SOUTH EASTERN KENYA UNIVERSITY

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF PHYSICAL SCIENCES

UNIT CODE: CSC 126 **UNIT NAME: PHYSICS OF COMPUTING SYSTEMS**

LECTURER: DR. NGUMBI **TEL: 0722918834**

JANUARY 2023

Purpose

To provide beginner students taking B.Sc in Computer Science with comprehensive understanding of the basic principles, concepts, definitions and elements in magnetism, electric circuits, electronic circuits and wave and explore their applications in computing systems. The course involves two main areas which include basic Physics and semiconductor Physics.

Expected Learning Outcomes

By the end of the course the learner should be able to:

- **(i)** Appreciate the basic principles, concepts, and elements in electricity, magnetism, semiconductors and waves.
- **(ii)** Describe applications of magnetism, electronics and waves in computing.
- **(iii)** Explain the impact of advances in semiconductor electronic technology on computing.

Course Content

Magnetism: Basic concepts of magnetism and electromagnetism. Basic laws applications in computing systems, e.g. magnetic storage media. **Electric circuits**: Review of Ohm's law. Elements of RLC circuits and their characteristics. RLC electric circuits analysis and synthesis.

Electronics: The P-N junction and its V-I characteristics; diode resistance. Conventional Photo and light emitting diodes. The bipolar junction transistor; common base, common emitter, common collector configurations and their characteristics. The transistor as a switch. Field effect transistor and their characteristics. Applications in basic logic circuits and in computer storage devices. **Waves**: General definitions, attributes and characteristics of waves. Sound waves. Electromagnetic waves. Sinusoidal component. The electromagnetic spectrum. Interference, reflection and refraction of waves. Application in computer communications.

Teaching/Learning Methodologies: Lectures; Tutorials; Class discussion.

Course Assessment

Examination - 70%; Continuous Assessments (Exercises and Tests) - 30%; **Total - 100%**

PART 1:	**ELECTRIC CIRCUITS**	**Week 1-3**
	1.1 Electrostatics	
	Basic Concept of Charge, Properties of Electric Charge, The Coulombs Law, The Electric Field, Direct Current Circuits (DC), The Ohm's Law, Kirchoff's Laws	
	1.2 Alternating Currents	**Week 4-5**
	• Elements of RLC circuits and their characteristics	
	• RLC electric circuits synthesis and analysis	
	CAT 1	**Week 6**
PART 2:	**MAGNETISM**	
	2.1 Basic concepts of magnetism and electromagnetism	**Week 7**
	2.2 Basic laws applications in computing systems, e.g. magnetic storage media	**Week 7**
PART 3:	**WAVES**	
	3.1 Periodic motions	**Week 8-10**
	3.2 Sinusoidal waves	
	3.3 Wave interactions: Interference, reflection and refraction of waves	
	3.4 Sound waves	
	3.5 Light as a wave	
	3.6 Huygens wave theory	
	3.7 Electromagnetic waves	

PART 4: ELECTRONICS

4.1 The P-N junction and its V-I characteristics; diode resistance **Week 11-12**

4.2 Conventional Photo and light emitting diodes

4.3 Transistors: bipolar junction transistor (common base, common emitter, common collector configurations and their characteristics, transistor as a switch); Field effect transistor and their characteristics.

4.4 Applications in basic logic circuits and in computer storage devices

PART 1: ELECTRIC CIRCUITS

1.1 Electrostatics

1.1.1 Basic concept of charge

The concept of electric charge is widely known and familiar through such phenomena as lightning and sparks from certain clothing. Some simple experiments that demonstrate the existence of electric forces and charges include

- running a comb through your hair on a dry day which makes the comb attracts bits of papers.
- An inflated balloon rubbed with a piece of wool. The balloon adheres to a wall or a room ceiling.

Materials presenting such a behavior are said to be electrified or to have become electrically charged. Such charges may vary from small scale to large scale.

Benjamin Franklin (1706 - 1790) performed a number of experiments and concluded that there are two kinds of electric charges. He named them positive (+ve) and negative (-ve) charge. Many bodies around us store charges in a balanced form hence are said to be electrically neutral. A body is said to be charged when there is a charge imbalance within it. Any two or more charged bodies exerts a force on each other.

Electrostatics is the study of charges that are either at rest with respect to each other or moving very slowly. The speed of flow of charges depends on whether the material is a conductor, a semiconductor or an insulator.

The electrostatic force between charged particles is one of the fundamental forces of nature. We begin by describing some of the basic properties of electric forces. We then discuss Coulomb's law, which is the fundamental law governing the force between any two charged particles.

1.1.2 Properties of electric charge

Various experiments show that: There exist two kinds of charges: a positive charge and a negative charge

- Unlike charges attract one another and like charges repel one another. This is the law of charges.

- **Electric charge is conserved**. i.e., when one object is rubbed against another, charge is not created in the process. The electrified state is due to a transfer of charge from one object to the other.
- **Electric charge is quantized**. i.e., Always occurs as some integral multiple of a fundamental amount of charge *e*. In modern terms, the electric charge q is said to be quantized. That is, electric charge exists as discrete "packets," and we can write $q = Ne$ where N is some integer. The electron has a charge ^{-}e and the proton has a charge of equal magnitude but opposite sign ^{+}e. Some particles, such as the neutron, have no charge. A neutral atom must contain as many protons as electrons.

