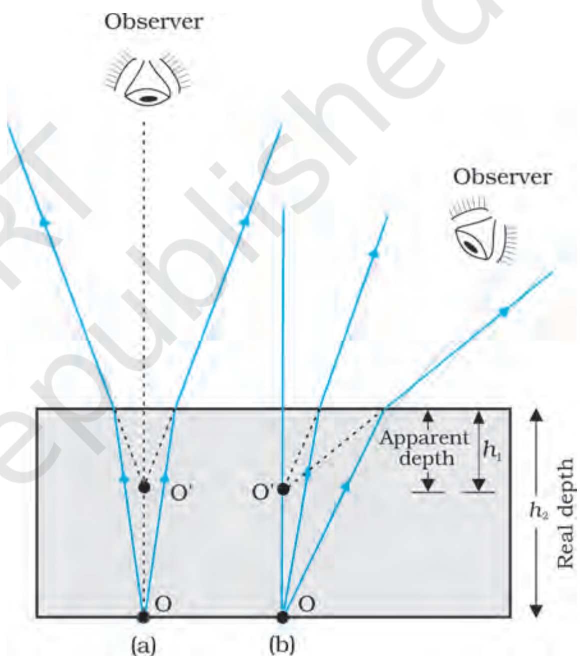


FIGURE 9.10 Apparent depth for (a) normal, and (b) oblique viewing.

9.4 Total Internal Reflection

When light travels from an optically denser medium to a rarer medium at the interface, it is partly reflected back into the same medium and partly refracted to the second medium. This reflection is called the *internal reflection*.

When a ray of light enters from a denser medium to a rarer medium, it bends away from the normal, for example, the ray AO_1B in Fig. 9.11. The incident ray AO_1 is partially reflected (O_1C) and partially transmitted (O_1B) or refracted, the angle of refraction (r) being larger than the angle of incidence (i). As the angle of incidence increases, so does the angle of

Reprint 2025-26

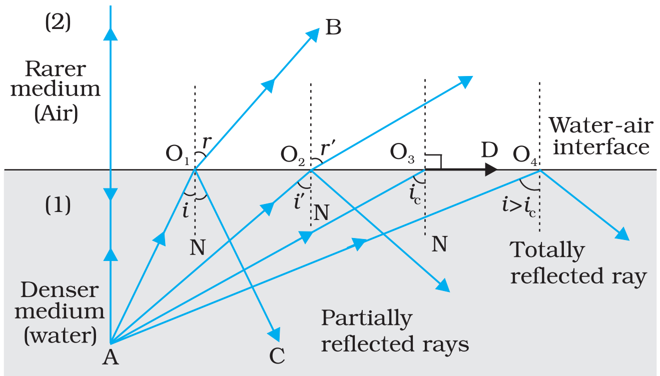


FIGURE 9.11 Refraction and internal reflection of rays from a point A in the denser medium (water) incident at different angles at the interface with a rarer medium (air).

refraction, till for the ray AO_3, the angle of refraction is $\pi/2$. The refracted ray is bent so much away from the normal that it grazes the surface at the interface between the two media. This is shown by the ray AO_3D in Fig. 9.11. If the angle of incidence is increased still further (e.g., the ray AO_4), refraction is not possible, and the incident ray is totally reflected. This is called *total internal reflection*. When light gets reflected by a surface, normally some fraction of it gets transmitted. The reflected ray, therefore, is always less intense than the incident ray, howsoever smooth the reflecting surface may be. In total internal reflection, on the other hand, no transmission of light takes place.

The angle of incidence corresponding to an angle of refraction 90°, say $\angle AO_3N$, is called the *critical angle* (i_c) for the given pair of media. We see from Snell's law [Eq. (9.10)] that if the relative refractive index of the refracting medium is less than one then, since the maximum value of sin r is unity, there is an upper limit to the value of sin i for which the law can be satisfied, that is, $i = i_c$ such that

$$\sin i_c = n_{21} \tag{9.12}$$

For values of i larger than i_c, Snell's law of refraction cannot be satisfied, and hence no refraction is possible.

The refractive index of denser medium 1 with respect to rarer medium 2 will be $n_{12} = 1/\sin i_c$. Some typical critical angles are listed in Table 9.1.

TABLE 9.1 CRITICAL ANGLE OF SOME TRANSPARENT MEDIA WITH RESPECT TO AIR

Substance medium	Refractive index	Critical angle
Water	1.33	48.75
Crown glass	1.52	41.14
Dense flint glass	1.62	37.31
Diamond	2.42	24.41

A demonstration for total internal reflection

All optical phenomena can be demonstrated very easily with the use of a laser torch or pointer, which is easily available nowadays. Take a glass beaker with clear water in it. Add a few drops of milk or any other suspension to water and stir so that water becomes a little turbid. Take a laser pointer and shine its beam through the turbid water. You will find that the path of the beam inside the water shines brightly.

Reprint 2025-26

Shine the beam from below the beaker such that it strikes at the upper water surface at the other end. Do you find that it undergoes partial reflection (which is seen as a spot on the table below) and partial refraction [which comes out in the air and is seen as a spot on the roof; Fig. 9.12(a)]? Now direct the laser beam from one side of the beaker such that it strikes the upper surface of water more obliquely [Fig. 9.12(b)]. Adjust the direction of laser beam until you find the angle for which the refraction above the water surface is totally absent and the beam is totally reflected back to water. This is total internal reflection at its simplest.

Pour this water in a long test tube and shine the laser light from top, as shown in Fig. 9.12(c). Adjust the direction of the laser beam such that it is totally internally reflected every time it strikes the walls of the tube. This is similar to what happens in optical fibres.

Take care not to look into the laser beam directly and not to point it at anybody's face.

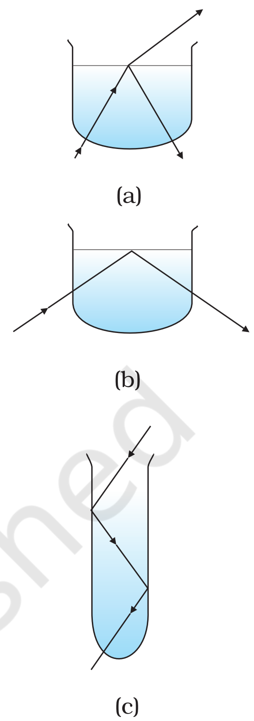


FIGURE 9.12 Observing total internal reflection in water with a laser beam (refraction due to glass of beaker neglected being very thin).

9.4.1 Total internal reflection in nature and its technological applications

(i) *Prism*: Prisms designed to bend light by 90 or by 180 make use of total internal reflection [Fig. 9.13(a) and (b)]. Such a prism is also used to invert images without changing their size [Fig. 9.13(c)].
In the first two cases, the critical angle i_c for the material of the prism must be less than 45 . We see from Table 9.1 that this is true for both crown glass and dense flint glass.

(ii) *Optical fibres*: Nowadays optical fibres are extensively used for transmitting audio and video signals through long distances. Optical fibres too make use of the phenomenon of total internal reflection. Optical fibres are fabricated with high quality composite glass/quartz fibres. Each fibre consists of a core and cladding. The refractive index of the material of the core is higher than that of the cladding.

When a signal in the form of light is directed at one end of the fibre at a suitable angle, it undergoes repeated total internal reflections along the length of the fibre and finally comes out at the other end (Fig. 9.14). Since light undergoes total internal reflection at each stage, there is no appreciable loss in the intensity of the light signal. Optical fibres are fabricated such that light reflected at one side of inner surface strikes the other at an angle larger than the critical angle. Even if the fibre is bent, light can easily travel along its length. Thus, an optical fibre can be used to act as an optical pipe.

A bundle of optical fibres can be put to several uses. Optical fibres are extensively used for transmitting and receiving

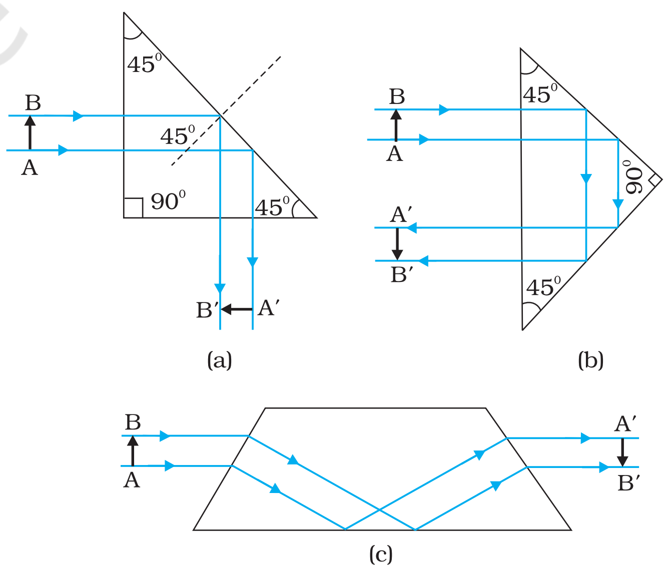


FIGURE 9.13 Prisms designed to bend rays by 90 and 180 or to invert image without changing its size make use of total internal reflection.

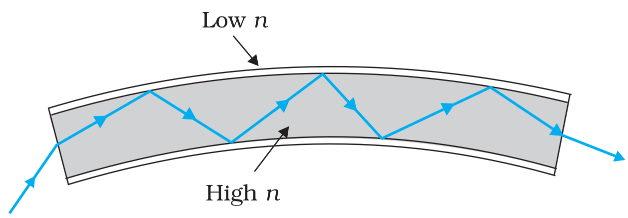


FIGURE 9.14 Light undergoes successive total internal reflections as it moves through an optical fibre.

electrical signals which are converted to light by suitable transducers. Obviously, optical fibres can also be used for transmission of optical signals. For example, these are used as a 'light pipe' to facilitate visual examination of internal organs like esophagus, stomach and intestines. You might have seen a commonly available decorative lamp with fine plastic fibres with their free ends forming a fountain like structure. The other end of the fibres is fixed over an electric lamp. When the lamp is switched on, the light travels from the bottom of each fibre and appears at the tip of its free end as a dot of light. The fibres in such decorative lamps are optical fibres.

The main requirement in fabricating optical fibres is that there should be very little absorption of light as it travels for long distances inside them. This has been achieved by purification and special preparation of materials such as quartz. In silica glass fibres, it is possible to transmit more than 95% of the light over a fibre length of 1 km. (Compare with what you expect for a block of ordinary window glass 1 km thick.)

9.5 Refraction at Spherical Surfaces and by Lenses

We have so far considered refraction at a plane interface. We shall now consider refraction at a spherical interface between two transparent media. An infinitesimal part of a spherical surface can be regarded as planar and the same laws of refraction can be applied at every point on the surface. Just as for reflection by a spherical mirror, the normal at the point of incidence is perpendicular to the tangent plane to the spherical surface at that point and, therefore, passes through its centre of curvature. We first consider refraction by a single spherical surface and follow it by thin lenses. A thin lens is a transparent optical medium bounded by two surfaces; at least one of which should be spherical. Applying the formula for image formation by a single spherical surface successively at the two surfaces of a lens, we shall obtain the lens maker's formula and then the lens formula.

9.5.1 Refraction at a spherical surface

Figure 9.15 shows the geometry of formation of image *I* of an object *O* on the principal axis of a spherical surface with centre of curvature C, and radius of curvature *R*. The rays are incident from a medium of refractive index n_1, to another of refractive index n_2. As before, we take the aperture (or the lateral size) of the surface to be small compared to other distances involved, so that small angle approximation can be made. In particular, NM will be taken to be nearly equal to the length of the perpendicular from the point N on the principal axis. We have, for small angles,

$$\tan \angle NOM = \frac{MN}{OM}$$

$$\tan \angle NCM = \frac{MN}{MC}$$

$$\tan \angle NIM = \frac{MN}{MI}$$

Now, for ΔNOC, i is the exterior angle. Therefore, $i = \angle NOM + \angle NCM$

$$i = \frac{MN}{OM} + \frac{MN}{MC} \qquad (9.13)$$

Similarly,

$$r = \angle NCM - \angle NIM$$

i.e., $$r = \frac{MN}{MC} - \frac{MN}{MI} \qquad (9.14)$$

Now, by Snell's law

$$n_1 \sin i = n_2 \sin r$$

or for small angles

$$n_1 i = n_2 r$$

Substituting i and r from Eqs. (9.13) and (9.14), we get

$$\frac{n_1}{OM} + \frac{n_2}{MI} = \frac{n_2 - n_1}{MC} \qquad (9.15)$$

Here, OM, MI and MC represent magnitudes of distances. Applying the Cartesian sign convention,

$$OM = -u,\ MI = +v,\ MC = +R$$

Substituting these in Eq. (9.15), we get

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R} \qquad (9.16)$$

Equation (9.16) gives us a relation between object and image distance in terms of refractive index of the medium and the radius of curvature of the curved spherical surface. It holds for any curved spherical surface.

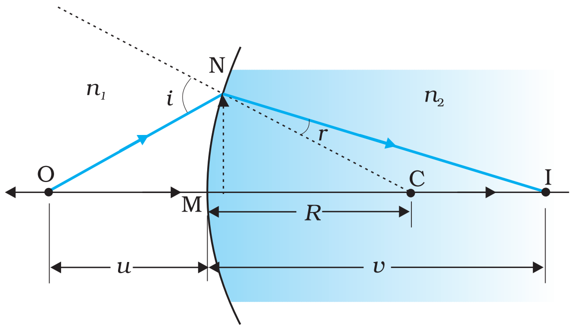


FIGURE 9.15 Refraction at a spherical surface separating two media.

Example 9.5 Light from a point source in air falls on a spherical glass surface (n = 1.5 and radius of curvature = 20 cm). The distance of the light source from the glass surface is 100 cm. At what position the image is formed?

Solution

We use the relation given by Eq. (9.16). Here

u = – 100 cm, v = ?, R = + 20 cm, n_1 = 1, and n_2 = 1.5.

We then have

$$\frac{1.5}{v} + \frac{1}{100} = \frac{0.5}{20}$$

or v = +100 cm

The image is formed at a distance of 100 cm from the glass surface, in the direction of incident light.

EXAMPLE 9.5

9.5.2 Refraction by a lens

Figure 9.16(a) shows the geometry of image formation by a double convex lens. The image formation can be seen in terms of two steps: (i) The first refracting surface forms the image I_1 of the object O [Fig. 9.16(b)]. The image I_1 acts as a virtual object for the second surface that forms the image at I [Fig. 9.16(c)]. Applying Eq. (9.15) to the first interface ABC, we get

$$\frac{n_1}{OB}+\frac{n_2}{BI_1}=\frac{n_2-n_1}{BC_1} \tag{9.17}$$

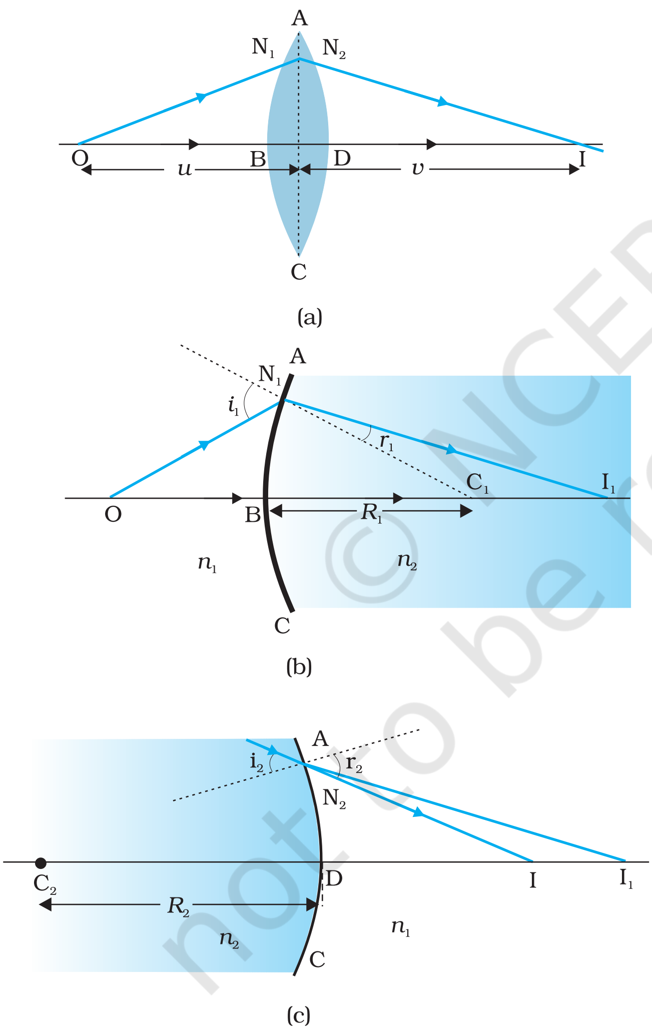


FIGURE 9.16 (a) The position of object, and the image formed by a double convex lens, (b) Refraction at the first spherical surface and (c) Refraction at the second spherical surface.

A similar procedure applied to the second interface* ADC gives,

$$-\frac{n_2}{DI_1}+\frac{n_1}{DI}=\frac{n_2-n_1}{DC_2} \tag{9.18}$$

For a thin lens, $BI_1 = DI_1$. Adding Eqs. (9.17) and (9.18), we get

$$\frac{n_1}{OB}+\frac{n_1}{DI}=(n_2-n_1)\left(\frac{1}{BC_1}+\frac{1}{DC_2}\right) \tag{9.19}$$

Suppose the object is at infinity, i.e., $OB \to \infty$ and $DI = f$, Eq. (9.19) gives

$$\frac{n_1}{f}=(n_2-n_1)\left(\frac{1}{BC_1}+\frac{1}{DC_2}\right) \tag{9.20}$$

The point where image of an object placed at infinity is formed is called the *focus* F, of the lens and the distance f gives its *focal length*. A lens has two foci, F and F′, on either side of it (Fig. 9.17). By the sign convention,

$BC_1 = + R_1$,

$DC_2 = -R_2$

So Eq. (9.20) can be written as

$$\frac{1}{f}=(n_{21}-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right) \quad \left[\because\ n_{21}=\frac{n_2}{n_1}\right] \tag{9.21}$$

Equation (9.21) is known as the *lens maker's formula*. It is useful to design lenses of desired focal length using surfaces of suitable radii of curvature. Note that the formula is true for a concave lens also. In that case R_1 is negative, R_2 positive and therefore, f is negative.

* Note that now the refractive index of the medium on the right side of ADC is n_1 while on its left it is n_2. Further DI_1 is negative as the distance is measured against the direction of incident light.

Reprint 2025-26

From Eqs. (9.19) and (9.20), we get

$$\frac{n_1}{\text{OB}}+\frac{n_1}{\text{DI}} = \frac{n_1}{f} \tag{9.22}$$

Again, in the thin lens approximation, B and D are both close to the optical centre of the lens. Applying the sign convention,

BO = – u, DI = +v, we get

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f} \tag{9.23}$$

Equation (9.23) is the familiar *thin lens formula*. Though we derived it for a real image formed by a convex lens, the formula is valid for both convex as well as concave lenses and for both real and virtual images.

It is worth mentioning that the two foci, F and F′, of a double convex or concave lens are equidistant from the optical centre. The focus on the side of the (original) source of light is called the *first focal point*, whereas the other is called the *second focal point*.

To find the image of an object by a lens, we can, in principle, take any two rays emanating from a point on an object; trace their paths using the laws of refraction and find the point where the refracted rays meet (or appear to meet). In practice, however, it is convenient to choose any two of the following rays:

(i) A ray emanating from the object parallel to the principal axis of the lens after refraction passes through the second principal focus F′ (in a convex lens) or appears to diverge (in a concave lens) from the first principal focus F.

(ii) A ray of light, passing through the optical centre of the lens, emerges without any deviation after refraction.

(iii) (a) A ray of light passing through the first principal focus of a convex lens [Fig. 9.17(a)] emerges parallel to the principal axis after refraction.

(b) A ray of light incident on a concave lens appearing to meet the principal axis at second focus point emerges parallel to the principal axis after refraction [Fig. 9.17(b)].

Figures 9.17(a) and (b) illustrate these rules for a convex and a concave lens, respectively. You should practice drawing similar ray diagrams for different positions of the object with respect to the lens and also verify that the lens formula, Eq. (9.23), holds good for all cases.

Here again it must be remembered that each point on an object gives out infinite number of rays. All these rays will pass through the same image point after refraction at the lens.

Magnification (m) produced by a lens is defined, like that for a mirror, as the ratio of the size of the image to that of the object. Proceeding

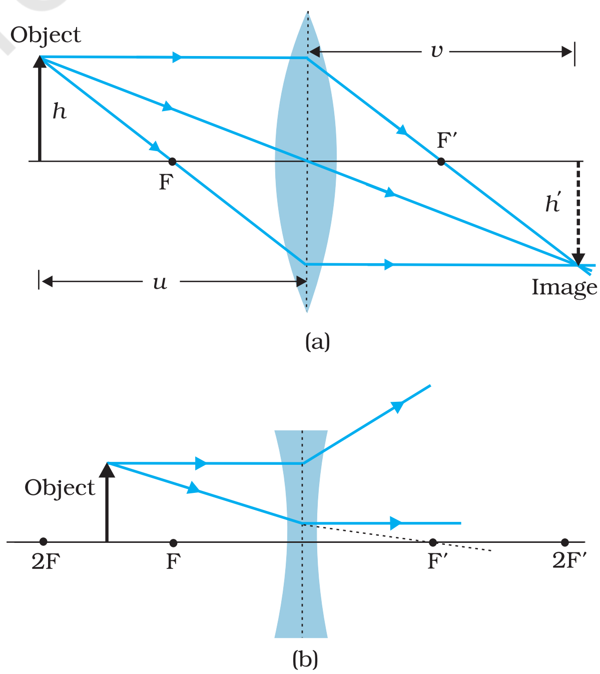


FIGURE 9.17 Tracing rays through (a) convex lens (b) concave lens.

in the same way as for spherical mirrors, it is easily seen that for a lens

$$m = \frac{h'}{h} = \frac{v}{u} \tag{9.24}$$

When we apply the sign convention, we see that, for erect (and virtual) image formed by a convex or concave lens, m is positive, while for an inverted (and real) image, m is negative.

EXAMPLE 9.6

Example 9.6 A magician during a show makes a glass lens with $n = 1.47$ disappear in a trough of liquid. What is the refractive index of the liquid? Could the liquid be water?

Solution
The refractive index of the liquid must be equal to 1.47 in order to make the lens disappear. This means $n_1 = n_2$. This gives $1/f = 0$ or $f \to \infty$. The lens in the liquid will act like a plane sheet of glass. No, the liquid is not water. It could be glycerine.

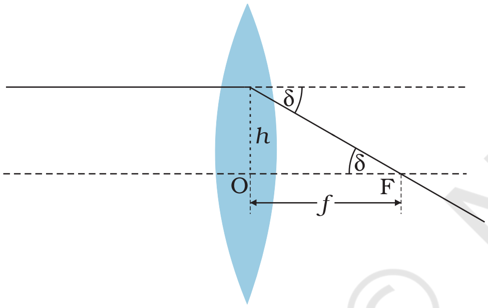


FIGURE 9.18 Power of a lens.

9.5.3 Power of a lens

Power of a lens is a measure of the convergence or divergence, which a lens introduces in the light falling on it. Clearly, a lens of shorter focal length bends the incident light more, while converging it in case of a convex lens and diverging it in case of a concave lens. The *power P* of a lens is defined as the tangent of the angle by which it converges or diverges a beam of light parallel to the principal axis falling at unit distance from the optical centre (Fig. 9.18).

$\tan\delta = \dfrac{h}{f}$; if $h = 1$, $\tan\delta = \dfrac{1}{f}$ or $\delta = \dfrac{1}{f}$ for small value of δ. Thus,

$$P = \frac{1}{f} \tag{9.25}$$

The SI unit for power of a lens is dioptre (D): 1D = 1m^{-1}. The power of a lens of focal length of 1 metre is one dioptre. Power of a lens is positive for a converging lens and negative for a diverging lens. Thus, when an optician prescribes a corrective lens of power + 2.5 D, the required lens is a convex lens of focal length + 40 cm. A lens of power of – 4.0 D means a concave lens of focal length – 25 cm.

EXAMPLE 9.7

Example 9.7 (i) If $f = 0.5$ m for a glass lens, what is the power of the lens? (ii) The radii of curvature of the faces of a double convex lens are 10 cm and 15 cm. Its focal length is 12 cm. What is the refractive index of glass? (iii) A convex lens has 20 cm focal length in air. What is focal length in water? (Refractive index of air-water = 1.33, refractive index for air-glass = 1.5.)

Solution

(i) Power = +2 dioptre.

(ii) Here, we have f = +12 cm, R_1 = +10 cm, R_2 = –15 cm.
Refractive index of air is taken as unity.
We use the lens formula of Eq. (9.22). The sign convention has to be applied for f, R_1 and R_2.
Substituting the values, we have

$$\frac{1}{12}=(n-1)\left(\frac{1}{10}-\frac{1}{-15}\right)$$

This gives n = 1.5.

(iii) For a glass lens in air, n_2 = 1.5, n_1 = 1, f = +20 cm. Hence, the lens formula gives

$$\frac{1}{20}=0.5\left[\frac{1}{R_1}-\frac{1}{R_2}\right]$$

For the same glass lens in water, n_2 = 1.5, n_1 = 1.33. Therefore,

$$\frac{1.33}{f}=(1.5-1.33)\left[\frac{1}{R_1}-\frac{1}{R_2}\right] \qquad (9.26)$$

Combining these two equations, we find f = + 78.2 cm.

EXAMPLE 9.7

9.5.4 Combination of thin lenses in contact

Consider two lenses A and B of focal length f_1 and f_2 placed in contact with each other. Let the object be placed at a point O beyond the focus of the first lens A (Fig. 9.19). The first lens produces an image at I_1. Since image I_1 is real, it serves as a virtual object for the second lens B, producing the final image at I. It must, however, be borne in mind that formation of image by the first lens is presumed only to facilitate determination of the position of the final image. In fact, the direction of rays emerging from the first lens gets modified in accordance with the angle at which they strike the second lens. Since the lenses are thin, we assume the optical centres of the lenses to be coincident. Let this central point be denoted by P.

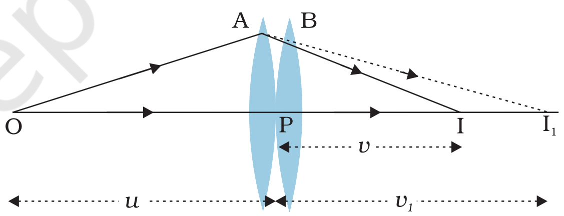


FIGURE 9.19 Image formation by a combination of two thin lenses in contact.

For the image formed by the first lens A, we get

$$\frac{1}{v_1}-\frac{1}{u}=\frac{1}{f_1} \qquad (9.27)$$

For the image formed by the second lens B, we get

$$\frac{1}{v}-\frac{1}{v_1}=\frac{1}{f_2} \qquad (9.28)$$

Adding Eqs. (9.27) and (9.28), we get

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f_1}+\frac{1}{f_2} \qquad (9.29)$$

If the two lens-system is regarded as equivalent to a single lens of focal length f, we have

Reprint 2025-26

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f}$$

so that we get

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \tag{9.30}$$

The derivation is valid for any number of thin lenses in contact. If several thin lenses of focal length $f_1, f_2, f_3,...$ are in contact, the effective focal length of their combination is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + ... \tag{9.31}$$

In terms of power, Eq. (9.31) can be written as

$$P = P_1 + P_2 + P_3 + ... \tag{9.32}$$

where P is the net power of the lens combination. Note that the sum in Eq. (9.32) is an algebraic sum of individual powers, so some of the terms on the right side may be positive (for convex lenses) and some negative (for concave lenses). Combination of lenses helps to obtain diverging or converging lenses of desired magnification. It also enhances sharpness of the image. Since the image formed by the first lens becomes the object for the second, Eq. (9.25) implies that the total magnification m of the combination is a product of magnification (m_1, m_2, m_3,...) of individual lenses

$$m = m_1\, m_2\, m_3\, ... \tag{9.33}$$

Such a system of combination of lenses is commonly used in designing lenses for cameras, microscopes, telescopes and other optical instruments.

Example 9.8 Find the position of the image formed by the lens combination given in the Fig. 9.20.

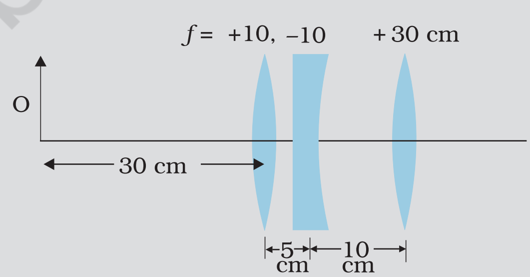


FIGURE 9.20

Solution Image formed by the first lens

$$\frac{1}{v_1} - \frac{1}{u_1} = \frac{1}{f_1}$$

$$\frac{1}{v_1} - \frac{1}{-30} = \frac{1}{10}$$

or $v_1 = 15$ cm

EXAMPLE 9.8

The image formed by the first lens serves as the object for the second. This is at a distance of (15 – 5) cm = 10 cm to the right of the second lens. Though the image is real, it serves as a virtual object for the second lens, which means that the rays appear to come from it for the second lens.

$$\frac{1}{v_2}-\frac{1}{10}=\frac{1}{-10}$$

or $v_2 = \infty$

The virtual image is formed at an infinite distance to the left of the second lens. This acts as an object for the third lens.

$$\frac{1}{v_3}-\frac{1}{u_3}=\frac{1}{f_3}$$

or $$\frac{1}{v_3}=\frac{1}{\infty}+\frac{1}{30}$$

or $v_3 = 30$ cm

The final image is formed 30 cm to the right of the third lens.

EXAMPLE 9.8

9.6 REFRACTION THROUGH A PRISM

Figure 9.21 shows the passage of light through a triangular prism ABC. The angles of incidence and refraction at the first face AB are i and r_1, while the angle of incidence (from glass to air) at the second face AC is r_2 and the angle of refraction or emergence e. The angle between the emergent ray RS and the direction of the incident ray PQ is called the *angle of deviation*, δ.

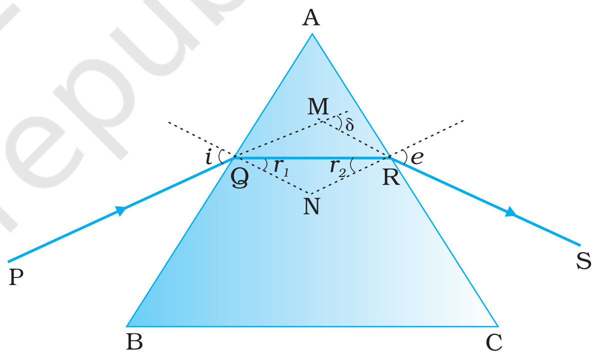


FIGURE 9.21 A ray of light passing through a triangular glass prism.

In the quadrilateral AQNR, two of the angles (at the vertices Q and R) are right angles. Therefore, the sum of the other angles of the quadrilateral is 180°.

$\angle A + \angle QNR = 180°$

From the triangle QNR,

$r_1 + r_2 + \angle QNR = 180°$

Comparing these two equations, we get

$$r_1 + r_2 = A \qquad (9.34)$$

The total deviation δ is the sum of deviations at the two faces,

$\delta = (i - r_1) + (e - r_2)$

that is,

$$\delta = i + e - A \qquad (9.35)$$

Thus, the angle of deviation depends on the angle of incidence. A plot between the angle of deviation and angle of incidence is shown in Fig. 9.22. You can see that, in general, any given value of δ, except for $i = e$, corresponds to two values i and hence of e. This, in fact, is expected from the symmetry of i and e in Eq. (9.35), i.e., δ remains the same if i

Reprint 2025-26

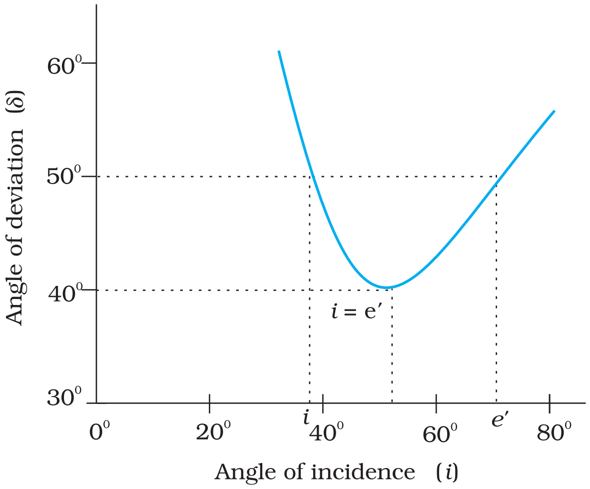


FIGURE 9.22 Plot of angle of deviation (δ) versus angle of incidence (i) for a triangular prism.

and e are interchanged. Physically, this is related to the fact that the path of ray in Fig. 9.21 can be traced back, resulting in the same angle of deviation. At the minimum deviation D_m, the refracted ray inside the prism becomes parallel to its base. We have

$\delta = D_m$, $i = e$ which implies $r_1 = r_2$.

Equation (9.34) gives

$$2r = A \text{ or } r = \frac{A}{2} \tag{9.36}$$

In the same way, Eq. (9.35) gives

$$D_m = 2i - A, \text{ or } i = (A + D_m)/2 \tag{9.37}$$

The refractive index of the prism is

$$n_{21} = \frac{n_2}{n_1} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \tag{9.38}$$

The angles A and D_m can be measured experimentally. Equation (9.38) thus provides a method of determining refractive index of the material of the prism.

For a small angle prism, i.e., a thin prism, D_m is also very small, and we get

$$n_{21} = \frac{\sin[(A + D_m)/2]}{\sin[A/2]} \simeq \frac{(A + D_m)/2}{A/2}$$

$$D_m = (n_{21} - 1)A$$

It implies that, thin prisms do not deviate light much.

9.7 Optical Instruments

A number of optical devices and instruments have been designed utilising reflecting and refracting properties of mirrors, lenses and prisms. Periscope, kaleidoscope, binoculars, telescopes, microscopes are some examples of optical devices and instruments that are in common use. Our eye is, of course, one of the most important optical device the nature has endowed us with. We have already studied about the human eye in Class X. We now go on to describe the principles of working of the microscope and the telescope.

9.7.1 The microscope

A simple magnifier or microscope is a converging lens of small focal length (Fig. 9.23). In order to use such a lens as a microscope, the lens is held near the object, one focal length away or less, and the eye is positioned close to the lens on the other side. The idea is to get an erect, magnified and virtual image of the object at a distance so that it can be viewed comfortably, i.e., at 25 cm or more. If the object is at a distance f, the image is at infinity. However, if the object is at a distance slightly less

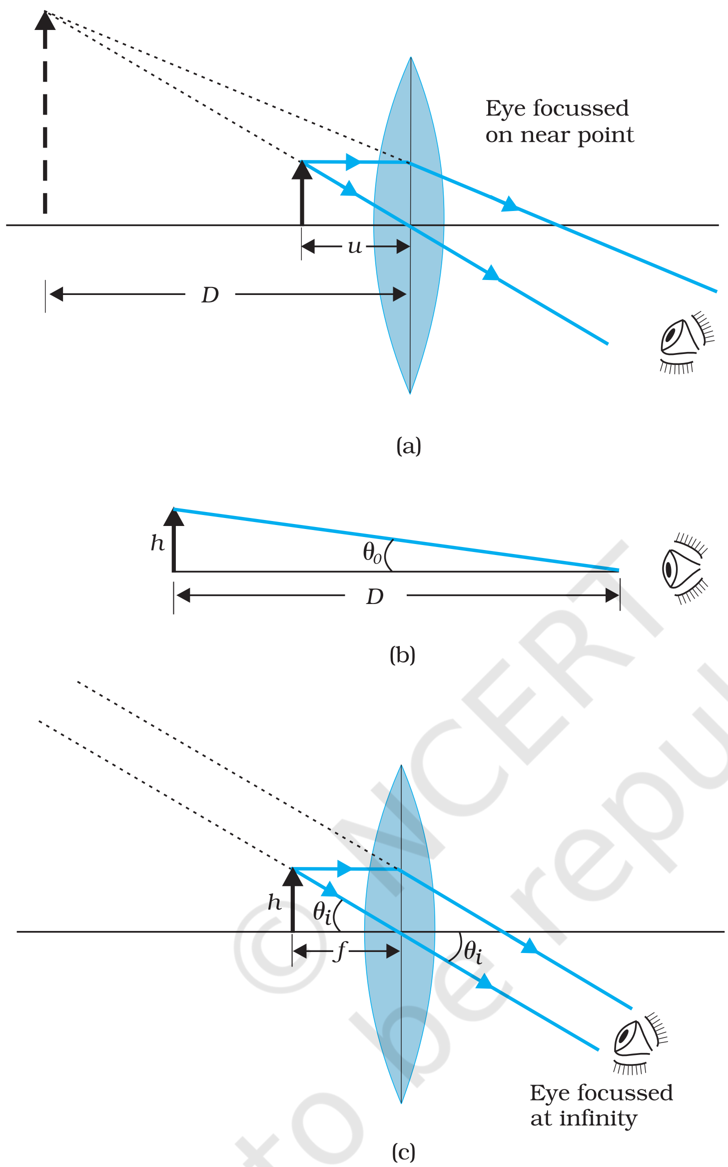


FIGURE 9.23 A simple microscope; (a) the magnifying lens is located such that the image is at the near point, (b) the angle subtended by the object, is the same as that at the near point, and (c) the object near the focal point of the lens; the image is far off but closer than infinity.

than the focal length of the lens, the image is virtual and closer than infinity. Although the closest comfortable distance for viewing the image is when it is at the near point (distance $D \cong 25$ cm), it causes some strain on the eye. Therefore, the image formed at infinity is often considered most suitable for viewing by the relaxed eye. We show both cases, the first in Fig. 9.23(a), and the second in Fig. 9.23(b) and (c).

The linear magnification m, for the image formed at the near point D, by a simple microscope can be obtained by using the relation

Reprint 2025-26

$$m = \frac{v}{u} = v\left(\frac{1}{v} - \frac{1}{f}\right) = \left(1 - \frac{v}{f}\right)$$

Now according to our sign convention, v is negative, and is equal in magnitude to D. Thus, the magnification is

$$m = \left(1 + \frac{D}{f}\right) \tag{9.39}$$

Since D is about 25 cm, to have a magnification of six, one needs a convex lens of focal length, f = 5 cm.

Note that $m = h'/h$ where h is the size of the object and h' the size of the image. This is also the ratio of the angle subtended by the image to that subtended by the object, if placed at D for comfortable viewing. (Note that this is not the angle actually subtended by the object at the eye, which is h/u.) What a single-lens simple magnifier achieves is that it allows the object to be brought closer to the eye than D.

We will now find the magnification when the image is at infinity. In this case we will have to obtained the *angular* magnification. Suppose the object has a height h. The maximum angle it can subtend, and be clearly visible (without a lens), is when it is at the near point, i.e., a distance D. The angle subtended is then given by

$$\tan \theta_o = \left(\frac{h}{D}\right) \approx \theta_o \tag{9.40}$$

We now find the angle subtended at the eye by the image when the object is at u. From the relations

$$\frac{h'}{h} = m = \frac{v}{u}$$

we have the angle subtended by the image

$\tan \theta_i = \frac{h'}{-v} = \frac{h}{-v} \cdot \frac{v}{u} = \frac{h}{-u} \approx \theta$. The angle subtended by the object, when it is at $u = -f$.

$$\theta_i = \left(\frac{h}{f}\right) \tag{9.41}$$

as is clear from Fig. 9.23(c). The angular magnification is, therefore

$$m = \left(\frac{\theta_i}{\theta_o}\right) = \frac{D}{f} \tag{9.42}$$

This is one less than the magnification when the image is at the near point, Eq. (9.39), but the viewing is more comfortable and the difference in magnification is usually small. In subsequent discussions of optical instruments (microscope and telescope) we shall assume the image to be at infinity.

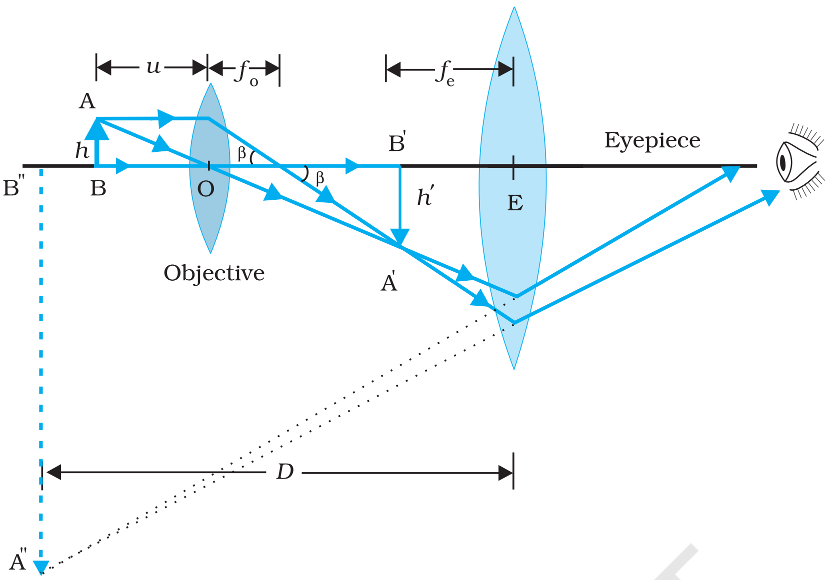


FIGURE 9.24 Ray diagram for the formation of image by a compound microscope.