1.1.3 The Coulomb's law

Charles Coulomb (1736–1806) measured the magnitudes of the electric forces between two small charged spheres. He found that the magnitude of the electrical force, F between two small charged bodies is directly proportional to the product of the charges q_1 q_2 on the two bodies and inversely proportional to the square of their separation distance r.

$$F \ \alpha \ \frac{q_1 q_2}{r^2} \qquad\qquad q_2 \circ \leftarrow r \rightarrow \circ q_1$$

$$\Rightarrow F = k\,\frac{q_1 q_2}{r^2} \quad (coulomb's \quad law)$$

where the constant $\boldsymbol{\varepsilon}_0$ is known as the *permittivity of free space* and has the value $\boldsymbol{8.8542 \times 10^{12}\ C^2/Nm^2}$.

$$\frac{1}{4\pi\varepsilon} = 8.99\times10^{9}\,Nm^2/C^2$$

$$\frac{1}{4\pi\varepsilon} = 8.99\times10^{9}\,Nm^2/C^2$$ This holds for point charges.

Coulomb's law states that:

(i) Electric force is inversely proportional to the square of separation, r between the two particles and directed along the line joining the particles

(ii) The force is proportional to the product of the charges q_1 and q_2 on the two particles

(iii) The force is attractive if the charges are of opposite sign and repulsive if the charges have the same sign.

This law is directly related to the Newton's law of gravitation. i.e. $F = \frac{Gm_1 m_2}{r^2}$

However, the two laws differ in that;

- F in gravitation always attract while for F in coulomb's law can attract or repel
- force is a vector while charge is a scalar

A coulomb is the amount of charge that flows through a given cross-section of a wire in one second if there is a steady current of 1 ampere in the wire. $Q = It$

Example

The electron and proton of a hydrogen atom are separated (on the average) by a distance of approximately $5.3\times10^{-11}\boldsymbol{m}$. Find the magnitudes of the electric force and the gravitational force between the two particles.

Solution

From Coulomb's law, we find that the attractive electric force has the magnitude

$$\boldsymbol{F_e} = \boldsymbol{k}\frac{|\boldsymbol{e}|^2}{\boldsymbol{r}^2} = \left(8.99\times10^9 \frac{\boldsymbol{N.m}^2}{\boldsymbol{C}^2}\right)\frac{\left(1.6\times10^{-9}\boldsymbol{C}\right)^2}{\left(5.3\times10^{-11}\boldsymbol{m}\right)^2} = 8.2\times10^{-8}\boldsymbol{N}$$

Using Newton's law of gravitation for the particle masses, we find that the gravitational force has the Magnitude.

$$\boldsymbol{F_g} = \boldsymbol{G}\frac{\boldsymbol{m_e m_p}}{\boldsymbol{r}^2} = \left(6.79\times10^{-11}\frac{\boldsymbol{N.m}^2}{\boldsymbol{kg}^2}\right)\frac{\left(9.11\times10^{-51}\boldsymbol{C}\right)\left(1.67\times10-27\boldsymbol{kg}\right)}{\left(5.3\times10^{-11}\boldsymbol{m}\right)^2} = 3.6\times10^{-47}\boldsymbol{N}$$

The ratio Thus, the gravitational force between charged atomic particles is negligible when compared with the electric force. Note the similarity of form of Newton's law of gravitation and Coulomb's law of electric forces. Other than magnitude, what is a fundamental difference between the two forces?

When the current in a wire is 1A, the amount of charge that flows past a given point of the wire in 1sec is 1C. The smallest unit of charge known in nature is the charge of an electron or proton. The charge of an electron or proton has magnitude $|e| = 1.602\times10^{-19}C$. Therefore, 1C of charge is equal to the charge of 6.3×10^{-18} electrons.

Charge and mass of electron, proton and neutron

Particle	Charge, C	Mass, kg
Electron, e	-1.602×10^{-19}	9.11×10^{-31}
Proton, p	$+1.602\times10^{-19}$	1.67×10^{-27}
Neutron, n	0	1.67×10^{-27}

For a system of more than two charge, one should calculate the force between any two charges as if the rest don't exist. To get the results, you get the vector sum of the forces for each pair or component.

The case of more than two charges the coulomb's law apply for each pair e.g. for $q_1, q_2, q_3......$ The force exerted on any one by all the others is;

$$\vec{F}_1 = \sum_{i=2}^{n} \vec{F}_{1i} = \vec{F}_{12} + \vec{F}_{13} + \vec{F}_{14}$$

Where $\vec{F}_{12}$ is the force exerted on q_1 by q_2

When more than two charges are present, the force between any pair of them

is given by

$$\boldsymbol{F}_{12} = \frac{1}{4\pi\varepsilon_0}\frac{\boldsymbol{q}_1\boldsymbol{q}_2}{\boldsymbol{r}^2}\hat{\boldsymbol{r}}$$

Therefore, the resultant force on any one of them equals the vector sum of the forces exerted by the various individual charges. For example, if four charges are present, then the resultant force exerted by particles

2, 3, and 4 on particle 1 is $\boldsymbol{F}_1 = \boldsymbol{F}_{21} + \boldsymbol{F}_{31} + \boldsymbol{F}_{41}$

Example

Consider three point charges located at the corners of a right triangle as shown, where $\boldsymbol{q}_1 = \boldsymbol{q}_3 = 5\boldsymbol{\mu C}$, $\boldsymbol{q}_2 = -2.0\boldsymbol{\mu C}$ and $\boldsymbol{a} = 0.10\boldsymbol{m}$ Find the resultant force exerted on q_3 .

Solution

First, note the direction of the individual forces exerted by q_1 and q_2 on q_3 . The force F_{23} exerted by q_2 on q_3 is attractive because q_2 and q_3 have opposite signs. The force F_{13} exerted by q1 on q3 is repulsive because both charges are positive.

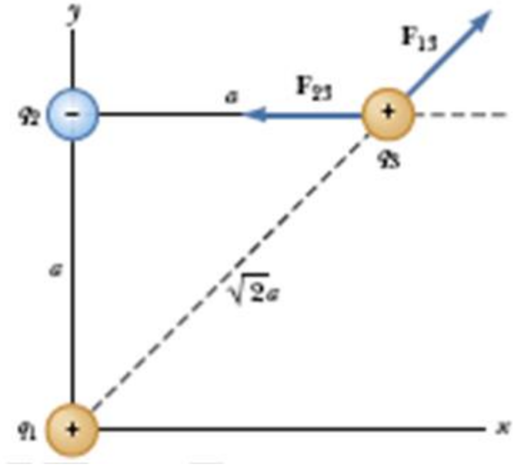


Figure 1.1

The magnitude F_{23} is

$$F_{23} = k_e \frac{|q_2||q_3|}{a^2} = \left(8.99\times10^9 N.m^2C^{-2}\right)\frac{\left(2.0\times10^{-6}C\right)\left(5.0\times10-6C\right)}{\left(0.10m\right)^2} = 9.0N$$

Note that because q_3 and q_2 have opposite signs, F_{23} is to the left,

$$F_{13} = k_e \frac{|q_1||q_3|}{\left(\sqrt{2}a\right)^2} = \left(8.99\times10^9 N.m^2C^{-2}\right)\frac{\left(2.0\times10^{-6}C\right)\left(5.0\times10-6C\right)}{2\left(0.10m\right)^2} = 11N$$

The force F_{13} is repulsive and makes an angle of 45° with the x axis. Therefore, the x and y components of F13 are equal, with magnitude given by;

F_{13} cos 45° = 7.9 N. The force F_{23} is in the negative x direction. Hence, the x and y components of the resultant force acting on q3 are

$$F_{3x} = F_{13x} + F_{23} = 7.9N - 9.0 = -1.1N$$

$$F_{3y} = F_{13y} = 7.9N$$

We can also express the resultant force acting on q_3 in unit vector form as:

$$F_3 = \left(-1.1i + 7.9j\right)N$$

NB: The coulomb's law describes the following:

(i) the force that binds the electrons of an atom to the nucleus

(ii) the forces that bind atoms to form molecules

(iii) the forces that bind atoms and molecules to form a solid or a liquid.

1.1.4 The Electric Field

Whenever a charge, q exists, it sets-up an electric field, E in the space around it. At any point, this electric field strength is defined in terms of the force F exerted on a positive test charge q_0 at a particular point. i.e

$$\vec{E} = \lim_{q_o = 0} \frac{\vec{F}}{q_0}$$

where $\vec{F}$ is the resultant force on the test charge q_0. The test charge should be very small so that its presence does not affect the original charge distribution. The electric field can therefore be simply defined as force per unit charge i.e., $\vec{E} = \dfrac{\vec{F}}{q_0}$

The direction of $\vec{E}$ is the direction of $\vec{F}$ .i.e., the direction to which a resting positive charge would tend to move when placed at that point. The force $\vec{F}$ on a charge q_0 due to a charge q_i is given by the relation;

$$\vec{F} = K\frac{q_i q_0}{r_{io}^2}\hat{r}_{io}, \qquad K = \frac{1}{4\pi\varepsilon_0} = 9.0\times 10^9\, Nm^2C^{-2}$$

The electric field at the position of q_0 is therefore given by

$$\vec{E} = \frac{\vec{F}}{q_0} = K\frac{q_i}{r_{io}^2}\hat{r}_{io}$$

Taking the position of q_i as the origin, then the field due to this charge at any position vector $\hat{r}$ is given by;

$$\vec{E} = K\frac{q_i}{r^2}\hat{r}$$

From Faradays law;

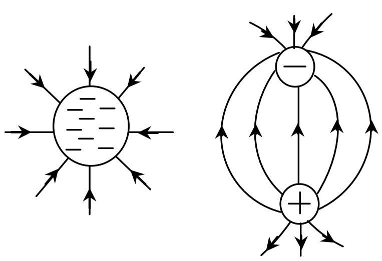


Figure 1.2

NB: The direction of E is the direction of F. i.e. the direction to which a resting positive charge placed at the point would tend to move.