A simple microscope has a limited maximum magnification (≤ 9) for realistic focal lengths. For much larger magnifications, one uses two lenses, one compounding the effect of the other. This is known as a *compound microscope*. A schematic diagram of a compound microscope is shown in Fig. 9.24. The lens nearest the object, called the *objective*, forms a real, inverted, magnified image of the object. This serves as the object for the second lens, the *eyepiece*, which functions essentially like a simple microscope or magnifier, produces the final image, which is enlarged and virtual. The first inverted image is thus near (at or within) the focal plane of the eyepiece, at a distance appropriate for final image formation at infinity, or a little closer for image formation at the near point. Clearly, the final image is inverted with respect to the original object.

We now obtain the magnification due to a compound microscope. The ray diagram of Fig. 9.24 shows that the (linear) magnification due to the objective, namely h'/h, equals

$$m_O = \frac{h'}{h} = \frac{L}{f_o} \tag{9.43}$$

where we have used the result

$$\tan\beta = \left(\frac{h}{f_o}\right) = \left(\frac{h'}{L}\right)$$

Here h' is the size of the first image, the object size being h and f_o being the focal length of the objective. The first image is formed near the focal point of the eyepiece. The distance L, i.e., the distance between the second focal point of the objective and the first focal point of the eyepiece (focal length f_e) is called the tube length of the compound microscope.

PHYSICS

The world's largest optical telescopes
http://astro.nineplanets.org/bigeyes.html

Reprint 2025-26

As the first inverted image is near the focal point of the eyepiece, we use the result from the discussion above for the simple microscope to obtain the (angular) magnification m_e due to it [Eq. (9.39)], when the final image is formed at the near point, is

$$m_e = \left(1 + \frac{D}{f_e}\right) \qquad [9.44(a)]$$

When the final image is formed at infinity, the angular magnification due to the eyepiece [Eq. (9.42)] is

$$m_e = (D/f_e) \qquad [9.44(b)]$$

Thus, the total magnification [(according to Eq. (9.33)], when the image is formed at infinity, is

$$m = m_o m_e = \left(\frac{L}{f_o}\right)\left(\frac{D}{f_e}\right) \qquad (9.45)$$

Clearly, to achieve a large magnification of a *small* object (hence the name microscope), the objective and eyepiece should have small focal lengths. In practice, it is difficult to make the focal length much smaller than 1 cm. Also large lenses are required to make L large.

For example, with an objective with f_o = 1.0 cm, and an eyepiece with focal length f_e = 2.0 cm, and a tube length of 20 cm, the magnification is

$$m = m_o m_e = \left(\frac{L}{f_o}\right)\left(\frac{D}{f_e}\right)$$

$$= \frac{20}{1} \times \frac{25}{2} = 250$$

Various other factors such as illumination of the object, contribute to the quality and visibility of the image. In modern microscopes, multi-component lenses are used for both the objective and the eyepiece to improve image quality by minimising various optical aberrations (defects) in lenses.

9.7.2 Telescope

The telescope is used to provide angular magnification of distant objects (Fig. 9.25). It also has an objective and an eyepiece. But here, the objective has a large focal length and a much larger aperture than the eyepiece. Light from a distant object enters the objective and a real image is formed in the tube at its second focal point. The eyepiece magnifies this image producing a final inverted image. The magnifying power m is the ratio of the angle β subtended at the eye by the final image to the angle α which the object subtends at the lens or the eye. Hence

$$m \approx \frac{\beta}{\alpha} \approx \frac{h}{f_e} \cdot \frac{f_o}{h} = \frac{f_o}{f_e} \qquad (9.46)$$

Reprint 2025-26

In this case, the length of the telescope tube is $f_o + f_e$.

Terrestrial telescopes have, in addition, a pair of inverting lenses to make the final image erect. Refracting telescopes can be used both for terrestrial and astronomical observations. For example, consider a telescope whose objective has a focal length of 100 cm and the eyepiece a focal length of 1 cm. The magnifying power of this telescope is m = 100/1 = 100.

Let us consider a pair of stars of actual separation 1′ (one minute of arc). The stars appear as though they are separated by an angle of 100 × 1′ = 100′ =1.67°.

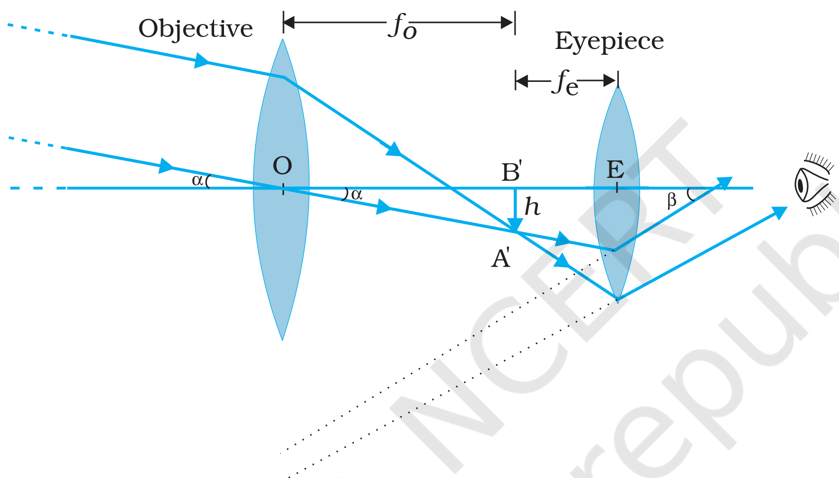


FIGURE 9.25 A refracting telescope.

The main considerations with an astronomical telescope are its light gathering power and its resolution or resolving power. The former clearly depends on the area of the objective. With larger diameters, fainter objects can be observed. The resolving power, or the ability to observe two objects distinctly, which are in very nearly the same direction, also depends on the diameter of the objective. So, the desirable aim in optical telescopes is to make them with objective of large diameter. The largest lens objective in use has a diameter of 40 inch (~1.02 m). It is at the Yerkes Observatory in Wisconsin, USA. Such big lenses tend to be very heavy and therefore, difficult to make and support by their edges. Further, it is rather difficult and expensive to make such large sized lenses which form images that are free from any kind of chromatic aberration and distortions.

For these reasons, modern telescopes use a concave mirror rather than a lens for the objective. Telescopes with mirror objectives are called *reflecting* telescopes. There is no chromatic aberration in a mirror. Mechanical support is much less of a problem since a mirror weighs much less than a lens of equivalent optical quality, and can be supported over its entire back surface, not just over its rim. One obvious problem with a reflecting telescope is that the objective mirror focusses light inside

Reprint 2025-26

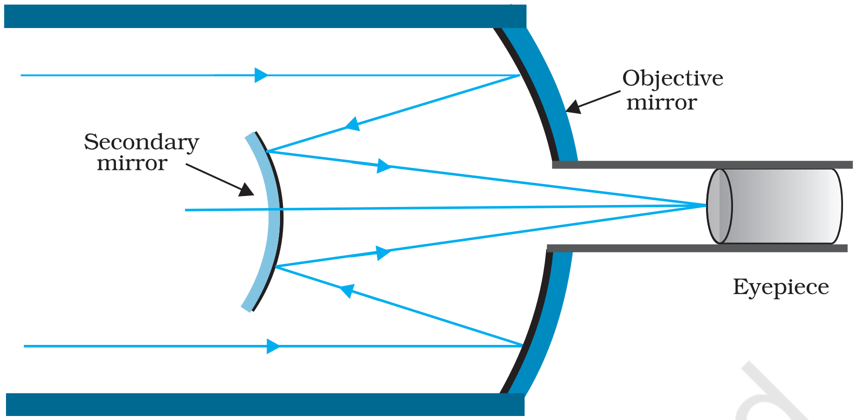


FIGURE 9.26 Schematic diagram of a reflecting telescope (Cassegrain).

the telescope tube. One must have an eyepiece and the observer right there, obstructing some light (depending on the size of the observer cage). This is what is done in the very large 200 inch (~5.08 m) diameters, Mt. Palomar telescope, California. The viewer sits near the focal point of the mirror, in a small cage. Another solution to the problem is to deflect the light being focussed by another mirror. One such arrangement using a convex secondary mirror to focus the incident light, which now passes through a hole in the objective primary mirror, is shown in Fig. 9.26. This is known as a *Cassegrain* telescope, after its inventor. It has the advantages of a large focal length in a short telescope. The largest telescope in India is in Kavalur, Tamil Nadu. It is a 2.34 m diameter reflecting telescope (Cassegrain). It was ground, polished, set up, and is being used by the Indian Institute of Astrophysics, Bangalore. The largest reflecting telescopes in the world are the pair of Keck telescopes in Hawaii, USA, with a reflector of 10 metre in diameter.

SUMMARY

1. Reflection is governed by the equation $\angle i = \angle r'$ and refraction by the Snell's law, $\sin i/\sin r = n$, where the incident ray, reflected ray, refracted ray and normal lie in the same plane. Angles of incidence, reflection and refraction are i, r' and r, respectively.
2. The *critical angle of incidence* i_c for a ray incident from a denser to rarer medium, is that angle for which the angle of refraction is 90°. For $i > i_c$, total internal reflection occurs. Multiple internal reflections in diamond ($i_c \cong 24.4°$), totally reflecting prisms and mirage, are some examples of total internal reflection. Optical fibres consist of glass fibres coated with a thin layer of material of *lower* refractive index. Light incident at an angle at one end comes out at the other, after multiple internal reflections, even if the fibre is bent.

3. *Cartesian sign convention*: Distances measured in the same direction as the incident light are positive; those measured in the opposite direction are negative. All distances are measured from the pole/optic centre of the mirror/lens on the principal axis. The heights measured upwards above x-axis and normal to the principal axis of the mirror/lens are taken as positive. The heights measured downwards are taken as negative.
4. *Mirror equation*:

$$\frac{1}{v}+\frac{1}{u}=\frac{1}{f}$$

where u and v are object and image distances, respectively and f is the focal length of the mirror. f is (approximately) half the radius of curvature R. f is negative for concave mirror; f is positive for a convex mirror.

5. For a prism of the angle A, of refractive index n_2 placed in a medium of refractive index n_1,

$$n_{21}=\frac{n_2}{n_1}=\frac{\sin\left[\left(A+D_m\right)/2\right]}{\sin\left(A/2\right)}$$

where D_m is the angle of minimum deviation.

6. *For refraction through a spherical interface* (from medium 1 to 2 of refractive index n_1 and n_2, respectively)

$$\frac{n_2}{v}-\frac{n_1}{u}=\frac{n_2-n_1}{R}$$

Thin lens formula

$$\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$$

Lens maker's formula

$$\frac{1}{f}=\frac{\left(n_2-n_1\right)}{n_1}\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$$

R_1 and R_2 are the radii of curvature of the lens surfaces. f is positive for a converging lens; f is negative for a diverging lens. The power of a lens $P = 1/f$.

The SI unit for power of a lens is dioptre (D): 1 D = 1 m^{-1}.

If several thin lenses of focal length f_1, f_2, f_3,.. are in contact, the effective focal length of their combination, is given by

$$\frac{1}{f}=\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}+\ldots$$

The total power of a combination of several lenses is

$P = P_1 + P_2 + P_3 + \ldots$

7. *Dispersion* is the splitting of light into its constituent colour.

8. *Magnifying power m of a simple microscope* is given by $m = 1 + (D/f)$, where D = 25 cm is the least distance of distinct vision and f is the focal length of the convex lens. If the image is at infinity, $m = D/f$. For a compound microscope, the magnifying power is given by $m = m_e \times m_0$ where $m_e = 1 + (D/f_e)$, is the magnification due to the eyepiece and m_o is the magnification produced by the objective. *Approximately,*

$$m = \frac{L}{f_o} \times \frac{D}{f_e}$$

where f_o and f_e are the focal lengths of the objective and eyepiece, respectively, and L is the distance between their focal points.

9. *Magnifying power m of a telescope* is the ratio of the angle β subtended at the eye by the image to the angle α subtended at the eye by the object.

$$m = \frac{\beta}{\alpha} = \frac{f_o}{f_e}$$

where f_0 and f_e are the focal lengths of the objective and eyepiece, respectively.

POINTS TO PONDER

1. The laws of reflection and refraction are true for all surfaces and pairs of media at the point of the incidence.
2. The real image of an object placed between f and $2f$ from a convex lens can be seen on a screen placed at the image location. If the screen is removed, is the image still there? This question puzzles many, because it is difficult to reconcile ourselves with an image suspended in air without a screen. But the image does exist. Rays from a given point on the object are converging to an image point in space and diverging away. The screen simply diffuses these rays, some of which reach our eye and we see the image. This can be seen by the images formed in air during a laser show.
3. Image formation needs regular reflection/refraction. In principle, all rays from a given point should reach the same image point. This is why you do not see your image by an irregular reflecting object, say the page of a book.
4. Thick lenses give coloured images due to dispersion. The variety in colour of objects we see around us is due to the constituent colours of the light incident on them. A monochromatic light may produce an entirely different perception about the colours on an object as seen in white light.
5. For a simple microscope, the angular size of the object equals the angular size of the image. Yet it offers magnification because we can keep the small object much closer to the eye than 25 cm and hence have it subtend a large angle. The image is at 25 cm which we can see. Without the microscope, you would need to keep the small object at 25 cm which would subtend a very small angle.

Reprint 2025-26

EXERCISES

9.1 A small candle, 2.5 cm in size is placed at 27 cm in front of a concave mirror of radius of curvature 36 cm. At what distance from the mirror should a screen be placed in order to obtain a sharp image? Describe the nature and size of the image. If the candle is moved closer to the mirror, how would the screen have to be moved?

9.2 A 4.5 cm needle is placed 12 cm away from a convex mirror of focal length 15 cm. Give the location of the image and the magnification. Describe what happens as the needle is moved farther from the mirror.

9.3 A tank is filled with water to a height of 12.5 cm. The apparent depth of a needle lying at the bottom of the tank is measured by a microscope to be 9.4 cm. What is the refractive index of water? If water is replaced by a liquid of refractive index 1.63 up to the same height, by what distance would the microscope have to be moved to focus on the needle again?

9.4 Figures 9.27(a) and (b) show refraction of a ray in air incident at 60° with the normal to a glass-air and water-air interface, respectively. Predict the angle of refraction in glass when the angle of incidence in water is 45° with the normal to a water-glass interface [Fig. 9.27(c)].

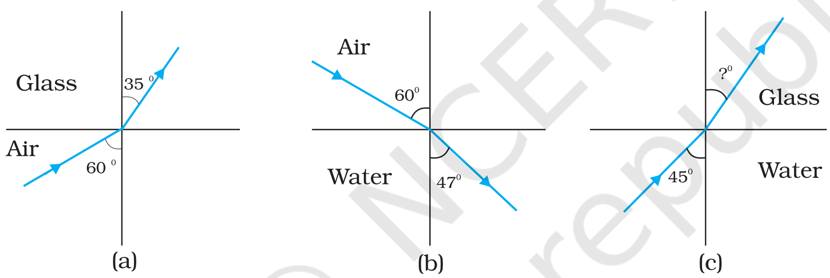


FIGURE 9.27

9.5 A small bulb is placed at the bottom of a tank containing water to a depth of 80cm. What is the area of the surface of water through which light from the bulb can emerge out? Refractive index of water is 1.33. (Consider the bulb to be a point source.)

9.6 A prism is made of glass of unknown refractive index. A parallel beam of light is incident on a face of the prism. The angle of minimum deviation is measured to be 40°. What is the refractive index of the material of the prism? The refracting angle of the prism is 60°. If the prism is placed in water (refractive index 1.33), predict the new angle of minimum deviation of a parallel beam of light.

9.7 Double-convex lenses are to be manufactured from a glass of refractive index 1.55, with both faces of the same radius of curvature. What is the radius of curvature required if the focal length is to be 20cm?

9.8 A beam of light converges at a point P. Now a lens is placed in the path of the convergent beam 12cm from P. At what point does the beam converge if the lens is (a) a convex lens of focal length 20cm, and (b) a concave lens of focal length 16cm?

9.9 An object of size 3.0cm is placed 14cm in front of a concave lens of focal length 21cm. Describe the image produced by the lens. What happens if the object is moved further away from the lens?

Reprint 2025-26

9.10 What is the focal length of a convex lens of focal length 30 cm in contact with a concave lens of focal length 20 cm? Is the system a converging or a diverging lens? Ignore thickness of the lenses.

9.11 A compound microscope consists of an objective lens of focal length 2.0 cm and an eyepiece of focal length 6.25 cm separated by a distance of 15 cm. How far from the objective should an object be placed in order to obtain the final image at (a) the least distance of distinct vision (25 cm), and (b) at infinity? What is the magnifying power of the microscope in each case?

9.12 A person with a normal near point (25 cm) using a compound microscope with objective of focal length 8.0 mm and an eyepiece of focal length 2.5 cm can bring an object placed at 9.0 mm from the objective in sharp focus. What is the separation between the two lenses? Calculate the magnifying power of the microscope,

9.13 A small telescope has an objective lens of focal length 144 cm and an eyepiece of focal length 6.0 cm. What is the magnifying power of the telescope? What is the separation between the objective and the eyepiece?

9.14 (a) A giant refracting telescope at an observatory has an objective lens of focal length 15 m. If an eyepiece of focal length 1.0 cm is used, what is the angular magnification of the telescope?

(b) If this telescope is used to view the moon, what is the diameter of the image of the moon formed by the objective lens? The diameter of the moon is 3.48×10^6 m, and the radius of lunar orbit is 3.8×10^8 m.

9.15 Use the mirror equation to deduce that:

(a) an object placed between f and $2f$ of a concave mirror produces a real image beyond $2f$.

(b) a convex mirror always produces a virtual image independent of the location of the object.

(c) the virtual image produced by a convex mirror is always diminished in size and is located between the focus and the pole.

(d) an object placed between the pole and focus of a concave mirror produces a virtual and enlarged image.

[*Note:* This exercise helps you deduce algebraically properties of images that one obtains from explicit ray diagrams.]

9.16 A small pin fixed on a table top is viewed from above from a distance of 50 cm. By what distance would the pin appear to be raised if it is viewed from the same point through a 15 cm thick glass slab held parallel to the table? Refractive index of glass = 1.5. Does the answer depend on the location of the slab?

9.17 (a) Figure 9.28 shows a cross-section of a 'light pipe' made of a glass fibre of refractive index 1.68. The outer covering of the pipe is made of a material of refractive index 1.44. What is the range of the angles of the incident rays with the axis of the pipe for which total reflections inside the pipe take place, as shown in the figure.

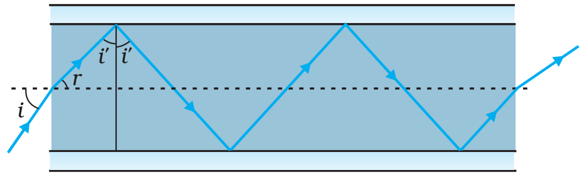


FIGURE 9.28

(b) What is the answer if there is no outer covering of the pipe?

9.18 The image of a small electric bulb fixed on the wall of a room is to be obtained on the opposite wall 3m away by means of a large convex lens. What is the maximum possible focal length of the lens required for the purpose?

9.19 A screen is placed 90cm from an object. The image of the object on the screen is formed by a convex lens at two different locations separated by 20cm. Determine the focal length of the lens.

9.20 (a) Determine the 'effective focal length' of the combination of the two lenses in Exercise 9.10, if they are placed 8.0cm apart with their principal axes coincident. Does the answer depend on which side of the combination a beam of parallel light is incident? Is the notion of effective focal length of this system useful at all?

(b) An object 1.5 cm in size is placed on the side of the convex lens in the arrangement (a) above. The distance between the object and the convex lens is 40cm. Determine the magnification produced by the two-lens system, and the size of the image.

9.21 At what angle should a ray of light be incident on the face of a prism of refracting angle 60° so that it just suffers total internal reflection at the other face? The refractive index of the material of the prism is 1.524.

9.22 A card sheet divided into squares each of size 1 mm^2 is being viewed at a distance of 9 cm through a magnifying glass (a converging lens of focal length 9 cm) held close to the eye.

(a) What is the magnification produced by the lens? How much is the area of each square in the virtual image?

(b) What is the angular magnification (magnifying power) of the lens?

(c) Is the magnification in (a) equal to the magnifying power in (b)? Explain.

9.23 (a) At what distance should the lens be held from the card sheet in Exercise 9.22 in order to view the squares distinctly with the maximum possible magnifying power?

(b) What is the magnification in this case?

(c) Is the magnification equal to the magnifying power in this case? Explain.

9.24 What should be the distance between the object in Exercise 9.23 and the magnifying glass if the virtual image of each square in the figure is to have an area of 6.25 mm^2. Would you be able to see the squares distinctly with your eyes very close to the magnifier?

[*Note:* Exercises 9.22 to 9.24 will help you clearly understand the difference between magnification in absolute size and the angular magnification (or magnifying power) of an instrument.]

9.25 Answer the following questions:

(a) The angle subtended at the eye by an object is equal to the angle subtended at the eye by the virtual image produced by a magnifying glass. In what sense then does a magnifying glass provide angular magnification?

(b) In viewing through a magnifying glass, one usually positions one's eyes very close to the lens. Does angular magnification change if the eye is moved back?

(c) Magnifying power of a simple microscope is inversely proportional to the focal length of the lens. What then stops us from using a convex lens of smaller and smaller focal length and achieving greater and greater magnifying power?

(d) Why must both the objective and the eyepiece of a compound microscope have short focal lengths?

(e) When viewing through a compound microscope, our eyes should be positioned not on the eyepiece but a short distance away from it for best viewing. Why? How much should be that short distance between the eye and eyepiece?

9.26 An angular magnification (magnifying power) of 30X is desired using an objective of focal length 1.25 cm and an eyepiece of focal length 5 cm. How will you set up the compound microscope?

9.27 A small telescope has an objective lens of focal length 140 cm and an eyepiece of focal length 5.0 cm. What is the magnifying power of the telescope for viewing distant objects when

(a) the telescope is in normal adjustment (i.e., when the final image is at infinity)?

(b) the final image is formed at the least distance of distinct vision (25 cm)?

9.28 (a) For the telescope described in Exercise 9.27 (a), what is the separation between the objective lens and the eyepiece?

(b) If this telescope is used to view a 100 m tall tower 3 km away, what is the height of the image of the tower formed by the objective lens?

(c) What is the height of the final image of the tower if it is formed at 25 cm?

9.29 A Cassegrain telescope uses two mirrors as shown in Fig. 9.26. Such a telescope is built with the mirrors 20 mm apart. If the radius of curvature of the large mirror is 220 mm and the small mirror is 140 mm, where will the final image of an object at infinity be?

9.30 Light incident normally on a plane mirror attached to a galvanometer coil retraces backwards as shown in Fig. 9.29. A current in the coil produces a deflection of 3.5° of the mirror. What is the displacement of the reflected spot of light on a screen placed 1.5 m away?

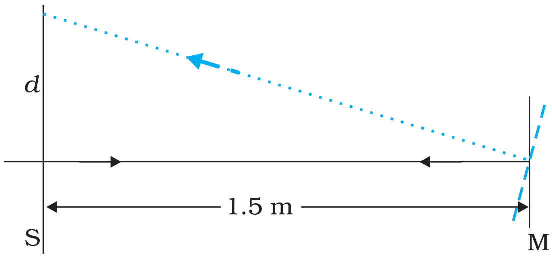


FIGURE 9.29

9.31 Figure 9.30 shows an equiconvex lens (of refractive index 1.50) in contact with a liquid layer on top of a plane mirror. A small needle with its tip on the principal axis is moved along the axis until its inverted image is found at the position of the needle. The distance of the needle from the lens is measured to be 45.0 cm. The liquid is removed and the experiment is repeated. The new distance is measured to be 30.0 cm. What is the refractive index of the liquid?

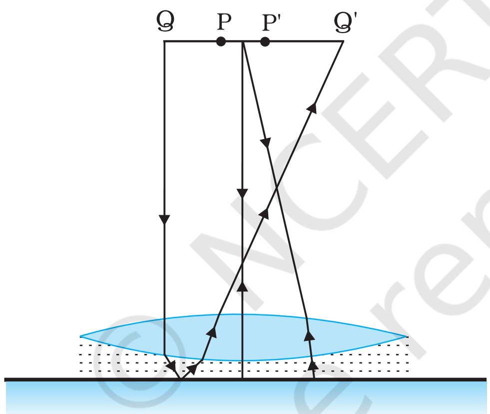


FIGURE 9.30

Notes

Reprint 2025-26

Chapter Ten

WAVE OPTICS

10.1 Introduction

In 1637 Descartes gave the corpuscular model of light and derived Snell's law. It explained the laws of reflection and refraction of light at an interface. The corpuscular model predicted that if the ray of light (on refraction) bends towards the normal then the speed of light would be greater in the second medium. This corpuscular model of light was further developed by Isaac Newton in his famous book entitled *OPTICKS* and because of the tremendous popularity of this book, the corpuscular model is very often attributed to Newton.

In 1678, the Dutch physicist Christiaan Huygens put forward the wave theory of light – it is this wave model of light that we will discuss in this chapter. As we will see, the wave model could satisfactorily explain the phenomena of reflection and refraction; however, it predicted that on refraction if the wave bends towards the normal then the speed of light would be less in the second medium. This is in contradiction to the prediction made by using the corpuscular model of light. It was much later confirmed by experiments where it was shown that the speed of light in water is less than the speed in air confirming the prediction of the wave model; Foucault carried out this experiment in 1850.

The wave theory was not readily accepted primarily because of Newton's authority and also because light could travel through vacuum and it was felt that a wave would always require a medium to propagate from one point to the other. However, when Thomas Young performed his famous interference experiment in 1801, it was firmly established that light is indeed a wave phenomenon. The wavelength of visible light was measured and found to be extremely small; for example, the wavelength of yellow light is about 0.6 μm. Because of the smallness of the wavelength of visible light (in comparison to the dimensions of typical mirrors and lenses), light can be assumed to approximately travel in straight lines. This is the field of geometrical optics, which we had discussed in the previous chapter. Indeed, the branch of optics in which one completely neglects the finiteness of the wavelength is called geometrical optics and a ray is defined as the path of energy propagation in the limit of wavelength tending to zero.

After the interference experiment of Young in 1801, for the next 40 years or so, many experiments were carried out involving the interference and diffraction of lightwaves; these experiments could only be satisfactorily explained by assuming a wave model of light. Thus, around the middle of the nineteenth century, the wave theory seemed to be very well established. The only major difficulty was that since it was thought that a wave required a medium for its propagation, how could light waves propagate through vacuum. This was explained when Maxwell put forward his famous electromagnetic theory of light. Maxwell had developed a set of equations describing the laws of electricity and magnetism and using these equations he derived what is known as the wave equation from which he *predicted* the existence of electromagnetic waves*. From the wave equation, Maxwell could calculate the speed of electromagnetic waves in free space and he found that the theoretical value was very close to the measured value of speed of light. From this, he propounded that *light must be an electromagnetic wave*. Thus, according to Maxwell, light waves are associated with changing electric and magnetic fields; changing electric field produces a time and space varying magnetic field and a changing magnetic field produces a time and space varying electric field. The changing electric and magnetic fields result in the propagation of electromagnetic waves (or light waves) even in vacuum.

In this chapter we will first discuss the original formulation of the *Huygens principle* and derive the laws of reflection and refraction. In Sections 10.4 and 10.5, we will discuss the phenomenon of interference which is based on the principle of superposition. In Section 10.6 we will discuss the phenomenon of diffraction which is based on Huygens-Fresnel principle. Finally in Section 10.7 we will discuss the phenomenon of polarisation which is based on the fact that the light waves are *transverse electromagnetic waves*.

---

* Maxwell had predicted the existence of electromagnetic waves around 1855; it was much later (around 1890) that Heinrich Hertz produced radiowaves in the laboratory. J.C. Bose and G. Marconi made practical applications of the *Hertzian waves*

10.2 Huygens Principle

We would first define a wavefront: when we drop a small stone on a calm pool of water, waves spread out from the point of impact. Every point on the surface starts oscillating with time. At any instant, a photograph of the surface would show circular rings on which the disturbance is maximum. Clearly, all points on such a circle are oscillating in phase because they are at the same distance from the source. Such a locus of points, which oscillate in phase is called a *wavefront*; thus *a wavefront is defined as a surface of constant phase*. The speed with which the wavefront moves outwards from the source is called the speed of the wave. The energy of the wave travels in a direction perpendicular to the wavefront.

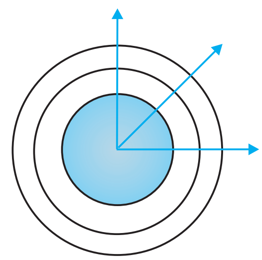


FIGURE 10.1 (a) A diverging spherical wave emanating from a point source. The wavefronts are spherical.

If we have a point source emitting waves uniformly in all directions, then the locus of points which have the same amplitude and vibrate in the same phase are spheres and we have what is known as a *spherical wave* as shown in Fig. 10.1(a). At a large distance from the source, a small portion of the sphere can be considered as a plane and we have what is known as a *plane wave* [Fig. 10.1(b)].

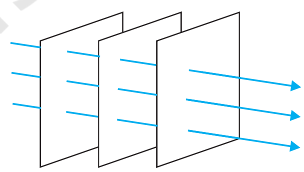


FIGURE 10.1 (b) At a large distance from the source, a small portion of the spherical wave can be approximated by a plane wave.

Now, if we know the shape of the wavefront at $t = 0$, then Huygens principle allows us to determine the shape of the wavefront at a later time τ. Thus, Huygens principle is essentially a geometrical construction, which given the shape of the wafefront at any time allows us to determine the shape of the wavefront at a later time. Let us consider a diverging wave and let F_1F_2 represent a portion of the spherical wavefront at $t = 0$ (Fig. 10.2). Now, according to Huygens principle, *each point of the wavefront is the source of a secondary disturbance and the wavelets emanating from these points spread out in all directions with the speed of the wave. These wavelets emanating from the wavefront are usually referred to as secondary wavelets and if we draw a common tangent to all these spheres, we obtain the new position of the wavefront at a later time.*

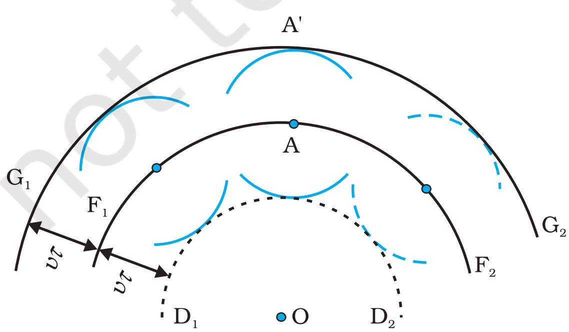


FIGURE 10.2 F_1F_2 represents the spherical wavefront (with O as centre) at $t = 0$. The envelope of the secondary wavelets emanating from F_1F_2 produces the forward moving wavefront G_1G_2. The backwave D_1D_2 does not exist.

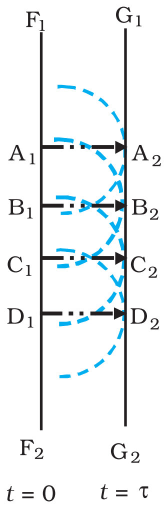


FIGURE 10.3 Huygens geometrical construction for a plane wave propagating to the right. F_1 F_2 is the plane wavefront at $t = 0$ and G_1G_2 is the wavefront at a later time τ. The lines A_1A_2, B_1B_2 ... etc., are normal to both F_1F_2 and G_1G_2 and represent rays.

Thus, if we wish to determine the shape of the wavefront at $t = \tau$, we draw spheres of radius $v\tau$ from each point on the spherical wavefront where v represents the speed of the waves in the medium. If we now draw a common tangent to all these spheres, we obtain the new position of the wavefront at $t = \tau$. The new wavefront shown as G_1G_2 in Fig. 10.2 is again spherical with point O as the centre.

The above model has one shortcoming: we also have a backwave which is shown as D_1D_2 in Fig. 10.2. Huygens argued that the amplitude of the secondary wavelets is maximum in the forward direction and zero in the backward direction; by making this adhoc assumption, Huygens could explain the absence of the backwave. However, this adhoc assumption is not satisfactory and the absence of the backwave is really justified from more rigorous wave theory.

In a similar manner, we can use Huygens principle to determine the shape of the wavefront for a plane wave propagating through a medium (Fig. 10.3).

10.3 Refraction and Reflection of Plane Waves using Huygens Principle

10.3.1 Refraction of a plane wave

We will now use Huygens principle to derive the laws of refraction. Let PP′ represent the surface separating medium 1 and medium 2, as shown in Fig. 10.4. Let v_1 and v_2 represent the speed of light in medium 1 and medium 2, respectively. We assume a plane wavefront AB propagating in the direction A′A incident on the interface at an angle i as shown in the figure. Let τ be the time taken by the wavefront to travel the distance BC. Thus,

$$BC = v_1 \tau$$

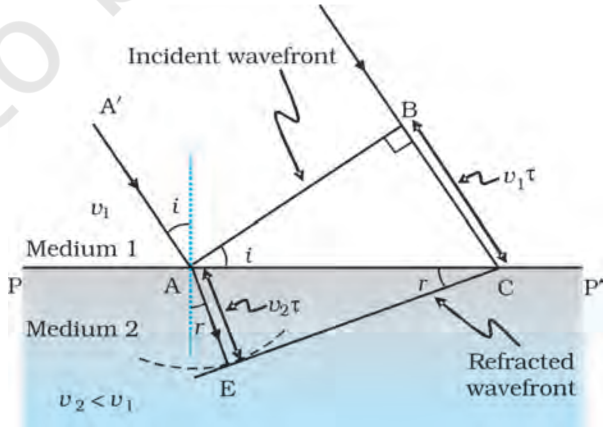


FIGURE 10.4 *A plane wave* AB is incident at an angle i on the surface PP′ separating medium 1 and medium 2. The plane wave undergoes refraction and CE represents the refracted wavefront. The figure corresponds to $v_2 < v_1$ so that the refracted waves bends towards the normal.

In order to determine the shape of the refracted wavefront, we draw a sphere of radius $v_2\tau$ from the point A in the second medium (the speed of the wave in the second medium is v_2). Let CE represent a tangent plane drawn from the point C on to the sphere. Then, AE = $v_2 \tau$ and CE would represent the refracted wavefront. If we now consider the triangles ABC and AEC, we readily obtain

$$\sin i = \frac{\text{BC}}{\text{AC}} = \frac{v_1 \tau}{\text{AC}} \tag{10.1}$$

and

$$\sin r = \frac{\text{AE}}{\text{AC}} = \frac{v_2 \tau}{\text{AC}} \tag{10.2}$$

where i and r are the angles of incidence and refraction, respectively. Thus we obtain

$$\frac{\sin i}{\sin r} = \frac{v_1}{v_2} \tag{10.3}$$

From the above equation, we get the important result that if $r < i$ (i.e., if the ray bends toward the normal), the speed of the light wave in the second medium (v_2) will be less then the speed of the light wave in the first medium (v_1). This prediction is opposite to the prediction from the corpuscular model of light and as later experiments showed, the prediction of the wave theory is correct. Now, if c represents the speed of light in vacuum, then,

$$n_1 = \frac{c}{v_1} \tag{10.4}$$

and

$$n_2 = \frac{c}{v_2} \tag{10.5}$$

are known as the refractive indices of medium 1 and medium 2, respectively. In terms of the refractive indices, Eq. (10.3) can be written as

$$n_1 \sin i = n_2 \sin r \tag{10.6}$$

This is the *Snell's law of refraction*. Further, if λ_1 and λ_2 denote the wavelengths of light in medium 1 and medium 2, respectively and if the distance BC is equal to λ_1 then the distance AE will be equal to λ_2 (because if the crest from B has reached C in time τ, then the crest from A should have also reached E in time τ); thus,

$$\frac{\lambda_1}{\lambda_2} = \frac{\text{BC}}{\text{AE}} = \frac{v_1}{v_2}$$

or

$$\frac{v_1}{\lambda_1} = \frac{v_2}{\lambda_2} \tag{10.7}$$

Christiaan Huygens (1629 – 1695) Dutch physicist, astronomer, mathematician and the founder of the wave theory of light. His book, *Treatise on light*, makes fascinating reading even today. He brilliantly explained the double refraction shown by the mineral calcite in this work in addition to reflection and refraction. He was the first to analyse circular and simple harmonic motion and designed and built improved clocks and telescopes. He discovered the true geometry of Saturn's rings.

CHRISTIAAN HUYGENS (1629 – 1695)

The above equation implies that when a wave gets refracted into a denser medium ($v_1 > v_2$) the wavelength and the speed of propagation decrease but the *frequency* ν (= v/λ) *remains the same.*

10.3.2 Refraction at a rarer medium

We now consider refraction of a plane wave at a rarer medium, i.e., $v_2 > v_1$. Proceeding in an exactly similar manner we can construct a refracted wavefront as shown in Fig. 10.5. The angle of refraction will now be greater than angle of incidence; however, we will still have $n_1 \sin i = n_2 \sin r$. We define an angle i_c by the following equation

$$\sin i_c = \frac{n_2}{n_1} \tag{10.8}$$

Thus, if $i = i_c$ then $\sin r = 1$ and $r = 90°$. Obviously, for $i > i_c$, there can not be any refracted wave. The angle i_c is known as the *critical angle* and for all angles of incidence greater than the critical angle, we will not have any refracted wave and the wave will undergo what is known as *total internal reflection*. The phenomenon of total internal reflection and its applications was discussed in Section 9.4.

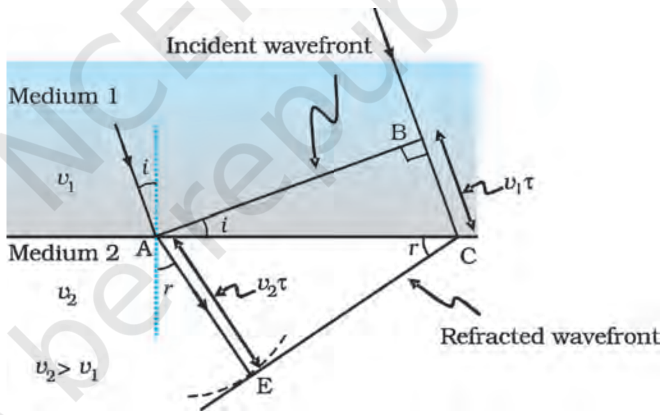


FIGURE 10.5 Refraction of a plane wave incident on a rarer medium for which $v_2 > v_1$. The plane wave bends away from the normal.

10.3.3 Reflection of a plane wave by a plane surface

We next consider a plane wave AB incident at an angle i on a reflecting surface MN. If v represents the speed of the wave in the medium and if τ represents the time taken by the wavefront to advance from the point B to C then the distance

$$BC = v\tau$$

In order to construct the reflected wavefront we draw a sphere of radius $v\tau$ from the point A as shown in Fig. 10.6. Let CE represent the tangent plane drawn from the point C to this sphere. Obviously

$$AE = BC = v\tau$$

Demonstration of interference, diffraction, refraction, resonance and Doppler effect
http://www.falstad.com/ripple/

PHYSICS

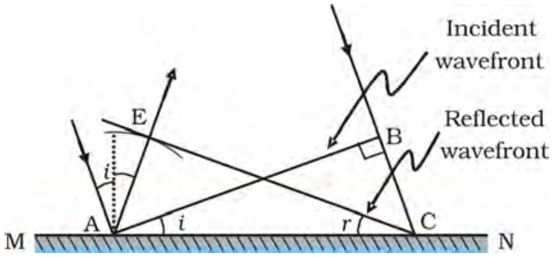


FIGURE 10.6 Reflection of a plane wave AB by the reflecting surface MN. AB and CE represent incident and reflected wavefronts.

If we now consider the triangles EAC and BAC we will find that they are congruent and therefore, the angles *i* and *r* (as shown in Fig. 10.6) would be equal. This is the *law of reflection*.

Once we have the laws of reflection and refraction, the behaviour of prisms, lenses, and mirrors can be understood. These phenomena were discussed in detail in Chapter 9 on the basis of rectilinear propagation of light. Here we just describe the behaviour of the wavefronts as they undergo reflection or refraction. In Fig. 10.7(a) we consider a plane wave passing through a thin prism. Clearly, since the speed of light waves is less in glass, the lower portion of the incoming wavefront (which travels through the greatest thickness of glass) will get delayed resulting in a tilt in the emerging wavefront as shown in the figure. In Fig. 10.7(b) we consider a plane wave incident on a thin convex lens; the central part of the incident plane wave traverses the thickest portion of the lens and is delayed the most. The emerging wavefront has a depression at the centre and therefore the wavefront becomes spherical and converges to the point F which is known as the *focus*. In Fig. 10.7(c) a plane wave is incident on a concave mirror and on reflection we have a spherical wave converging to the focal point F. In a similar manner, we can understand refraction and reflection by concave lenses and convex mirrors.

From the above discussion it follows that the total time taken from a point on the object to the corresponding point on the image is the same measured along any ray. For example, when a convex lens focusses light to form a real image, although the ray going through the centre traverses a shorter path, but because of the slower speed in glass, the time taken is the same as for rays travelling near the edge of the lens.

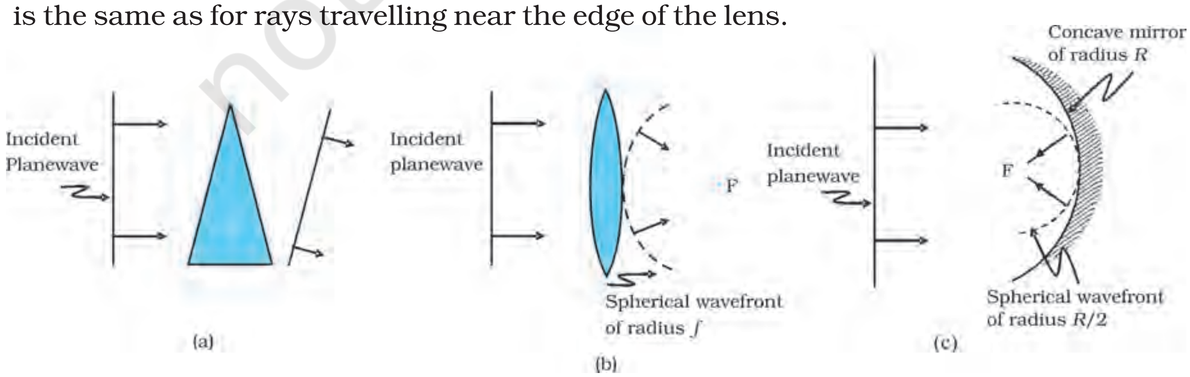


FIGURE 10.7 Refraction of a plane wave by (a) a thin prism, (b) a convex lens. (c) Reflection of a plane wave by a concave mirror.