The tangent to the line of force gives the direction of the electric field at that point. The lines of force are drawn so that the number of line per unit cross-section area is proportional to the magnitude of **E**.

$$\because \ \vec{F} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_0}{r^2}$$

$$\Rightarrow \ \vec{E} = \frac{\vec{F}}{q_0} = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r^2} \text{ hence}$$

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 = \sum \vec{E}_n, \quad n = 1, 2, 3...$$

i.e. The total electric field at a point is equal to the vector sum of the fields due to all the individual charges. This is the principle of superposition.

Questions

1. Two charges $q_1 = 1\mu C$ and $q_2 = 2\mu C$ are located on a straight line of length 10 cm. At what point along the line joining the two charges is the electric field equal to zero? (*Hint: see figure below*)

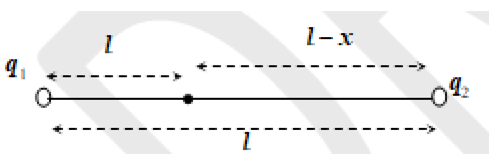


Figure 1.3

2. Determine the electric force on a proton placed in an electric field of 2.0×10^4 N/C directed along the positive x axis. **Ans 3.2×10^{-15} N**
3. A charge $q_1 = 7.0\mu C$ is located at the origin and the second charge $q_2 = -5.0\mu C$ is located on the x axis 0.30 m from the origin. Find the electric field at the point P, which has coordinates (0, 0.4)m. **Ans [(1.1×10^5)i+(2.46×10^5)j] N/C**

1.1.5 Motion of Charged Particles in a Uniform Electric Field

When a particle of charge q and mass m is placed in an electric field E, the electric force exerted on the charge is given as;

$$\vec{F} = q\vec{E}$$

If this is the only force exerted on the charge, then applying Newton's second law to the charge gives;

$$\vec{F} = m\vec{a} = q\vec{E}$$

Therefore, the particle acceleration is;

$$\vec{a} = \frac{q}{m}\vec{E}$$

If the field $\vec{E}$ is uniform (constant in magnitude and direction), then the acceleration is constant for the motion. If the charge is positive charge, this acceleration is in the direction of the electric field. On the other hand, if the charge is negative the acceleration is in this direction opposite the electric field.

Example

A positive point charge q of mass m is released from rest in a uniform electric field E directed along the x axis as shown in figure on the left. Describe its motion.

Solution

The constant acceleration $\vec{a} = \frac{q}{m}\vec{E}$. The motion of charge is simple linear motion along the x axis. Applying the kinematic equation in one dimension;

(i) $x - x_o = v_o t + \frac{1}{2}at^2$

(ii) $v = v_o + at$

(iii) $v^2 = v_o^2 + 2a(x - x_o)$

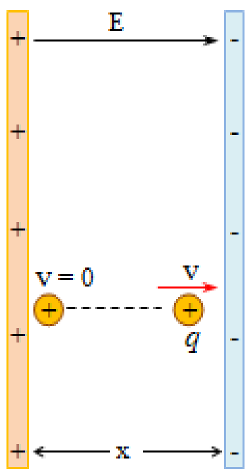


Figure 1.4

Take at $x = x_o = 0$, $v_o = 0$

Then, $x = \frac{1}{2}at^2 = \frac{1}{2}\left(\frac{qE}{m}\right)t^2$ and $v = at$

$\therefore v = at = \frac{qE}{m}t$ or $v^2 = 2ax = \left(\frac{2qE}{m}\right)x$

This means the kinetic energy of the charge after it has moved a distance x is

$k = \frac{1}{2}mv^2$ or $k = \frac{1}{2}m\left(\frac{2qE}{m}\right)x = qEx$

NB: Same results can be obtained from the work=-kinetic energy theorem since the work done by an electric force is $F_e x = qEx$ and $W = \Delta K$.

1.1.6 Electric Flux and Gauss Law

Electric flux is represented by the number of electric field lines penetrating the surface. Consider an electric filed that is uniform in both magnitude and direction as shown below.

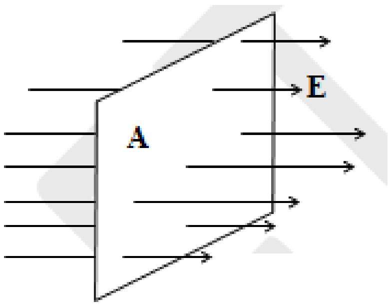


Figure 1.5

The electric field lines will penetrate the rectangular surface of area A, which is perpendicular to the field. With the number of lines per unit area being proportional to the magnitude of the electric field, then the number of lines penetrating the surface is proportional; to the product $\vec{E} \cdot A$.

The product of the electric field strength E and a surface area A perpendicular to the field is called the electric flux, Φ.

$$\Phi = \vec{E} \cdot A \text{ units N.m}^2\text{/C.}$$

It is a measure of the electric field perpendicular to a surface. If the surface under consideration is not perpendicular to the field as shown below, the number of lines (the flux) through it must be less than that $\vec{E} \cdot A$ hence given by;

$$\Phi = \vec{E} \cdot A \ \cos\theta$$

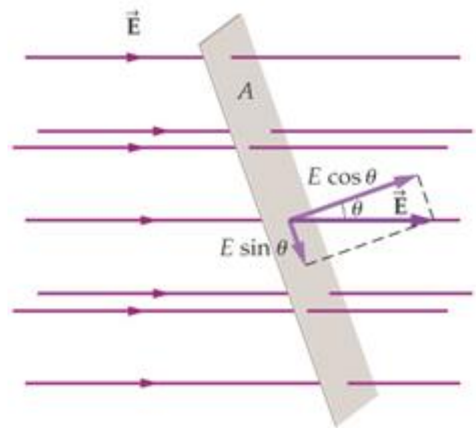


Figure 1.6

We note that, whenever a surface is being penetrated it encloses some net charge and the net number of lines that go through the surface is proportional to the net charge within the surface. Also, the number of lines counted is independent of the shape of the surface enclosing the charge. This is the Gauss's law. It is a general law that applies to any closed surface.

Gauss's law states that: the electric flux through a closed surface is proportional to the charge enclosed by the surface;

$$\Phi = \frac{q}{\varepsilon_o}, \ SI\ unit: N.m^2/C$$

$$\varepsilon_o = \frac{1}{4\pi\varepsilon_o} = 8.85\times10^{-12} C^2/N\cdot m^2$$

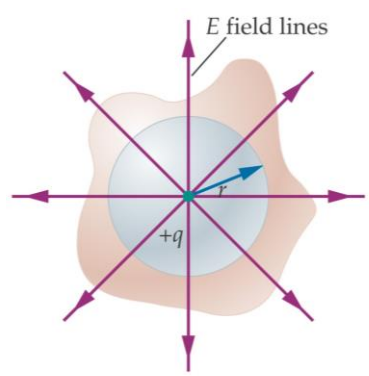


Figure 1.7

It is an important tool since it permits the assessment of the amount of enclosed charge by mapping the field on a surface outside the charge distribution. For geometries of sufficient symmetry, it simplifies the calculation of the electric field.

Example

An electric field of 500 V/m makes an angle of 30.00 with a surface vector. If its magnitude is 0.500 m^2, calculate the electric flux that passes through this surface.

Solution: The electric flux through the surface is given by:

$$\Phi = \vec{E} \cdot A \cos\theta$$

$$\Phi = (500\, V/m)(0.500\, m^2) \cos 30^o, \quad \Phi = 217\, Vm$$

Example

Explain why electric field does not exist inside a spherical shell?

Solution: within a spherical shell, the enclosed charge q will be zero, since the surface charge density is dispersed outside the surface. Therefore there is no charge inside the spherical shell hence, E = 0.

Question

A uniform electric field has a magnitude, $E = 3\times10^3 \hat{i}\ N/C$. Calculate the flux of this field through a square surface of 10 cm on a side whose plane is parallel to the yz plane? [**Ans: 30 Nm^2/C**].

1.1.7 Direct Current Circuits (DC)

A direct current (DC) circuit- Is one in which the flow of charge is in only one direction.

I. Electric current

An electric current is a flow of charge.

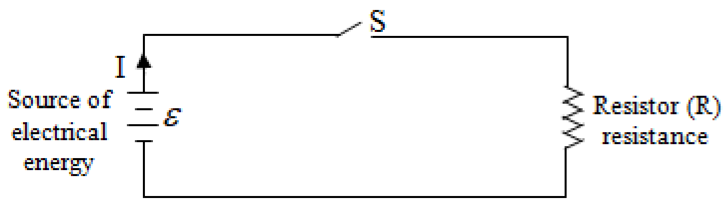


Figure 1.10

A source of electromotive force

This is a device that transforms one source of energy into a source of electric energy. The electric current (I) - Is defined as the amount of charge per second that pass through a cross-section of the conductor.

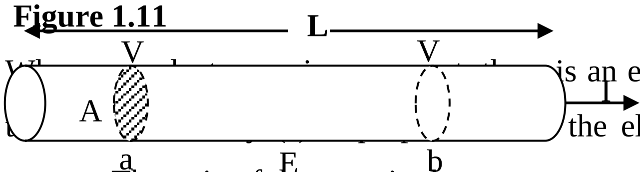
$$I = \frac{\Delta Q \quad (coulomb)}{\Delta t \quad (second)}$$

The SI units of current (I) is the coulomb per second called Ampere (A).

The direction of the current is defined as the direction that the positive charge would move in response to the electric field

Ohm's Law

Figure 1.11



is an electric field inside the conductor. In many conductors, the electric field (E) in the conductor which produces the current. The ratio of the magnitude

$$\therefore \quad \sigma = \frac{J}{E} \quad or \quad J = \sigma E \text{...}1$$

$$i.e.\ \sigma = \frac{current\ \ density(J)}{electrical\ \ \text{int}ensity(E)}$$

Materials which obey this law are said to be Ohmic

The latter figure shows a segment of a wire of a length L and cross section area A carrying current I. We choose the segment short enough to ensure that the electric field does not vary appreciably over the distance L. If the electric field is directed from point a to b, the potential is lower at point b than a by amount $i.e.\ V = V_a - V_b = EL$ Where E is the electric field. The current in the wire is the current density x the cross-sectional area:

$i.e.\ I = JA$ but $J = \sigma E$ $\quad \therefore \quad I = \sigma EA$

But $\quad E = \dfrac{V}{L} \quad \therefore \quad I = \sigma A \dfrac{V}{L}$ $\quad$ or $\quad V = \dfrac{L}{\sigma A} I$

The quantity $\dfrac{V}{I} = \dfrac{L}{\sigma A}$ is called the Resistance R of the wire segment

$$\therefore R = \frac{V}{I} = \frac{L}{\sigma A} \text{...}2$$

SI units of R= Ohm $\quad 1\Omega = 1\dfrac{Volt}{Ampere}$

Written in another way,

$$V = IR \text{.............................}ohms\ \ law\text{...}3$$

The Ohm's law states that: The potential difference V between two points of a conductor is directly proportional to the current flowing in it provided that temperature is constant.
NB: Resistance is independent of the voltage and current.