Example 10.1

(a) When monochromatic light is incident on a surface separating two media, the reflected and refracted light both have the same frequency as the incident frequency. Explain why?

(b) When light travels from a rarer to a denser medium, the speed decreases. Does the reduction in speed imply a reduction in the energy carried by the light wave?

(c) In the wave picture of light, intensity of light is determined by the square of the amplitude of the wave. What determines the intensity of light in the photon picture of light.

Solution

(a) Reflection and refraction arise through interaction of incident light with the atomic constituents of matter. Atoms may be viewed as oscillators, which take up the frequency of the external agency (light) causing forced oscillations. The frequency of light emitted by a charged oscillator equals its frequency of oscillation. Thus, the frequency of scattered light equals the frequency of incident light.

(b) No. Energy carried by a wave depends on the amplitude of the wave, not on the speed of wave propagation.

(c) For a given frequency, intensity of light in the photon picture is determined by the number of photons crossing an unit area per unit time.

EXAMPLE 10.1

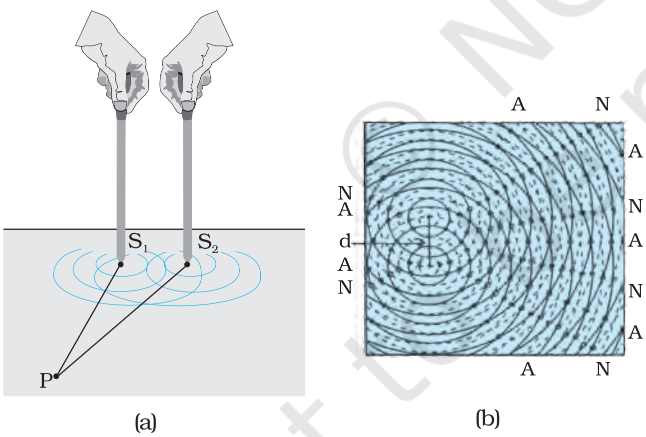


FIGURE 10.8 (a) Two needles oscillating in phase in water represent two coherent sources. (b) The pattern of displacement of water molecules at an instant on the surface of water showing nodal N (no displacement) and antinodal A (maximum displacement) lines.

10.4 Coherent and Incoherent Addition of Waves

In this section we will discuss the interference pattern produced by the superposition of two waves. You may recall that we had discussed the superposition principle in Chapter 14 of your Class XI textbook. Indeed the entire field of interference is based on the *superposition principle* according to which *at a particular point in the medium, the resultant displacement produced by a number of waves is the vector sum of the displacements produced by each of the waves.*

Consider two needles S_1 and S_2 moving periodically up and down in an identical fashion in a trough of water [Fig. 10.8(a)]. They produce two water waves, and at a particular point, the phase difference between the displacements produced by each of the waves does not change with time; when this happens the two sources are said to be *coherent*. Figure 10.8(b) shows the position of crests (solid circles) and troughs (dashed circles) at a given instant of time. Consider a point P for which

$$S_1 P = S_2 P$$

Reprint 2025-26

Since the distances S_1 P and S_2 P are equal, waves from S_1 and S_2 will take the same time to travel to the point P and waves that emanate from S_1 and S_2 in phase will also arrive, at the point P, in phase.

Thus, if the displacement produced by the source S_1 at the point P is given by

$$y_1 = a \cos \omega t$$

then, the displacement produced by the source S_2 (at the point P) will also be given by

$$y_2 = a \cos \omega t$$

Thus, the resultant of displacement at P would be given by

$$y = y_1 + y_2 = 2\, a \cos \omega t$$

Since the intensity is proportional to the square of the amplitude, the resultant intensity will be given by

$$I = 4\, I_0$$

where I_0 represents the intensity produced by each one of the individual sources; I_0 is proportional to a^2. In fact at any point on the perpendicular bisector of S_1S_2, the intensity will be $4I_0$. The two sources are said to interfere constructively and we have what is referred to as *constructive interference*. We next consider a point *Q* [Fig. 10.9(a)] for which

$$S_2Q - S_1Q = 2\lambda$$

The waves emanating from S_1 will arrive exactly two cycles earlier than the waves from S_2 and will again be in phase [Fig. 10.9(a)]. Thus, if the displacement produced by S_1 is given by

$$y_1 = a \cos \omega t$$

then the displacement produced by S_2 will be given by

$$y_2 = a \cos (\omega t - 4\pi) = a \cos \omega t$$

where we have used the fact that a path difference of 2λ corresponds to a phase difference of 4π. The two displacements are once again in phase and the intensity will again be $4\, I_0$ giving rise to constructive interference. In the above analysis we have assumed that the distances S_1Q and S_2Q are much greater than *d* (which represents the distance between S_1 and S_2) so that although S_1Q and S_2Q are not equal, the amplitudes of the displacement produced by each wave are very nearly the same.

We next consider a point R [Fig. 10.9(b)] for which

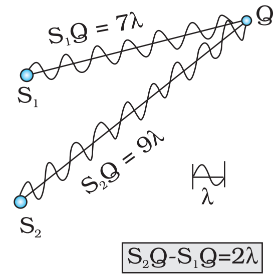
$$S_2R - S_1R = -2.5\lambda$$

The waves emanating from S_1 will arrive exactly two and a half cycles later than the waves from S_2 [Fig. 10.10(b)]. Thus if the displacement produced by S_1 is given by

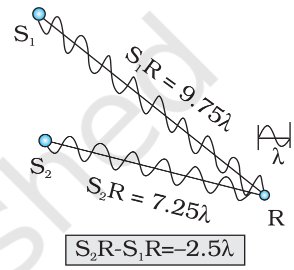
$$y_1 = a \cos \omega t$$

then the displacement produced by S_2 will be given by

$$y_2 = a \cos (\omega t + 5\pi) = -\, a \cos \omega t$$



(a)



(b)

FIGURE 10.9 (a) Constructive interference at a point Q for which the path difference is 2λ. (b) Destructive interference at a point R for which the path difference is $2.5\,\lambda$.

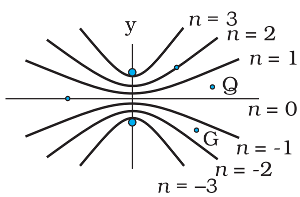


FIGURE 10.10 Locus of points for which $S_1P - S_2P$ is equal to zero, $\pm\lambda$, $\pm 2\lambda$, $\pm 3\lambda$.

where we have used the fact that a path difference of 2.5λ corresponds to a phase difference of 5π. The two displacements are now out of phase and the two displacements will cancel out to give zero intensity. This is referred to as *destructive interference*.

To summarise: If we have two coherent sources S_1 and S_2 vibrating in phase, then for an arbitrary point P whenever the path difference,

$$S_1P \sim S_2P = n\lambda \quad (n = 0, 1, 2, 3,...) \tag{10.9}$$

we will have constructive interference and the resultant intensity will be $4I_0$; the sign ~ between S_1P and S_2 P represents the difference between S_1P and S_2 P. On the other hand, if the point P is such that the path difference,

$$S_1P \sim S_2P = (n+\frac{1}{2})\lambda \quad (n = 0, 1, 2, 3, ...) \tag{10.10}$$

we will have *destructive interference* and the resultant intensity will be zero. Now, for any other arbitrary point G (Fig. 10.10) let the phase difference between the two displacements be ϕ. Thus, if the displacement produced by S_1 is given by

$$y_1 = a \cos \omega t$$

then, the displacement produced by S_2 would be

$$y_2 = a \cos (\omega t + \phi)$$

and the resultant displacement will be given by

$$\begin{aligned} y &= y_1 + y_2 \\ &= a [\cos \omega t + \cos (\omega t + \phi)] \\ &= 2\, a \cos (\phi/2) \cos (\omega t + \phi/2) \end{aligned}$$

$$\left[\because \cos A + \cos B = 2\cos\left(\frac{A+B}{2}\right)\cos\left(\frac{A-B}{2}\right)\right]$$

The amplitude of the resultant displacement is $2a \cos (\phi/2)$ and therefore the intensity at that point will be

$$I = 4\, I_0 \cos^2 (\phi/2) \tag{10.11}$$

If $\phi = 0, \pm 2\pi, \pm 4\pi,...$ which corresponds to the condition given by Eq. (10.9) we will have constructive interference leading to maximum intensity. On the other hand, if $\phi = \pm\pi, \pm 3\pi, \pm 5\pi$ ... [which corresponds to the condition given by Eq. (10.10)] we will have destructive interference leading to zero intensity.

Now if the two sources are coherent (i.e., if the two needles are going up and down regularly) then the phase difference ϕ at any point will not change with time and we will have a stable interference pattern; i.e., the positions of maxima and minima will not change with time. However, if the two needles do not maintain a constant phase difference, then the interference pattern will also change with time and, if the phase difference changes very rapidly with time, the positions of maxima and minima will also vary rapidly with time and we will see a "time-averaged" intensity distribution. When this happens, we will observe an average intensity that will be given by

$$I = 2\, I_0 \tag{10.12}$$

at all points.

PHYSICS

Ripple Tank experiments on wave interference
http://phet.colorado.edu/en/simulation/legacy/wave-interference

When the phase difference between the two vibrating sources changes rapidly with time, we say that the two sources are incoherent and when this happens the intensities just add up. This is indeed what happens when two separate light sources illuminate a wall.

10.5 Interference of Light Waves and Young's Experiment

We will now discuss interference using light waves. If we use two sodium lamps illuminating two pinholes (Fig. 10.11) we will not observe any interference fringes. This is because of the fact that the light wave emitted from an ordinary source (like a sodium lamp) undergoes abrupt phase changes in times of the order of 10^{-10} seconds. Thus the light waves coming out from two independent sources of light will not have any fixed phase relationship and would be incoherent, when this happens, as discussed in the previous section, the intensities on the screen will add up.

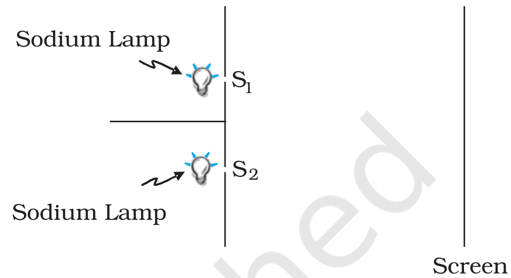


FIGURE 10.11 If two sodium lamps illuminate two pinholes S_1 and S_2, the intensities will add up and no interference fringes will be observed on the screen.

The British physicist Thomas Young used an ingenious technique to "lock" the phases of the waves emanating from S_1 and S_2. He made two pinholes S_1 and S_2 (very close to each other) on an opaque screen [Fig. 10.12(a)]. These were illuminated by another pinholes that was in turn, lit by a bright source. Light waves spread out from S and fall on both S_1 and S_2. S_1 and S_2 then behave like two coherent sources because light waves coming out from S_1 and S_2 are derived from the same original source and any abrupt phase change in S will manifest in exactly similar phase changes in the light coming out from S_1 and S_2. Thus, the two sources S_1 and S_2 will be *locked* in phase; i.e., they will be coherent like the two vibrating needle in our water wave example [Fig. 10.8(a)].

The spherical waves emanating from S_1 and S_2 will produce interference fringes on the screen GG′, as shown in Fig. 10.12(b). The positions of maximum and minimum intensities can be calculated by using the analysis given in Section 10.4.

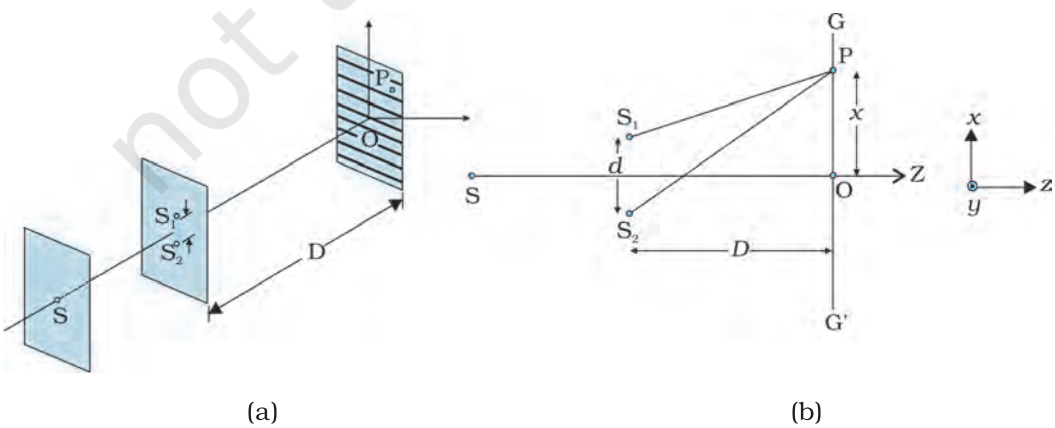


FIGURE 10.12 Young's arrangement to produce interference pattern.

Reprint 2025-26

THOMAS YOUNG (1773 – 1829)

Thomas Young (1773 – 1829) English physicist, physician and Egyptologist. Young worked on a wide variety of scientific problems, ranging from the structure of the eye and the mechanism of vision to the decipherment of the Rosetta stone. He revived the wave theory of light and recognised that interference phenomena provide proof of the wave properties of light.

We will have constructive interference resulting in a bright region when $\frac{xd}{D} = n\lambda$. That is,

$$x = x_n = \frac{n\lambda D}{d};\ n = 0, \pm 1, \pm 2, \dots \qquad (10.13)$$

On the other hand, we will have destructive interference resulting in a dark region when $\frac{xd}{D} = (n+\frac{1}{2})\lambda$ that is

$$x = x_n = (n+\frac{1}{2})\frac{\lambda D}{d};\quad n = 0, \pm 1, \pm 2 \qquad (10.14)$$

Thus dark and bright bands appear on the screen, as shown in Fig. 10.13. Such bands are called *fringes*. Equations (10.13) and (10.14) show that dark and bright fringes are equally spaced.

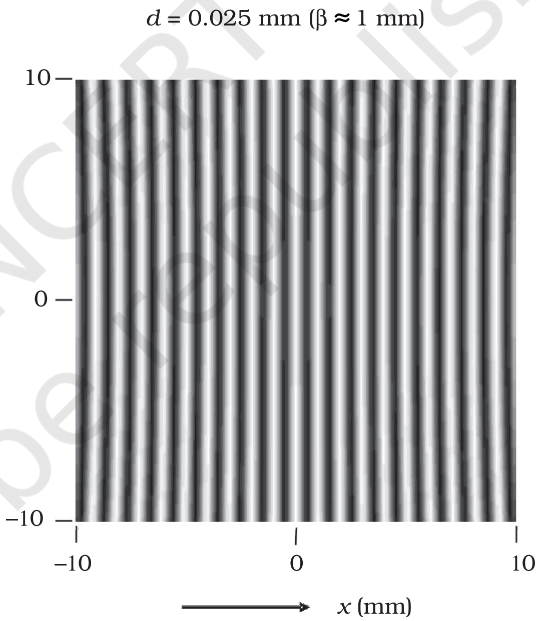


FIGURE 10.13 Computer generated fringe pattern produced by two point source S_1 and S_2 on the screen GG′ (Fig. 10.12); d = 0.025 mm, D = 5 cm and $\lambda = 5 \times 10^{-5}$ cm.) (Adopted from OPTICS by A. Ghatak, Tata McGraw Hill Publishing Co. Ltd., New Delhi, 2000.)

10.6 Diffraction

If we look clearly at the shadow cast by an opaque object, close to the region of geometrical shadow, there are alternate dark and bright regions just like in interference. This happens due to the phenomenon of diffraction. Diffraction is a general characteristic exhibited by all types of waves, be it sound waves, light waves, water waves or matter waves. Since the wavelength of light is much smaller than the dimensions of most obstacles; we do not encounter diffraction effects of light in everyday observations. However, the finite resolution of our eye or of optical instruments such as telescopes or microscopes is limited due to the phenomenon of diffraction. Indeed the colours that you see when a CD is viewed is due to diffraction effects. We will now discuss the phenomenon of diffraction.

10.6.1 The single slit

In the discussion of Young's experiment, we stated that a single narrow slit acts as a new source from which light spreads out. Even before Young, early experimenters – including Newton – had noticed that light spreads out from narrow holes and slits. It seems to turn around corners and enter regions where we would expect a shadow. These effects, known as *diffraction*, can only be properly understood using wave ideas. After all, you are hardly surprised to hear sound waves from someone talking around a corner!

When the double slit in Young's experiment is replaced by a single narrow slit (illuminated by a monochromatic source), a broad pattern with a central bright region is seen. On both sides, there are alternate dark and bright regions, the intensity becoming weaker away from the centre (Fig. 10.15). To understand this, go to Fig. 10.14, which shows a parallel beam of light falling normally on a single slit LN of width a. The diffracted light goes on to meet a screen. The midpoint of the slit is M.

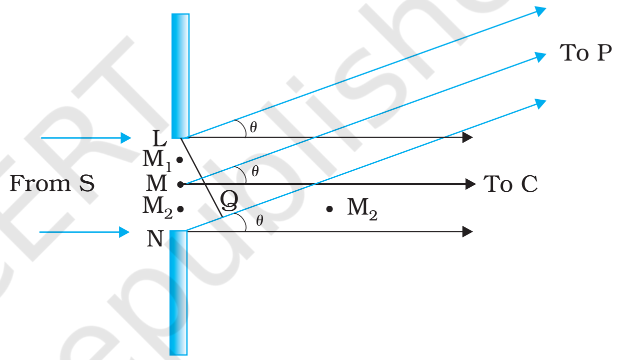


FIGURE 10.14 The geometry of path differences for diffraction by a single slit.

A straight line through M perpendicular to the slit plane meets the screen at C. We want the intensity at any point P on the screen. As before, straight lines joining P to the different points L,M,N, etc., can be treated as parallel, making an angle θ with the normal MC.

The basic idea is to divide the slit into much smaller parts, and add their contributions at P with the proper phase differences. We are treating different parts of the wavefront at the slit as secondary sources. Because the incoming wavefront is parallel to the plane of the slit, these sources are in phase.

It is observed that the intensity has a central maximum at $\theta = 0$ and other secondary maxima at $\theta \approx (n+1/2)\ \lambda/a$, which go on becoming weaker and weaker with increasing n. The minima (zero intensity) are at $\theta \approx n\lambda/a$, $n = \pm1, \pm2, \pm3, ...$

The photograph and intensity pattern corresponding to it is shown in Fig. 10.15.

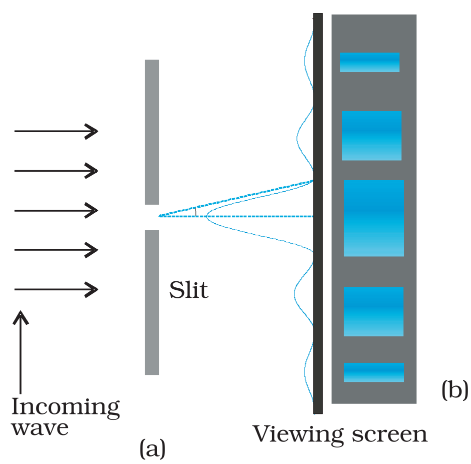


FIGURE 10.15 Intensity distribution and photograph of fringes due to diffraction at single slit.

There has been prolonged discussion about difference between intereference and diffraction among scientists since the discovery of these phenomena. In this context, it is interesting to note what Richard Feynman* has said in his famous Feynman Lectures on Physics:

> *No one has ever been able to define the difference between interference and diffraction satisfactorily. It is just a question of usage, and there is no specific, important physical difference between them. The best we can do is, roughly speaking, is to say that when there are only a few sources, say two interfering sources, then the result is usually called interference, but if there is a large number of them, it seems that the word diffraction is more often used.*

In the double-slit experiment, we must note that the pattern on the screen is actually a superposition of single-slit diffraction from each slit or hole, and the double-slit interference pattern.

10.6.2 Seeing the single slit diffraction pattern

It is surprisingly easy to see the single-slit diffraction pattern for oneself. The equipment needed can be found in most homes — two razor blades and one clear glass electric bulb preferably with a straight filament. One has to hold the two blades so that the edges are parallel and have a narrow slit in between. This is easily done with the thumb and forefingers (Fig. 10.16).

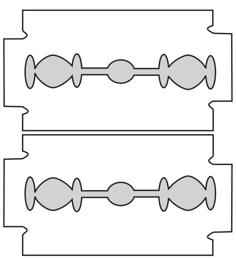


FIGURE 10.16 Holding two blades to form a single slit. A bulb filament viewed through this shows clear diffraction bands.

Keep the slit parallel to the filament, right in front of the eye. Use spectacles if you normally do. With slight adjustment of the width of the slit and the parallelism of the edges, the pattern should be seen with its bright and dark bands. Since the position of all the bands (except the central one) depends on wavelength, they will show some colours. Using a filter for red or blue will make the fringes clearer. With both filters available, the wider fringes for red compared to blue can be seen.

In this experiment, the filament plays the role of the first slit S in Fig. 10.15. The lens of the eye focuses the pattern on the screen (the retina of the eye).

With some effort, one can cut a double slit in an aluminium foil with a blade. The bulb filament can be viewed as before to repeat Young's experiment. In daytime, there is another suitable bright source subtending a small angle at the eye. This is the reflection of the Sun in any shiny convex surface (e.g., a cycle bell). Do not try direct sunlight – it can damage the eye and will not give fringes anyway as the Sun subtends an angle of (1/2) .

In interference and diffraction, light energy is redistributed. If it reduces in one region, producing a dark fringe, it increases in another region, producing a bright fringe. There is no gain or loss of energy, which is consistent with the principle of conservation of energy.

---

* Richard Feynman was one of the recipients of the 1965 Nobel Prize in Physics for his fundamental work in quantum electrodynamics.

10.7 Polarisation

Consider holding a long string that is held horizontally, the other end of which is assumed to be fixed. If we move the end of the string up and down in a periodic manner, we will generate a wave propagating in the $+x$ direction (Fig. 10.17). Such a wave could be described by the following equation

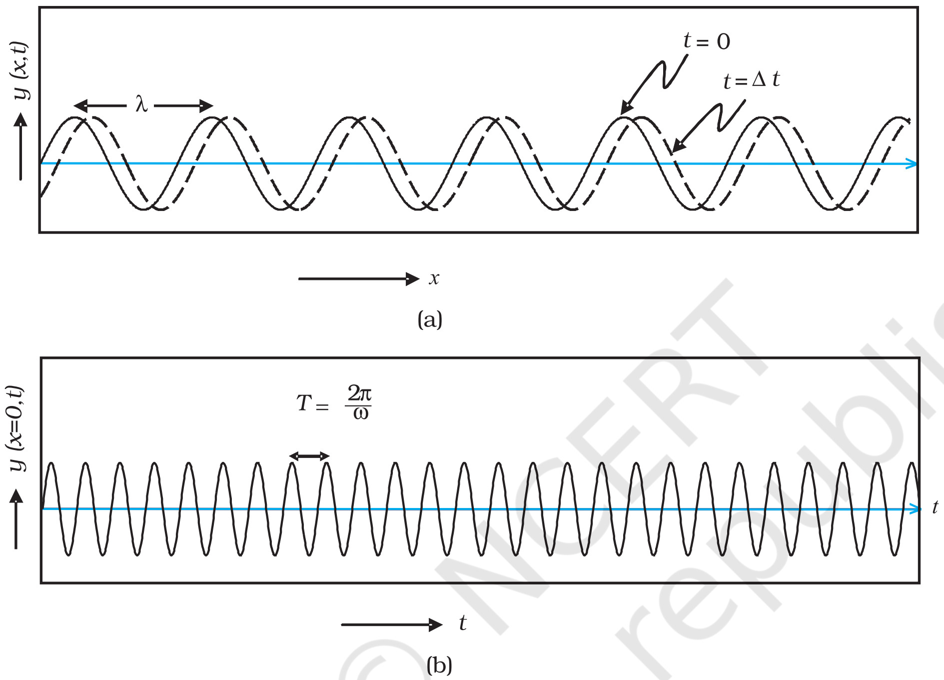


FIGURE 10.17 (a) The curves represent the displacement of a string at $t = 0$ and at $t = \Delta t$, respectively when a sinusoidal wave is propagating in the $+x$-direction. (b) The curve represents the time variation of the displacement at $x = 0$ when a sinusoidal wave is propagating in the $+x$-direction. At $x = \Delta x$, the time variation of the displacement will be slightly displaced to the right.

$$y(x,t) = a \sin(kx - \omega t) \tag{10.15}$$

where a and $\omega(= 2\pi\nu)$ represent the amplitude and the angular frequency of the wave, respectively; further,

$$\lambda = \frac{2\pi}{k} \tag{10.16}$$

represents the wavelength associated with the wave. We had discussed propagation of such waves in Chapter 14 of Class XI textbook. Since the displacement (which is along the y direction) is at right angles to the direction of propagation of the wave, we have what is known as a *transverse wave*. Also, since the displacement is in the y direction, it is often referred to as a y-polarised wave. Since each point on the string moves on a straight line, the wave is also referred to as a linearly polarised wave. Further, the string always remains confined to the *x-y* plane and therefore it is also referred to as a *plane polarised wave*.

In a similar manner we can consider the vibration of the string in the *x-z* plane generating a *z*-polarised wave whose displacement will be given by

$$z(x,t) = a \sin(kx - \omega t) \tag{10.17}$$

It should be mentioned that the linearly polarised waves [described by Eqs. (10.15) and (10.17)] are all transverse waves; i.e., the displacement of each point of the string is always at right angles to the direction of propagation of the wave. Finally, if the plane of vibration of the string is changed randomly in very short intervals of time, then we have what is known as an *unpolarised wave*. Thus, for an unpolarised wave the displacement will be randomly changing with time though it will always be perpendicular to the direction of propagation.

Light waves are transverse in nature; i.e., the electric field associated with a propagating light wave is always at right angles to the direction of propagation of the wave. This can be easily demonstrated using a simple polaroid. You must have seen thin plastic like sheets, which are called *polaroids*. A polaroid consists of long chain molecules aligned in a particular direction. The electric vectors (associated with the propagating light wave) along the direction of the aligned molecules get absorbed. Thus, if an unpolarised light wave is incident on such a polaroid then the light wave will get linearly polarised with the electric vector oscillating along a direction perpendicular to the aligned molecules; this direction is known as the *pass-axis* of the polaroid.

Thus, if the light from an ordinary source (like a sodium lamp) passes through a polaroid sheet P_1, it is observed that its intensity is reduced by half. Rotating P_1 has no effect on the transmitted beam and transmitted intensity remains constant. Now, let an identical piece of polaroid P_2 be placed before P_1. As expected, the light from the lamp is reduced in intensity on passing through P_2 alone. But now rotating P_1 has a dramatic effect on the light coming from P_2. In one position, the intensity transmitted by P_2 followed by P_1 is nearly zero. When turned by 90° from this position, P_1 transmits nearly the full intensity emerging from P_2 (Fig. 10.18).

The experiment at figure 10.18 can be easily understood by assuming that light passing through the polaroid P_2 gets polarised along the pass-axis of P_2. If the pass-axis of P_2 makes an angle θ with the pass-axis of P_1, then when the polarised beam passes through the polaroid P_2, the component $E \cos \theta$ (along the pass-axis of P_2) will pass through P_2. Thus, as we rotate the polaroid P_1 (or P_2), the intensity will vary as:

$$I = I_0 \cos^2\theta \tag{10.18}$$

where I_0 is the intensity of the polarized light after passing through P_1. This is known as *Malus' law*. The above discussion shows that the

Reprint 2025-26

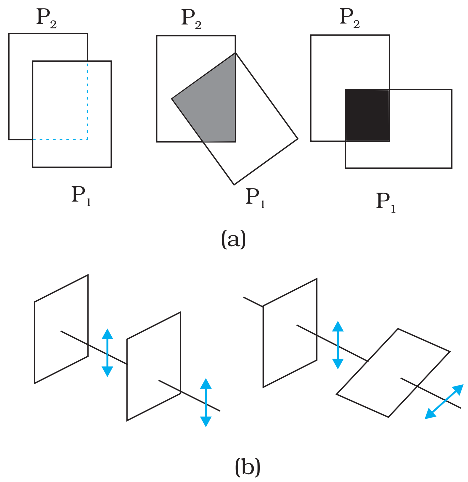


FIGURE 10.18 (a) Passage of light through two polaroids P_2 and P_1. The transmitted fraction falls from 1 to 0 as the angle between them varies from 0° to 90°. Notice that the light seen through a single polaroid P_1 does not vary with angle. (b) Behaviour of the electric vector when light passes through two polaroids. The transmitted polarisation is the component parallel to the polaroid axis. The double arrows show the oscillations of the electric vector.

intensity coming out of a single polaroid is half of the incident intensity. By putting a second polaroid, the intensity can be further controlled from 50% to zero of the incident intensity by adjusting the angle between the pass-axes of two polaroids.

Polaroids can be used to control the intensity, in sunglasses, windowpanes, etc. Polaroids are also used in photographic cameras and 3D movie cameras.

Example 10.2 Discuss the intensity of transmitted light when a polaroid sheet is rotated between two crossed polaroids?

Solution Let I_0 be the intensity of polarised light after passing through the first polariser P_1. Then the intensity of light after passing through second polariser P_2 will be

$$I = I_0 \cos^2\theta,$$

where θ is the angle between pass axes of P_1 and P_2. Since P_1 and P_3 are crossed the angle between the pass axes of P_2 and P_3 will be $(\pi/2-\theta)$. Hence the intensity of light emerging from P_3 will be

$$I = I_0 \cos^2\theta \cos^2\left(\frac{\pi}{2}-\theta\right)$$

$$= I_0 \cos^2\theta \sin^2\theta = (I_0/4)\sin^2 2\theta$$

Therefore, the transmitted intensity will be maximum when $\theta = \pi/4$.

EXAMPLE 10.2

SUMMARY

1. Huygens' principle tells us that each point on a wavefront is a source of secondary waves, which add up to give the wavefront at a later time.
2. Huygens' construction tells us that the new wavefront is the forward envelope of the secondary waves. When the speed of light is independent of direction, the secondary waves are spherical. The rays are then perpendicular to both the wavefronts and the time of travel is the same measured along any ray. This principle leads to the well known laws of reflection and refraction.
3. The principle of superposition of waves applies whenever two or more sources of light illuminate the same point. When we consider the intensity of light due to these sources at the given point, there is an interference term in addition to the sum of the individual intensities. But this term is important only if it has a non-zero average, which occurs only if the sources have the same frequency and a stable phase difference.
4. Young's double slit of separation d gives equally spaced interference fringes.
5. A single slit of width a gives a diffraction pattern with a central maximum. The intensity falls to zero at angles of $\pm\frac{\lambda}{a}, \pm\frac{2\lambda}{a}$, etc., with successively weaker secondary maxima in between.
6. Natural light, e.g., from the sun is unpolarised. This means the electric vector takes all possible directions in the transverse plane, rapidly and randomly, during a measurement. A polaroid transmits only one component (parallel to a special axis). The resulting light is called linearly polarised or plane polarised. When this kind of light is viewed through a second polaroid whose axis turns through 2π, two maxima and minima of intensity are seen.

POINTS TO PONDER

1. Waves from a point source spread out in all directions, while light was seen to travel along narrow rays. It required the insight and experiment of Huygens, Young and Fresnel to understand how a wave theory could explain all aspects of the behaviour of light.
2. The crucial new feature of waves is interference of amplitudes from different sources which can be both constructive and destructive, as shown in Young's experiment.
3. Diffraction phenomena define the limits of ray optics. The limit of the ability of microscopes and telescopes to distinguish very close objects is set by the wavelength of light.
4. Most interference and diffraction effects exist even for longitudinal waves like sound in air. But polarisation phenomena are special to transverse waves like light waves.

Reprint 2025-26

EXERCISES

10.1 Monochromatic light of wavelength 589 nm is incident from air on a water surface. What are the wavelength, frequency and speed of (a) reflected, and (b) refracted light? Refractive index of water is 1.33.

10.2 What is the shape of the wavefront in each of the following cases:

(a) Light diverging from a point source.

(b) Light emerging out of a convex lens when a point source is placed at its focus.

(c) The portion of the wavefront of light from a distant star intercepted by the Earth.

10.3 (a) The refractive index of glass is 1.5. What is the speed of light in glass? (Speed of light in vacuum is 3.0×10^8 m s^{-1})

(b) Is the speed of light in glass independent of the colour of light? If not, which of the two colours red and violet travels slower in a glass prism?

10.4 In a Young's double-slit experiment, the slits are separated by 0.28 mm and the screen is placed 1.4 m away. The distance between the central bright fringe and the fourth bright fringe is measured to be 1.2 cm. Determine the wavelength of light used in the experiment.

10.5 In Young's double-slit experiment using monochromatic light of wavelength λ, the intensity of light at a point on the screen where path difference is λ, is K units. What is the intensity of light at a point where path difference is $\lambda/3$?

10.6 A beam of light consisting of two wavelengths, 650 nm and 520 nm, is used to obtain interference fringes in a Young's double-slit experiment.

(a) Find the distance of the third bright fringe on the screen from the central maximum for wavelength 650 nm.

(b) What is the least distance from the central maximum where the bright fringes due to both the wavelengths coincide?

Reprint 2025-26

Chapter Eleven

DUAL NATURE OF RADIATION AND MATTER

11.1 Introduction

The Maxwell's equations of electromagnetism and Hertz experiments on the generation and detection of electromagnetic waves in 1887 strongly established the wave nature of light. Towards the same period at the end of 19th century, experimental investigations on conduction of electricity (electric discharge) through gases at low pressure in a discharge tube led to many historic discoveries. The discovery of X-rays by Roentgen in 1895, and of electron by J. J. Thomson in 1897, were important milestones in the understanding of atomic structure. It was found that at sufficiently low pressure of about 0.001 mm of mercury column, a discharge took place between the two electrodes on applying the electric field to the gas in the discharge tube. A fluorescent glow appeared on the glass opposite to cathode. The colour of glow of the glass depended on the type of glass, it being yellowish-green for soda glass. The cause of this fluorescence was attributed to the radiation which appeared to be coming from the cathode. These *cathode rays* were discovered, in 1870, by William Crookes who later, in 1879, suggested that these rays consisted of streams of fast moving negatively charged particles. The British physicist J. J. Thomson (1856-1940) confirmed this hypothesis. By applying mutually perpendicular electric and magnetic fields across the discharge tube, J. J. Thomson was the first to determine experimentally the speed and the specific charge [charge to mass ratio (e/m)] of the cathode ray particles. They were found to travel with speeds ranging from about 0.1 to 0.2 times the speed of light (3×10^8 m/s). The presently accepted value of e/m is 1.76×10^{11} C/kg. Further, the value of e/m was found to be independent of the nature of the material/metal used as the cathode (emitter), or the gas introduced in the discharge tube. This observation suggested the universality of the cathode ray particles.

Around the same time, in 1887, it was found that certain metals, when irradiated by ultraviolet light, emitted negatively charged particles having small speeds. Also, certain metals when heated to a high temperature were found to emit negatively charged particles. The value of e/m of these particles was found to be the same as that for cathode ray particles. These observations thus established that all these particles, although produced under different conditions, were identical in nature. J. J. Thomson, in 1897, named these particles as *electrons*, and suggested that they were fundamental, universal constituents of matter. For his epoch-making discovery of electron, through his theoretical and experimental investigations on conduction of electricity by gasses, he was awarded the Nobel Prize in Physics in 1906. In 1913, the American physicist R. A. Millikan (1868-1953) performed the pioneering oil-drop experiment for the precise measurement of the charge on an electron. He found that the charge on an oil-droplet was always an integral multiple of an elementary charge, 1.602×10^{-19} C. Millikan's experiment established that *electric charge is quantised.* From the values of charge (e) and specific charge (e/m), the mass (m) of the electron could be determined.

11.2 Electron Emission

We know that metals have free electrons (negatively charged particles) that are responsible for their conductivity. However, the free electrons cannot normally escape out of the metal surface. If an electron attempts to come out of the metal, the metal surface acquires a positive charge and pulls the electron back to the metal. The free electron is thus held inside the metal surface by the attractive forces of the ions. Consequently, the electron can come out of the metal surface only if it has got sufficient energy to overcome the attractive pull. A certain minimum amount of energy is required to be given to an electron to pull it out from the surface of the metal. This minimum energy required by an electron to escape from the metal surface is called the *work function* of the metal. It is generally denoted by ϕ_0 and measured in eV (electron volt). One electron volt is the energy gained by an electron when it has been accelerated by a potential difference of 1 volt, so that 1 eV = 1.602×10^{-19} J.

This unit of energy is commonly used in atomic and nuclear physics. The work function (ϕ_0) depends on the properties of the metal and the nature of its surface.

The minimum energy required for the electron emission from the metal surface can be supplied to the free electrons by any one of the following physical processes:

(i) *Thermionic emission:* By suitably heating, sufficient thermal energy can be imparted to the free electrons to enable them to come out of the metal.

(ii) *Field emission*: By applying a very strong electric field (of the order of 10^8 V m^{-1}) to a metal, electrons can be pulled out of the metal, as in a spark plug.

(iii) *Photoelectric emission*: When light of suitable frequency illuminates a metal surface, electrons are emitted from the metal surface. These photo(light)-generated electrons are called *photoelectrons*.

11.3 Photoelectric Effect

11.3.1 Hertz's observations

The phenomenon of photoelectric emission was discovered in 1887 by Heinrich Hertz (1857-1894), during his electromagnetic wave experiments. In his experimental investigation on the production of electromagnetic waves by means of a spark discharge, Hertz observed that high voltage sparks across the detector loop were enhanced when the emitter plate was illuminated by ultraviolet light from an arc lamp.

Light shining on the metal surface somehow facilitated the escape of free, charged particles which we now know as electrons. When light falls on a metal surface, some electrons near the surface absorb enough energy from the incident radiation to overcome the attraction of the positive ions in the material of the surface. After gaining sufficient energy from the incident light, the electrons escape from the surface of the metal into the surrounding space.

11.3.2 Hallwachs' and Lenard's observations

Wilhelm Hallwachs and Philipp Lenard investigated the phenomenon of photoelectric emission in detail during 1886-1902.

Lenard (1862-1947) observed that when ultraviolet radiations were allowed to fall on the emitter plate of an evacuated glass tube enclosing two electrodes (metal plates), current flows in the circuit (Fig. 11.1). As soon as the ultraviolet radiations were stopped, the current flow also stopped. These observations indicate that when ultraviolet radiations fall on the emitter plate C, electrons are ejected from it which are attracted towards the positive, collector plate A by the electric field. The electrons flow through the evacuated glass tube, resulting in the current flow. Thus, light falling on the surface of the emitter causes current in the external circuit. Hallwachs and Lenard studied how this photo current varied with collector plate potential, and with frequency and intensity of incident light.

Hallwachs, in 1888, undertook the study further and connected a negatively charged zinc plate to an electroscope. He observed that the zinc plate lost its charge when it was illuminated by ultraviolet light. Further, the uncharged zinc plate became positively charged when it was irradiated by ultraviolet light. Positive charge on a positively charged zinc plate was found to be further enhanced when it was illuminated by ultraviolet light. From these observations he concluded that negatively charged particles were emitted from the zinc plate under the action of ultraviolet light.

After the discovery of the electron in 1897, it became evident that the incident light causes electrons to be emitted from the emitter plate. Due to negative charge, the emitted electrons are pushed towards the collector plate by the electric field. Hallwachs and Lenard also observed that when ultraviolet light fell on the emitter plate, no electrons were emitted at all when the frequency of the incident light was smaller than a certain minimum value, called the *threshold frequency*. This minimum frequency depends on the nature of the material of the emitter plate.

It was found that certain metals like zinc, cadmium, magnesium, etc., responded only to ultraviolet light, having short wavelength, to cause electron emission from the surface. However, some alkali metals such as lithium, sodium, potassium, caesium and rubidium were sensitive even to visible light. All these *photosensitive substances* emit electrons when they are illuminated by light. After the discovery of electrons, these electrons were termed as *photoelectrons*. The phenomenon is called *photoelectric effect*.

11.4 Experimental Study of Photoelectric Effect

Figure 11.1 depicts a schematic view of the arrangement used for the experimental study of the photoelectric effect. It consists of an evacuated glass/quartz tube having a thin photosensitive plate C and another metal plate A. Monochromatic light from the source S of sufficiently short wavelength passes through the window W and falls on the photosensitive plate C (emitter). A transparent quartz window is sealed on to the glass tube, which permits ultraviolet radiation to pass through it and irradiate the photosensitive plate C. The electrons are emitted by the plate C and are collected by the plate A (collector), by the electric field created by the battery. The battery maintains the potential difference between the plates C and A, that can be varied. The polarity of the plates C and A can be reversed by a commutator. Thus, the plate A can be maintained at a desired positive or negative potential with respect to emitter C. When the collector plate A is positive with respect to the emitter plate C, the electrons are attracted to it. The emission of electrons causes flow of electric current in the circuit. The potential difference between the emitter and collector plates is measured by a voltmeter (V) whereas the resulting photo current flowing in the circuit is measured by a microammeter (μA). The photoelectric current can be increased or decreased by varying the potential of collector plate A with respect to the emitter plate C. The intensity and frequency of the incident light can be varied, as can the potential difference V between the emitter C and the collector A.

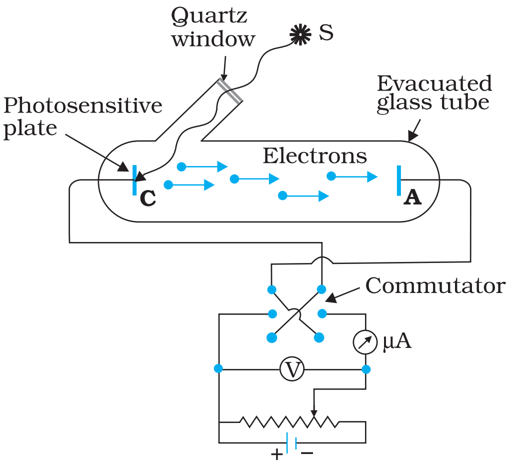


FIGURE 11.1 Experimental arrangement for study of photoelectric effect.

We can use the experimental arrangement of Fig. 11.1 to study the variation of photocurrent with (a) intensity of radiation, (b) frequency of incident radiation, (c) the potential difference between the plates A and C, and (d) the nature of the material of plate C. Light of different frequencies can be used by putting appropriate coloured filter or coloured glass in the path of light falling

Reprint 2025-26

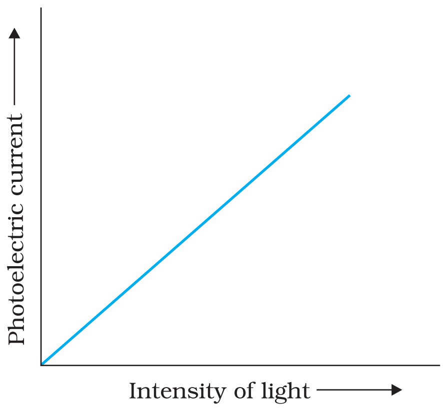


FIGURE 11.2 Variation of Photoelectric current with intensity of light.

on the emitter C. The intensity of light is varied by changing the distance of the light source from the emitter.

11.4.1 Effect of intensity of light on photocurrent

The collector A is maintained at a positive potential with respect to emitter C so that electrons ejected from C are attracted towards collector A. Keeping the frequency of the incident radiation and the potential fixed, the intensity of light is varied and the resulting photoelectric current is measured each time. It is found that the photocurrent increases linearly with intensity of incident light as shown graphically in Fig. 11.2. The photocurrent is directly proportional to the number of photoelectrons emitted per second. This implies that *the number of photoelectrons emitted per second is directly proportional to the intensity of incident radiation.*

11.4.2 Effect of potential on photoelectric current

We first keep the plate A at some positive potential with respect to the plate C and illuminate the plate C with light of fixed frequency ν and fixed intensity I_1. We next vary the positive potential of plate A gradually and measure the resulting photocurrent each time. It is found that the photoelectric current increases with increase in positive (accelerating) potential. At some stage, for a certain positive potential of plate A, all the emitted electrons are collected by the plate A and the photoelectric current becomes maximum or saturates. If we increase the accelerating potential of plate A further, the photocurrent does not increase. This maximum value of the photoelectric current is called *saturation current.* Saturation current corresponds to the case when all the photoelectrons emitted by the emitter plate C reach the collector plate A.

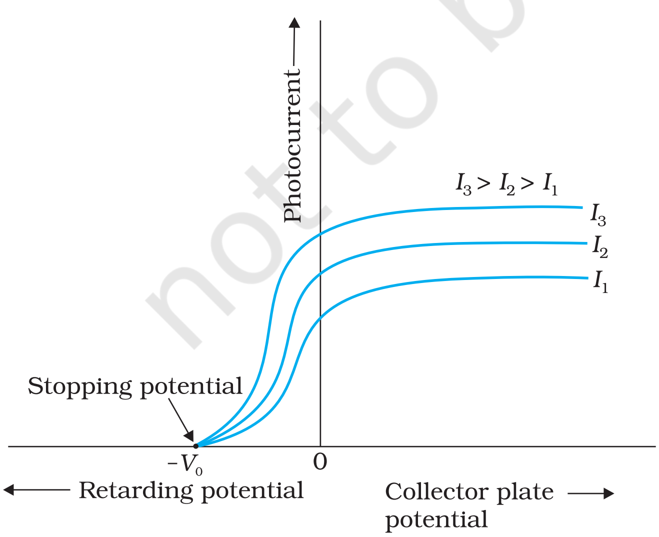


FIGURE 11.3 Variation of photocurrent with collector plate potential for different intensity of incident radiation.

We now apply a negative (retarding) potential to the plate A with respect to the plate C and make it increasingly negative gradually. When the polarity is reversed, the electrons are repelled and only the sufficiently energetic electrons are able to reach the collector A. The photocurrent is found to decrease rapidly until it drops to zero at a certain sharply defined, critical value of the negative potential V_0 on the plate A. For a particular frequency of incident radiation, *the minimum negative (retarding) potential V_0 given to the plate A for which the photocurrent stops or becomes zero is called the cut-off or stopping potential.*

The interpretation of the observation in terms of photoelectrons is straightforward. All the photoelectrons emitted from the metal do not have the same energy. Photoelectric current is zero when the stopping potential is sufficient to repel even the most energetic photoelectrons, with the maximum kinetic energy (K_{max}), so that

$$K_{max} = e\,V_0 \qquad (11.1)$$

We can now repeat this experiment with incident radiation of the same frequency but of higher intensity I_2 and I_3 ($I_3 > I_2 > I_1$). We note that the saturation currents are now found to be at higher values. This shows that more electrons are being emitted per second, proportional to the intensity of incident radiation. But the stopping potential remains the same as that for the incident radiation of intensity I_1, as shown graphically in Fig. 11.3. Thus, *for a given frequency of the incident radiation, the stopping potential is independent of its intensity*. In other words, the maximum kinetic energy of photoelectrons depends on the light source and the emitter plate material, but is independent of intensity of incident radiation.

11.4.3 Effect of frequency of incident radiation on stopping potential

We now study the relation between the frequency ν of the incident radiation and the stopping potential V_0. We suitably adjust the same intensity of light radiation at various frequencies and study the variation of photocurrent with collector plate potential. The resulting variation is shown in Fig. 11.4. We obtain different values of stopping potential but the same value of the saturation current for incident radiation of different frequencies. The energy of the emitted electrons depends on the frequency of the incident radiations. The stopping potential is more negative for higher frequencies of incident radiation. Note from Fig. 11.4 that the stopping potentials are in the order $V_{03} > V_{02} > V_{01}$ if the frequencies are in the order $\nu_3 > \nu_2 > \nu_1$. This implies that greater the frequency of incident light, greater is the maximum kinetic energy of the photoelectrons. Consequently, we need greater retarding potential to stop them completely. If we plot a graph between the frequency of incident radiation and the corresponding stopping potential for different metals we get a straight line, as shown in Fig. 11.5.

The graph shows that

(i) the stopping potential V_0 varies linearly with the frequency of incident radiation for a given photosensitive material.

(ii) there exists a certain minimum cut-off frequency ν_0 for which the stopping potential is zero.

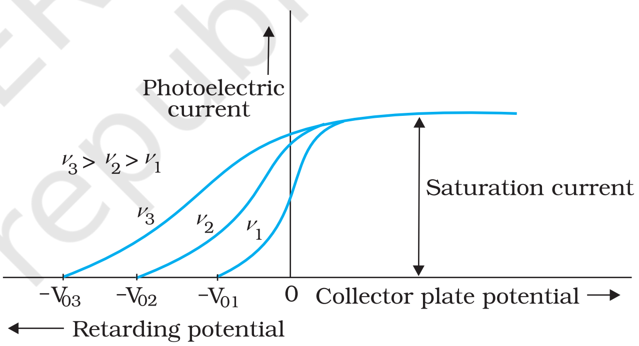


FIGURE 11.4 Variation of photoelectric current with collector plate potential for different frequencies of incident radiation.

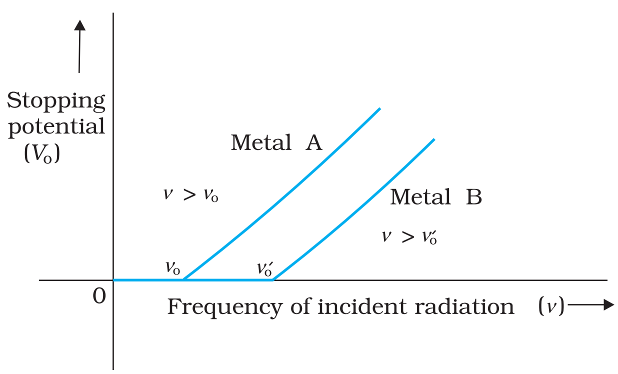


FIGURE 11.5 Variation of stopping potential V_0 with frequency ν of incident radiation for a given photosensitive material.

These observations have two implications:

(i) *The maximum kinetic energy of the photoelectrons varies linearly with the frequency of incident radiation, but is independent of its intensity.*

(ii) *For a frequency* ν *of incident radiation, lower than the cut-off frequency* ν_0, *no photoelectric emission is possible even if the intensity is large.*

This minimum, cut-off frequency ν_0, is called the *threshold frequency*. It is different for different metals.

Different photosensitive materials respond differently to light. Selenium is more sensitive than zinc or copper. The same photosensitive substance gives different response to light of different wavelengths. For example, ultraviolet light gives rise to photoelectric effect in copper while green or red light does not.

Note that in all the above experiments, it is found that, if frequency of the incident radiation exceeds the threshold frequency, the photoelectric emission starts instantaneously without any apparent time lag, even if the incident radiation is very dim. It is now known that emission starts in a time of the order of 10^{-9} s or less.

We now summarise the experimental features and observations described in this section.

(i) For a given photosensitive material and frequency of incident radiation (above the threshold frequency), the photoelectric current is directly proportional to the intensity of incident light (Fig. 11.2).

(ii) For a given photosensitive material and frequency of incident radiation, saturation current is found to be proportional to the intensity of incident radiation whereas the stopping potential is independent of its intensity (Fig. 11.3).

(iii) For a given photosensitive material, there exists a certain minimum cut-off frequency of the incident radiation, called the *threshold frequency*, below which no emission of photoelectrons takes place, no matter how intense the incident light is. Above the threshold frequency, the stopping potential or equivalently the maximum kinetic energy of the emitted photoelectrons increases linearly with the frequency of the incident radiation, but is independent of its intensity (Fig. 11.5).

(iv) The photoelectric emission is an instantaneous process without any apparent time lag ($\sim 10^{-9}$s or less), even when the incident radiation is made exceedingly dim.

11.5 Photoelectric Effect and Wave Theory of Light

The wave nature of light was well established by the end of the nineteenth century. The phenomena of interference, diffraction and polarisation were explained in a natural and satisfactory way by the wave picture of light. According to this picture, light is an electromagnetic wave consisting of electric and magnetic fields with continuous distribution of energy over the region of space over which the wave is extended. Let us now see if this wave picture of light can explain the observations on photoelectric emission given in the previous section.

According to the wave picture of light, the free electrons at the surface of the metal (over which the beam of radiation falls) absorb the radiant energy continuously. The greater the intensity of radiation, the greater are the amplitude of electric and magnetic fields. Consequently, the greater the intensity, the greater should be the energy absorbed by each electron. In this picture, the maximum kinetic energy of the photoelectrons on the surface is then expected to increase with increase in intensity. Also, no matter what the frequency of radiation is, a sufficiently intense beam of radiation (over sufficient time) should be able to impart enough energy to the electrons, so that they exceed the minimum energy needed to escape from the metal surface . A threshold frequency, therefore, should not exist. These expectations of the wave theory directly contradict observations (i), (ii) and (iii) given at the end of sub-section 11.4.3.

Further, we should note that in the wave picture, the absorption of energy by electron takes place continuously over the entire wavefront of the radiation. Since a large number of electrons absorb energy, the energy absorbed per electron per unit time turns out to be small. Explicit calculations estimate that it can take hours or more for a single electron to pick up sufficient energy to overcome the work function and come out of the metal. This conclusion is again in striking contrast to observation (iv) that the photoelectric emission is instantaneous. In short, the wave picture is unable to explain the most basic features of photoelectric emission.

11.6 Einstein's Photoelectric Equation: Energy Quantum of Radiation

In 1905, Albert Einstein (1879-1955) proposed a radically new picture of electromagnetic radiation to explain photoelectric effect. In this picture, photoelectric emission does not take place by continuous absorption of energy from radiation. Radiation energy is built up of discrete units – the so called *quanta of energy of radiation*. Each quantum of radiant energy has energy $h\nu$, where h is Planck's constant and ν the frequency of light. In photoelectric effect, an electron absorbs a quantum of energy ($h\nu$) of radiation. If this quantum of energy absorbed exceeds the minimum energy needed for the electron to escape from the metal surface (work function ϕ_0), the electron is emitted with maximum kinetic energy

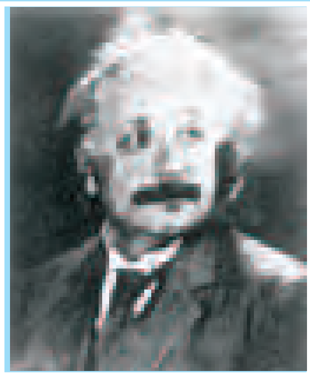
$$K_{\max} = h\nu - \phi_0 \tag{11.2}$$

More tightly bound electrons will emerge with kinetic energies less than the maximum value. Note that the intensity of light of a given frequency is determined by the number of photons incident per second. Increasing the intensity will increase the number of emitted electrons per second. However, the maximum kinetic energy of the emitted photoelectrons is determined by the energy of each photon.

Equation (11.2) is known as *Einstein's photoelectric equation*. We now see how this equation accounts in a simple and elegant manner all the observations on photoelectric effect given at the end of sub-section 11.4.3.

Reprint 2025-26

ALBERT EINSTEIN (1879 – 1955)



Albert Einstein (1879 – 1955) Einstein, one of the greatest physicists of all time, was born in Ulm, Germany. In 1905, he published three path-breaking papers. In the first paper, he introduced the notion of light quanta (now called photons) and used it to explain the features of photoelectric effect. In the second paper, he developed a theory of Brownian motion, confirmed experimentally a few years later and provided a convincing evidence of the atomic picture of matter. The third paper gave birth to the special theory of relativity. In 1916, he published the general theory of relativity. Some of Einstein's most significant later contributions are: the notion of stimulated emission introduced in an alternative derivation of Planck's blackbody radiation law, static model of the universe which started modern cosmology, quantum statistics of a gas of massive bosons, and a critical analysis of the foundations of quantum mechanics. In 1921, he was awarded the Nobel Prize in physics for his contribution to theoretical physics and the photoelectric effect.

- According to Eq. (11.2), K_{max} depends linearly on ν, and is independent of intensity of radiation, in agreement with observation. This has happened because in Einstein's picture, photoelectric effect arises from the absorption of a single quantum of radiation by a single electron. The intensity of radiation (that is proportional to the number of energy quanta per unit area per unit time) is irrelevant to this basic process.
- Since K_{max} must be non-negative, Eq. (11.2) implies that photoelectric emission is possible only if
  $h\nu > \phi_0$
  or $\nu > \nu_0$, where
  $$\nu_0 = \frac{\phi_0}{h} \tag{11.3}$$
  Equation (11.3) shows that the greater the work function ϕ_0, the higher the minimum or threshold frequency ν_0 needed to emit photoelectrons. Thus, there exists a threshold frequency ν_0 (= ϕ_0/h) for the metal surface, below which no photoelectric emission is possible, no matter how intense the incident radiation may be or how long it falls on the surface.
- In this picture, intensity of radiation as noted above, is proportional to the number of energy quanta per unit area per unit time. The greater the number of energy quanta available, the greater is the number of electrons absorbing the energy quanta and greater, therefore, is the number of electrons coming out of the metal (for $\nu > \nu_0$). This explains why, for $\nu > \nu_0$, photoelectric current is proportional to intensity.
- In Einstein's picture, the basic elementary process involved in photoelectric effect is the absorption of a light quantum by an electron. This process is instantaneous. Thus, whatever may be the intensity i.e., the number of quanta of radiation per unit area per unit time, photoelectric emission is instantaneous. Low intensity does not mean delay in emission, since the basic elementary process is the same. Intensity only determines how many electrons are able to participate in the elementary process (absorption of a light quantum by a single electron) and, therefore, the photoelectric current.

Using Eq. (11.1), the photoelectric equation, Eq. (11.2), can be written as

$$e V_0 = h\nu - \phi_0; \text{ for } \nu \geq \nu_0$$

$$\text{or } V_0 = \frac{h}{e}\nu - \frac{\phi_0}{e} \tag{11.4}$$

This is an important result. It predicts that the V_0 versus ν curve is a straight line with slope = (h/e), independent of the nature of the material. During 1906-1916, Millikan performed a series of experiments on photoelectric effect, aimed at disproving Einstein's photoelectric equation. He measured the slope of the straight line obtained for sodium, similar to that shown in Fig. 11.5. Using the known value of e, he determined the value of Planck's constant h. This value was close to the value of Planck's contant (= 6.626×10^{-34}J s) determined in an entirely different context. In this way, in 1916, Millikan proved the validity of Einstein's photoelectric equation, instead of disproving it.

The successful explanation of photoelectric effect using the hypothesis of light quanta and the experimental determination of values of h and ϕ_0, in agreement with values obtained from other experiments, led to the acceptance of Einstein's picture of photoelectric effect. Millikan verified photoelectric equation with great precision, for a number of alkali metals over a wide range of radiation frequencies.

11.7 Particle Nature of Light: The Photon

Photoelectric effect thus gave evidence to the strange fact that light in interaction with matter behaved as if it was made of quanta or packets of energy, each of energy $h\nu$.

Is the light quantum of energy to be associated with a particle? Einstein arrived at the important result, that the light quantum can also be associated with momentum ($h\nu/c$). A definite value of energy as well as momentum is a strong sign that the light quantum can be associated with a particle. This particle was later named *photon*. The particle-like behaviour of light was further confirmed, in 1924, by the experiment of A.H. Compton (1892-1962) on scattering of X-rays from electrons. In 1921, Einstein was awarded the Nobel Prize in Physics for his contribution to theoretical physics and the photoelectric effect. In 1923, Millikan was awarded the Nobel Prize in physics for his work on the elementary charge of electricity and on the photoelectric effect.

We can summarise the photon picture of electromagnetic radiation as follows:

(i) In interaction of radiation with matter, radiation behaves as if it is made up of particles called photons.

(ii) Each photon has energy E (=$h\nu$) and momentum p (= $h\nu/c$), and speed c, the speed of light.

(iii) All photons of light of a particular frequency ν, or wavelength λ, have the same energy E (=$h\nu = hc/\lambda$) and momentum p (= $h\nu/c = h/\lambda$), whatever the intensity of radiation may be. By increasing the intensity of light of given wavelength, there is only an increase in the number of photons per second crossing a given area, with each photon having the same energy. Thus, photon energy is independent of intensity of radiation.

(iv) Photons are electrically neutral and are not deflected by electric and magnetic fields.

(v) In a photon-particle collision (such as photon-electron collision), the total energy and total momentum are conserved. However, the number of photons may not be conserved in a collision. The photon may be absorbed or a new photon may be created.

Reprint 2025-26

EXAMPLE 11.1

Example 11.1 Monochromatic light of frequency 6.0×10^{14} Hz is produced by a laser. The power emitted is 2.0×10^{-3} W. (a) What is the energy of a photon in the light beam? (b) How many photons per second, on an average, are emitted by the source?

Solution

(a) Each photon has an energy

$E = h\nu = (6.63 \times 10^{-34}$ J s) $(6.0 \times 10^{14}$ Hz)

$= 3.98 \times 10^{-19}$ J

(b) If N is the number of photons emitted by the source per second, the power P transmitted in the beam equals N times the energy per photon E, so that $P = N E$. Then

$$N = \frac{P}{E} = \frac{2.0 \times 10^{-3}\,\text{W}}{3.98 \times 10^{-19}\,\text{J}}$$

$= 5.0 \times 10^{15}$ photons per second.

EXAMPLE 11.2

Example 11.2 The work function of caesium is 2.14 eV. Find (a) the threshold frequency for caesium, and (b) the wavelength of the incident light if the photocurrent is brought to zero by a stopping potential of 0.60 V.

Solution

(a) For the cut-off or threshold frequency, the energy $h\nu_0$ of the incident radiation must be equal to work function ϕ_0, so that

$$\nu_0 = \frac{\phi_0}{h} = \frac{2.14\,\text{eV}}{6.63 \times 10^{-34}\,\text{J s}}$$

$$= \frac{2.14 \times 1.6 \times 10^{-19}\,\text{J}}{6.63 \times 10^{-34}\,\text{J s}} = 5.16 \times 10^{14}\ \text{Hz}$$

Thus, for frequencies less than this threshold frequency, no photoelectrons are ejected.

(b) Photocurrent reduces to zero, when maximum kinetic energy of the emitted photoelectrons equals the potential energy $e V_0$ by the retarding potential V_0. Einstein's Photoelectric equation is

$$eV_0 = h\nu - \phi_0 = \frac{hc}{\lambda} - \phi_0$$

or, $\lambda = hc/(eV_0 + \phi_0)$

$$= \frac{(6.63 \times 10^{-34}\,\text{J s}) \times (3 \times 10^{8}\,\text{m/s})}{(0.60\,\text{eV} + 2.14\,\text{eV})}$$

$$= \frac{19.89 \times 10^{-26}\,\text{J m}}{(2.74\,\text{eV})}$$

$$\lambda = \frac{19.89 \times 10^{-26}\,\text{J m}}{2.74 \times 1.6 \times 10^{-19}\,\text{J}} = 454\ \text{nm}$$

11.8 WAVE NATURE OF MATTER

The dual (wave-particle) nature of light (electromagnetic radiation, in general) comes out clearly from what we have learnt in this and the preceding chapters. The wave nature of light shows up in the phenomena of interference, diffraction and polarisation. On the other hand, in photoelectric effect and Compton effect which involve energy and momentum transfer, radiation behaves as if it is made up of a bunch of particles – the photons. Whether a particle or wave description is best suited for understanding an experiment depends on the nature of the experiment. For example, in the familiar phenomenon of seeing an object by our eye, both descriptions are important. The gathering and focussing mechanism of light by the eye-lens is well described in the wave picture. But its absorption by the rods and cones (of the retina) requires the photon picture of light.

A natural question arises: If radiation has a dual (wave-particle) nature, might not the particles of nature (the electrons, protons, etc.) also exhibit wave-like character? In 1924, the French physicist Louis Victor de Broglie (pronounced as de Broy) (1892-1987) put forward the bold hypothesis that moving particles of matter should display wave-like properties under suitable conditions. He reasoned that nature was symmetrical and that the two basic physical entities – matter and energy, must have symmetrical character. If radiation shows dual aspects, so should matter. De Broglie proposed that the wave length λ associated with a particle of momentum p is given as

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{11.5}$$

LOUIS VICTOR DE BROGLIE (1892 – 1987)

Louis Victor de Broglie (1892 – 1987) French physicist who put forth revolutionary idea of wave nature of matter. This idea was developed by Erwin Schródinger into a full-fledged theory of quantum mechanics commonly known as wave mechanics. In 1929, he was awarded the Nobel Prize in Physics for his discovery of the wave nature of electrons.

where m is the mass of the particle and v its speed. Equation (11.5) is known as the *de Broglie relation* and the wavelength λ of the *matter wave* is called *de Broglie wavelength*. The dual aspect of matter is evident in the de Broglie relation. On the left hand side of Eq. (11.5), λ is the attribute of a wave while on the right hand side the momentum p is a typical attribute of a particle. Planck's constant h relates the two attributes.

Equation (11.5) for a material particle is basically a hypothesis whose validity can be tested only by experiment. However, it is interesting to see that it is satisfied also by a photon. For a photon, as we have seen,

$$p = h\nu / c \tag{11.6}$$

Therefore,

$$\frac{h}{p} = \frac{c}{\nu} = \lambda \tag{11.7}$$

That is, the de Broglie wavelength of a photon given by Eq. (11.5) equals the wavelength of electromagnetic radiation of which the photon is a quantum of energy and momentum.

Clearly, from Eq. (11.5), λ is smaller for a heavier particle (large m) or more energetic particle (large v). For example, the de Broglie wavelength of a ball of mass 0.12 kg moving with a speed of 20 m s^{-1} is easily calculated:

$p = m\,v = 0.12 \text{ kg} \times 20 \text{ m s}^{-1} = 2.40 \text{ kg m s}^{-1}$

$$\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34}\text{ J s}}{2.40 \text{ kg m s}^{-1}} = 2.76 \times 10^{-34} \text{ m}$$

This wavelength is so small that it is beyond any measurement. This is the reason why macroscopic objects in our daily life do not show wave-like properties. On the other hand, in the sub-atomic domain, the wave character of particles is significant and measurable.

Example 11.3 What is the de Broglie wavelength associated with (a) an electron moving with a speed of 5.4×10^6 m/s, and (b) a ball of mass 150 g travelling at 30.0 m/s?

Solution

(a) For the electron:
Mass $m = 9.11 \times 10^{-31}$ kg, speed $v = 5.4 \times 10^6$ m/s. Then, momentum
$p = m\,v = 9.11 \times 10^{-31}$ (kg) $\times$ 5.4 $\times$ 10^6 (m/s)
$p = 4.92 \times 10^{-24}$ kg m/s
de Broglie wavelength, $\lambda = h/p$

$$= \frac{6.63 \times 10^{-34}\text{ J s}}{4.92 \times 10^{-24}\text{ kg m/s}}$$

$\lambda = 0.135$ nm

(b) For the ball:
Mass $m' = 0.150$ kg, speed $v' = 30.0$ m/s.
Then momentum $p' = m'\,v' = 0.150$ (kg) $\times$ 30.0 (m/s)
$p' = 4.50$ kg m/s
de Broglie wavelength $\lambda' = h/p'$.

$$= \frac{6.63 \times 10^{-34}\text{ J s}}{4.50 \times \text{ kg m/s}}$$

$\lambda' = 1.47 \times 10^{-34}$ m

The de Broglie wavelength of electron is comparable with X-ray wavelengths. However, for the ball it is about 10^{-19} times the size of the proton, quite beyond experimental measurement.

EXAMPLE 11.3

SUMMARY

1. The minimum energy needed by an electron to come out from a metal surface is called the work function of the metal. Energy (greater than the work function (ϕ_o) required for electron emission from the metal surface can be supplied by suitably heating or applying strong electric field or irradiating it by light of suitable frequency.
2. Photoelectric effect is the phenomenon of emission of electrons by metals when illuminated by light of suitable frequency. Certain metals respond to ultraviolet light while others are sensitive even to the visible light. Photoelectric effect involves conversion of light energy into electrical energy. It follows the law of conservation of energy. The photoelectric emission is an instantaneous process and possesses certain special features.

3. Photoelectric current depends on (i) the intensity of incident light, (ii) the potential difference applied between the two electrodes, and (iii) the nature of the emitter material.
4. The stopping potential (V_o) depends on (i) the frequency of incident light, and (ii) the nature of the emitter material. For a given frequency of incident light, it is independent of its intensity. The stopping potential is directly related to the maximum kinetic energy of electrons emitted: $e\,V_0 = (1/2)\,m\,v^2_{max} = K_{max}$.
5. Below a certain frequency (threshold frequency) ν_0, characteristic of the metal, no photoelectric emission takes place, no matter how large the intensity may be.
6. The classical wave theory could not explain the main features of photoelectric effect. Its picture of continuous absorption of energy from radiation could not explain the independence of K_{max} on intensity, the existence of ν_o and the instantaneous nature of the process. Einstein explained these features on the basis of photon picture of light. According to this, light is composed of discrete packets of energy called quanta or photons. Each photon carries an energy E (= $h\,\nu$) and momentum p (= h/λ), which depend on the frequency (ν) of incident light and not on its intensity. Photoelectric emission from the metal surface occurs due to absorption of a photon by an electron.
7. Einstein's photoelectric equation is in accordance with the energy conservation law as applied to the photon absorption by an electron in the metal. The maximum kinetic energy $(1/2)m\,v^2_{max}$ is equal to the photon energy ($h\nu$) minus the work function ϕ_0 (= $h\nu_0$) of the target metal:

$$\frac{1}{2}\,m\,v^2_{max} = V_0\,e = h\nu - \phi_0 = h\,(\nu - \nu_0)$$

   This photoelectric equation explains all the features of the photoelectric effect. Millikan's first precise measurements confirmed the Einstein's photoelectric equation and obtained an accurate value of Planck's constant h. This led to the acceptance of particle or photon description (nature) of electromagnetic radiation, introduced by Einstein.
8. Radiation has dual nature: wave and particle. The nature of experiment determines whether a wave or particle description is best suited for understanding the experimental result. Reasoning that radiation and matter should be symmetrical in nature, Louis Victor de Broglie attributed a wave-like character to matter (material particles). The waves associated with the moving material particles are called matter waves or de Broglie waves.
9. The de Broglie wavelength (λ) associated with a moving particle is related to its momentum p as: $\lambda = h/p$. The dualism of matter is inherent in the de Broglie relation which contains a wave concept (λ) and a particle concept (p). The de Broglie wavelength is independent of the charge and nature of the material particle. It is significantly measurable (of the order of the atomic-planes spacing in crystals) only in case of sub-atomic particles like electrons, protons, etc. (due to smallness of their masses and hence, momenta). However, it is indeed very small, quite beyond measurement, in case of macroscopic objects, commonly encountered in everyday life.

Reprint 2025-26

Physical Quantity	Symbol	Dimensions	Unit	Remarks
Planck's constant	h	$[ML^2T^{-1}]$	J s	$E = h\nu$
Stopping potential	V_0	$[ML^2T^{-3}A^{-1}]$	V	$e\,V_0 = K_{max}$
Work function	ϕ_0	$[ML^2T^{-2}]$	J; eV	$K_{max} = E - \phi_0$
Threshold frequency	ν_0	$[T^{-1}]$	Hz	$\nu_0 = \phi_0 / h$
de Broglie wavelength	λ	[L]	m	$\lambda = h/p$

POINTS TO PONDER

1. Free electrons in a metal are free in the sense that they move inside the metal in a constant potential (This is only an approximation). They are not free to move out of the metal. They need additional energy to get out of the metal.
2. Free electrons in a metal do not all have the same energy. Like molecules in a gas jar, the electrons have a certain energy distribution at a given temperature. This distribution is different from the usual Maxwell's distribution that you have learnt in the study of kinetic theory of gases. You will learn about it in later courses, but the difference has to do with the fact that electrons obey Pauli's exclusion principle.
3. Because of the energy distribution of free electrons in a metal, the energy required by an electron to come out of the metal is different for different electrons. Electrons with higher energy require less additional energy to come out of the metal than those with lower energies. Work function is the least energy required by an electron to come out of the metal.
4. Observations on photoelectric effect imply that in the event of matter-light interaction, *absorption of energy takes place in discrete units of hν*. This is not quite the same as saying that light consists of particles, each of energy $h\nu$.
5. Observations on the stopping potential (its independence of intensity and dependence on frequency) are the crucial discriminator between the wave-picture and photon-picture of photoelectric effect.
6. The wavelength of a matter wave given by $\lambda = \dfrac{h}{p}$ has physical significance; its phase velocity v_p has no physical significance. However, the group velocity of the matter wave is physically meaningful and equals the velocity of the particle.

EXERCISES

11.1 Find the
(a) maximum frequency, and
(b) minimum wavelength of X-rays produced by 30 kV electrons.

11.2 The work function of caesium metal is 2.14 eV. When light of frequency 6×10^{14}Hz is incident on the metal surface, photoemission of electrons occurs. What is the

(a) maximum kinetic energy of the emitted electrons,

(b) Stopping potential, and

(c) maximum speed of the emitted photoelectrons?

11.3 The photoelectric cut-off voltage in a certain experiment is 1.5 V. What is the maximum kinetic energy of photoelectrons emitted?

11.4 Monochromatic light of wavelength 632.8 nm is produced by a helium-neon laser. The power emitted is 9.42 mW.

(a) Find the energy and momentum of each photon in the light beam,

(b) How many photons per second, on the average, arrive at a target irradiated by this beam? (Assume the beam to have uniform cross-section which is less than the target area), and

(c) How fast does a hydrogen atom have to travel in order to have the same momentum as that of the photon?

11.5 In an experiment on photoelectric effect, the slope of the cut-off voltage versus frequency of incident light is found to be 4.12×10^{-15} V s. Calculate the value of Planck's constant.

11.6 The threshold frequency for a certain metal is 3.3×10^{14} Hz. If light of frequency 8.2×10^{14} Hz is incident on the metal, predict the cut-off voltage for the photoelectric emission.

11.7 The work function for a certain metal is 4.2 eV. Will this metal give hotoelectric emission for incident radiation of wavelength 330 nm?

11.8 Light of frequency 7.21×10^{14} Hz is incident on a metal surface. Electrons with a maximum speed of 6.0×10^{5} m/s are ejected from the surface. What is the threshold frequency for photoemission of electrons?

11.9 Light of wavelength 488 nm is produced by an argon laser which is used in the photoelectric effect. When light from this spectral line is incident on the emitter, the stopping (cut-off) potential of photoelectrons is 0.38 V. Find the work function of the material from which the emitter is made.

11.10 What is the de Broglie wavelength of

(a) a bullet of mass 0.040 kg travelling at the speed of 1.0 km/s,

(b) a ball of mass 0.060 kg moving at a speed of 1.0 m/s, and

(c) a dust particle of mass 1.0×10^{-9} kg drifting with a speed of 2.2 m/s?

11.11 Show that the wavelength of electromagnetic radiation is equal to the de Broglie wavelength of its quantum (photon).

Reprint 2025-26

Chapter Twelve

ATOMS

12.1 Introduction

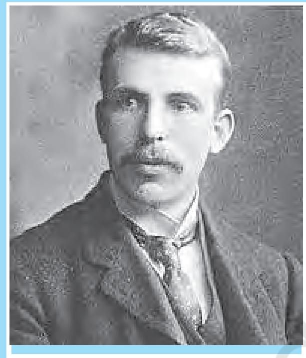
By the nineteenth century, enough evidence had accumulated in favour of atomic hypothesis of matter. In 1897, the experiments on electric discharge through gases carried out by the English physicist J. J. Thomson (1856 – 1940) revealed that atoms of different elements contain negatively charged constituents (electrons) that are identical for all atoms. However, atoms on a whole are electrically neutral. Therefore, an atom must also contain some positive charge to neutralise the negative charge of the electrons. But what is the arrangement of the positive charge and the electrons inside the atom? In other words, what is the structure of an atom?

The first model of atom was proposed by J. J. Thomson in 1898. According to this model, the positive charge of the atom is uniformly distributed throughout the volume of the atom and the negatively charged electrons are embedded in it like seeds in a watermelon. This model was picturesquely called *plum pudding model* of the atom. However subsequent studies on atoms, as described in this chapter, showed that the distribution of the electrons and positive charges are very different from that proposed in this model.

We know that condensed matter (solids and liquids) and dense gases at all temperatures emit electromagnetic radiation in which a continuous distribution of several wavelengths is present, though with different intensities. This radiation is considered to be due to oscillations of atoms and molecules, governed by the interaction of each atom or molecule with its neighbours. *In contrast*, light emitted from rarefied gases heated in a flame, or excited electrically in a glow tube such as the familiar neon sign or mercury vapour light has only certain discrete wavelengths. The spectrum appears as a series of bright lines. In such gases, the average spacing between atoms is large. Hence, the radiation emitted can be considered due to individual atoms rather than because of interactions between atoms or molecules.

In the early nineteenth century it was also established that each element is associated with a characteristic spectrum of radiation, for example, hydrogen always gives a set of lines with fixed relative position between the lines. This fact suggested an intimate relationship between the internal structure of an atom and the spectrum of radiation emitted by it. In 1885, Johann Jakob Balmer (1825 – 1898) obtained a simple empirical formula which gave the wavelengths of a group of lines emitted by atomic hydrogen. Since hydrogen is simplest of the elements known, we shall consider its spectrum in detail in this chapter.

Ernst Rutherford (1871–1937), a former research student of J. J. Thomson, was engaged in experiments on α-particles emitted by some radioactive elements. In 1906, he proposed a classic experiment of scattering of these α-particles by atoms to investigate the atomic structure. This experiment was later performed around 1911 by Hans Geiger (1882–1945) and Ernst Marsden (1889–1970, who was 20 year-old student and had not yet earned his bachelor's degree). The details are discussed in Section 12.2. The explanation of the results led to the birth of Rutherford's planetary model of atom (also called the *nuclear model of the atom*). According to this the entire positive charge and most of the mass of the atom is concentrated in a small volume called the nucleus with electrons revolving around the nucleus just as planets revolve around the sun.



ERNST RUTHERFORD (1871 – 1937)

Ernst Rutherford (1871 – 1937) New Zealand born, British physicist who did pioneering work on radioactive radiation. He discovered alpha-rays and beta-rays. Along with Federick Soddy, he created the modern theory of radioactivity. He studied the 'emanation' of thorium and discovered a new noble gas, an isotope of radon, now known as thoron. By scattering alpha-rays from the metal foils, he discovered the atomic nucleus and proposed the planetary model of the atom. He also estimated the approximate size of the nucleus.

Rutherford's nuclear model was a major step towards how we see the atom today. However, it could not explain why atoms emit light of only discrete wavelengths. How could an atom as simple as hydrogen, consisting of a single electron and a single proton, emit a complex spectrum of specific wavelengths? In the classical picture of an atom, the electron revolves round the nucleus much like the way a planet revolves round the sun. However, we shall see that there are some serious difficulties in accepting such a model.

12.2 Alpha-particle Scattering and Rutherford's Nuclear Model of Atom

At the suggestion of Ernst Rutherford, in 1911, H. Geiger and E. Marsden performed some experiments. In one of their experiments, as shown in

Reprint 2025-26

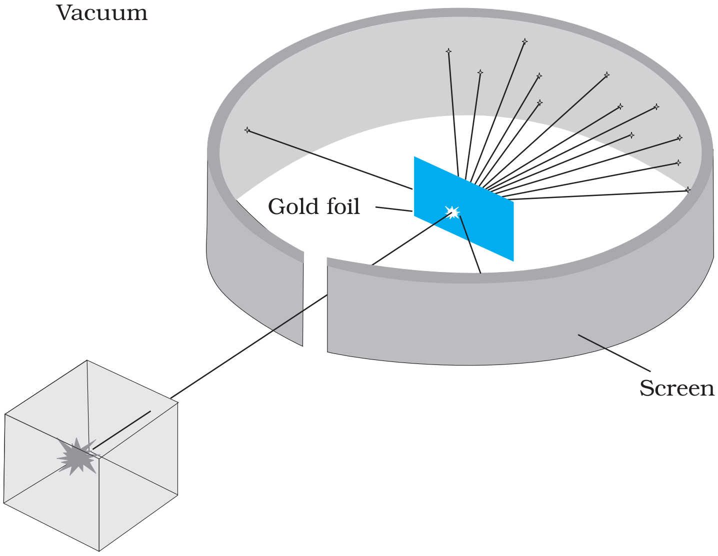


FIGURE 12.1 Geiger-Marsden scattering experiment. The entire apparatus is placed in a vacuum chamber (not shown in this figure).

Fig. 12.1, they directed a beam of 5.5 MeV α-particles emitted from a $^{214}_{83}\text{Bi}$ radioactive source at a thin metal foil made of gold. Figure 12.2 shows a schematic diagram of this experiment. Alpha-particles emitted by a $^{214}_{83}\text{Bi}$ radioactive source were collimated into a narrow beam by their passage through lead bricks. The beam was allowed to fall on a thin foil of gold of thickness 2.1×10^{-7} m. The scattered alpha-particles were observed through a rotatable detector consisting of zinc sulphide screen and a microscope. The scattered alpha-particles on striking the screen produced brief light flashes or scintillations. These flashes may be viewed through a microscope and the distribution of the number of scattered particles may be studied as a function of angle of scattering.

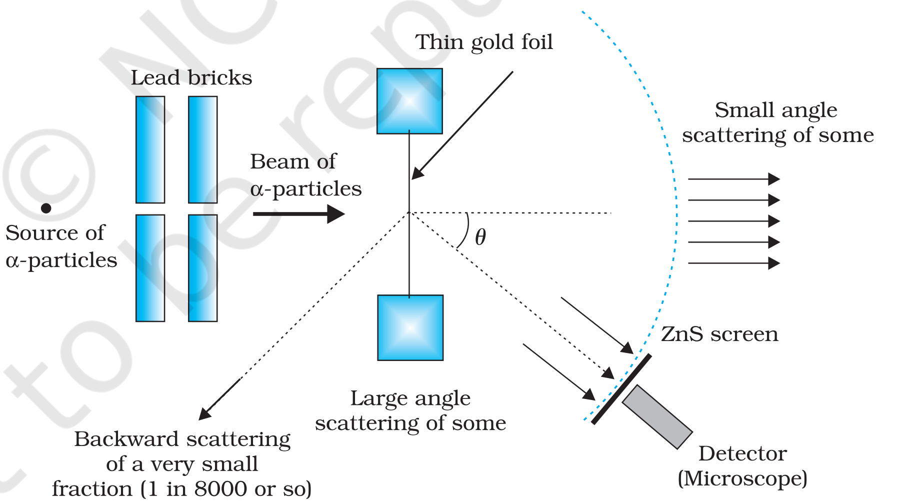


FIGURE 12.2 Schematic arrangement of the Geiger-Marsden experiment.