From the equation , $R = \frac{L}{\sigma A}$, the resistance of a wire is proportional to the length (L) and inversely proportional to the cross-sectional area (A) of the wire. The constant of proportionality is called resistivity; ρ .the resistivity is the reciprocal of conductivity, σ .

i.e $\rho = \frac{1}{\sigma}$. SI unit of $\rho = \left[\frac{RA}{L}\right] = ohm.metre$

NB: Resistivity depends on temperature.

Example

Calculate the resistance of an aluminum cylinder that is 10.0 cm long and has a cross sectional area of $2.50 \times 10^{-4} m^2$ and resistivity of $2.82 \times 10^{-8} \Omega m$. Repeat the calculation for a glass cylinder with a resistivity 3.0×10^{-10} Ωm.

Solution

$$R = \rho \frac{L}{A} = 2.82 \times 10^{-8} \Omega m \left(\frac{0.100m}{2.50 \times 10^{-4} m^2} \right) = 1.13 \times 10^{-5} \Omega$$

$$R = \rho \frac{L}{A} = 3.0 \times 10^{10} \Omega m \left(\frac{0.100m}{2.50 \times 10^{-4} m^2} \right) = 1.2 \times 10^{13} \Omega$$

The resistance varies with temperature. Therefore if a wire has a resistance R_o at a temperature T_o then its resistance R at a temperature T is;

$$R = R_o + \alpha R_o (T - T_o)$$

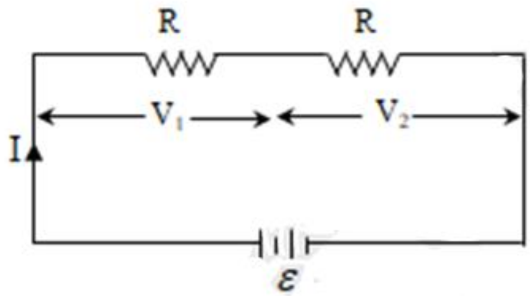
where α is the temperature coefficient of resistance of the material of the wire and it varies with temperature and so its application is over short ranges (Units: K^{-1} or $^{O}C^{-1}$).

Similarly for resistivity; $\rho = \rho_o + \alpha \rho_o (T - T_o)$

Question: A copper conductor of square cross section 1 mm^2 on a side carries a constant current of 20 A. The density of free electrons is 8×10^{28} electron per cubic meter. Determine the current density and the drift velocity.

Resistors in series and parallel.

- **Resistors in series**



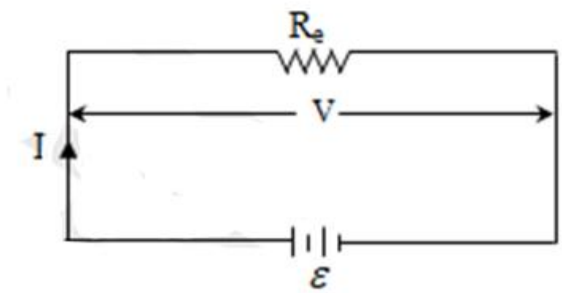


Figure 1.12 (a) (b)

$$V = V_1 + V_2$$

Each has same current.

From Ohm's law

$$V = IR$$

$$IR_{eq} = IR_1 + IR_2 \;,\; R_{eq} = R_1 + R_2$$

$$\therefore \quad IR_{eq} = V = I(R_1 + R_2)$$

Hence: $R_{ser} = R_1 + R_2$

- **Resistors in parallel.**

We apply the concept of conservation of charge

$$I = I_1 + I_2$$

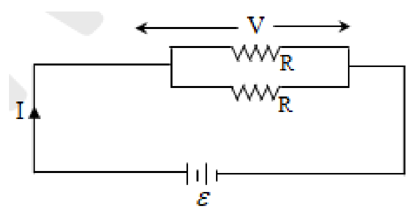


Figure 1.12 (c)

From Ohm's law,

$$I = \frac{V}{R} \qquad \therefore \quad I_{R_{eq}} = I_{1(R_1)} + I_{2(R_2)}$$

$$\frac{V}{R_{eq}} = \frac{V}{R_1} + \frac{V}{R_2}$$

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$\therefore \quad \frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots$$

NB **1:** The combined resistance of n equal resistances R in parallel is given by

$$\frac{1}{R_{par}} = \frac{n}{R} \Rightarrow R_{par} = \frac{R}{n}$$

2: For two resistors only in parallel, rearrangement gives; $R_{par} = \dfrac{R_1 R_2}{R_1 + R_2}$

1.1.8 Electrical Energy and Power

Consider the circuit shown the following figure. Let a positive charge ΔQ moving around the circuit from point ***a*** through the battery, the resistor and back to ***a***. As the charge moves from ***a*** to ***b*** its potential energy increases by some amount $\Delta U = \Delta V \Delta Q$ (where V is the potential difference between point ***b*** and ***a***) while the chemical potential energy in the battery decreases by the same amount (From $\Delta U = q\Delta V$).