A typical graph of the total number of α-particles scattered at different angles, in a given interval of time, is shown in Fig. 12.3. The dots in this figure represent the data points and the solid curve is the theoretical prediction based on the assumption that the target atom has a small, dense, positively charged nucleus. Many of the α-particles pass through the foil. It means that they do not suffer any collisions. Only about 0.14% of the incident α-particles scatter by more than 1°; and about 1 in 8000 deflect by more than 90°. Rutherford argued that, to deflect the α-particle backwards, it must experience a large repulsive force. This force could be provided if the greater part of the mass of the atom and its positive charge were concentrated tightly at its centre. Then the incoming α-particle could get very close to the positive charge without penetrating it, and such a close encounter would result in a large deflection. This agreement supported the hypothesis of the nuclear atom. This is why Rutherford is credited with the *discovery* of the nucleus.

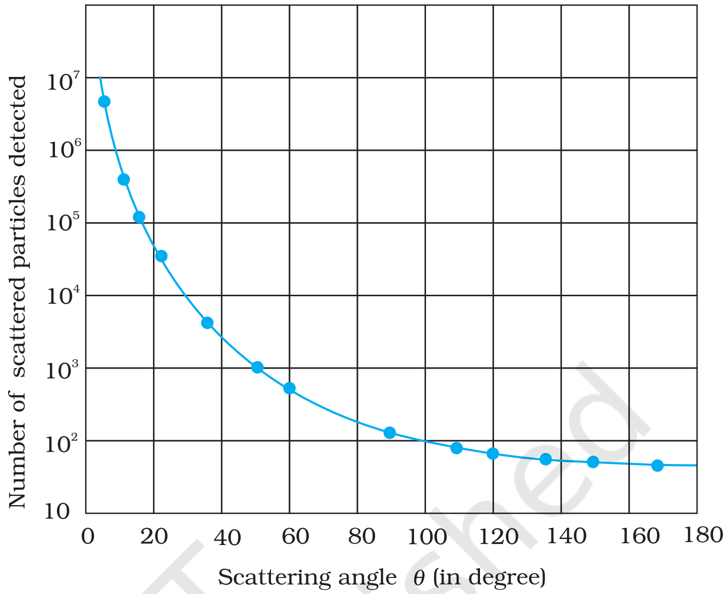


FIGURE 12.3 Experimental data points (shown by dots) on scattering of α-particles by a thin foil at different angles obtained by Geiger and Marsden using the setup shown in Figs. 12.1 and 12.2. Rutherford's nuclear model predicts the solid curve which is seen to be in good agreement with experiment.

In Rutherford's nuclear model of the atom, the entire positive charge and most of the mass of the atom are concentrated in the nucleus with the electrons some distance away. The electrons would be moving in orbits about the nucleus just as the planets do around the sun. Rutherford's experiments suggested the size of the nucleus to be about 10^{-15} m to 10^{-14} m. From kinetic theory, the size of an atom was known to be 10^{-10} m, about 10,000 to 100,000 times larger than the size of the nucleus. Thus, the electrons would seem to be at a distance from the nucleus of about 10,000 to 100,000 times the size of the nucleus itself. Thus, most of an atom is empty space. With the atom being largely empty space, it is easy to see why most α-particles go right through a thin metal foil. However, when α-particle happens to come near a nucleus, the intense electric field there scatters it through a large angle. The atomic electrons, being so light, do not appreciably affect the α-particles.

The scattering data shown in Fig. 12.3 can be analysed by employing Rutherford's nuclear model of the atom. As the gold foil is very thin, it can be assumed that α-particles will suffer not more than one scattering during their passage through it. Therefore, computation of the trajectory of an alpha-particle scattered by a single nucleus is enough. Alpha-particles are nuclei of helium atoms and, therefore, carry two units, $2e$, of positive charge and have the mass of the helium atom. The charge of the gold nucleus is Ze, where Z is the atomic number of the atom; for gold $Z = 79$. Since the nucleus of gold is about 50 times heavier than an α-particle, it is reasonable to assume that it remains stationary throughout the scattering process. Under these assumptions, the trajectory of an alpha-particle can be computed employing Newton's second law of motion and the Coulomb's law for electrostatic force of repulsion between the alpha-particle and the positively charged nucleus.

The magnitude of this force is

$$F = \frac{1}{4\pi\varepsilon_0}\frac{(2e)(Ze)}{r^2} \tag{12.1}$$

where r is the distance between the α-particle and the nucleus. The force is directed along the line joining the α-particle and the nucleus. The magnitude and direction of the force on an α-particle continuously changes as it approaches the nucleus and recedes away from it.

12.2.1 Alpha-particle trajectory

The trajectory traced by an α-particle depends on the impact parameter, b of collision. The *impact parameter* is the perpendicular distance of the initial velocity vector of the α-particle from the centre of the nucleus (Fig. 12.4). A given beam of α-particles has a distribution of impact parameters b, so that the beam is scattered in various directions with different probabilities (Fig. 12.4). (In a beam, all particles have nearly same kinetic energy.) It is seen that an α-particle close to the nucleus (small impact parameter) suffers large scattering. In case of head-on collision, the impact parameter is minimum and the α-particle rebounds back ($\theta \cong \pi$). For a large impact parameter, the α-particle goes nearly undeviated and has a small deflection ($\theta \cong 0$).

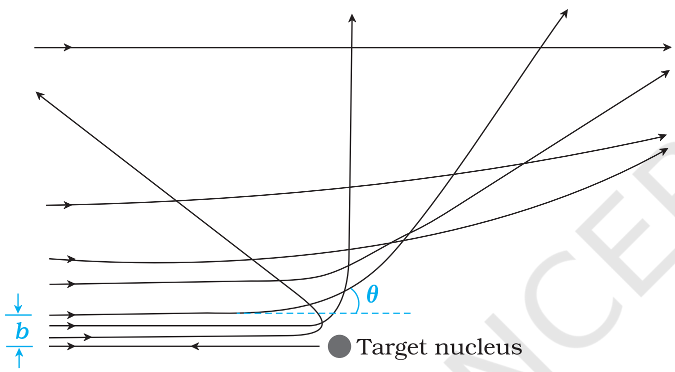


FIGURE 12.4 Trajectory of α-particles in the coulomb field of a target nucleus. The impact parameter, b and scattering angle θ are also depicted.

The fact that only a small fraction of the number of incident particles rebound back indicates that the number of α-particles undergoing head on collision is small. This, in turn, implies that the mass and positive charge of the atom is concentrated in a small volume. Rutherford scattering therefore, is a powerful way to determine an upper limit to the size of the nucleus.

Example 12.1 In the Rutherford's nuclear model of the atom, the nucleus (radius about 10^{-15} m) is analogous to the sun about which the electron move in orbit (radius ≈ 10^{-10} m) like the earth orbits around the sun. If the dimensions of the solar system had the same proportions as those of the atom, would the earth be closer to or farther away from the sun than actually it is? The radius of earth's orbit is about 1.5×10^{11} m. The radius of sun is taken as 7×10^{8} m.

Solution The ratio of the radius of electron's orbit to the radius of nucleus is $(10^{-10}\text{ m})/(10^{-15}\text{ m}) = 10^5$, that is, the radius of the electron's orbit is 10^5 times larger than the radius of nucleus. If the radius of the earth's orbit around the sun were 10^5 times larger than the radius of the sun, the radius of the earth's orbit would be $10^5 \times 7 \times 10^8$ m = 7×10^{13} m. This is more than 100 times greater than the actual orbital radius of earth. Thus, the earth would be much farther away from the sun.

It implies that an atom contains a much greater fraction of empty space than our solar system does.

EXAMPLE 12.1

Example 12.2 In a Geiger-Marsden experiment, what is the distance of closest approach to the nucleus of a 7.7 MeV α-particle before it comes momentarily to rest and reverses its direction?

Solution The key idea here is that throughout the scattering process, the total mechanical energy of the system consisting of an α-particle and a gold nucleus is conserved. The system's initial mechanical energy is E_i, before the particle and nucleus interact, and it is equal to its mechanical energy E_f when the α-particle momentarily stops. The initial energy E_i is just the kinetic energy K of the incoming α- particle. The final energy E_f is just the electric potential energy U of the system. The potential energy U can be calculated from Eq. (12.1).

Let d be the centre-to-centre distance between the α-particle and the gold nucleus when the α-particle is at its stopping point. Then we can write the conservation of energy $E_i = E_f$ as

$$K = \frac{1}{4\pi\varepsilon_0}\frac{(2e)(Ze)}{d} = \frac{2Ze^2}{4\pi\varepsilon_0 d}$$

Thus the distance of closest approach d is given by

$$d = \frac{2Ze^2}{4\pi\varepsilon_0 K}$$

The maximum kinetic energy found in α-particles of natural origin is 7.7 MeV or 1.2×10^{-12} J. Since $1/4\pi\varepsilon_0 = 9.0 \times 10^9$ N m^2/C^2. Therefore with $e = 1.6 \times 10^{-19}$ C, we have,

$$d = \frac{(2)(9.0\times10^9\,\mathrm{Nm^2}/C^2)(1.6\times10^{-19}\,C)^2 Z}{1.2\times10^{-12}\,\mathrm{J}}$$

$$= 3.84 \times 10^{-16}\, Z\,\mathrm{m}$$

The atomic number of foil material gold is $Z = 79$, so that

d (Au) = 3.0×10^{-14} m = 30 fm. (1 fm (i.e. fermi) = 10^{-15} m.)

The radius of gold nucleus is, therefore, less than 3.0×10^{-14} m. This is not in very good agreement with the observed result as the actual radius of gold nucleus is 6 fm. The cause of discrepancy is that the distance of closest approach is considerably larger than the sum of the radii of the gold nucleus and the α-particle. Thus, the α-particle reverses its motion without ever actually *touching* the gold nucleus.

EXAMPLE 12.2

12.2.2 Electron orbits

The Rutherford nuclear model of the atom which involves classical concepts, pictures the atom as an electrically neutral sphere consisting of a very small, massive and positively charged nucleus at the centre surrounded by the revolving electrons in their respective dynamically stable orbits. The electrostatic force of attraction, F_e between the revolving electrons and the nucleus provides the requisite centripetal force (F_c) to keep them in their orbits. Thus, for a dynamically stable orbit in a hydrogen atom

$$F_e = F_c$$

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2} = \frac{mv^2}{r} \qquad (12.2)$$

Thus the relation between the orbit radius and the electron velocity is

$$r = \frac{e^2}{4\pi\varepsilon_0 m v^2} \tag{12.3}$$

The kinetic energy (K) and electrostatic potential energy (U) of the electron in hydrogen atom are

$$K = \frac{1}{2}mv^2 = \frac{e^2}{8\pi\varepsilon_0 r} \text{ and } U = -\frac{e^2}{4\pi\varepsilon_0 r}$$

(The negative sign in U signifies that the electrostatic force is in the $-r$ direction.) Thus the total energy E of the electron in a hydrogen atom is

$$E = K + U = \frac{e^2}{8\pi\varepsilon_0 r} - \frac{e^2}{4\pi\varepsilon_0 r}$$

$$= -\frac{e^2}{8\pi\varepsilon_0 r} \tag{12.4}$$

The total energy of the electron is negative. This implies the fact that the electron is bound to the nucleus. If E were positive, an electron will not follow a closed orbit around the nucleus.

Example 12.3 It is found experimentally that 13.6 eV energy is required to separate a hydrogen atom into a proton and an electron. Compute the orbital radius and the velocity of the electron in a hydrogen atom.

Solution Total energy of the electron in hydrogen atom is -13.6 eV = $-13.6 \times 1.6 \times 10^{-19}$ J = -2.2×10^{-18} J. Thus from Eq. (12.4), we have

$$E = -\frac{e^2}{8\pi\varepsilon_0 r} = -2.2 \times 10^{-18} \text{ J}$$

This gives the orbital radius

$$r = -\frac{e^2}{8\pi\varepsilon_0 E} = -\frac{(9 \times 10^9 \text{ N m}^2/\text{C}^2)(1.6 \times 10^{-19}\text{C})^2}{(2)(-2.2 \times 10^{-18} \text{ J})}$$

$= 5.3 \times 10^{-11}$ m.

The velocity of the revolving electron can be computed from Eq. (12.3) with $m = 9.1 \times 10^{-31}$ kg,

$$v = \frac{e}{\sqrt{4\pi\varepsilon_0 m r}} = 2.2 \times 10^6 \text{ m/s.}$$

EXAMPLE 12.3

12.3 Atomic Spectra

As mentioned in Section 12.1, each element has a characteristic spectrum of radiation, which it emits. When an atomic gas or vapour is excited at low pressure, usually by passing an electric current through it, the emitted radiation has a spectrum which contains certain specific wavelengths only. A spectrum of this kind is termed as emission line spectrum and it consists of bright lines on a dark background. The spectrum emitted by atomic hydrogen is shown in Fig. 12.5. Study of emission line spectra of a material can therefore serve as a type of "fingerprint" for identification of the gas. When white light passes through a gas and we analyse the transmitted light using a spectrometer we find some dark lines in the spectrum. These dark lines correspond precisely to those wavelengths which were found in the emission line spectrum of the gas. This is called the *absorption spectrum* of the material of the gas.

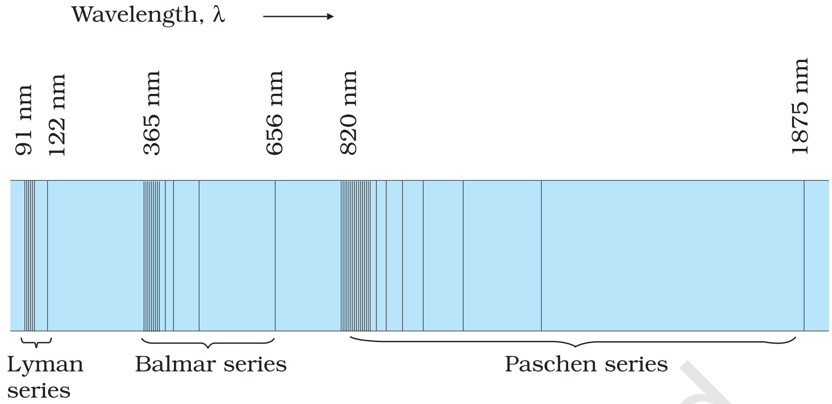


FIGURE 12.5 Emission lines in the spectrum of hydrogen.

12.4 Bohr Model of the Hydrogen Atom

The model of the atom proposed by Rutherford assumes that the atom, consisting of a central nucleus and revolving electron is stable much like sun-planet system which the model imitates. However, there are some fundamental differences between the two situations. While the planetary system is held by gravitational force, the nucleus-electron system being charged objects, interact by Coulomb's Law of force. We know that an object which moves in a circle is being constantly accelerated – the acceleration being centripetal in nature. According to classical electromagnetic theory, an accelerating charged particle emits radiation in the form of electromagnetic waves. The energy of an accelerating electron should therefore, continuously decrease. The electron would spiral inward and eventually fall into the nucleus (Fig. 12.6). Thus, such an atom can not be stable. Further, according to the classical electromagnetic theory, the frequency of the electromagnetic waves emitted by the revolving electrons is equal to the frequency of revolution. As the electrons spiral inwards, their angular velocities and hence their frequencies would change continuously, and so will the frequency of the light emitted. Thus, they would emit a continuous spectrum, in contradiction to the line spectrum actually observed. Clearly Rutherford model tells only a part of the story implying that the classical ideas are not sufficient to explain the atomic structure.

Niels Henrik David Bohr (1885 – 1962) Danish physicist who explained the spectrum of hydrogen atom based on quantum ideas. He gave a theory of nuclear fission based on the liquid-drop model of nucleus. Bohr contributed to the clarification of conceptual problems in quantum mechanics, in particular by proposing the complementary principle.

NIELS HENRIK DAVID BOHR (1885 – 1962)

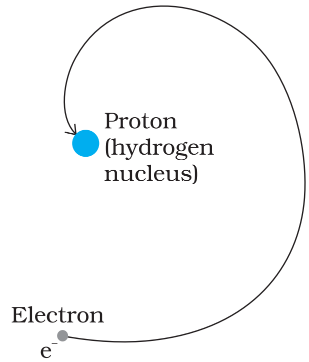


FIGURE 12.6 An accelerated atomic electron must spiral into the nucleus as it loses energy.

EXAMPLE 12.4

Example 12.4 According to the classical electromagnetic theory, calculate the initial frequency of the light emitted by the electron revolving around a proton in hydrogen atom.

Solution From Example 12.3 we know that velocity of electron moving around a proton in hydrogen atom in an orbit of radius 5.3×10^{-11} m is 2.2×10^{-6} m/s. Thus, the frequency of the electron moving around the proton is

$$\nu = \frac{v}{2\pi r} = \frac{2.2 \times 10^{6}\ \text{m s}^{-1}}{2\pi\left(5.3 \times 10^{-11}\ \text{m}\right)}$$

$$\approx 6.6 \times 10^{15}\ \text{Hz}.$$

According to the classical electromagnetic theory we know that the frequency of the electromagnetic waves emitted by the revolving electrons is equal to the frequency of its revolution around the nucleus. Thus the initial frequency of the light emitted is 6.6×10^{15} Hz.

It was Niels Bohr (1885 – 1962) who made certain modifications in this model by adding the ideas of the newly developing quantum hypothesis. Niels Bohr studied in Rutherford's laboratory for several months in 1912 and he was convinced about the validity of Rutherford nuclear model. Faced with the dilemma as discussed above, Bohr, in 1913, concluded that in spite of the success of electromagnetic theory in explaining large-scale phenomena, it could not be applied to the processes at the atomic scale. It became clear that a fairly radical departure from the established principles of classical mechanics and electromagnetism would be needed to understand the structure of atoms and the relation of atomic structure to atomic spectra. Bohr combined classical and early quantum concepts and gave his theory in the form of three postulates. These are :

(i) Bohr's first postulate was that *an electron in an atom could revolve in certain stable orbits without the emission of radiant energy*, contrary to the predictions of electromagnetic theory. According to this postulate, each atom has certain definite stable states in which it

Reprint 2025-26

can exist, and each possible state has definite total energy. These are called the stationary states of the atom.

(ii) Bohr's second postulate defines these stable orbits. This postulate states that the *electron* revolves around the nucleus *only in those orbits for which the angular momentum is some integral multiple of* $h/2\pi$ where h is the Planck's constant (= 6.6×10^{-34} J s). Thus the angular momentum (L) of the orbiting electron is quantised. That is

$$L = nh/2\pi \qquad (12.5)$$

(iii) Bohr's third postulate incorporated into atomic theory the early quantum concepts that had been developed by Planck and Einstein. It states that *an electron might make a transition from one of its specified non-radiating orbits to another of lower energy. When it does so, a photon is emitted having energy equal to the energy difference between the initial and final states. The frequency of the emitted photon is then given by*

$$h\nu = E_i - E_f \qquad (12.6)$$

where E_i and E_f are the energies of the initial and final states and $E_i > E_f$.

For a hydrogen atom, Eq. (12.4) gives the expression to determine the energies of different energy states. But then this equation requires the radius r of the electron orbit. To calculate r, Bohr's second postulate about the angular momentum of the electron–the quantisation condition – is used.

The radius of n[th] possible orbit thus found is

$$r_n = \left(\frac{n^2}{m}\right)\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\varepsilon_0}{e^2} \qquad (12.7)$$

The total energy of the electron in the stationary states of the hydrogen atom can be obtained by substituting the value of orbital radius in Eq. (12.4) as

$$E_n = -\left(\frac{e^2}{8\pi\varepsilon_0}\right)\left(\frac{m}{n^2}\right)\left(\frac{2\pi}{h}\right)^2\left(\frac{e^2}{4\pi\varepsilon_0}\right)$$

or $$E_n = -\frac{me^4}{8n^2\varepsilon_0^2h^2} \qquad (12.8)$$

Substituting values, Eq. (12.8) yields

$$E_n = -\frac{2.18 \times 10^{-18}}{n^2}\,\mathrm{J} \qquad (12.9)$$

Atomic energies are often expressed in electron volts (eV) rather than joules. Since 1 eV = 1.6×10^{-19} J, Eq. (12.9) can be rewritten as

$$E_n = -\frac{13.6}{n^2}\ \mathrm{eV} \qquad (12.10)$$

The negative sign of the total energy of an electron moving in an orbit means that the electron is bound with the nucleus. Energy will thus be required to remove the electron from the hydrogen atom to a distance infinitely far away from its nucleus (or proton in hydrogen atom).

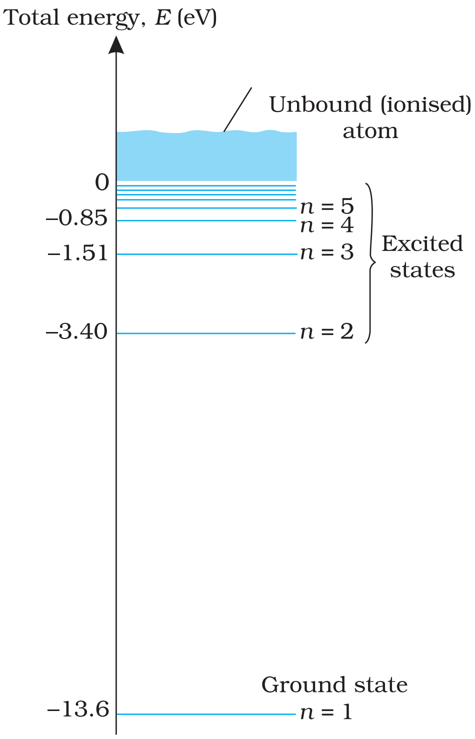


FIGURE 12.7 The energy level diagram for the hydrogen atom. The electron in a hydrogen atom at room temperature spends most of its time in the ground state. To ionise a hydrogen atom an electron from the ground state, 13.6 eV of energy must be supplied. (The horizontal lines specify the presence of allowed energy states.)

12.4.1 Energy levels

The energy of an atom is the *least* (largest negative value) when its electron is revolving in an orbit closest to the nucleus i.e., the one for which $n = 1$. For $n = 2, 3, ...$ the absolute value of the energy E is smaller, hence the energy is progressively larger in the outer orbits. The *lowest* state of the atom, called the *ground state*, is that of the lowest energy, with the electron revolving in the orbit of smallest radius, the Bohr radius, a_0. The energy of this state ($n = 1$), E_1 is –13.6 eV. Therefore, the minimum energy required to free the electron from the ground state of the hydrogen atom is 13.6 eV. It is called the *ionisation energy* of the hydrogen atom. This prediction of the Bohr's model is in excellent agreement with the experimental value of ionisation energy.

At room temperature, most of the hydrogen atoms are in *ground state*. When a hydrogen atom receives energy by processes such as electron collisions, the atom may acquire sufficient energy to raise the electron to higher energy states. The atom is then said to be in an *excited state*. From Eq. (12.10), for $n = 2$; the energy E_2 is –3.40 eV. It means that the energy required to excite an electron in hydrogen atom to its first excited state, is an energy equal to $E_2 - E_1 = –3.40$ eV – (–13.6) eV = 10.2 eV. Similarly, $E_3 = –1.51$ eV and $E_3 - E_1 = 12.09$ eV, or to excite the hydrogen atom from its ground state ($n = 1$) to second excited state ($n = 3$), 12.09 eV energy is required, and so on. From these excited states the electron can then fall back to a state of lower energy, emitting a photon in the process. Thus, as the excitation of hydrogen atom increases (that is as n increases) the value of minimum energy required to free the electron from the excited atom decreases.

The energy level diagram* for the stationary states of a hydrogen atom, computed from Eq. (12.10), is given in Fig. 12.7. The principal quantum number n labels the stationary states in the ascending order of energy. In this diagram, the highest energy state corresponds to $n = \infty$ in Eq, (12.10) and has an energy of 0 eV. This is the energy of the atom when the electron is completely removed ($r = \infty$) from the nucleus and is at rest. Observe how the energies of the excited states come closer and closer together as n increases.

12.5 The Line Spectra of the Hydrogen Atom

According to the third postulate of Bohr's model, when an atom makes a transition from the higher energy state with quantum number n_i to the lower energy state with quantum number n_f ($n_f < n_i$), the difference of energy is carried away by a photon of frequency ν_{if} such that

* An electron can have any total energy above $E = 0$ eV. In such situations the electron is free. Thus there is a continuum of energy states above $E = 0$ eV, as shown in Fig. 12.7.

$$h\nu_{if} = E_{ni} - E_{nf} \tag{12.11}$$

Since both n_f and n_i are integers, this immediately shows that in transitions between different atomic levels, light is radiated in various discrete frequencies.

The various lines in the atomic spectra are produced when electrons jump from higher energy state to a lower energy state and photons are emitted. These spectral lines are called emission lines. But when an atom absorbs a photon that has precisely the same energy needed by the electron in a lower energy state to make transitions to a higher energy state, the process is called absorption. Thus if photons with a continuous range of frequencies pass through a rarefied gas and then are analysed with a spectrometer, a series of dark spectral absorption lines appear in the continuous spectrum. The dark lines indicate the frequencies that have been absorbed by the atoms of the gas.

The explanation of the hydrogen atom spectrum provided by Bohr's model was a brilliant achievement, which greatly stimulated progress towards the modern quantum theory. In 1922, Bohr was awarded Nobel Prize in Physics.

12.6 De Broglie's Explanation of Bohr's Second Postulate of Quantisation

Of all the postulates, Bohr made in his model of the atom, perhaps the most puzzling is his second postulate. It states that the angular momentum of the electron orbiting around the nucleus is quantised (that is, $L_n = nh/2\pi$; n = 1, 2, 3 ...). Why should the angular momentum have only those values that are integral multiples of $h/2\pi$? The French physicist Louis de Broglie explained this puzzle in 1923, ten years after Bohr proposed his model.

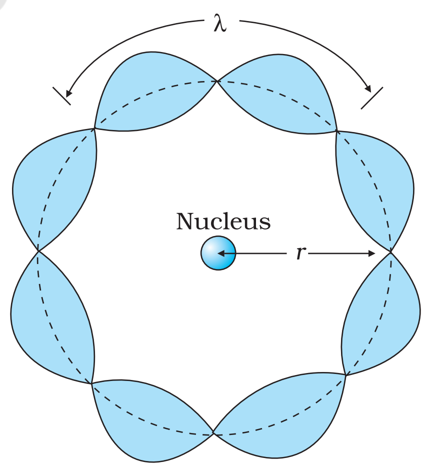


FIGURE 12.8 A standing wave is shown on a circular orbit where four de Broglie wavelengths fit into the circumference of the orbit.

We studied, in Chapter 11, about the de Broglie's hypothesis that material particles, such as electrons, also have a wave nature. C. J. Davisson and L. H. Germer later experimentally verified the wave nature of electrons in 1927. Louis de Broglie argued that the electron in its circular orbit, as proposed by Bohr, must be seen as a particle wave. In analogy to waves travelling on a string, particle waves too can lead to standing waves under resonant conditions. From Chapter 14 of Class XI Physics textbook, we know that when a string is plucked, a vast number of wavelengths are excited. However only those wavelengths survive which have nodes at the ends and form the standing wave in the string. It means that in a string, standing waves are formed when the total distance travelled by a wave down the string and back is one wavelength, two wavelengths, or any integral number of wavelengths. Waves with other wavelengths interfere with themselves upon reflection and their amplitudes quickly drop to zero. For an electron moving in n^{th} circular orbit of radius r_n, the total distance is the circumference of the orbit, $2\pi r_n$. Thus

$$2\pi\, r_n = n\lambda, \qquad n = 1, 2, 3... \tag{12.12}$$

Figure 12.8 illustrates a standing particle wave on a circular orbit for $n = 4$, i.e., $2\pi r_n = 4\lambda$, where λ is the de Broglie wavelength of the electron moving in n^{th} orbit. From Chapter 11, we have $\lambda = h/p$, where p is the magnitude of the electron's momentum. If the speed of the electron is much less than the speed of light, the momentum is mv_n. Thus, $\lambda = h/mv_n$. From Eq. (12.12), we have

$$2\pi\, r_n = n\, h/mv_n \quad \text{or} \quad m\, v_n r_n = nh/2\pi$$

This is the quantum condition proposed by Bohr for the angular momentum of the electron [Eq. (12.15)]. In Section 12.5, we saw that this equation is the basis of explaining the discrete orbits and energy levels in hydrogen atom. Thus de Broglie hypothesis provided an explanation for Bohr's second postulate for the quantisation of angular momentum of the orbiting electron. The quantised electron orbits and energy states are due to the wave nature of the electron and only resonant standing waves can persist.

Bohr's model, involving classical trajectory picture (planet-like electron orbiting the nucleus), correctly predicts the gross features of the hydrogenic atoms*, in particular, the frequencies of the radiation emitted or selectively absorbed. This model however has many limitations. Some are:

(i) The Bohr model is applicable to hydrogenic atoms. It cannot be extended even to mere two electron atoms such as helium. The analysis of atoms with more than one electron was attempted on the lines of Bohr's model for hydrogenic atoms but did not meet with any success. Difficulty lies in the fact that each electron interacts not only with the positively charged nucleus but also with all other electrons.

The formulation of Bohr model involves electrical force between positively charged nucleus and electron. It does not include the electrical forces between electrons which necessarily appear in multi-electron atoms.

(ii) While the Bohr's model correctly predicts the frequencies of the light emitted by hydrogenic atoms, the model is unable to explain the relative intensities of the frequencies in the spectrum. In emission spectrum of hydrogen, some of the visible frequencies have weak intensity, others strong. Why? Experimental observations depict that some transitions are more favoured than others. Bohr's model is unable to account for the intensity variations.

Bohr's model presents an elegant picture of an atom and cannot be generalised to complex atoms. For complex atoms we have to use a new and radical theory based on Quantum Mechanics, which provides a more complete picture of the atomic structure.

* Hydrogenic atoms are the atoms consisting of a nucleus with positive charge +Ze and a single electron, where Z is the proton number. Examples are hydrogen atom, singly ionised helium, doubly ionised lithium, and so forth. In these atoms more complex electron-electron interactions are nonexistent.

SUMMARY

1. Atom, as a whole, is electrically neutral and therefore contains equal amount of positive and negative charges.
2. In *Thomson's model*, an atom is a spherical cloud of positive charges with electrons embedded in it.
3. In *Rutherford's model*, most of the mass of the atom and all its positive charge are concentrated in a tiny nucleus (typically one by ten thousand the size of an atom), and the electrons revolve around it.
4. Rutherford nuclear model has two main difficulties in explaining the structure of atom: (a) It predicts that atoms are unstable because the accelerated electrons revolving around the nucleus must spiral into the nucleus. This contradicts the stability of matter. (b) It cannot explain the characteristic line spectra of atoms of different elements.
5. Atoms of most of the elements are stable and emit characteristic spectrum. The spectrum consists of a set of isolated parallel lines termed as line spectrum. It provides useful information about the atomic structure.
6. To explain the line spectra emitted by atoms, as well as the stability of atoms, Niel's Bohr proposed a model for hydrogenic (single elctron) atoms. He introduced three postulates and laid the foundations of quantum mechanics:
   (a) In a hydrogen atom, an electron revolves in certain stable orbits (called stationary orbits) without the emission of radiant energy.
   (b) The stationary orbits are those for which the angular momentum is some integral multiple of $h/2\pi$. (Bohr's quantisation condition.) That is $L = nh/2\pi$, where n is an integer called the principal quantum number.
   (c) The third postulate states that an electron might make a transition from one of its specified non-radiating orbits to another of lower energy. When it does so, a photon is emitted having energy equal to the energy difference between the initial and final states. The frequency (ν) of the emitted photon is then given by
   $h\nu = E_i - E_f$
   An atom absorbs radiation of the same frequency the atom emits, in which case the electron is transferred to an orbit with a higher value of n.
   $E_i + h\nu = E_f$
7. As a result of the quantisation condition of angular momentum, the electron orbits the nucleus at only specific radii. For a hydrogen atom it is given by
$$r_n = \left(\frac{n^2}{m}\right)\left(\frac{h}{2\pi}\right)^2 \frac{4\pi\varepsilon_0}{e^2}$$
The total energy is also quantised:
$$E_n = -\frac{me^4}{8n^2\varepsilon_0^2h^2}$$
$$= -13.6 \text{ eV}/n^2$$
The $n = 1$ state is called ground state. In hydrogen atom the ground state energy is -13.6 eV. Higher values of n correspond to excited states ($n > 1$). Atoms are excited to these higher states by collisions with other atoms or electrons or by absorption of a photon of right frequency.

8. de Broglie's hypothesis that electrons have a wavelength $\lambda = h/mv$ gave an explanation for Bohr's quantised orbits by bringing in the wave-particle duality. The orbits correspond to circular standing waves in which the circumference of the orbit equals a whole number of wavelengths.
9. Bohr's model is applicable only to hydrogenic (single electron) atoms. It cannot be extended to even two electron atoms such as helium. This model is also unable to explain for the relative intensities of the frequencies emitted even by hydrogenic atoms.

POINTS TO PONDER

1. Both the Thomson's as well as the Rutherford's models constitute an unstable system. Thomson's model is unstable electrostatically, while Rutherford's model is unstable because of electromagnetic radiation of orbiting electrons.
2. What made Bohr quantise angular momentum (second postulate) and not some other quantity? Note, h has dimensions of angular momentum, and for circular orbits, angular momentum is a very relevant quantity. The second postulate is then so natural!
3. The orbital picture in Bohr's model of the hydrogen atom was inconsistent with the uncertainty principle. It was replaced by modern quantum mechanics in which Bohr's orbits are regions where the electron may be found with large probability.
4. Unlike the situation in the solar system, where planet-planet gravitational forces are very small as compared to the gravitational force of the sun on each planet (because the mass of the sun is so much greater than the mass of any of the planets), the electron-electron electric force interaction is comparable in magnitude to the electron-nucleus electrical force, because the charges and distances are of the same order of magnitude. This is the reason why the Bohr's model with its planet-like electron is not applicable to many electron atoms.
5. Bohr laid the foundation of the quantum theory by postulating specific orbits in which electrons do not radiate. Bohr's model include only one quantum number n. The new theory called quantum mechanics supports Bohr's postulate. However in quantum mechanics (more generally accepted), a given energy level may not correspond to just one quantum state. For example, a state is characterised by four quantum numbers (n, l, m, and s), but for a pure Coulomb potential (as in hydrogen atom) the energy depends only on n.
6. In Bohr model, contrary to ordinary classical expectation, the frequency of revolution of an electron in its orbit is not connected to the frequency of spectral line. The later is the difference between two orbital energies divided by h. For transitions between large quantum numbers (n to n – 1, n very large), however, the two coincide as expected.
7. Bohr's semiclassical model based on some aspects of classical physics and some aspects of modern physics also does not provide a true picture of the simplest hydrogenic atoms. The true picture is quantum mechanical affair which differs from Bohr model in a number of fundamental ways. But then if the Bohr model is not strictly correct, why do we bother about it? The reasons which make Bohr's model still useful are:

(i) The model is based on just three postulates but accounts for almost all the general features of the hydrogen spectrum.

(ii) The model incorporates many of the concepts we have learnt in classical physics.

(iii) The model demonstrates how a theoretical physicist occasionally must quite literally ignore certain problems of approach in hopes of being able to make some predictions. If the predictions of the theory or model agree with experiment, a theoretician then must somehow hope to explain away or rationalise the problems that were ignored along the way.

EXERCISES

12.1 Choose the correct alternative from the clues given at the end of the each statement:

(a) The size of the atom in Thomson's model is .... the atomic size in Rutherford's model. (much greater than/no different from/much less than.)

(b) In the ground state of .... electrons are in stable equilibrium, while in .... electrons always experience a net force. (Thomson's model/ Rutherford's model.)

(c) A *classical* atom based on .... is doomed to collapse. (Thomson's model/ Rutherford's model.)

(d) An atom has a nearly continuous mass distribution in a .... but has a highly non-uniform mass distribution in .... (Thomson's model/ Rutherford's model.)

(e) The positively charged part of the atom possesses most of the mass in .... (Rutherford's model/both the models.)

12.2 Suppose you are given a chance to repeat the alpha-particle scattering experiment using a thin sheet of solid hydrogen in place of the gold foil. (Hydrogen is a solid at temperatures below 14 K.) What results do you expect?

12.3 A difference of 2.3 eV separates two energy levels in an atom. What is the frequency of radiation emitted when the atom make a transition from the upper level to the lower level?

12.4 The ground state energy of hydrogen atom is –13.6 eV. What are the kinetic and potential energies of the electron in this state?

12.5 A hydrogen atom initially in the ground level absorbs a photon, which excites it to the $n = 4$ level. Determine the wavelength and frequency of photon.

12.6 (a) Using the Bohr's model calculate the speed of the electron in a hydrogen atom in the $n = 1$, 2, and 3 levels. (b) Calculate the orbital period in each of these levels.

12.7 The radius of the innermost electron orbit of a hydrogen atom is 5.3×10^{-11} m. What are the radii of the $n = 2$ and $n = 3$ orbits?

12.8 A 12.5 eV electron beam is used to bombard gaseous hydrogen at room temperature. What series of wavelengths will be emitted?

12.9 In accordance with the Bohr's model, find the quantum number that characterises the earth's revolution around the sun in an orbit of radius 1.5×10^{11} m with orbital speed 3×10^{4} m/s. (Mass of earth $= 6.0 \times 10^{24}$ kg.)

Chapter Thirteen

NUCLEI

13.1 Introduction

In the previous chapter, we have learnt that in every atom, the positive charge and mass are densely concentrated at the centre of the atom forming its nucleus. The overall dimensions of a nucleus are much smaller than those of an atom. Experiments on scattering of α-particles demonstrated that the radius of a nucleus was smaller than the radius of an atom by a factor of about 10^4. This means the volume of a nucleus is about 10^{-12} times the volume of the atom. In other words, an atom is almost empty. If an atom is enlarged to the size of a classroom, the nucleus would be of the size of pinhead. Nevertheless, the nucleus contains most (more than 99.9%) of the mass of an atom.

Does the nucleus have a structure, just as the atom does? If so, what are the constituents of the nucleus? How are these held together? In this chapter, we shall look for answers to such questions. We shall discuss various properties of nuclei such as their size, mass and stability, and also associated nuclear phenomena such as radioactivity, fission and fusion.

13.2 Atomic Masses and Composition of Nucleus

The mass of an atom is very small, compared to a kilogram; for example, the mass of a carbon atom, ^{12}C, is 1.992647×10^{-26} kg. Kilogram is not a very convenient unit to measure such small quantities. Therefore, a different mass unit is used for expressing atomic masses. This unit is the atomic mass unit (u), defined as $1/12^{th}$ of the mass of the carbon (^{12}C) atom. According to this definition

$$1u = \frac{\text{mass of one }^{12}\text{C atom}}{12}$$

$$= \frac{1.992647 \times 10^{-26}\ \text{kg}}{12}$$

$$= 1.660539 \times 10^{-27}\ \text{kg} \qquad (13.1)$$

The atomic masses of various elements expressed in atomic mass unit (u) are close to being integral multiples of the mass of a hydrogen atom. There are, however, many striking exceptions to this rule. For example, the atomic mass of chlorine atom is 35.46 u.

Accurate measurement of atomic masses is carried out with a mass spectrometer, The measurement of atomic masses reveals the existence of different types of atoms of the same element, which exhibit the same chemical properties, but differ in mass. Such atomic species of the same element differing in mass are called *isotopes*. (In Greek, isotope means the same place, i.e. they occur in the same place in the periodic table of elements.) It was found that practically every element consists of a mixture of several isotopes. The relative abundance of different isotopes differs from element to element. Chlorine, for example, has two isotopes having masses 34.98 u and 36.98 u, which are nearly integral multiples of the mass of a hydrogen atom. The relative abundances of these isotopes are 75.4 and 24.6 per cent, respectively. Thus, the average mass of a chlorine atom is obtained by the weighted average of the masses of the two isotopes, which works out to be

$$= \frac{75.4 \times 34.98 + 24.6 \times 36.98}{100}$$

$$= 35.47\ \text{u}$$

which agrees with the atomic mass of chlorine.

Even the lightest element, hydrogen has three isotopes having masses 1.0078 u, 2.0141 u, and 3.0160 u. The nucleus of the lightest atom of hydrogen, which has a relative abundance of 99.985%, is called the proton. The mass of a proton is

$$m_p = 1.00727\ \text{u} = 1.67262 \times 10^{-27}\ \text{kg} \qquad (13.2)$$

This is equal to the mass of the hydrogen atom (= 1.00783u), minus the mass of a single electron (m_e = 0.00055 u). The other two isotopes of hydrogen are called deuterium and tritium. Tritium nuclei, being unstable, do not occur naturally and are produced artificially in laboratories.

The positive charge in the nucleus is that of the protons. A proton carries one unit of fundamental charge and is stable. It was earlier thought that the nucleus may contain electrons, but this was ruled out later using arguments based on quantum theory. All the electrons of an atom are outside the nucleus. We know that the number of these electrons outside the nucleus of the atom is *Z*, the atomic number. The total charge of the atomic electrons is thus (–Ze), and since the atom is neutral, the charge of the nucleus is (+Ze). The number of protons in the nucleus of the atom is, therefore, exactly Z, the atomic number.

Discovery of Neutron

Since the nuclei of deuterium and tritium are isotopes of hydrogen, they must contain only one proton each. But the masses of the nuclei of hydrogen, deuterium and tritium are in the ratio of 1:2:3. Therefore, the nuclei of deuterium and tritium must contain, in addition to a proton, some neutral matter. The amount of neutral matter present in the nuclei of these isotopes, expressed in units of mass of a proton, is approximately equal to one and two, respectively. This fact indicates that the nuclei of atoms contain, in addition to protons, neutral matter in multiples of a basic unit. This hypothesis was verified in 1932 by James Chadwick who observed emission of neutral radiation when beryllium nuclei were bombarded with alpha-particles (α-particles are helium nuclei, to be discussed in a later section). It was found that this neutral radiation could knock out protons from light nuclei such as those of helium, carbon and nitrogen. The only neutral radiation known at that time was photons (electromagnetic radiation). Application of the principles of conservation of energy and momentum showed that if the neutral radiation consisted of photons, the energy of photons would have to be much higher than is available from the bombardment of beryllium nuclei with α-particles. The clue to this puzzle, which Chadwick satisfactorily solved, was to assume that the neutral radiation consists of a new type of neutral particles called *neutrons*. From conservation of energy and momentum, he was able to determine the mass of new particle 'as very nearly the same as mass of proton'.