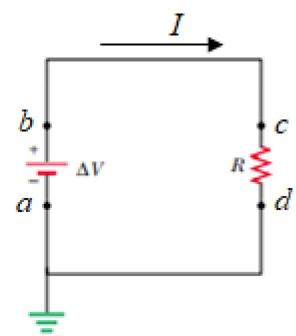


Figure 1.13

The charge loses this electrical potential energy as it undergoes collision with atoms in the resistor as moves from ***c*** to ***d*** the resistor. The rate at which the charge ΔQ loses potential energy in going through the resistor or simply the electrical power delivered by an energy source as it carries a charge ***q*** through potential ΔV in time ***t*** is;

$$P = \frac{\Delta U}{\Delta t} = \frac{\Delta Q}{\Delta t}\Delta V$$

Hence;

$$P = I\,\Delta V$$

where *I* is the current in the circuit. However, the charge regains this energy when it passes through the battery.

In this case, a battery supplies the power to the resistor. However, to determine the power transferred to any device carrying a current *I* and having a potential difference ΔV between its terminals, the above equation ($P = I\,\Delta V$) can be used. Since for a resistor, $\Delta V = IR$, we can express the power delivered to the resistor alternatively as;

$$P = I^2 R = \frac{(\Delta V)^2}{R}$$

If I is in amperes, ΔV in volts, and *R* in ohms, then the SI unit of power is the watt. The power lost as internal energy in a conductor of resistance *R* is called ***joule heating***. This transformation is also often referred to as an I^2R loss. A device that supplies electrical energy (e.g. a battery, is called either a *source of electromotive force* or simply an *emf source*.

Example

An automated electric heating system is constructed by applying a potential difference of 120 V to a Nichrome wire that has a total resistance of $8.00\ \Omega$. Calculate the current carried by the wire and the power rating of the heater.

Solution: Since $\Delta V = IR$; $I = \dfrac{\Delta V}{R} = \dfrac{120\,V}{8.00\ \Omega} = 15.0\ A$

The power rating is given by; $P = I^2R = (15.0\ A)^2(8.00\ \Omega) = 1.80\ kW$

QN: If we doubled the applied potential difference, how would the current and power be affected?

Question: A light bulb is rated at 120 v/ 75W. The bulb is powered by a 120 v. Find the current in the bulb and its resistance.

1.1.9 Kirchoff's Laws

Ohm's law is unable to say about the current in a complicated circuit. Kirchoff, in 1842 gave two general laws which are known after him. These laws add up to nothing! However, they completely characterize circuit behavior

i. Kirchoff's Current Law (KCL) – Its a consequence of the law of conservation of charge.
States that, the algebraic sum of the currents at any junction/node in a circuit is zero.

i.e. $\sum I = 0$ meaning $I_1 + I_2 + ... + I_n = 0$

(This is the branch theorem and follows the law of charge conservation. This is due t o the fact that charge or current cannot accumulate at the junction point. NB: current entering are considered to be positive while those leaving are negative).

ii. Kirchoff's Voltage Law (KVL)
It states that, in any closed circuit, the algebraic sum of the product of the currents and the corresponding resistance in respective branches are equal to the total *e.m.f* of the circuit.

i.e. $\sum IR = \sum E$

also written as: the algebraic sum of the voltage increases and decreases around any closed loop is equal to zero. i.e $\sum V = 0$

(This is the loop theorem and follows the law of energy conservation).

Applications of D.C circuits

The application of Kirchoff's rules and Ohm's law are geared towards the measurement of voltage, current and resistance. There are number of devices used to measure the above parameters for the circuit. e.g. ammeter, voltmeter, wheatstone bridge and potentiometer.

1.2 Alternating Currents

1.2.1 Alternating currents signals

Most of the energy source we use in our electric or electronic devices is from AC source. This is sometimes converted into DC source.

- The term ***alternating*** indicates only that the waveform alternates between two prescribed levels in a set time sequence. An alternating current is produced in a circuit when the potential difference (which we call *voltage*) between the terminals of the emf source (e.g. AC generator) in the circuit changes sign from moment to moment.
- The voltage V between the terminals of an AC generator is given by

$$v = V_o \sin(2\pi ft)$$

where V_0 is the magnitude of the maximum voltage (the voltage *amplitude*), f is the frequency (in cycles per second) and t is the time in seconds. As we shall see, such a voltage is very easily produced by a rotating armature. The voltage V produced between the terminals of an AC generator fluctuates sinusoidally with time.

If an AC generator is placed across a resistor R the current in the circuit is given by:

$$i = \left(\frac{V_o}{R}\right)\sin(2\pi ft) = I_o \sin(2\pi ft)$$

where $I_0 = \dfrac{V_0}{R}$ is the current amplitude.