The mass of a neutron is now known to a high degree of accuracy. It is

$$m_n = 1.00866 \text{ u} = 1.6749\times10^{-27} \text{ kg} \tag{13.3}$$

Chadwick was awarded the 1935 Nobel Prize in Physics for his discovery of the neutron.

A free neutron, unlike a free proton, is unstable. It decays into a proton, an electron and a antineutrino (another elementary particle), and has a mean life of about 1000s. It is, however, stable inside the nucleus.

The composition of a nucleus can now be described using the following terms and symbols:

Z - *atomic number* = number of protons [13.4(a)]

N - *neutron number* = number of neutrons [13.4(b)]

A - *mass number* = $Z + N$

= total number of protons and neutrons [13.4(c)]

One also uses the term nucleon for a proton or a neutron. Thus the number of nucleons in an atom is its mass number A.

Nuclear species or nuclides are shown by the notation ${}_{Z}^{A}\text{X}$ where X is the chemical symbol of the species. For example, the nucleus of gold is denoted by ${}_{79}^{197}\text{Au}$. It contains 197 nucleons, of which 79 are protons and the rest118 are neutrons.

The composition of isotopes of an element can now be readily explained. The nuclei of isotopes of a given element contain the same number of protons, but differ from each other in their number of neutrons. Deuterium, ${}^{2}_{1}\mathrm{H}$, which is an isotope of hydrogen, contains one proton and one neutron. Its other isotope tritium, ${}^{3}_{1}\mathrm{H}$, contains one proton and two neutrons. The element gold has 32 isotopes, ranging from $A = 173$ to $A = 204$. We have already mentioned that chemical properties of elements depend on their electronic structure. As the atoms of isotopes have identical electronic structure they have identical chemical behaviour and are placed in the same location in the periodic table.

All nuclides with same mass number A are called *isobars*. For example, the nuclides ${}^{3}_{1}\mathrm{H}$ and ${}^{3}_{2}\mathrm{He}$ are isobars. Nuclides with same neutron number N but different atomic number Z, for example ${}^{198}_{80}\mathrm{Hg}$ and ${}^{197}_{79}\mathrm{Au}$, are called *isotones*.

13.3 Size of the Nucleus

As we have seen in Chapter 12, Rutherford was the pioneer who postulated and established the existence of the atomic nucleus. At Rutherford's suggestion, Geiger and Marsden performed their classic experiment: on the scattering of α-particles from thin gold foils. Their experiments revealed that the distance of closest approach to a gold nucleus of an α-particle of kinetic energy 5.5 MeV is about 4.0×10^{-14} m. The scattering of α-particle by the gold sheet could be understood by Rutherford by assuming that the coulomb repulsive force was solely responsible for scattering. Since the positive charge is confined to the nucleus, the actual size of the nucleus has to be less than 4.0×10^{-14} m.

If we use α-particles of higher energies than 5.5 MeV, the distance of closest approach to the gold nucleus will be smaller and at some point the scattering will begin to be affected by the short range nuclear forces, and differ from Rutherford's calculations. Rutherford's calculations are based on pure coulomb repulsion between the positive charges of the α-particle and the gold nucleus. From the distance at which deviations set in, nuclear sizes can be inferred.

By performing scattering experiments in which fast electrons, instead of α-particles, are projectiles that bombard targets made up of various elements, the sizes of nuclei of various elements have been accurately measured.

It has been found that a nucleus of mass number A has a radius

$$R = R_0 A^{1/3} \tag{13.5}$$

where $R_0 = 1.2 \times 10^{-15}$ m (=1.2 fm; 1 fm = 10^{-15} m). This means the volume of the nucleus, which is proportional to R^3 is proportional to A. Thus the density of nucleus is a constant, independent of A, for all nuclei. Different nuclei are like a drop of liquid of constant density. The density of nuclear matter is approximately 2.3×10^{17} kg m^{-3}. This density is very large compared to ordinary matter, say water, which is 10^3 kg m^{-3}. This is understandable, as we have already seen that most of the atom is empty. Ordinary matter consisting of atoms has a large amount of empty space.

EXAMPLE 13.1

Example 13.1 Given the mass of iron nucleus as 55.85u and A=56, find the nuclear density?

Solution

$m_{Fe} = 55.85, \quad u = 9.27 \times 10^{-26}$ kg

$$\text{Nuclear density} = \frac{\text{mass}}{\text{volume}} = \frac{9.27\times10^{-26}}{(4\pi/3)(1.2\times10^{-15})^3}\times\frac{1}{56}$$

$$= 2.29 \times 10^{17}\ \text{kg m}^{-3}$$

The density of matter in neutron stars (an astrophysical object) is comparable to this density. This shows that matter in these objects has been compressed to such an extent that they resemble a *big nucleus*.

13.4 Mass-Energy and Nuclear Binding Energy

13.4.1 Mass – Energy

Einstein showed from his theory of special relativity that it is necessary to treat mass as another form of energy. Before the advent of this theory of special relativity it was presumed that mass and energy were conserved separately in a reaction. However, Einstein showed that mass is another form of energy and one can convert mass-energy into other forms of energy, say kinetic energy and vice-versa.

Einstein gave the famous mass-energy equivalence relation

$$E = mc^2 \tag{13.6}$$

Here the energy equivalent of mass *m* is related by the above equation and *c* is the velocity of light in vacuum and is approximately equal to 3×10^8 m s^{-1}.

EXAMPLE 13.2

Example 13.2 Calculate the energy equivalent of 1 g of substance.

Solution

Energy, $E = 10^{-3} \times (3 \times 10^8)^2$ J

$E = 10^{-3} \times 9 \times 10^{16} = 9 \times 10^{13}$ J

Thus, if one gram of matter is converted to energy, there is a release of enormous amount of energy.

Experimental verification of the Einstein's mass-energy relation has been achieved in the study of nuclear reactions amongst nucleons, nuclei, electrons and other more recently discovered particles. In a reaction the conservation law of energy states that the initial energy and the final energy are equal provided the energy associated with mass is also included. This concept is important in understanding nuclear masses and the interaction of nuclei with one another. They form the subject matter of the next few sections.

13.4.2 Nuclear binding energy

In Section 13.2 we have seen that the nucleus is made up of neutrons and protons. Therefore it may be expected that the mass of the nucleus is equal to the total mass of its individual protons and neutrons. However, the nuclear mass M is found to be always less than this. For example, let us consider $^{16}_{8}\text{O}$; a nucleus which has 8 neutrons and 8 protons. We have

Mass of 8 neutrons = 8 × 1.00866 u

Mass of 8 protons = 8 × 1.00727 u

Mass of 8 electrons = 8 × 0.00055 u

Therefore the expected mass of $^{16}_{8}\text{O}$ nucleus
= 8 × 2.01593 u = 16.12744 u.

The atomic mass of $^{16}_{8}\text{O}$ found from mass spectroscopy experiments is seen to be 15.99493 u. Substracting the mass of 8 electrons (8 × 0.00055 u) from this, we get the experimental mass of $^{16}_{8}\text{O}$ nucleus to be 15.99053 u.

Thus, we find that the mass of the $^{16}_{8}\text{O}$ nucleus is less than the total mass of its constituents by 0.13691u. The difference in mass of a nucleus and its constituents, ΔM, is called the *mass defect*, and is given by

$$\Delta M = [Zm_p + (A - Z)m_n] - M \tag{13.7}$$

What is the meaning of the mass defect? It is here that Einstein's equivalence of mass and energy plays a role. Since the mass of the oxygen nucleus is less that the sum of the masses of its constituents (8 protons and 8 neutrons, in the unbound state), the equivalent energy of the oxygen nucleus is less than that of the sum of the equivalent energies of its constituents. If one wants to break the oxygen nucleus into 8 protons and 8 neutrons, this extra energy $\Delta M c^2$, has to be supplied. This energy required E_b is related to the mass defect by

$$E_b = \Delta M c^2 \tag{13.8}$$

Example 13.3 Find the energy equivalent of one atomic mass unit, first in Joules and then in MeV. Using this, express the mass defect of $^{16}_{8}\text{O}$ in MeV/c^2.

Solution

1u = 1.6605 × 10^{-27} kg

To convert it into energy units, we multiply it by c^2 and find that

energy equivalent = 1.6605 × 10^{-27} × (2.9979 × 10^8)^2 kg m^2/s^2

$$= 1.4924 \times 10^{-10}\ \text{J}$$

$$= \frac{1.4924 \times 10^{-10}}{1.602 \times 10^{-19}}\ \text{eV}$$

$$= 0.9315 \times 10^{9}\ \text{eV}$$

$$= 931.5\ \text{MeV}$$

or, 1u = 931.5 MeV/c^2

For $^{16}_{8}\text{O}$, ΔM = 0.13691 u = 0.13691×931.5 MeV/c^2

= 127.5 MeV/c^2

The energy needed to separate $^{16}_{8}\text{O}$ into its constituents is thus 127.5 MeV/c^2.

EXAMPLE 13.3

If a certain number of neutrons and protons are brought together to form a nucleus of a certain charge and mass, an energy E_b will be released in the process. The energy E_b is called the *binding energy* of the nucleus. If we separate a nucleus into its nucleons, we would have to supply a total energy equal to E_b, to those particles. Although we cannot tear apart a nucleus in this way, the nuclear binding energy is still a convenient measure of how well a nucleus is held together. A more useful measure of the binding between the constituents of the nucleus is the *binding energy per nucleon*, E_{bn}, which is the ratio of the binding energy E_b of a nucleus to the number of the nucleons, A, in that nucleus:

$$E_{bn} = E_b / A \tag{13.9}$$

We can think of binding energy per nucleon as the average energy per nucleon needed to separate a nucleus into its individual nucleons.

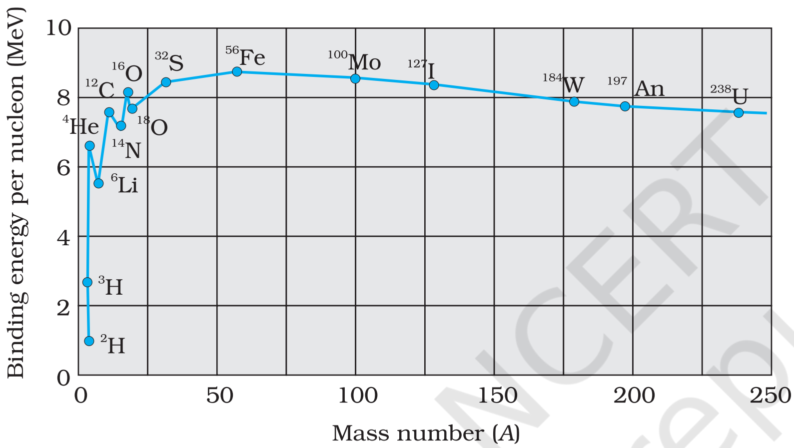


FIGURE 13.1 The binding energy per nucleon as a function of mass number.

Figure 13.1 is a plot of the binding energy per nucleon E_{bn} versus the mass number A for a large number of nuclei. We notice the following main features of the plot:

(i) the binding energy per nucleon, E_{bn}, is practically constant, i.e. practically independent of the atomic number for nuclei of middle mass number ($30 < A < 170$). The curve has a maximum of about 8.75 MeV for $A = 56$ and has a value of 7.6 MeV for $A = 238$.

(ii) E_{bn} is lower for both light nuclei ($A<30$) and heavy nuclei ($A>170$).

We can draw some conclusions from these two observations:

(i) The force is attractive and sufficiently strong to produce a binding energy of a few MeV per nucleon.

(ii) The constancy of the binding energy in the range $30 < A < 170$ is a consequence of the fact that the nuclear force is short-ranged. Consider a particular nucleon inside a sufficiently large nucleus. It will be under the influence of only some of its neighbours, which come within the range of the nuclear force. If any other nucleon is at a distance more than the range of the nuclear force from the particular nucleon it will have no influence on the binding energy of the nucleon under consideration. If a nucleon can have a maximum of p neighbours within the range of nuclear force, its binding energy would be proportional to p. Let the binding energy of the nucleus be pk, where k is a constant having the dimensions of energy. If we increase A by adding nucleons they will not change the binding energy of a nucleon inside. Since most of the nucleons in a large nucleus reside inside it and not on the surface, the change in binding energy per nucleon would be small. The binding energy per nucleon is a constant and is approximately equal to pk. The property that a given nucleon influences only nucleons close to it is also referred to as saturation property of the nuclear force.

(iii) A very heavy nucleus, say A = 240, has lower binding energy per nucleon compared to that of a nucleus with A = 120. Thus if a nucleus A = 240 breaks into two A = 120 nuclei, nucleons get more tightly bound. This implies energy would be released in the process. It has very important implications for energy production through *fission*, to be discussed later in Section 13.7.1.

(iv) Consider two very light nuclei ($A \leq 10$) joining to form a heavier nucleus. The binding energy per nucleon of the fused heavier nuclei is more than the binding energy per nucleon of the lighter nuclei. This means that the final system is more tightly bound than the initial system. Again energy would be released in such a process of *fusion*. This is the energy source of sun, to be discussed later in Section 13.7.2.

13.5 Nuclear Force

The force that determines the motion of atomic electrons is the familiar Coulomb force. In Section 13.4, we have seen that for average mass nuclei the binding energy per nucleon is approximately 8 MeV, which is much larger than the binding energy in atoms. Therefore, to bind a nucleus together there must be a strong attractive force of a totally different kind. It must be strong enough to overcome the repulsion between the (positively charged) protons and to bind both protons and neutrons into the tiny nuclear volume. We have already seen that the constancy of binding energy per nucleon can be understood in terms of its short-range. Many features of the nuclear binding force are summarised below. These are obtained from a variety of experiments carried out during 1930 to 1950.

(i) The nuclear force is much stronger than the Coulomb force acting between charges or the gravitational forces between masses. The nuclear binding force has to dominate over the Coulomb repulsive force between protons inside the nucleus. This happens only because the nuclear force is much stronger than the coulomb force. The gravitational force is much weaker than even Coulomb force.

(ii) The nuclear force between two nucleons falls rapidly to zero as their distance is more than a few femtometres. This leads to *saturation of forces* in a medium or a large-sized nucleus, which is the reason for the constancy of the binding energy per nucleon.

A rough plot of the potential energy between two nucleons as a function of distance is shown in the Fig. 13.2. The potential energy is a minimum at a distance r_0 of about 0.8 fm. This means that the force is attractive for distances larger than 0.8 fm and repulsive if they are separated by distances less than 0.8 fm.

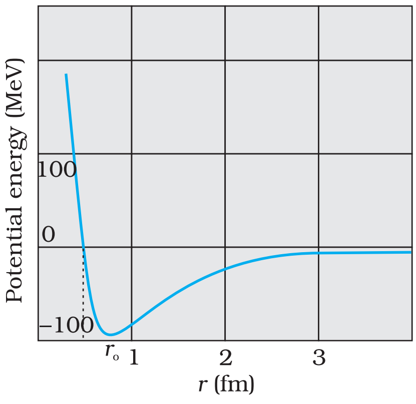


FIGURE 13.2 Potential energy of a pair of nucleons as a function of their separation. For a separation greater than r_0, the force is attractive and for separations less than r_0, the force is strongly repulsive.

(iii) The nuclear force between neutron-neutron, proton-neutron and proton-proton is approximately the same. The nuclear force does not depend on the electric charge.

Unlike Coulomb's law or the Newton's law of gravitation there is no simple mathematical form of the nuclear force.

13.6 Radioactivity

A. H. Becquerel discovered radioactivity in 1896 purely by accident. While studying the fluorescence and phosphorescence of compounds irradiated with visible light, Becquerel observed an interesting phenomenon. After illuminating some pieces of uranium-potassium sulphate with visible light, he wrapped them in black paper and separated the package from a photographic plate by a piece of silver. When, after several hours of exposure, the photographic plate was developed, it showed blackening due to something that must have been emitted by the compound and was able to penetrate both black paper and the silver.

Experiments performed subsequently showed that radioactivity was a nuclear phenomenon in which an unstable nucleus undergoes a decay. This is referred to as *radioactive decay*. Three types of radioactive decay occur in nature :

(i) α-decay in which a helium nucleus ${}^4_2\text{He}$ is emitted;

(ii) β-decay in which electrons or positrons (particles with the same mass as electrons, but with a charge exactly opposite to that of electron) are emitted;

(iii) γ-decay in which high energy (hundreds of keV or more) photons are emitted.

13.7 Nuclear Energy

The curve of binding energy per nucleon E_{bn}, given in Fig. 13.1, has a long flat middle region between A = 30 and A = 170. In this region the binding energy per nucleon is nearly constant (8.0 MeV). For the lighter nuclei region, A < 30, and for the heavier nuclei region, A > 170, the binding energy per nucleon is less than 8.0 MeV, as we have noted earlier. Now, the greater the binding energy, the less is the total mass of a bound system, such as a nucleus. Consequently, if nuclei with less total binding energy transform to nuclei with greater binding energy, there will be a net energy release. This is what happens when a heavy nucleus decays into two or more intermediate mass fragments (*fission*) or when light nuclei fuse into a havier nucleus (*fusion*.)

Exothermic chemical reactions underlie conventional energy sources such as coal or petroleum. Here the energies involved are in the range of electron volts. On the other hand, in a nuclear reaction, the energy release is of the order of MeV. Thus for the same quantity of matter, nuclear sources produce a million times more energy than a chemical source. Fission of 1 kg of uranium, for example, generates 10^{14} J of energy; compare it with burning of 1 kg of coal that gives 10^{7} J.

13.7.1 Fission

New possibilities emerge when we go beyond natural radioactive decays and study nuclear reactions by bombarding nuclei with other nuclear particles such as proton, neutron, α-particle, etc.

A most important neutron-induced nuclear reaction is fission. An example of fission is when a uranium isotope ${}^{235}_{92}U$ bombarded with a neutron breaks into two intermediate mass nuclear fragments

$${}^{1}_{0}n + {}^{235}_{92}U \rightarrow {}^{236}_{92}U \rightarrow {}^{144}_{56}Ba + {}^{89}_{36}Kr + 3\,{}^{1}_{0}n \quad (13.10)$$

The same reaction can produce other pairs of intermediate mass fragments

$${}^{1}_{0}n + {}^{235}_{92}U \rightarrow {}^{236}_{92}U \rightarrow {}^{133}_{51}Sb + {}^{99}_{41}Nb + 4\,{}^{1}_{0}n \quad (13.11)$$

Or, as another example,

$${}^{1}_{0}n + {}^{235}_{92}U \rightarrow {}^{140}_{54}Xe + {}^{94}_{38}Sr + 2\,{}^{1}_{0}n \quad (13.12)$$

The fragment products are radioactive nuclei; they emit β particles in succession to achieve stable end products.

The energy released (the Q value) in the fission reaction of nuclei like uranium is of the order of 200 MeV per fissioning nucleus. This is estimated as follows:

Let us take a nucleus with $A = 240$ breaking into two fragments each of $A = 120$. Then

E_{bn} for $A = 240$ nucleus is about 7.6 MeV,

E_{bn} for the two $A = 120$ fragment nuclei is about 8.5 MeV.

$\therefore$ Gain in binding energy for nucleon is about 0.9 MeV.

Hence the total gain in binding energy is 240×0.9 or 216 MeV.

The disintegration energy in fission events first appears as the kinetic energy of the fragments and neutrons. Eventually it is transferred to the surrounding matter appearing as heat. The source of energy in nuclear reactors, which produce electricity, is nuclear fission. The enormous energy released in an atom bomb comes from uncontrolled nuclear fission.

13.7.2 Nuclear fusion – energy generation in stars

When two light nuclei fuse to form a larger nucleus, energy is released, since the larger nucleus is more tightly bound, as seen from the binding energy curve in Fig.13.1. Some examples of such energy liberating nuclear fusion reactions are :

$${}^{1}_{1}H + {}^{1}_{1}H \rightarrow {}^{2}_{1}H + e^{+} + \nu + 0.42 \text{ MeV} \quad [13.13(a)]$$

$${}^{2}_{1}H + {}^{2}_{1}H \rightarrow {}^{3}_{2}He + n + 3.27 \text{ MeV} \quad [13.13(b)]$$

$${}^{2}_{1}H + {}^{2}_{1}H \rightarrow {}^{3}_{1}H + {}^{1}_{1}H + 4.03 \text{ MeV} \quad [13.13(c)]$$

In the first reaction, two protons combine to form a deuteron and a positron with a release of 0.42 MeV energy. In reaction [13.13(b)], two deuterons combine to form the light isotope of helium. In reaction (13.13c), two deuterons combine to form a triton and a proton. For fusion to take place, the two nuclei must come close enough so that attractive short-range nuclear force is able to affect them. However, since they are both positively charged particles, they experience coulomb repulsion. They, therefore, must have enough energy to overcome this coulomb barrier. The height of the barrier depends on the charges and radii of the two interacting nuclei. It can be shown, for example, that the barrier height for two protons is ~ 400 keV, and is higher for nuclei with higher charges. We can estimate the temperature at which two protons in a proton gas would (averagely) have enough energy to overcome the coulomb barrier:

$(3/2)k\,T = K \simeq 400$ keV, which gives $T \sim 3 \times 10^9$ K.

When fusion is achieved by raising the temperature of the system so that particles have enough kinetic energy to overcome the coulomb repulsive behaviour, it is called *thermonuclear fusion*.

Thermonuclear fusion is the source of energy output in the interior of stars. The interior of the sun has a temperature of 1.5×10^7 K, which is considerably less than the estimated temperature required for fusion of particles of average energy. Clearly, fusion in the sun involves protons whose energies are much above the average energy.

The fusion reaction in the sun is a multi-step process in which the hydrogen is burned into helium. Thus, the fuel in the sun is the hydrogen in its core. The *proton-proton* (p, p) *cycle* by which this occurs is represented by the following sets of reactions:

$${}_1^1\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_1^2\mathrm{H} + e^+ + \nu + 0.42\ \mathrm{MeV} \qquad \text{(i)}$$

$$e^+ + e^- \rightarrow \gamma + \gamma + 1.02\ \mathrm{MeV} \qquad \text{(ii)}$$

$${}_1^2\mathrm{H} + {}_1^1\mathrm{H} \rightarrow {}_2^3\mathrm{He} + \gamma + 5.49\ \mathrm{MeV} \qquad \text{(iii)}$$

$${}_2^3\mathrm{He} + {}_2^3\mathrm{He} \rightarrow {}_2^4\mathrm{He} + {}_1^1\mathrm{H} + {}_1^1\mathrm{H} + 12.86\ \mathrm{MeV} \qquad \text{(iv)} \qquad (13.14)$$

For the fourth reaction to occur, the first three reactions must occur twice, in which case two light helium nuclei unite to form ordinary helium nucleus. If we consider the combination 2(i) + 2(ii) + 2(iii) +(iv), the net effect is

$$4\,{}_1^1\mathrm{H} + 2e^- \rightarrow {}_2^4\mathrm{He} + 2\nu + 6\gamma + 26.7\ \mathrm{MeV}$$

$$\text{or} \quad (4\,{}_1^1\mathrm{H} + 4e^-) \rightarrow ({}_2^4\mathrm{He} + 2e^-) + 2\nu + 6\gamma + 26.7\ \mathrm{MeV} \qquad (13.15)$$

Thus, four hydrogen atoms combine to form an ${}_2^4\mathrm{He}$ atom with a release of 26.7 MeV of energy.

Helium is not the only element that can be synthesized in the interior of a star. As the hydrogen in the core gets depleted and becomes helium, the core starts to cool. The star begins to collapse under its own gravity which increases the temperature of the core. If this temperature increases to about 10^8 K, fusion takes place again, this time of helium nuclei into carbon. This kind of process can generate through fusion higher and higher mass number elements. But elements more massive than those near the peak of the binding energy curve in Fig. 13.1 cannot be so produced.

The age of the sun is about 5×10^9 y and it is estimated that there is enough hydrogen in the sun to keep it going for another 5 billion years. After that, the hydrogen burning will stop and the sun will begin to cool and will start to collapse under gravity, which will raise the core temperature. The outer envelope of the sun will expand, turning it into the so called *red giant*.

13.7.3 Controlled thermonuclear fusion

The natural thermonuclear fusion process in a star is replicated in a thermonuclear fusion device. In controlled fusion reactors, the aim is to generate steady power by heating the nuclear fuel to a temperature in the range of 10^8 K. At these temperatures, the fuel is a mixture of positive ions and electrons (plasma). The challenge is to confine this plasma, since no container can stand such a high temperature. Several countries around the world including India are developing techniques in this connection. If successful, fusion reactors will hopefully supply almost unlimited power to humanity.

Example 13.4 Answer the following questions:

(a) Are the equations of nuclear reactions (such as those given in Section 13.7) 'balanced' in the sense a chemical equation (e.g., $2H_2 + O_2 \rightarrow 2\ H_2O$) is? If not, in what sense are they balanced on both sides?

(b) If both the number of protons and the number of neutrons are conserved in each nuclear reaction, in what way is mass converted into energy (or vice-versa) in a nuclear reaction?

(c) A general impression exists that mass-energy interconversion takes place only in nuclear reaction and never in chemical reaction. This is strictly speaking, incorrect. Explain.

Solution

(a) A chemical equation is balanced in the sense that the number of atoms of each element is the same on both sides of the equation. A chemical reaction merely alters the original combinations of atoms. In a nuclear reaction, elements may be transmuted. Thus, the number of atoms of each element is not necessarily conserved in a nuclear reaction. However, the number of protons and the number of neutrons are both separately conserved in a nuclear reaction. [Actually, even this is not strictly true in the realm of very high energies – what is strictly conserved is the total charge and total 'baryon number'. We need not pursue this matter here.]

In nuclear reactions (e.g., Eq. 13.10), the number of protons and the number of neutrons are the same on the two sides of the equation.

(b) We know that the binding energy of a nucleus gives a negative contribution to the mass of the nucleus (mass defect). Now, since proton number and neutron number are conserved in a nuclear reaction, the total rest mass of neutrons and protons is the same on either side of a reaction. But the total binding energy of nuclei on the left side need not be the same as that on the right hand side. The difference in these binding energies appears as energy released or absorbed in a nuclear reaction. Since binding energy

EXAMPLE 13.4

EXAMPLE 13.4

contributes to mass, we say that the difference in the total mass of nuclei on the two sides get converted into energy or vice-versa. It is in these sense that a nuclear reaction is an example of mass-energy interconversion.

(c) From the point of view of mass-energy interconversion, a chemical reaction is similar to a nuclear reaction *in principle.* The energy released or absorbed in a chemical reaction can be traced to the difference in chemical (not nuclear) binding energies of atoms and molecules on the two sides of a reaction. Since, strictly speaking, chemical binding energy also gives a negative contribution (mass defect) to the total mass of an atom or molecule, we can equally well say that the difference in the total mass of atoms or molecules, on the two sides of the chemical reaction gets converted into energy or vice-versa. However, the mass defects involved in a chemical reaction are almost a million times smaller than those in a nuclear reaction.This is the reason for the general impression, (which is *incorrect*) that mass-energy interconversion does not take place in a chemical reaction.

SUMMARY

1. An atom has a nucleus. The nucleus is positively charged. The radius of the nucleus is smaller than the radius of an atom by a factor of 10^4. More than 99.9% mass of the atom is concentrated in the nucleus.
2. On the atomic scale, mass is measured in atomic mass units (u). By definition, 1 atomic mass unit (1u) is $1/12^{th}$ mass of one atom of ^{12}C;
   $1u = 1.660563 \times 10^{-27}$ kg.
3. A nucleus contains a neutral particle called neutron. Its mass is almost the same as that of proton
4. The atomic number Z is the number of protons in the atomic nucleus of an element. The mass number A is the total number of protons and neutrons in the atomic nucleus; $A = Z+N$; Here N denotes the number of neutrons in the nucleus.

   A nuclear species or a nuclide is represented as $^A_Z X$, where X is the chemical symbol of the species.

   Nuclides with the same atomic number Z, but different neutron number N are called *isotopes*. Nuclides with the same A are *isobars* and those with the same N are *isotones*.

   Most elements are mixtures of two or more isotopes. The atomic mass of an element is a weighted average of the masses of its isotopes and calculated in accordance to the relative abundances of the isotopes.
5. A nucleus can be considered to be spherical in shape and assigned a radius. Electron scattering experiments allow determination of the nuclear radius; it is found that radii of nuclei fit the formula

   $R = R_0 A^{1/3}$,

   where R_0 = a constant = 1.2 fm. This implies that the nuclear density is independent of A. It is of the order of 10^{17} kg/m^3.
6. Neutrons and protons are bound in a nucleus by the short-range strong nuclear force. The nuclear force does not distinguish between neutron and proton.

Reprint 2025-26

7. The nuclear mass M is always less than the total mass, Σm, of its constituents. The difference in mass of a nucleus and its constituents is called the *mass defect*,
   $\Delta M = (Z\,m_p + (A - Z)m_n) - M$
   Using Einstein's mass energy relation, we express this mass difference in terms of energy as
   $\Delta E_b = \Delta M\,c^2$
   The energy ΔE_b represents the *binding energy* of the nucleus. In the mass number range A = 30 to 170, the binding energy per nucleon is nearly constant, about 8 MeV/nucleon.
8. Energies associated with nuclear processes are about a million times larger than chemical process.
9. The Q-value of a nuclear process is
   Q = final kinetic energy – initial kinetic energy.
   Due to conservation of mass-energy, this is also,
   Q = (sum of initial masses – sum of final masses)c^2
10. Radioactivity is the phenomenon in which nuclei of a given species transform by giving out α or β or γ rays; α-rays are helium nuclei; β-rays are electrons. γ-rays are electromagnetic radiation of wavelengths shorter than X-rays.
11. Energy is released when less tightly bound nuclei are transmuted into more tightly bound nuclei. In fission, a heavy nucleus like ${}^{235}_{92}\mathrm{U}$ breaks into two smaller fragments, e.g., ${}^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{133}_{51}\mathrm{Sb} + {}^{99}_{41}\mathrm{Nb} + 4\,{}^{1}_{0}\mathrm{n}$
12. In fusion, lighter nuclei combine to form a larger nucleus. Fusion of hydrogen nuclei into helium nuclei is the source of energy of all stars including our sun.

Physical Quantity	Symbol	Dimensions	Units	Remarks
Atomic mass unit		[M]	u	Unit of mass for expressing atomic or nuclear masses. One atomic mass unit equals 1/12th of the mass of ^{12}C atom.
Disintegration or decay constant	λ	[T^{-1}]	s^{-1}	
Half-life	$T_{1/2}$	[T]	s	Time taken for the decay of one-half of the initial number of nuclei present in a radioactive sample.
Mean life	τ	[T]	s	Time at which number of nuclei has been reduced to e^{-1} of its initial value
Activity of a radio-active sample	R	[T^{-1}]	Bq	Measure of the activity of a radioactive source.

POINTS TO PONDER

1. The density of nuclear matter is independent of the size of the nucleus. The mass density of the atom does not follow this rule.
2. The radius of a nucleus determined by electron scattering is found to be slightly different from that determined by alpha-particle scattering. This is because electron scattering senses the charge distribution of the nucleus, whereas alpha and similar particles sense the nuclear matter.
3. After Einstein showed the equivalence of mass and energy, $E = mc^2$, we cannot any longer speak of separate laws of conservation of mass and conservation of energy, but we have to speak of a unified law of conservation of mass and energy. The most convincing evidence that this principle operates in nature comes from nuclear physics. It is central to our understanding of nuclear energy and harnessing it as a source of power. Using the principle, Q of a nuclear process (decay or reaction) can be expressed also in terms of initial and final masses.
4. The nature of the binding energy (per nucleon) curve shows that exothermic nuclear reactions are possible, when two light nuclei fuse or when a heavy nucleus undergoes fission into nuclei with intermediate mass.
5. For fusion, the light nuclei must have sufficient initial energy to overcome the coulomb potential barrier. That is why fusion requires very high temperatures.
6. Although the binding energy (per nucleon) curve is smooth and slowly varying, it shows peaks at nuclides like ^{4}He, ^{16}O etc. This is considered as evidence of atom-like shell structure in nuclei.
7. Electron and positron are a particle-antiparticle pair. They are identical in mass; their charges are equal in magnitude and opposite. (It is found that when an electron and a positron come together, they annihilate each other giving energy in the form of gamma-ray photons.)
8. Radioactivity is an indication of the instability of nuclei. Stability requires the ratio of neutron to proton to be around 1:1 for light nuclei. This ratio increases to about 3:2 for heavy nuclei. (More neutrons are required to overcome the effect of repulsion among the protons.) Nuclei which are away from the stability ratio, i.e., nuclei which have an excess of neutrons or protons are unstable. In fact, only about 10% of knon isotopes (of all elements), are stable. Others have been either artificially produced in the laboratory by bombarding α, p, d, n or other particles on targets of stable nuclear species or identified in astronomical observations of matter in the universe.

EXERCISES

You may find the following data useful in solving the exercises:

$e = 1.6\times10^{-19}$C $\quad$ $N = 6.023\times10^{23}$ per mole

$1/(4\pi\varepsilon_0) = 9\times10^9$ N m^2/C^2 $\quad$ $k = 1.381\times10^{-23}$J K^{-1}

1 MeV $= 1.6\times10^{-13}$J $\quad$ 1 u = 931.5 MeV/c^2

1 year $= 3.154\times10^7$ s

m_H = 1.007825 u $\quad$ m_n = 1.008665 u

$m({}^4_2\text{He})$ = 4.002603 u $\quad$ m_e = 0.000548 u

13.1 Obtain the binding energy (in MeV) of a nitrogen nucleus $\left({}^{14}_{7}\text{N}\right)$, given $m\left({}^{14}_{7}\text{N}\right)$=14.00307 u

13.2 Obtain the binding energy of the nuclei ${}^{56}_{26}\text{Fe}$ and ${}^{209}_{83}\text{Bi}$ in units of MeV from the following data:

$m\,({}^{56}_{26}\text{Fe})$ = 55.934939 u $\quad$ $m\,({}^{209}_{83}\text{Bi})$ = 208.980388 u

13.3 A given coin has a mass of 3.0 g. Calculate the nuclear energy that would be required to separate all the neutrons and protons from each other. For simplicity assume that the coin is entirely made of ${}^{63}_{29}\text{Cu}$ atoms (of mass 62.92960 u).

13.4 Obtain approximately the ratio of the nuclear radii of the gold isotope ${}^{197}_{79}\text{Au}$ and the silver isotope ${}^{107}_{47}\text{Ag}$.

13.5 The Q value of a nuclear reaction $A + b \rightarrow C + d$ is defined by

$Q = [\, m_A + m_b - m_C - m_d]c^2$

where the masses refer to the respective nuclei. Determine from the given data the Q-value of the following reactions and state whether the reactions are exothermic or endothermic.

(i) ${}^1_1\text{H}+{}^3_1\text{H} \rightarrow {}^2_1\text{H}+{}^2_1\text{H}$

(ii) ${}^{12}_{6}\text{C}+{}^{12}_{6}\text{C} \rightarrow {}^{20}_{10}\text{Ne}+{}^4_2\text{He}$

Atomic masses are given to be

$m\,({}^2_1\text{H})$ = 2.014102 u

$m\,({}^3_1\text{H})$ = 3.016049 u

$m\,({}^{12}_{6}\text{C})$ = 12.000000 u

$m\,({}^{20}_{10}\text{Ne})$ = 19.992439 u

13.6 Suppose, we think of fission of a ${}^{56}_{26}\text{Fe}$ nucleus into two equal fragments, ${}^{28}_{13}\text{Al}$. Is the fission energetically possible? Argue by working out Q of the process. Given $m\,({}^{56}_{26}\text{Fe})$ = 55.93494 u and $m\,({}^{28}_{13}\text{Al})$ = 27.98191 u.

13.7 The fission properties of ${}^{239}_{94}\mathrm{Pu}$ are very similar to those of ${}^{235}_{92}\mathrm{U}$. The average energy released per fission is 180 MeV. How much energy, in MeV, is released if all the atoms in 1 kg of pure ${}^{239}_{94}\mathrm{Pu}$ undergo fission?

13.8 How long can an electric lamp of 100W be kept glowing by fusion of 2.0 kg of deuterium? Take the fusion reaction as

$${}^{2}_{1}\mathrm{H}+{}^{2}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He}+\mathrm{n}+3.27 \text{ MeV}$$

13.9 Calculate the height of the potential barrier for a head on collision of two deuterons. (Hint: The height of the potential barrier is given by the Coulomb repulsion between the two deuterons when they just touch each other. Assume that they can be taken as hard spheres of radius 2.0 fm.)

13.10 From the relation $R = R_0 A^{1/3}$, where R_0 is a constant and A is the mass number of a nucleus, show that the nuclear matter density is nearly constant (i.e. independent of A).

Chapter Fourteen

SEMICONDUCTOR ELECTRONICS: MATERIALS, DEVICES AND SIMPLE CIRCUITS

14.1 Introduction

Devices in which a controlled flow of electrons can be obtained are the basic *building blocks* of all the electronic circuits. Before the discovery of transistor in 1948, such devices were mostly vacuum tubes (also called valves) like the vacuum diode which has two electrodes, viz., anode (often called plate) and cathode; triode which has three electrodes – cathode, plate and grid; tetrode and pentode (respectively with 4 and 5 electrodes). In a vacuum tube, the electrons are supplied by a heated cathode and the controlled flow of these electrons *in vacuum* is obtained by varying the voltage between its different electrodes. Vacuum is required in the inter-electrode space; otherwise the moving electrons may lose their energy on collision with the air molecules in their path. In these devices the electrons can flow only from the cathode to the anode (i.e., only in one direction). Therefore, such devices are generally referred to as *valves*. These vacuum tube devices are bulky, consume high power, operate generally at high voltages (~100 V) and have limited life and low reliability. The seed of the development of modern *solid-state semiconductor electronics* goes back to 1930's when it was realised that some solid-state semiconductors and their junctions offer the possibility of controlling the number and the direction of flow of charge carriers through them. Simple excitations like light, heat or small applied voltage can change the number of mobile charges in a semiconductor. Note that the supply and flow of charge carriers in the semiconductor devices are *within the solid itself*, while in the earlier vacuum tubes/valves, the mobile electrons were obtained from a heated cathode and they were made to flow in an *evacuated* space or vacuum. No external heating or large evacuated space is required by the semiconductor devices. They are small in size, consume low power, operate at low voltages and have long life and high reliability. Even the Cathode Ray Tubes (CRT) used in television and computer monitors which work on the principle of vacuum tubes are being replaced by Liquid Crystal Display (LCD) monitors with supporting solid state electronics. Much before the full implications of the semiconductor devices was formally understood, a naturally occurring crystal of *galena* (Lead sulphide, PbS) with a metal point contact attached to it was used as *detector* of radio waves.

In the following sections, we will introduce the basic concepts of semiconductor physics and discuss some semiconductor devices like junction diodes (a 2-electrode device) and bipolar junction transistor (a 3-electrode device). A few circuits illustrating their applications will also be described.

14.2 Classification of Metals, Conductors and Semiconductors

On the basis of conductivity

On the basis of the relative values of electrical conductivity (σ) or resistivity ($\rho = 1/\sigma$), the solids are broadly classified as:

(i) ***Metals:*** They possess very low resistivity (or high conductivity).

$\rho \sim 10^{-2} - 10^{-8}\ \Omega$ m

$\sigma \sim 10^{2} - 10^{8}$ S m^{-1}

(ii) ***Semiconductors:*** They have resistivity or conductivity intermediate to metals and insulators.

$\rho \sim 10^{-5} - 10^{6}\ \Omega$ m

$\sigma \sim 10^{5} - 10^{-6}$ S m^{-1}

(iii) ***Insulators:*** They have high resistivity (or low conductivity).

$\rho \sim 10^{11} - 10^{19}\ \Omega$ m

$\sigma \sim 10^{-11} - 10^{-19}$ S m^{-1}

The values of ρ and σ given above are indicative of magnitude and could well go outside the ranges as well. Relative values of the resistivity are not the only criteria for distinguishing metals, insulators and semiconductors from each other. There are some other differences, which will become clear as we go along in this chapter.

Our interest in this chapter is in the study of semiconductors which could be:

(i) *Elemental semiconductors*: Si and Ge

(ii) *Compound semiconductors*: Examples are:

- Inorganic: CdS, GaAs, CdSe, InP, etc.
- Organic: anthracene, doped pthalocyanines, etc.
- Organic polymers: polypyrrole, polyaniline, polythiophene, etc.

Most of the currently available semiconductor devices are based on elemental semiconductors Si or Ge and compound *inorganic* semiconductors. However, after 1990, a few semiconductor devices using

Reprint 2025-26

organic semiconductors and semiconducting polymers have been developed signalling the birth of a futuristic technology of polymer-electronics and molecular-electronics. In this chapter, we will restrict ourselves to the study of inorganic semiconductors, particularly elemental semiconductors Si and Ge. The general concepts introduced here for discussing the elemental semiconductors, by-and-large, apply to most of the compound semiconductors as well.

On the basis of energy bands

According to the Bohr atomic model, in an *isolated atom* the energy of any of its electrons is decided by the orbit in which it revolves. But when the atoms come together to form a solid they are close to each other. So the outer orbits of electrons from neighbouring atoms would come very close or could even overlap. This would make the nature of electron motion in a solid very different from that in an isolated atom.

Inside the crystal each electron has a unique position and no two electrons see exactly the same pattern of surrounding charges. Because of this, each electron will have a different *energy level*. These different energy levels with continuous energy variation form what are called *energy bands*. The energy band which includes the energy levels of the valence electrons is called the *valence band*. The energy band above the valence band is called the *conduction band*. With no external energy, all the valence electrons will reside in the valence band. If the lowest level in the conduction band happens to be lower than the highest level of the valence band, the electrons from the valence band can easily move into the conduction band. Normally the conduction band is empty. But when it overlaps on the valence band electrons can move freely into it. This is the case with metallic conductors.

If there is some gap between the conduction band and the valence band, electrons in the valence band all remain bound and no free electrons are available in the conduction band. This makes the material an insulator. But some of the electrons from the valence band may gain external energy to cross the gap between the conduction band and the valence band. Then these electrons will move into the conduction band. At the same time they will create vacant energy levels in the valence band where other valence electrons can move. Thus the process creates the possibility of conduction due to electrons in conduction band as well as due to vacancies in the valence band.

Let us consider what happens in the case of Si or Ge crystal containing N atoms. For Si, the outermost orbit is the third orbit ($n = 3$), while for Ge it is the fourth orbit ($n = 4$). The number of electrons in the outermost orbit is 4 (2s and 2p electrons). Hence, the total number of outer electrons in the crystal is 4N. The maximum possible number of electrons in the outer orbit is 8 (2s + 6p electrons). So, for the 4N valence electrons there are 8N available energy states. These 8N discrete energy levels can either form a continuous band or they may be grouped in different bands depending upon the distance between the atoms in the crystal (see box on Band Theory of Solids).

At the distance between the atoms in the crystal lattices of Si and Ge, the energy band of these 8N states is split apart into two which are separated by an *energy gap* E_g (Fig. 14.1). The lower band which is completely occupied by the 4N valence electrons at temperature of absolute zero is the *valence band*. The other band consisting of 4N energy states, called the *conduction band*, is completely empty at absolute zero.

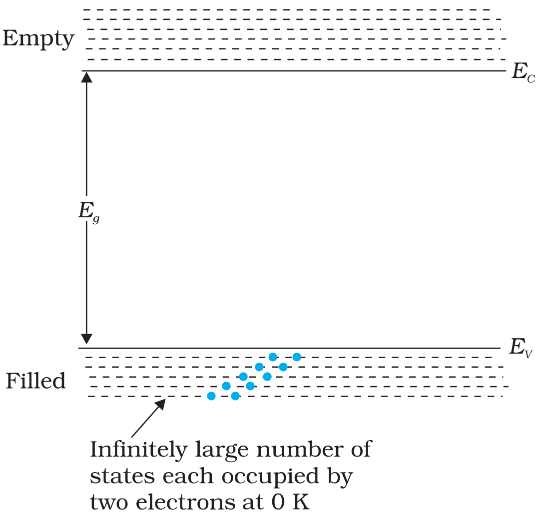


FIGURE 14.1 The energy band positions in a semiconductor at 0 K. The upper band, called the conduction band, consists of infinitely large number of closely spaced energy states. The lower band, called the valence band, consists of closely spaced completely filled energy states.

The lowest energy level in the conduction band is shown as E_C and highest energy level in the valence band is shown as E_V. Above E_C and below E_V there are a large number of closely spaced energy levels, as shown in Fig. 14.1.

The gap between the top of the valence band and bottom of the conduction band is called the *energy band gap* (Energy gap E_g). It may be large, small, or zero, depending upon the material. These different situations, are depicted in Fig. 14.2 and discussed below:

Case I: This refers to a situation, as shown in Fig. 14.2(a). One can have a metal either when the conduction band is partially filled and the balanced band is partially empty or when the conduction and valance bands overlap. When there is overlap electrons from valence band can easily move into the conduction band. This situation makes a large number of electrons available for electrical conduction. When the valence band is partially empty, electrons from its lower level can move to higher level making conduction possible. Therefore, the resistance of such materials is low or the conductivity is high.

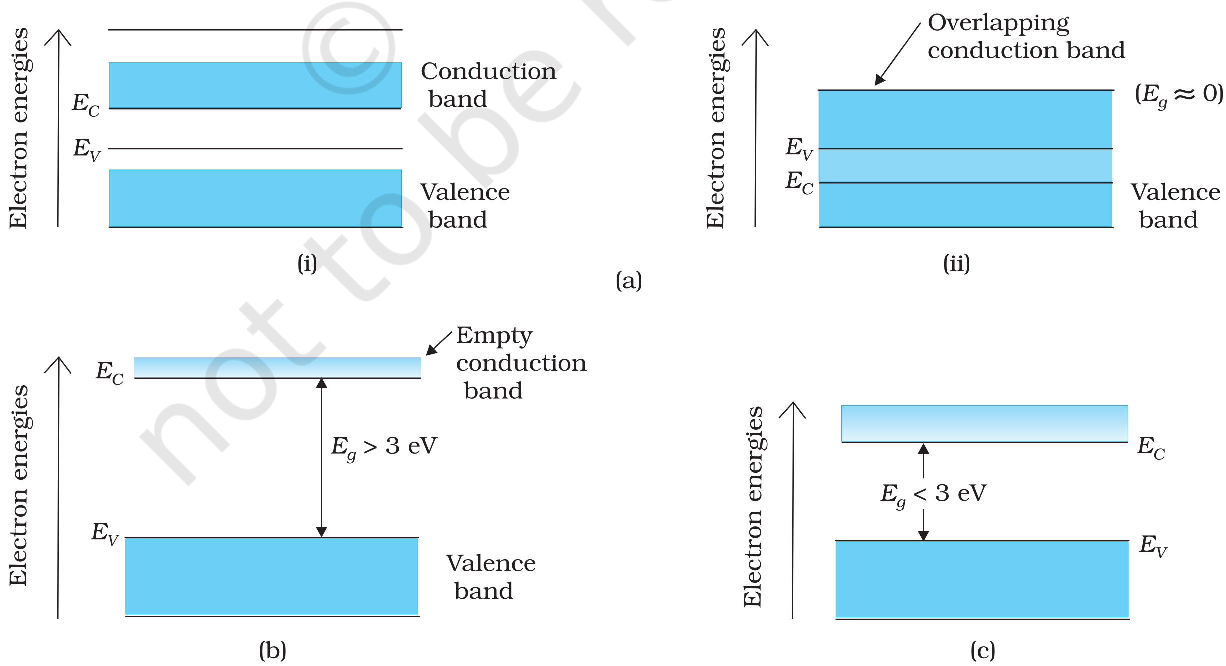


FIGURE 14.2 Difference between energy bands of (a) metals, (b) insulators and (c) semiconductors.

Case II: In this case, as shown in Fig. 14.2(b), a large band gap E_g exists (E_g > 3 eV). There are no electrons in the conduction band, and therefore no electrical conduction is possible. Note that the energy gap is so large that electrons cannot be excited from the valence band to the conduction band by thermal excitation. This is the case of *insulators*.

Case III: This situation is shown in Fig. 14.2(c). Here a finite but small band gap (E_g < 3 eV) exists. Because of the small band gap, at room temperature some electrons from valence band can acquire enough energy to cross the energy gap and enter the *conduction band*. These electrons (though small in numbers) can move in the conduction band. Hence, the resistance of *semiconductors* is not as high as that of the insulators.

In this section we have made a broad classification of metals, conductors and semiconductors. In the section which follows you will learn the conduction process in semiconductors.

14.3 Intrinsic Semiconductor

We shall take the most common case of Ge and Si whose lattice structure is shown in Fig. 14.3. These structures are called the diamond-like structures. Each atom is surrounded by four nearest neighbours. We know that Si and Ge have four valence electrons. In its crystalline structure, every Si or Ge atom tends *to share* one of its four valence electrons with each of its four nearest neighbour atoms, and also *to take share* of one electron from each such neighbour. These shared electron pairs are referred to as forming a *covalent bond* or simply a *valence bond*. The two shared electrons can be assumed to shuttle back-and-forth between the associated atoms holding them together strongly. Figure 14.4 schematically shows the 2-dimensional representation of Si or Ge structure shown in Fig. 14.3 which overemphasises the covalent bond. It shows an idealised picture in which no bonds are broken (all bonds are intact). Such a situation arises at low temperatures. As the temperature increases, more thermal energy becomes available to these electrons and some of these electrons may break–away (becoming *free* electrons contributing to conduction). The thermal energy effectively ionises only a few atoms in the crystalline lattice and creates a *vacancy* in the bond as shown in Fig. 14.5(a). The neighbourhood, from which the free electron (with charge $-q$) has come out leaves a vacancy with an effective charge ($+q$). This *vacancy* with the effective positive electronic charge is called a *hole*. The hole behaves as an *apparent free particle* with effective positive charge.

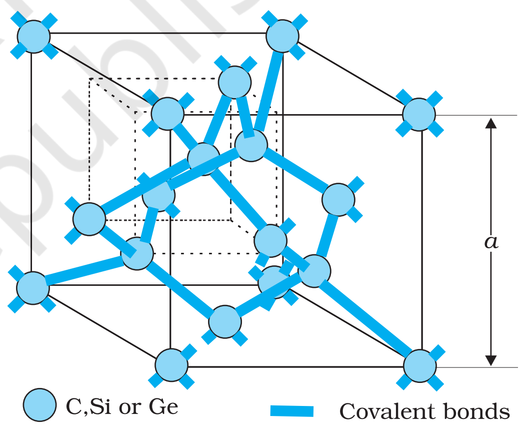


FIGURE 14.3 Three-dimensional diamond-like crystal structure for Carbon, Silicon or Germanium with respective lattice spacing a equal to 3.56, 5.43 and 5.66 Å.

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. That is

$$n_e = n_h = n_i \qquad (14.1)$$

where n_i is called intrinsic carrier concentration.

Semiconductors posses the unique property in which, apart from electrons, the holes also move. Suppose there is a hole at site 1 as shown

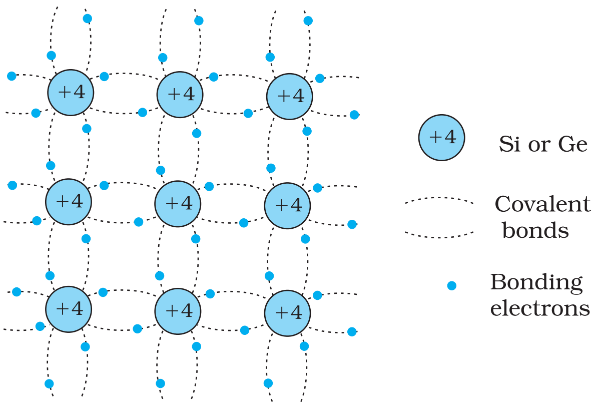


FIGURE 14.4 Schematic two-dimensional representation of Si or Ge structure showing covalent bonds at low temperature (all bonds intact). +4 symbol indicates inner cores of Si or Ge.

in Fig. 14.5(a). The movement of holes can be visualised as shown in Fig. 14.5(b). An electron from the covalent bond at site 2 may jump to the vacant site 1 (hole). Thus, after such a jump, the hole is at site 2 and the site 1 has now an electron. Therefore, apparently, the hole has moved from site 1 to site 2. Note that the electron originally set free [Fig. 14.5(a)] is not involved in this process of hole motion. The free electron moves completely independently as conduction electron and gives rise to an electron current, I_e under an applied electric field. Remember that the motion of hole is only a convenient way of describing the actual motion of *bound* electrons, whenever there is an empty bond anywhere in the crystal. Under the action of an electric field, these holes move towards negative potential giving the hole current, I_h. The total current, I is thus the sum of the electron current I_e and the hole current I_h:

$$I = I_e + I_h \tag{14.2}$$

It may be noted that apart from the *process of generation* of conduction electrons and holes, a simultaneous *process of recombination* occurs in which the electrons *recombine* with the holes. At equilibrium, the rate of generation is equal to the rate of recombination of charge carriers. The recombination occurs due to an electron colliding with a hole.

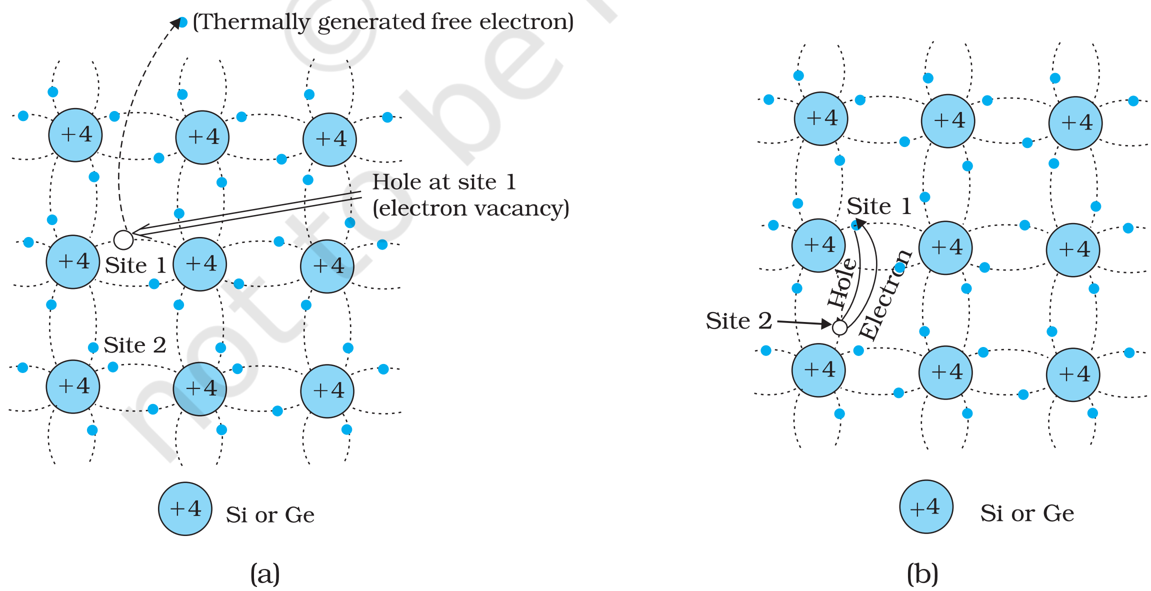


FIGURE 14.5 (a) Schematic model of generation of hole at site 1 and conduction electron due to thermal energy at moderate temperatures. (b) Simplified representation of possible thermal motion of a hole. The electron from the lower left hand covalent bond (site 2) goes to the earlier hole site1, leaving a hole at its site indicating an apparent movement of the hole from site 1 to site 2.

An intrinsic semiconductor will behave like an insulator at $T = 0$ K as shown in Fig. 14.6(a). It is the thermal energy at higher temperatures ($T > 0$K), which excites some electrons from the valence band to the conduction band. These thermally excited electrons at $T > 0$ K, partially occupy the conduction band. Therefore, the energy-band diagram of an intrinsic semiconductor will be as shown in Fig. 14.6(b). Here, some electrons are shown in the conduction band. These have come from the valence band leaving equal number of holes there.

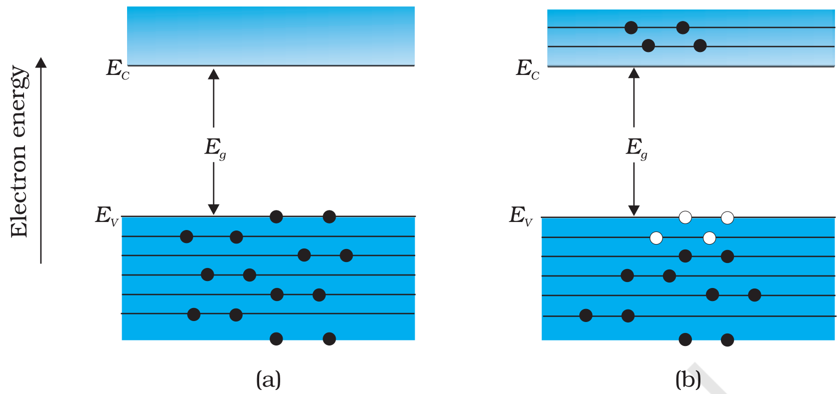


FIGURE 14.6 (a) An intrinsic semiconductor at $T = 0$ K behaves like insulator. (b) At $T > 0$ K, four thermally generated electron-hole pairs. The filled circles (•) represent electrons and empty circles (○) represent holes.

Example 14.1 C, Si and Ge have same lattice structure. Why is C insulator while Si and Ge intrinsic semiconductors?

Solution The 4 bonding electrons of C, Si or Ge lie, respectively, in the second, third and fourth orbit. Hence, energy required to take out an electron from these atoms (i.e., ionisation energy E_g) will be least for Ge, followed by Si and highest for C. Hence, number of free electrons for conduction in Ge and Si are significant but negligibly small for C.

EXAMPLE 14.1

14.4 Extrinsic Semiconductor

The conductivity of an intrinsic semiconductor depends on its temperature, but at room temperature its conductivity is very low. As such, no important electronic devices can be developed using these semiconductors. Hence there is a necessity of improving their conductivity. This can be done by making use of impurities.

When a small amount, say, a few parts per million (ppm), of a suitable impurity is added to the pure semiconductor, the conductivity of the semiconductor is increased manifold. Such materials are known as *extrinsic semiconductors* or *impurity semiconductors*. The deliberate addition of a desirable impurity is called *doping* and the impurity atoms are called *dopants*. Such a material is also called a *doped semiconductor*. The dopant has to be such that it does not distort the original pure semiconductor lattice. It occupies only a very few of the original semiconductor atom sites in the crystal. A necessary condition to attain this is that the sizes of the dopant and the semiconductor atoms should be nearly the same.

There are two types of dopants used in doping the tetravalent Si or Ge:

(i) Pentavalent (valency 5); like Arsenic (As), Antimony (Sb), Phosphorous (P), etc.

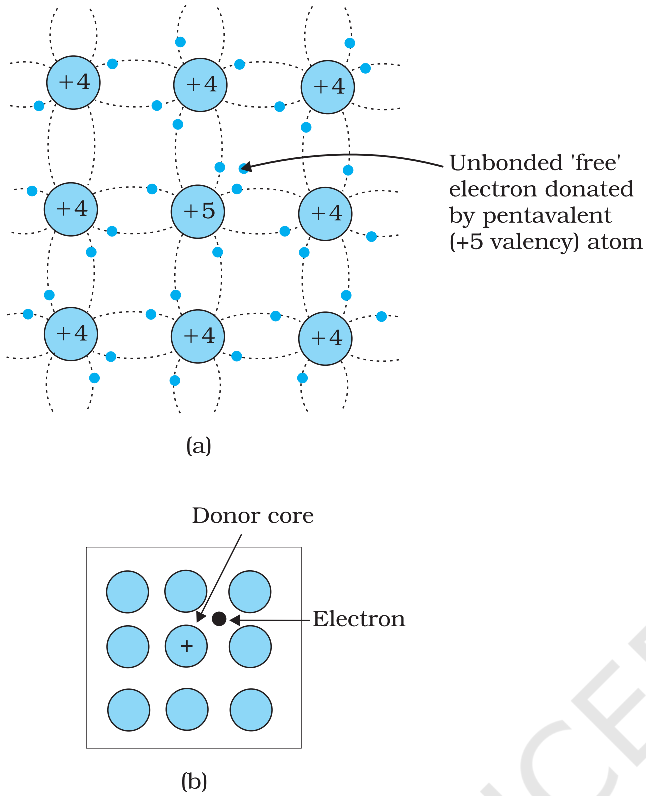


FIGURE 14.7 (a) Pentavalent donor atom (As, Sb, P, etc.) doped for tetravalent Si or Ge giving n-type semiconductor, and (b) Commonly used schematic representation of n-type material which shows only the fixed cores of the substituent donors with one additional effective positive charge and its associated extra electron.

(ii) Trivalent (valency 3); like Indium (In), Boron (B), Aluminium (Al), etc.

We shall now discuss how the doping changes the number of charge carriers (and hence the conductivity) of semiconductors. Si or Ge belongs to the fourth group in the Periodic table and, therefore, we choose the dopant element from nearby fifth or third group, expecting and taking care that the size of the dopant atom is nearly the same as that of Si or Ge. Interestingly, the pentavalent and trivalent dopants in Si or Ge give two entirely different types of semiconductors as discussed below.

(i) n-type semiconductor

Suppose we dope Si or Ge with a pentavalent element as shown in Fig. 14.7. When an atom of +5 valency element occupies the position of an atom in the crystal lattice of Si, four of its electrons bond with the four silicon neighbours while the fifth remains very weakly bound to its parent atom. This is because the four electrons participating in bonding are seen as part of the effective core of the atom by the fifth electron. As a result the ionisation energy required to set this electron free is very small and even at room temperature it will be free to move in the lattice of the semiconductor. For example, the energy required is ~ 0.01 eV for germanium, and 0.05 eV for silicon, to separate this electron from its atom. This is in contrast to the energy required to jump the forbidden band (about 0.72 eV for germanium and about 1.1 eV for silicon) at room temperature in the intrinsic semiconductor. Thus, the pentavalent dopant is donating one extra electron for conduction and hence is known as *donor* impurity. The number of electrons made available for conduction by dopant atoms depends strongly upon the doping level and is independent of any increase in ambient temperature. On the other hand, the number of free electrons (with an equal number of holes) generated by Si atoms, increases weakly with temperature.

In a doped semiconductor the total number of conduction electrons n_e is due to the electrons contributed by donors and those generated intrinsically, while the total number of holes n_h is only due to the holes from the intrinsic source. But the rate of recombination of holes would increase due to the increase in the number of electrons. As a result, the number of holes would get reduced further.

Thus, with proper level of doping the number of conduction electrons can be made much larger than the number of holes. Hence in an extrinsic semiconductor doped with pentavalent impurity, electrons become the *majority carriers* and holes the *minority carriers*. These semiconductors are, therefore, known as n-*type semiconductors*. For n-type semiconductors, we have,

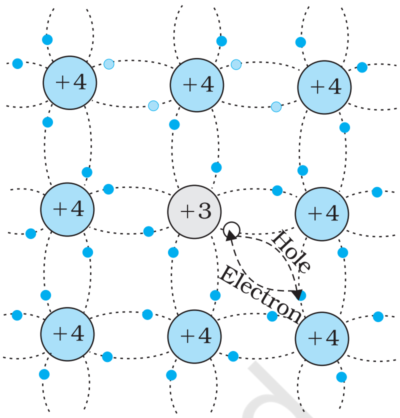
$$n_e >> n_h \tag{14.3}$$

(ii) p-type semiconductor

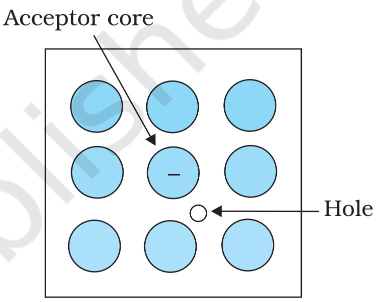
This is obtained when Si or Ge is doped with a trivalent impurity like Al, B, In, etc. The dopant has one valence electron less than Si or Ge and, therefore, this atom can form covalent bonds with neighbouring three Si atoms but does not have any electron to offer to the fourth Si atom. So the bond between the fourth neighbour and the trivalent atom has a vacancy or hole as shown in Fig. 14.8. Since the neighbouring Si atom in the lattice wants an electron in place of a hole, an electron in the outer orbit of an atom in the neighbourhood may jump to fill this vacancy, leaving a vacancy or hole at its own site. Thus the *hole* is available for conduction. Note that the trivalent foreign atom becomes effectively negatively charged when it shares fourth electron with neighbouring Si atom. Therefore, the dopant atom of p-type material can be treated as *core of one negative charge* along with its associated hole as shown in Fig. 14.8(b). It is obvious that one *acceptor* atom gives one *hole*. These holes are in addition to the intrinsically generated holes while the source of conduction electrons is only intrinsic generation. Thus, for such a material, the holes are the majority carriers and electrons are minority carriers. Therefore, extrinsic semiconductors doped with trivalent impurity are called p-*type semiconductors*. For p-type semiconductors, the recombination process will reduce the number (n_i)of intrinsically generated electrons to n_e. We have, for p-type semiconductors

$$n_h >> n_e \tag{14.4}$$

Note that *the crystal maintains an overall charge neutrality as the charge of additional charge carriers is just equal and opposite to that of the ionised cores in the lattice.*



(a)



(b)

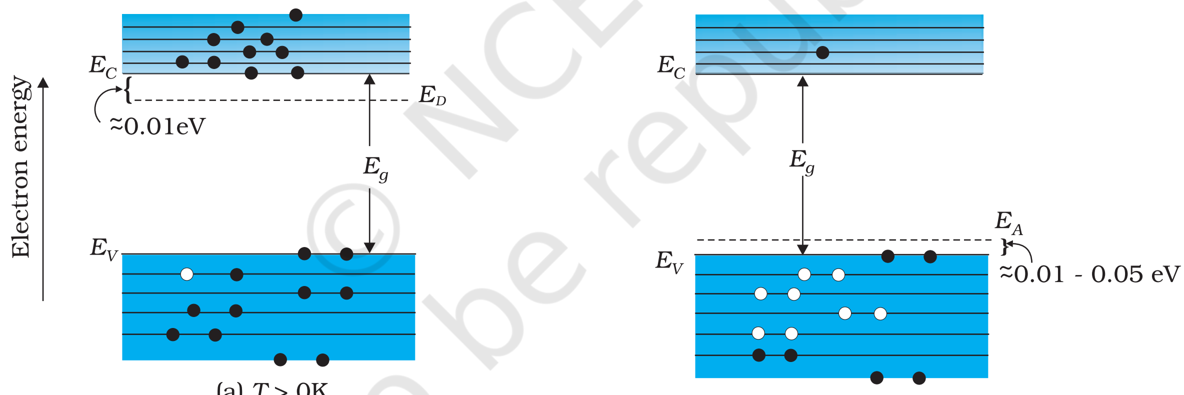
FIGURE 14.8 (a) Trivalent acceptor atom (In, Al, B etc.) doped in tetravalent Si or Ge lattice giving p-type semiconductor. (b) Commonly used schematic representation of p-type material which shows only the fixed core of the substituent acceptor with one effective additional negative charge and its associated hole.

In extrinsic semiconductors, because of the abundance of majority current carriers, the minority carriers produced thermally have more chance of meeting majority carriers and thus getting destroyed. Hence, the dopant, by adding a large number of current carriers of one type, which become the majority carriers, indirectly helps to reduce the intrinsic concentration of minority carriers.

The semiconductor's energy band structure is affected by doping. In the case of extrinsic semiconductors, additional energy states due to donor impurities (E_D) and acceptor impurities (E_A) also exist. In the energy band diagram of n-type Si semiconductor, the donor energy level E_D is slightly below the bottom E_C of the conduction band and electrons from this level move into the conduction band with very small supply of energy. At room temperature, most of the donor atoms get ionised but very few (~10^{12}) atoms of Si get ionised. So the conduction band will have most electrons coming from the donor impurities, as shown in Fig. 14.9(a). Similarly, for p-type semiconductor, the acceptor energy level E_A is slightly above the top E_V of the valence band as shown in Fig. 14.9(b). With very small supply of energy an electron from the valence band can jump to the level E_A and ionise the acceptor negatively. (Alternately, we can also say that with very small supply of energy the hole from level E_A sinks down into the valence band. Electrons rise up and holes fall down when they gain external energy.) At room temperature, most of the acceptor atoms get ionised leaving holes in the valence band. Thus at room temperature the density of holes in the valence band is predominantly due to impurity in the extrinsic semiconductor. The electron and hole concentration in a semiconductor *in thermal equilibrium* is given by

$$n_e n_h = n_i^2 \tag{14.5}$$

Though the above description is grossly approximate and hypothetical, it helps in understanding the difference between metals, insulators and semiconductors (extrinsic and intrinsic) in a simple manner. The difference in the resistivity of C, Si and Ge depends upon the energy gap between their conduction and valence bands. For C (diamond), Si and Ge, the energy gaps are 5.4 eV, 1.1 eV and 0.7 eV, respectively. Sn also is a group IV element but it is a metal because the energy gap in its case is 0 eV.



(a) T > 0K
one thermally generated electron-hole pair + 9 electrons from donor atoms

(b) T > 0K

FIGURE 14.9 Energy bands of (a) n-type semiconductor at T > 0K, (b) p-type semiconductor at T > 0K.

EXAMPLE 14.2

Example 14.2 Suppose a pure Si crystal has 5×10^{28} atoms m^{-3}. It is doped by 1 ppm concentration of pentavalent As. Calculate the number of electrons and holes. Given that $n_i = 1.5 \times 10^{16}\ m^{-3}$.

Solution Note that thermally generated electrons ($n_i \sim 10^{16}\ m^{-3}$) are negligibly small as compared to those produced by doping.

Therefore, $n_e \approx N_D$.

Since $n_e n_h = n_i^2$, The number of holes

$n_h = (2.25 \times 10^{32})/(5 \times 10^{22})$

$\sim 4.5 \times 10^9\ m^{-3}$

14.5 p-n Junction

A p-n junction is the basic building block of many semiconductor devices like diodes, transistor, etc. A clear understanding of the junction behaviour is important to analyse the working of other semiconductor devices. We will now try to understand how a junction is formed and how the junction behaves under the influence of external applied voltage (also called *bias*).

14.5.1 p-n junction formation

Consider a thin p-type silicon (p-Si) semiconductor wafer. By adding precisely a small quantity of pentavelent impurity, part of the p-Si wafer can be converted into n-Si. There are several processes by which a semiconductor can be formed. The wafer now contains p-region and n-region and a metallurgical junction between p-, and n- region.

Two important processes occur during the formation of a p-n junction: *diffusion* and *drift*. We know that in an n-type semiconductor, the concentration of electrons (number of electrons per unit volume) is more compared to the concentration of holes. Similarly, in a p-type semiconductor, the concentration of holes is more than the concentration of electrons. During the formation of p-n junction, and due to the concentration gradient across p-, and n- sides, holes diffuse from p-side to n-side ($p \rightarrow n$) and electrons diffuse from n-side to p-side ($n \rightarrow p$). This motion of charge carries gives rise to diffusion current across the junction.

Formation and working of p-n junction diode
http://hyperphysics.phy-astr.gsu.edu/hbase/solids/pnjun.html

When an electron diffuses from $n \rightarrow p$, it leaves behind an ionised donor on n-side. This ionised donor (positive charge) is immobile as it is bonded to the surrounding atoms. As the electrons continue to diffuse from $n \rightarrow p$, a layer of positive charge (or positive space-charge region) on n-side of the junction is developed.

Similarly, when a hole diffuses from $p \rightarrow n$ due to the concentration gradient, it leaves behind an ionised acceptor (negative charge) which is immobile. As the holes continue to diffuse, a layer of negative charge (or negative space-charge region) on the p-side of the junction is developed. This space-charge region on either side of the junction together is known as *depletion region* as the electrons and holes taking part in the initial movement across the junction *depleted* the region of its free charges (Fig. 14.10). The thickness of depletion region is of the order of one-tenth of a micrometre. Due to the positive space-charge region on n-side of the junction and negative space charge region on p-side of the junction, an electric field directed from positive charge towards negative charge develops. Due to this field, an electron on p-side of the junction moves to n-side and a hole on n-side of the junction moves to p-side. The motion of charge carriers due to the electric field is called drift. Thus a drift current, which is opposite in direction to the diffusion current (Fig. 14.10) starts.

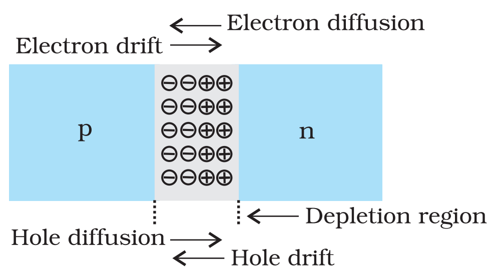


FIGURE 14.10 p-n junction formation process.

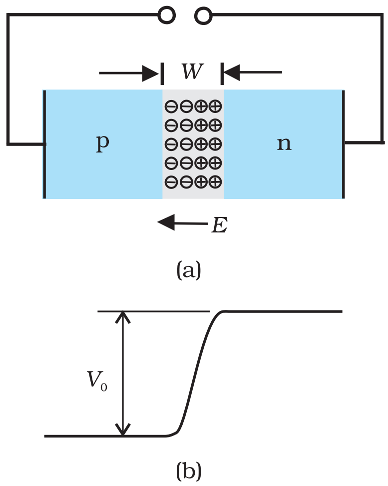


FIGURE 14.11 (a) Diode under equilibrium (V = 0), (b) Barrier potential under no bias.

Initially, diffusion current is large and drift current is small. As the diffusion process continues, the space-charge regions on either side of the junction extend, thus increasing the electric field strength and hence drift current. This process continues until the diffusion current equals the drift current. Thus a p-n junction is formed. In a p-n junction under equilibrium there is *no net* current.

The loss of electrons from the n-region and the gain of electron by the p-region causes a difference of potential across the junction of the two regions. The polarity of this potential is such as to oppose further flow of carriers so that a condition of equilibrium exists. Figure 14.11 shows the p-n junction at equilibrium and the potential across the junction. The n-material has lost electrons, and p material has acquired electrons. The n material is thus positive relative to the p material. Since this potential tends to prevent the movement of electron from the n region into the p region, it is often called a *barrier potential*.

EXAMPLE 14.3

Example 14.3 Can we take one slab of p-type semiconductor and physically join it to another n-type semiconductor to get p-n junction?

Solution No! Any slab, howsoever flat, will have roughness much larger than the inter-atomic crystal spacing (~2 to 3 Å) and hence *continuous contact* at the atomic level will not be possible. The junction will behave as a *discontinuity* for the flowing charge carriers.

14.6 Semiconductor Diode

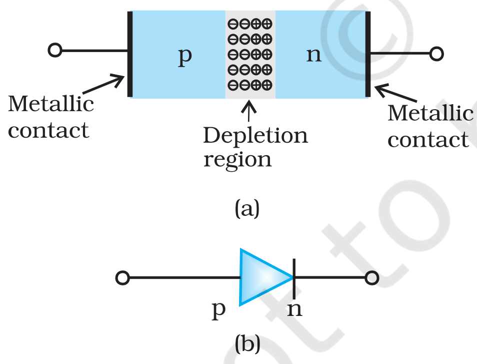


FIGURE 14.12 (a) Semiconductor diode, (b) Symbol for p-n junction diode.

A semiconductor diode [Fig. 14.12(a)] is basically a p-n junction with metallic contacts provided at the ends for the application of an external voltage. It is a two terminal device. A p-n junction diode is symbolically represented as shown in Fig. 14.12(b).

The direction of arrow indicates the conventional direction of current (when the diode is under forward bias). The equilibrium barrier potential can be altered by applying an external voltage V across the diode. The situation of p-n junction diode under equilibrium (without bias) is shown in Fig. 14.11(a) and (b).

14.6.1 p-n junction diode under forward bias

When an external voltage V is applied across a semiconductor diode such that p-side is connected to the positive terminal of the battery and n-side to the negative terminal [Fig. 14.13(a)], it is said to be *forward biased*.

The applied voltage mostly drops across the depletion region and the voltage drop across the p-side and n-side of the junction is negligible. (This is because the resistance of the depletion region – a region where there are no charges – is very high compared to the resistance of n-side and p-side.) The direction of the applied voltage (V) is opposite to the built-in potential V_0. As a result, the depletion layer width decreases and the barrier height is reduced [Fig. 14.13(b)]. The effective barrier height under forward bias is $(V_0 - V)$.

If the applied voltage is small, the barrier potential will be reduced only slightly below the equilibrium value, and only a small number of carriers in the material—those that happen to be in the uppermost energy levels—will possess enough energy to cross the junction. So the current will be small. If we increase the applied voltage significantly, the barrier height will be reduced and more number of carriers will have the required energy. Thus the current increases.

Due to the applied voltage, electrons from n-side cross the depletion region and reach p-side (where they are minority carries). Similarly, holes from p-side cross the junction and reach the n-side (where they are minority carries). This process under forward bias is known as minority carrier injection. At the junction boundary, on each side, the minority carrier concentration increases significantly compared to the locations far from the junction.

Due to this concentration gradient, the injected electrons on p-side diffuse from the junction edge of p-side to the other end of p-side. Likewise, the injected holes on n-side diffuse from the junction edge of n-side to the other end of n-side (Fig. 14.14). This motion of charged carriers on either side gives rise to current. The total diode forward current is sum of hole diffusion current and conventional current due to electron diffusion. The magnitude of this current is usually in mA.

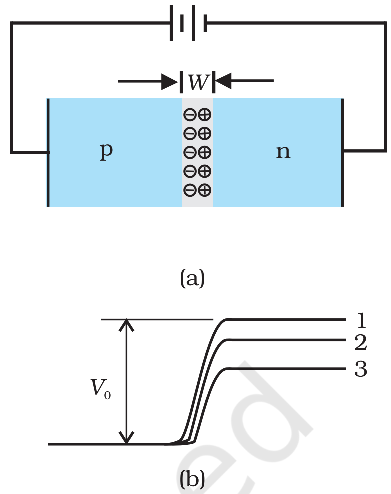


FIGURE 14.13 (a) p-n junction diode under forward bias, (b) Barrier potential (1) without battery, (2) Low battery voltage, and (3) High voltage battery.

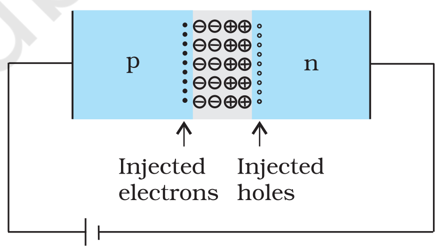


FIGURE 14.14 Forward bias minority carrier injection.

14.6.2 p-n junction diode under reverse bias

When an external voltage (V) is applied across the diode such that n-side is positive and p-side is negative, it is said to be *reverse biased* [Fig.14.15(a)]. The applied voltage mostly drops across the depletion region. The direction of applied voltage is same as the direction of barrier potential. As a result, the barrier height increases and the depletion region widens due to the change in the electric field. The effective barrier height under reverse bias is $(V_0 + V)$, [Fig. 14.15(b)]. This suppresses the flow of electrons from n $\rightarrow$ p and holes from p $\rightarrow$ n. Thus, diffusion current, decreases enormously compared to the diode under forward bias.

The electric field direction of the junction is such that if electrons on p-side or holes on n-side in their random motion come close to the junction, they will be swept to its majority zone. This drift of carriers gives rise to current. The drift current is of the order of a few μA. This is quite low because it is due to the motion of carriers from their minority side to their majority side across the junction. The drift current is also there under forward bias but it is negligible (μA) when compared with current due to injected carriers which is usually in mA.

The diode reverse current is not very much dependent on the applied voltage. Even a small voltage is sufficient to sweep the minority carriers from one side of the junction to the other side of the junction. The current

Reprint 2025-26

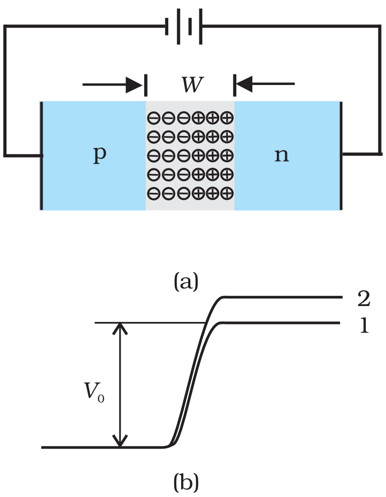


FIGURE 14.15 (a) Diode under reverse bias, (b) Barrier potential under reverse bias.

is not limited by the magnitude of the applied voltage but is limited due to the concentration of the minority carrier on either side of the junction.

The current under reverse bias is essentially voltage independent upto a critical reverse bias voltage, known as breakdown voltage (V_{br}). When $V = V_{br}$, the diode reverse current increases sharply. Even a slight increase in the bias voltage causes large change in the current. If the reverse current is not limited by an external circuit below the rated value (specified by the manufacturer) the p-n junction will get destroyed. Once it exceeds the rated value, the diode gets destroyed due to overheating. This can happen even for the diode under forward bias, if the forward current exceeds the rated value.

The circuit arrangement for studying the *V-I* characteristics of a diode, (i.e., the variation of current as a function of applied voltage) are shown in Fig. 14.16(a) and (b). The battery is connected to the diode through a potentiometer (or reheostat) so that the applied voltage to the diode can be changed. For different values of voltages, the value of the current is noted. A graph between *V* and *I* is obtained as in Fig. 14.16(c). Note that in forward bias measurement, we use a milliammeter since the expected current is large (as explained in the earlier section) while a micrometer is used in reverse bias to measure the current. You can see in Fig. 14.16(c) that in forward

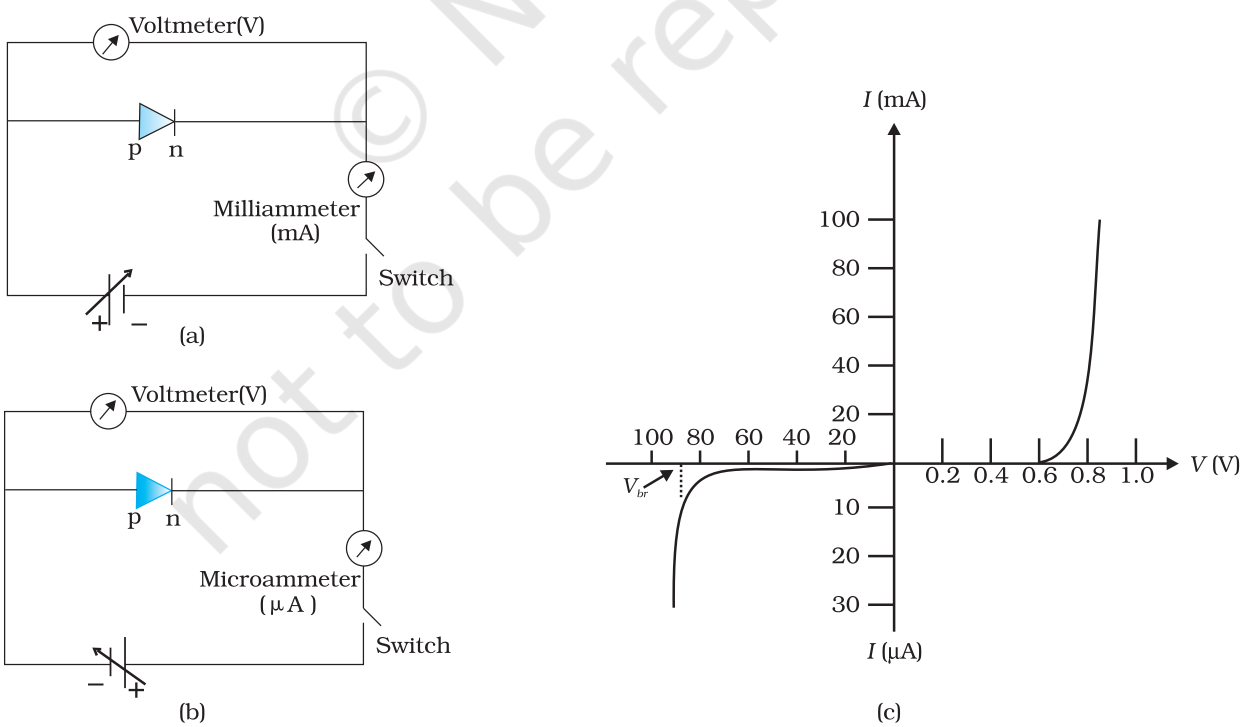


FIGURE 14.16 Experimental circuit arrangement for studying *V-I* characteristics of a p-n junction diode (a) in forward bias, (b) in reverse bias. (c) Typical *V-I* characteristics of a silicon diode.

bias, the current first increases very slowly, almost negligibly, till the voltage across the diode crosses a certain value. After the characteristic voltage, the diode current increases significantly (exponentially), even for a very small increase in the diode bias voltage. This voltage is called the *threshold voltage* or cut-in voltage (~0.2V for germanium diode and ~0.7 V for silicon diode).

For the diode in reverse bias, the current is very small (~μA) and almost remains constant with change in bias. It is called *reverse saturation current*. However, for special cases, at very high reverse bias (break down voltage), the current suddenly increases. This special action of the diode is discussed later in Section 14.8. The general purpose diode are not used beyond the reverse saturation current region.

The above discussion shows that the p-n junction diode primerly allows the flow of current only in one direction (forward bias). The forward bias resistance is low as compared to the reverse bias resistance. This property is used for rectification of ac voltages as discussed in the next section. For diodes, we define a quantity called *dynamic resistance* as the ratio of small change in voltage ΔV to a small change in current ΔI:

$$r_d = \frac{\Delta V}{\Delta I} \tag{14.6}$$

Example 14.4 The *V-I* characteristic of a silicon diode is shown in the Fig. 14.17. Calculate the resistance of the diode at (a) I_D = 15 mA and (b) V_D = –10 V.

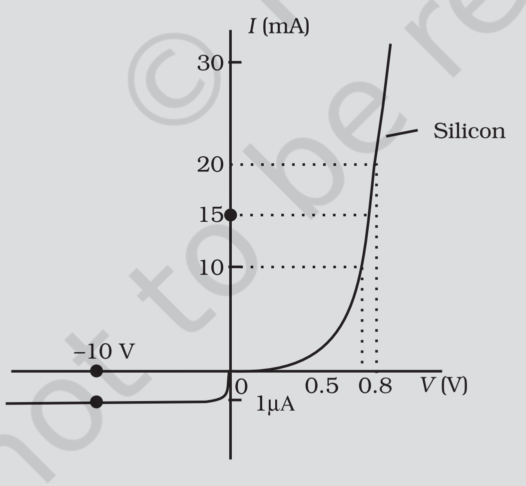


FIGURE 14.17

Solution Considering the diode characteristics as a straight line between I = 10 mA to I = 20 mA passing through the origin, we can calculate the resistance using Ohm's law.

(a) From the curve, at I = 20 mA, V = 0.8 V; I = 10 mA, V = 0.7 V
$r_{fb} = \Delta V/\Delta I$ = 0.1V/10 mA = 10 Ω

(b) From the curve at V = –10 V, I = –1 μA,
Therefore,
r_{rb} = 10 V/1μA= 1.0 × 10^7 Ω

EXAMPLE 14.4

14.7 Application of Junction Diode as a Rectifier

From the *V-I* characteristic of a junction diode we see that it allows current to pass only when it is forward biased. So if an alternating voltage is applied across a diode the current flows only in that part of the cycle when the diode is forward biased. This property is used *to rectify* alternating voltages and the circuit used for this purpose is called a *rectifier*.

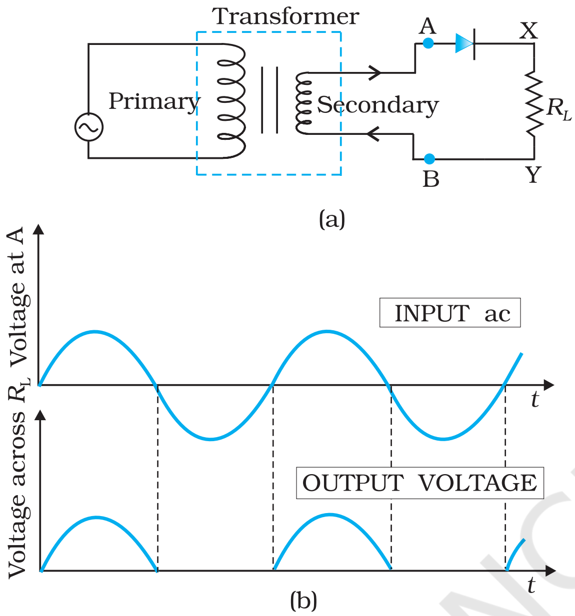


FIGURE 14.18 (a) Half-wave rectifier circuit, (b) Input ac voltage and output voltage waveforms from the rectifier circuit.

If an alternating voltage is applied across a diode in series with a load, a pulsating voltage will appear across the load only during the half cycles of the ac input during which the diode is forward biased. Such rectifier circuit, as shown in Fig. 14.18, is called a *half-wave rectifier*. The secondary of a transformer supplies the desired ac voltage across terminals A and B. When the voltage at A is positive, the diode is forward biased and it conducts. When A is negative, the diode is reverse-biased and it does not conduct. The reverse saturation current of a diode is negligible and can be considered equal to zero for practical purposes. (The reverse breakdown voltage of the diode must be sufficiently higher than the peak ac voltage at the secondary of the transformer to protect the diode from reverse breakdown.)

Therefore, in the positive *half-cycle* of ac there is a current through the load resistor R_L and we get an output voltage, as shown in Fig. 14.18(b), whereas there is no current in the negative half-cycle. In the next positive half-cycle, again we get the output voltage. Thus, the output voltage, though still varying, is restricted to *only one direction* and is said to be *rectified*. Since the rectified output of this circuit is only for half of the input ac wave it is called as *half-wave rectifier*.

The circuit using two diodes, shown in Fig. 14.19(a), gives output rectified voltage corresponding to both the positive as well as negative half of the ac cycle. Hence, it is known as *full-wave rectifier*. Here the p-side of the two diodes are connected to the ends of the secondary of the transformer. The n-side of the diodes are connected together and the output is taken between this common point of diodes and the midpoint of the secondary of the transformer. So for a full-wave rectifier the secondary of the transformer is provided with a centre tapping and so it is called *centre-tap transformer*. As can be seen from Fig.14.19(c) the voltage rectified by each diode is only half the total secondary voltage. Each diode rectifies only for half the cycle, but the two do so for alternate cycles. Thus, the output between their common terminals and the centre-tap of the transformer becomes a full-wave rectifier output. (Note that there is another circuit of full wave rectifier which does not need a centre-tap transformer but needs four diodes.) Suppose the input voltage to A

Reprint 2025-26

with respect to the centre tap at any instant is positive. It is clear that, at that instant, voltage at B being out of phase will be negative as shown in Fig.14.19(b). So, diode D_1 gets forward biased and conducts (while D_2 being reverse biased is not conducting). Hence, during this positive half cycle we get an output current (and a output voltage across the load resistor R_L) as shown in Fig.14.19(c). In the course of the ac cycle when the voltage at A becomes negative with respect to centre tap, the voltage at B would be positive. In this part of the cycle diode D_1 would not conduct but diode D_2 would, giving an output current and output voltage (across R_L) during the negative half cycle of the input ac. Thus, we get output voltage during both the positive as well as the negative half of the cycle. Obviously, this is a more efficient circuit for getting rectified voltage or current than the half-wave rectifier.

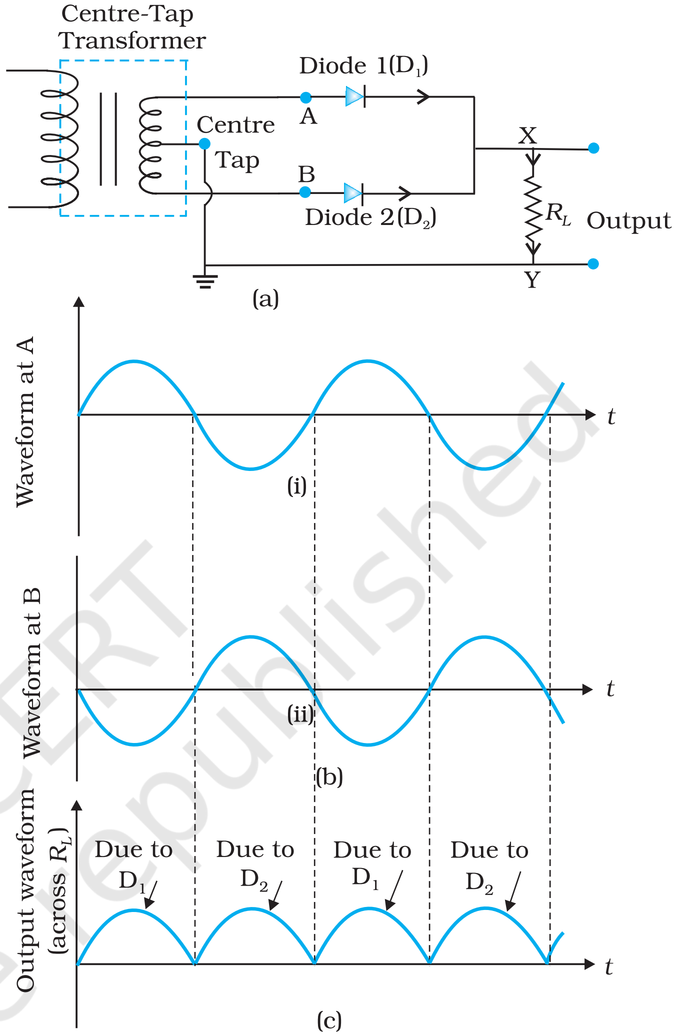


FIGURE 14.19 (a) A Full-wave rectifier circuit; (b) Input wave forms given to the diode D_1 at A and to the diode D_2 at B; (c) Output waveform across the load R_L connected in the full-wave rectifier circuit.

The rectified voltage is in the form of pulses of the shape of half sinusoids. Though it is unidirectional it does not have a steady value. To get steady dc output from the pulsating voltage normally a capacitor is connected across the output terminals (parallel to the load R_L). One can also use an inductor in series with R_L for the same purpose. Since these additional circuits appear to *filter* out the *ac ripple* and give a *pure dc* voltage, so they are called filters.

Now we shall discuss the role of capacitor in filtering. When the voltage across the capacitor is rising, it gets charged. If there is no external load, it remains charged to the peak voltage of the rectified output. When there is a load, it gets discharged through the load and the voltage across it begins to fall. In the next half-cycle of rectified output it again gets charged to the peak value (Fig. 14.20). The rate of fall of the voltage across the capacitor depends inversely upon the product of capacitance C and the effective resistance R_L used in the circuit and is called the *time constant*. To make the time constant large value of C should be large. So capacitor input filters use large capacitors. The *output voltage* obtained by using capacitor input filter is nearer to the *peak voltage* of the rectified voltage. This type of filter is most widely used in power supplies.

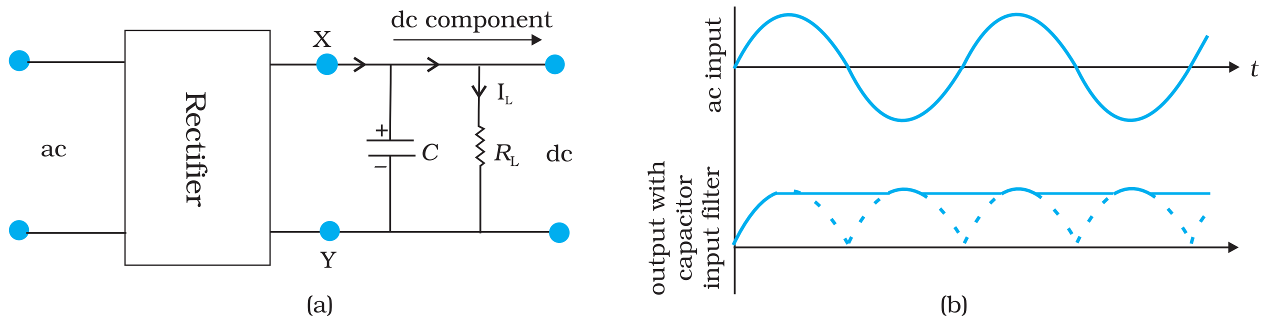


FIGURE 14.20 (a) A full-wave rectifier with capacitor filter, (b) Input and output voltage of rectifier in (a).

SUMMARY

1. Semiconductors are the basic materials used in the present solid state electronic devices like diode, transistor, ICs, etc.
2. Lattice structure and the atomic structure of constituent elements decide whether a particular material will be insulator, metal or semiconductor.
3. Metals have low resistivity (10^{-2} to 10^{-8} Ωm), insulators have very high resistivity (>10^8 Ω m^{-1}), while semiconductors have intermediate values of resistivity.
4. Semiconductors are elemental (Si, Ge) as well as compound (GaAs, CdS, etc.).
5. Pure semiconductors are called 'intrinsic semiconductors'. The presence of charge carriers (electrons and holes) is an 'intrinsic' property of the material and these are obtained as a result of thermal excitation. The number of electrons (n_e) is equal to the number of holes (n_h) in intrinsic conductors. Holes are essentially electron vacancies with an effective positive charge.
6. The number of charge carriers can be changed by 'doping' of a suitable impurity in pure semiconductors. Such semiconductors are known as extrinsic semiconductors. These are of two types (n-type and p-type).
7. In n-type semiconductors, $n_e >> n_h$ while in p-type semiconductors $n_h >> n_e$.
8. n-type semiconducting Si or Ge is obtained by doping with pentavalent atoms (donors) like As, Sb, P, etc., while p-type Si or Ge can be obtained by doping with trivalent atom (acceptors) like B, Al, In etc.
9. $n_e n_h = n_i^2$ in all cases. Further, the material possesses an *overall charge neutrality*.
10. There are two distinct band of energies (called valence band and conduction band) in which the electrons in a material lie. Valence band energies are low as compared to conduction band energies. All energy levels in the valence band are filled while energy levels in the conduction band may be fully empty or partially filled. The electrons in the conduction band are free to move in a solid and are responsible for the conductivity. The extent of conductivity depends upon the energy gap (E_g) between the top of valence band (E_V) and the bottom of the conduction band E_C. The electrons from valence band can be excited by

heat, light or electrical energy to the conduction band and thus, produce a change in the current flowing in a semiconductor.

11. For insulators $E_g > 3$ eV, for semiconductors E_g is 0.2 eV to 3 eV, while for metals $E_g \approx 0$.
12. p-n junction is the 'key' to all semiconductor devices. When such a junction is made, a 'depletion layer' is formed consisting of immobile ion-cores devoid of their electrons or holes. This is responsible for a junction potential barrier.
13. By changing the external applied voltage, junction barriers can be changed. In forward bias (n-side is connected to negative terminal of the battery and p-side is connected to the positive), the barrier is decreased while the barrier increases in reverse bias. Hence, forward bias current is more (mA) while it is very small (μA) in a p-n junction diode.
14. Diodes can be used for rectifying an ac voltage (restricting the ac voltage to one direction). With the help of a capacitor or a suitable filter, a dc voltage can be obtained.

POINTS TO PONDER

1. The energy bands (E_C or E_V) in the semiconductors are space delocalised which means that these are not located in any specific place inside the solid. The energies are the overall averages. When you see a picture in which E_C or E_V are drawn as straight lines, then they should be respectively taken simply as the *bottom* of conduction band energy levels and *top* of valence band energy levels.
2. In elemental semiconductors (Si or Ge), the n-type or p-type semiconductors are obtained by introducing 'dopants' as defects. In compound semiconductors, the change in relative stoichiometric ratio can also change the type of semiconductor. For example, in ideal GaAs the ratio of Ga:As is 1:1 but in Ga-rich or As-rich GaAs it could respectively be $Ga_{1.1}\,As_{0.9}$ or $Ga_{0.9}\,As_{1.1}$. In general, the presence of defects control the properties of semiconductors in many ways.

EXERCISES

14.1 In an n-type silicon, which of the following statement is true:
(a) Electrons are majority carriers and trivalent atoms are the dopants.
(b) Electrons are minority carriers and pentavalent atoms are the dopants.
(c) Holes are minority carriers and pentavalent atoms are the dopants.
(d) Holes are majority carriers and trivalent atoms are the dopants.

14.2 Which of the statements given in Exercise 14.1 is true for p-type semiconductos.

14.3 Carbon, silicon and germanium have four valence electrons each. These are characterised by valence and conduction bands separated by energy band gap respectively equal to $(E_g)_C$, $(E_g)_{Si}$ and $(E_g)_{Ge}$. Which of the following statements is true?

(a) $(E_g)_{Si} < (E_g)_{Ge} < (E_g)_C$

(b) $(E_g)_C < (E_g)_{Ge} > (E_g)_{Si}$

(c) $(E_g)_C > (E_g)_{Si} > (E_g)_{Ge}$

(d) $(E_g)_C = (E_g)_{Si} = (E_g)_{Ge}$

14.4 In an unbiased p-n junction, holes diffuse from the p-region to n-region because

(a) free electrons in the n-region attract them.

(b) they move across the junction by the potential difference.

(c) hole concentration in p-region is more as compared to n-region.

(d) All the above.

14.5 When a forward bias is applied to a p-n junction, it

(a) raises the potential barrier.

(b) reduces the majority carrier current to zero.

(c) lowers the potential barrier.

(d) None of the above.

14.6 In half-wave rectification, what is the output frequency if the input frequency is 50 Hz. What is the output frequency of a full-wave rectifier for the same input frequency.

Reprint 2025-26

Notes

APPENDICES

APPENDIX A 1
THE GREEK ALPHABET

Alpha	Α	α	Iota	Ι	ι	Rho	Ρ	ρ
Beta	Β	β	Kappa	Κ	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	Τ	τ
Delta	Δ	δ	Mu	Μ	μ	Upsilon	Υ	υ
Epsilon	Ε	ε	Nu	Ν	ν	Phi	Φ	ϕ, φ
Zeta	Ζ	ς	Xi	Ξ	ξ	Chi	Χ	χ
Eta	Η	η	Omicron	Ο	ο	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

APPENDIX A 2
COMMON SI PREFIXES AND SYMBOLS FOR MULTIPLES AND SUB-MULTIPLES

Multiple			Sub-Multiple		
Factor	**Prefix**	**Symbol**	**Factor**	**Prefix**	**symbol**
10^{18}	Exa	E	10^{-18}	atto	a
10^{15}	Peta	P	10^{-15}	femto	f
10^{12}	Tera	T	10^{-12}	pico	p
10^{9}	Giga	G	10^{-9}	nano	n
10^{6}	Mega	M	10^{-6}	micro	μ
10^{3}	kilo	k	10^{-3}	milli	m
10^{2}	Hecto	h	10^{-2}	centi	c
10^{1}	Deca	da	10^{-1}	deci	d

APPENDIX A 3
SOME IMPORTANT CONSTANTS

Name	Symbol	Value
Speed of light in vacuum	c	2.9979×10^{8} m s^{-1}
Charge of electron	e	1.602×10^{-19} C
Gravitational constant	G	6.673×10^{-11} N m^{2} kg^{-2}
Planck constant	h	6.626×10^{-34} J s
Boltzmann constant	k	1.381×10^{-23} J K^{-1}
Avogadro number	N_A	6.022×10^{23} mol^{-1}
Universal gas constant	R	8.314 J mol^{-1}K^{-1}
Mass of electron	m_e	9.110×10^{-31} kg
Mass of neutron	m_n	1.675×10^{-27} kg
Mass of proton	m_p	1.673×10^{-27} kg
Electron-charge to mass ratio	e/m_e	1.759×10^{11} C/kg
Faraday constant	F	9.648×10^{4} C/mol
Rydberg constant	R	1.097×10^{7} m^{-1}
Bohr radius	a_0	5.292×10^{-11} m
Stefan-Boltzmann constant	σ	5.670×10^{-8} Wm^{-2}K^{-4}
Wien's Constant	b	2.898×10^{-3} mK
Permittivity of free space	ε_0 $1/4\pi\,\varepsilon_0$	8.854×10^{-12} C^{2} N^{-1}m^{-2} 8.987×10^{9} N m^{2} C^{-2}
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ T m A^{-1} $\cong 1.257 \times 10^{-6}$ Wb A^{-1} m^{-1}

OTHER USEFUL CONSTANTS

Name	Symbol	Value
Mechanical equivalent of heat	J	4.186 J cal^{-1}
Standard atmospheric pressure	1 atm	1.013×10^{5} Pa
Absolute zero	0 K	−273.15 °C
Electron volt	1 eV	1.602×10^{-19} J
Unified Atomic mass unit	1 u	1.661×10^{-27} kg
Electron rest energy	mc^{2}	0.511 MeV
Energy equivalent of 1 u	1 uc^{2}	931.5 MeV
Volume of ideal gas (0 °C and 1atm)	V	22.4 L mol^{-1}
Acceleration due to gravity (sea level, at equator)	g	9.78049 m s^{-2}

ANSWERS

CHAPTER 9

9.1 v = –54 cm. The image is real, inverted and magnified. The size of the image is 5.0 cm. As $u \to f$, $v \to \infty$; for $u < f$, image is virtual.

9.2 v = 6.7 cm. Magnification = 5/9, i.e., the size of the image is 2.5 cm. As $u \to \infty$; $v \to f$ (but never beyond) while $m \to 0$.

9.3 1.33; 1.7 cm

9.4 n_{ga} = 1.51; n_{wa} = 1.32; n_{gw} = 1.144; which gives sin r = 0.6181 i.e., $r \simeq 38°$.

9.5 $r = 0.8 \times \tan i_c$ and $\sin i_c = 1/1.33 \cong 0.75$, where r is the radius (in m) of the largest circle from which light comes out and i_c is the critical angle for water-air interface, Area = 2.6 m^2

9.6 $n \cong$ 1.53 and D_m for prism in water $\cong$ 10°

9.7 R = 22 cm

9.8 Here the object is virtual and the image is real. u = +12 cm (object on right; virtual)

(a) f = +20 cm. Image is real and at 7.5 cm from the lens on its right side.

(b) f = –16 cm. Image is real and at 48 cm from the lens on its right side.

9.9 v = 8.4 cm, image is erect and virtual. It is diminished to a size 1.8 cm. As $u \to \infty$, $v \to f$ (but never beyond f while $m \to 0$).

Note that when the object is placed at the focus of the concave lens (21 cm), the image is located at 10.5 cm (not at infinity as one might wrongly think).

9.10 A diverging lens of focal length 60 cm

9.11 (a) v_e = –25 cm and f_e = 6.25 cm give u_e = –5 cm; v_O = (15 – 5) cm = 10 cm,

$f_O = u_O$ = – 2.5 cm; Magnifying power = 20

(b) u_O = – 2.59 cm.

Magnifying power = 13.5.

9.12 Angular magnification of the eye-piece for image at 25 cm

$$= \frac{25}{2.5} + 1 = 11; \quad |u_e| = \frac{25}{11}\text{cm} = 2.27\text{cm}; \quad v_O = 7.2\,\text{cm}$$

Separation = 9.47 cm; Magnifying power = 88

9.13 24; 150 cm

9.14 (a) Angular magnification = 1500

(b) Diameter of the image = 13.7 cm.

9.15 Apply mirror equation and the condition:

(a) $f < 0$ (concave mirror); $u < 0$ (object on left)

(b) $f > 0$; $u < 0$

(c) $f > 0$ (convex mirror) and $u < 0$

(d) $f < 0$ (concave mirror); $f < u < 0$

to deduce the desired result.

9.16 The pin appears raised by 5.0 cm. It can be seen with an explicit ray diagram that the answer is independent of the location of the slab (for small angles of incidence).

9.17 (a) $\sin i_c' = 1.44/1.68$ which gives $i_c' = 59°$. Total internal reflection takes place when $i > 59°$ or when $r < r_{max} = 31°$. Now, $(\sin i_{max}/\sin r_{max}) = 1.68$, which gives $i_{max} \simeq 60°$. Thus, all incident rays of angles in the range $0 < i < 60°$ will suffer total internal reflections in the pipe. (If the length of the pipe is finite, which it is in practice, there will be a lower limit on i determined by the ratio of the diameter to the length of the pipe.)

(b) If there is no outer coating, $i_c' = \sin^{-1}(1/1.68) = 36.5°$. Now, $i = 90°$ will have $r = 36.5°$ and $i' = 53.5°$ which is greater than i_c'. Thus, *all* incident rays (in the range $53.5° < i < 90°$) will suffer total internal reflections.

9.18 For fixed distance s between object and screen, the lens equation does not give a real solution for u or v if f is greater than $s/4$.

Therefore, $f_{max} = 0.75$ m.

9.19 21.4 cm

9.20 (a) (i) Let a parallel beam be the incident from the left on the convex lens first.

$f_1 = 30$ cm and $u_1 = -\infty$, give $v_1 = +30$ cm. This image becomes a virtual object for the second lens.

$f_2 = -20$ cm, $u_2 = +(30 - 8)$ cm $= +22$ cm which gives, $v_2 = -220$ cm. The parallel incident beam appears to diverge from a point 216 cm from the centre of the two-lens system.

(ii) Let the parallel beam be incident from the left on the concave lens first: $f_1 = -20$ cm, $u_1 = -\infty$, give $v_1 = -20$ cm. This image becomes a real object for the second lens: $f_2 = +30$ cm, $u_2 = -(20 + 8)$ cm $= -28$ cm which gives, $v_2 = -420$ cm. The parallel incident beam appears to diverge from a point 416 cm on the left of the centre of the two-lens system.

Clearly, the answer depends on which side of the lens system the parallel beam is incident. *Further we do not have a simple lens equation true for all u (and v) in terms of a definite constant of the system (the constant being determined by f_1 and f_2, and the separation between the lenses).* The notion of effective focal length, therefore, does not seem to be meaningful for this system.

(b) $u_1 = -40$ cm, $f_1 = 30$ cm, gives $v_1 = 120$ cm.

Magnitude of magnification due to the first (convex) lens is 3.

$u_2 = +(120 - 8)$ cm $= +112$ cm (object virtual);

$f_2 = -20$ cm which gives $v_2 = -\dfrac{112 \times 20}{92}$ cm

Magnitude of magnification due to the second (concave) lens = 20/92.

Net magnitude of magnification = 0.652

Size of the image = 0.98 cm

9.21 If the refracted ray in the prism is incident on the second face at the critical angle i_c, the angle of refraction r at the first face is $(60°–i_c)$.

Now, $i_c = \sin^{-1}(1/1.524) \simeq 41°$

Therefore, $r = 19°$

$\sin i = 0.4962$; $i \simeq 30°$

9.22 (a) $\frac{1}{v} + \frac{1}{9} = \frac{1}{10}$

i.e., $v = -90$ cm,

Magnitude of magnification = 90/9 = 10.

Each square in the virtual image has an area $10 \times 10 \times 1\ mm^2 = 100\ mm^2 = 1\ cm^2$

(b) Magnifying power = 25/9 = 2.8

(c) No, magnification of an image by a lens and angular magnification (or magnifying power) of an optical instrument are two separate things. The latter is the ratio of the angular size of the object (which is equal to the angular size of the image even if the image is magnified) to the angular size of the object if placed at the near point (25 cm). Thus, magnification magnitude is $|(v/u)|$ and magnifying power is $(25/|u|)$. Only when the image is located at the near point $|v|$ = 25 cm, are the two quantities equal.

9.23 (a) Maximum magnifying power is obtained when the image is at the near point (25 cm)

$u = -7.14$ cm.

(b) Magnitude of magnification = $(25/|u|)$ = 3.5.

(c) Magnifying power = 3.5

Yes, the magnifying power (when the image is produced at 25 cm) is equal to the magnitude of magnification.

9.24 Magnification $= \sqrt{(6.25/1)} = 2.5$

$v = +2.5u$

$$+\frac{1}{2.5u} - \frac{1}{u} = \frac{1}{10}$$

i.e., $u = -6$ cm

$|v| = 15$ cm

The virtual image is closer than the normal near point (25 cm) and cannot be seen by the eye distinctly.

9.25 (a) Even though the absolute image size is bigger than the object size, the angular size of the image is equal to the angular size of the object. The magnifier helps in the following way: without it object would be placed no closer than 25 cm; with it the object can be placed much closer. The closer object has larger angular size than the same object at 25 cm. It is in this sense that angular magnification is achieved.

(b) Yes, it decreases a little because the angle subtended at the eye is then slightly less than the angle subtended at the lens. The effect is negligible if the image is at a very large distance away. [*Note:* When the eye is separated from the lens, the angles subtended at the eye by the first object and its image are not equal.]

(c) First, grinding lens of very small focal length is not easy. More important, if you decrease focal length, aberrations (both spherical and chromatic) become more pronounced. So, in practice, you cannot get a magnifying power of more than 3 or so with a simple convex lens. However, using an aberration corrected lens system, one can increase this limit by a factor of 10 or so.

(d) Angular magnification of eye-piece is $[(25/f_e) + 1]$ (f_e in cm) which increases if f_e is smaller. Further, magnification of the objective is given by $\dfrac{v_O}{|u_O|} = \dfrac{1}{(|u_O|/f_O)-1}$

which is large when $|u_O|$ is slightly greater than f_O. The microscope is used for viewing very close object. So $|u_O|$ is small, and so is f_O.

(e) The image of the objective in the eye-piece is known as 'eye-ring'. All the rays from the object refracted by objective go through the eye-ring. Therefore, it is an ideal position for our eyes for viewing. If we place our eyes too close to the eye-piece, we shall not collect much of the light and also reduce our field of view. If we position our eyes on the eye-ring and the area of the pupil of our eye is greater or equal to the area of the eye-ring, our eyes will collect all the light refracted by the objective. The precise location of the eye-ring naturally depends on the separation between the objective and the eye-piece. When you view through a microscope by placing your eyes on one end,the ideal distance between the eyes and eye-piece is usually built-in the design of the instrument.

9.26 Assume microscope in normal use i.e., image at 25 cm. Angular magnification of the eye-piece

$$= \frac{25}{5} + 1 = 6$$

Magnification of the objective

$$= \frac{30}{6} = 5$$

$$\frac{1}{5u_O} - \frac{1}{u_O} = \frac{1}{1.25}$$

which gives $u_O = -1.5$ cm; $v_0 = 7.5$ cm. $|u_e| = (25/6)$ cm $= 4.17$ cm. The separation between the objective and the eye-piece should be (7.5 + 4.17) cm = 11.67 cm. Further the object should be placed 1.5 cm from the objective to obtain the desired magnification.

9.27 (a) $m = (f_O/f_e) = 28$

(b) $m = \dfrac{f_O}{f_e}\left[1 + \dfrac{f_O}{25}\right] = 33.6$

9.28 (a) $f_o + f_e = 145$ cm

(b) Angle subtended by the tower = (100/3000) = (1/30) rad.

Angle subtended by the image produced by the objective

$$= \frac{h}{f_o} = \frac{h}{140}$$

Equating the two, $h = 4.7$ cm.

(c) Magnification (magnitude) of the eye-piece = 6. Height of the final image (magnitude) = 28 cm.

9.29 The image formed by the larger (concave) mirror acts as virtual object for the smaller (convex) mirror. Parallel rays coming from the object at infinity will focus at a distance of 110 mm from the larger mirror. The distance of virtual object for the smaller mirror = (110 –20) = 90 mm. The focal length of smaller mirror is 70 mm. Using the mirror formula, image is formed at 315 mm from the smaller mirror.

9.30 The reflected rays get deflected by twice the angle of rotation of the mirror. Therefore, $d/1.5 = \tan 7°$. Hence $d = 18.4$ cm.

9.31 $n = 1.33$

CHAPTER 10

10.1 (a) Reflected light: (wavelength, frequency, speed same as incident light)

$\lambda = 589$ nm, $\nu = 5.09 \times 10^{14}$ Hz, $c = 3.00 \times 10^{8}$ m s^{-1}

(b) Refracted light: (frequency same as the incident frequency)
$\nu = 5.09 \times 10^{14}$ Hz
$v = (c/n) = 2.26 \times 10^{8}$ m s^{-1}, $\lambda = (v/\nu) = 444$ nm

10.2 (a) Spherical
(b) Plane
(c) Plane (a small area on the surface of a large sphere is nearly planar).

10.3 (a) 2.0×10^{8} m s^{-1}
(b) No. The refractive index, and hence the speed of light in a medium, depends on wavelength. [When no particular wavelength or colour of light is specified, we may take the given refractive index to refer to yellow colour.] Now we know violet colour deviates more than red in a glass prism, i.e. $n_v > n_r$. Therefore, the violet component of white light travels slower than the red component.

10.4 $\lambda = \dfrac{1.2\times10^{-2}\times0.28\times10^{-3}}{4\times1.4}$ m = 600 nm

10.5 K/4

10.6 (a) 1.17 mm (b) 1.56 mm

10.7 0.15°

10.8 $\tan^{-1}(1.5) \simeq 56.3°$

10.9 5000 Å, 6×10^{14} Hz; 45°

10.10 40m

Chapter 11

11.1 (a) 7.24×10^{18} Hz (b) 0.041 nm
11.2 (a) 0.34 eV = 0.54×10^{-19} J (b) 0.34 V (c) 344 km/s
11.3 1.5 eV = 2.4×10^{-19} J
11.4 (a) 3.14×10^{-19} J, 1.05×10^{-27} kg m/s (b) 3×10^{16} photons/s
(c) 0.63 m/s
11.5 6.59×10^{-34} J s
11.6 2.0 V
11.7 No, because $\nu < \nu_o$
11.8 4.73×10^{14} Hz
11.9 2.16 eV = 3.46×10^{-19} J
11.10 (a) 1.7×10^{-35} m (b) 1.1×10^{-32} m (c) 3.0×10^{-23} m
11.11 $\lambda = h/p = h/(h\nu/c) = c/\nu$

Chapter 12

12.1 (a) No different from
(b) Thomson's model; Rutherford's model
(c) Rutherford's model
(d) Thomson's model; Rutherford's model
(e) Both the models
12.2 The nucleus of a hydrogen atom is a proton. The mass of it is 1.67×10^{-27} kg, whereas the mass of an incident α-particle is 6.64×10^{-27} kg. Because the scattering particle is more massive than the target nuclei (proton), the α-particle won't bounce back in even in a head-on collision. It is similar to a football colliding with a tenis ball at rest. Thus, there would be no large-angle scattering.
12.3 5.6×10^{14} Hz
12.4 13.6 eV; –27.2 eV
12.5 9.7×10^{-8} m; 3.1×10^{15} Hz.
12.6 (a) 2.18×10^{6} m/s; 1.09×10^{6} m/s; 7.27×10^{5} m/s
(b) 1.52×10^{-16} s; 1.22×10^{-15} s; 4.11×10^{-15} s.
12.7 2.12×10^{-10} m; 4.77×10^{-10} m
12.8 Lyman series: 103 nm and 122 nm; Balmer series: 656 nm.
12.9 2.6×10^{74}

Chapter 13

13.1 104.7 MeV
13.2 8.79 MeV, 7.84 MeV
13.3 1.584×10^{25} MeV or 2.535×10^{12} J
13.4 1.23

13.5 (i) $Q = -4.03$ MeV; endothermic

(ii) $Q = 4.62$ MeV; exothermic

13.6 $Q = m\left({}^{56}_{26}\text{Fe}\right) - 2m\left({}^{28}_{13}\text{Al}\right) = 26.90$ MeV; not possible.

13.7 4.536×10^{26} MeV

13.8 About 4.9×10^{4} y

13.9 360 KeV

Chapter 14

14.1 (c)

14.2 (d)

14.3 (c)

14.4 (c)

14.5 (c)

14.6 50 Hz for half-wave, 100 Hz for full-wave

BIBLIOGRAPHY

TEXTBOOKS

For additional reading on the topics covered in this book, you may like to consult one or more of the following books. Some of these books however are more advanced and contain many more topics than this book.

1 **Ordinary Level Physics**, A.F. Abbott, Arnold-Heinemann (1984).
2 **Advanced Level Physics**, M. Nelkon and P. Parker, 6th Edition, Arnold-Heinemann (1987).
3 **Advanced Physics**, Tom Duncan, John Murray (2000).
4 **Fundamentals of Physics**, David Halliday, Robert Resnick and Jearl Walker, 7th Edition John Wily (2004).
5 **University Physics** (Sears and Zemansky's), H.D. Young and R.A. Freedman, 11th Edition, Addison—Wesley (2004).
6 **Problems in Elementary Physics**, B. Bukhovtsa, V. Krivchenkov, G. Myakishev and V. Shalnov, MIR Publishers, (1971).
7 **Lectures on Physics** (3 volumes), R.P. Feynman, Addision – Wesley (1965).
8 **Berkeley Physics Course** (5 volumes) McGraw Hill (1965).
   a. Vol. 1 – Mechanics: (Kittel, Knight and Ruderman)
   b. Vol. 2 – Electricity and Magnetism (E.M. Purcell)
   c. Vol. 3 – Waves and Oscillations (Frank S. Crawford)
   d. Vol. 4 – Quantum Physics (Wichmann)
   e. Vol. 5 – Statistical Physics (F. Reif)
9 **Fundamental University Physics**, M. Alonso and E. J. Finn, Addison – Wesley (1967).
10 **College Physics**, R.L. Weber, K.V. Manning, M.W. White and G.A. Weygand, Tata McGraw Hill (1977).
11 **Physics**: **Foundations and Frontiers**, G. Gamow and J.M. Cleveland, Tata McGraw Hill (1978).
12 **Physics for the Inquiring Mind**, E.M. Rogers, Princeton University Press (1960).
13 **PSSC Physics Course**, DC Heath and Co. (1965) Indian Edition, NCERT (1967).
14 **Physics Advanced Level**, Jim Breithampt, Stanley Thornes Publishers (2000).
15 **Physics,** Patrick Fullick, Heinemann (2000).
16 **Conceptual Physics,** Paul G. Hewitt, Addision—Wesley (1998).
17 **College Physics,** Raymond A. Serway and Jerry S. Faughn, Harcourt Brace and Co. (1999).
18 **University Physics,** Harris Benson, John Wiley (1996).
19 **University Physics,** William P. Crummet and Arthur B. Western, Wm.C. Brown (1994).
20 **General Physics,** Morton M. Sternheim and Joseph W. Kane, John Wiley (1988).
21 **Physics,** Hans C. Ohanian, W.W. Norton (1989).

22 **Advanced Physics,** Keith Gibbs, Cambridge University Press (1996).
23 **Understanding Basic Mechanics,** F. Reif, John Wiley (1995).
24 **College Physics,** Jerry D. Wilson and Anthony J. Buffa, Prentice Hall (1997).
25 **Senior Physics, Part – I,** I.K. Kikoin and A.K. Kikoin, MIR Publishers (1987).
26 **Senior Physics, Part – II,** B. Bekhovtsev, MIR Publishers (1988).
27 **Understanding Physics,** K. Cummings, Patrick J. Cooney, Priscilla W. Laws and Edward F. Redish, John Wiley (2005).
28 **Essentials of Physics,** John D. Cutnell and Kenneth W. Johnson, John Wiley (2005).

GENERAL BOOKS

For instructive and entertaining general reading on science, you may like to read some of the following books. Remember however, that many of these books are written at a level far beyond the level of the present book.

1 **Mr. Tompkins** in paperback, G. Gamow, Cambridge University Press (1967).
2 **The Universe and Dr. Einstein**, C. Barnett, Time Inc. New York (1962).
3 **Thirty years that Shook Physics**, G. Gamow, Double Day, New York (1966).
4 **Surely You're Joking, Mr. Feynman**, R.P. Feynman, Bantam books (1986).
5 **One, Two, Three... Infinity**, G. Gamow, Viking Inc. (1961).
6 **The Meaning of Relativity**, A. Einstein, (Indian Edition) Oxford and IBH Pub. Co. (1965).
7 **Atomic Theory and the Description of Nature**, Niels Bohr, Cambridge (1934).
8 **The Physical Principles of Quantum Theory**, W. Heisenberg, University of Chicago Press (1930).
9 **The Physics—Astronomy Frontier**, F. Hoyle and J.V. Narlikar, W.H. Freeman (1980).
10 **The Flying Circus of Physics with Answer**, J. Walker, John Wiley and Sons (1977).
11 **Physics for Everyone** (series), L.D. Landau and A.I. Kitaigorodski, MIR Publisher (1978).
Book 1: Physical Bodies
Book 2: Molecules
Book 3: Electrons
Book 4: Photons and Nuclei.
12 **Physics can be Fun**, Y. Perelman, MIR Publishers (1986).
13 **Power of Ten**, Philip Morrison and Eames, W.H. Freeman (1985).
14 **Physics in your Kitchen Lab.**, I.K. Kikoin, MIR Publishers (1985).
15 **How Things Work: The Physics of Everyday Life**, Louis A. Bloomfield, John Wiley (2005).
16 **Physics Matters: An Introduction to Conceptual Physics,** James Trefil and Robert M. Hazen, John Wiley (2004).

Reprint 2025-26