Whatever the source of origin, the electric current is fundamentally the same in all cases, but the manner in which it varies with time may be very different.

NB: Electricity is produced by generators at power stations and then distributed by a vast network of *transmission lines* (called the National Grid system) to industry and for domestic use.

- It is easier and cheaper to generate alternating current (a.c.) than direct current (d.c).
- The a.c. is more conveniently distributed than d.c. since its voltage can be readily altered using transformers (step-up or step-down).
- Whenever d.c. is needed in preference to a.c., devices called rectifiers are used for conversion.

1.2.2 A.C as a wave-form

- The voltage v produced between the terminals of an AC generator fluctuates sinusoidally with time.
- An AC voltage (or ac current) varies ***sinusoidally w***ith time, as shown below. This is a ***periodic*** voltage since it varies with time such that it continually repeats.
- ***Period (T)*****:** The time interval (in seconds) between successive repetitions of a periodic waveform. This can also be called the Periodic Time of the waveform for sine waves, or the Pulse Width for square waves.
- ***Cycle:*** The portion of a waveform contained in one period of time.
- ***Frequency (Hertz):*** is the number of times the waveform repeats itself within a one second time period. Frequency is the reciprocal of the time period, (f = 1/T).
- ***The Amplitude (A):*** is the magnitude or intensity of the signal waveform measured in volts or amps.
- ***Instantaneous value:*** The magnitude (values) of the alternating quantities (waveform) at any instant of time; They are represented by small letters, i, v, e etc.
- ***Peak amplitude*****:** The maximum value of the waveform as measured from its average (or mean) value. Also referred to as **maximum value** or the **crest value** or the **amplitude** of the waveform. Such values are represented by uppercase letters Vm, Im or Vo, Io or Vp, Ip.
- ***A peak-to-peak*** **value of e.m.f:** is the difference between the maximum and minimum values in a cycle.

1.2.3 Power in A.C signals

Given that $I = I_0 \sin(2\pi ft)$ and $V = V_0 \sin(2\pi ft)$, Using $P = IV$ we get for power: $P = I_0 V_0 \sin^2(2\pi ft)$. Since the power also fluctuates with time, it is convenient to measure the *average power* over one cycle. This is obtained from the average value of $\sin^2(2\pi ft)$ over one cycle: ½.

Thus we can write $\bar{P} = \frac{1}{2} I_0 V_0$ where $\bar{P}$ denotes average power over one cycle. Since $I_0 = \dfrac{V_0}{R}$ we can also write $\bar{P} = \frac{1}{2}\dfrac{V_0^2}{R}$ or $\bar{P} = \dfrac{V_{\text{rms}}^2}{R}$ where $V_{\text{rms}} = \dfrac{V_0}{\sqrt{2}}$ is the *root mean square* voltage.

The root mean square (or rms) value is a way of expressing an average value of the magnitude of a quantity that varies sinusoidally with time over one cycle. Since the current in an AC circuit also varies sinusoidally with time we can also write $I_{\text{rms}} = \dfrac{I_0}{\sqrt{2}}$ where I_0 is the current amplitude. Hence we can also write $\bar{P} = I_{\text{rms}} V_{\text{rms}}$.

Note the difference between *average power* $\bar{P}$ and *peak power* P_0:

$$\bar{P} = I_{\text{rms}} V_{\text{rms}} \text{ or } \bar{P} = \frac{V_{\text{rms}}^2}{R};\ P_0 = I_0 V_0 \text{ or } P_0 = \frac{V_0^2}{R} \quad \text{and that} \quad P_0 = 2\bar{P}$$

(We assume that power is being delivered to a *resistive load*.)

Example: *A light bulb consumes 60 watts of power when connected to 120 v_{ac}.*

a) *What is the resistance of the light bulb?*

$$R = \frac{(120\ \text{V})^2}{60\ \text{W}};\quad R = 240\ \frac{\text{V}(\cancel{\text{J}}/\text{C})}{\cancel{\text{J}}/\text{s}}$$

$$\bar{P} = \frac{V_{\text{rms}}^2}{R};\quad R = \frac{V_{\text{rms}}^2}{\bar{P}};\quad \boxed{R = 240\ \text{ohms}}$$

b) *What current does the light bulb draw?*

Current refers to *rms current*:

$$I_{\text{rms}} = \frac{V_{\text{rms}}}{R};\ I_{\text{rms}} = \frac{120\ \text{V}}{240\ \Omega};\ \boxed{I_{\text{rms}} = 0.50\ \text{A}}$$

$$\Rightarrow I_{\text{rms}} = \frac{\bar{P}}{V_{\text{rms}}}:\ I_{\text{rms}} = \frac{60\ \text{W}}{120\ \text{V}};\ \boxed{I_{\text{rms}} = 0.50\ \text{A}}$$

c) *What is the peak power consumed by the light bulb?*

$$V_0=\sqrt{2}V_{\text{rms}}: \quad V_0=\sqrt{2}(120\text{ V}); \quad V_0=170\text{ V}$$

$$I_0=\sqrt{2}I_{\text{rms}}: \quad I_0=\sqrt{2}(0.50\text{ A}); \quad I_0=0.707\text{ A}$$

$$P_0=I_0V_0: \quad P_0=(0.707\text{ A})(170\text{ V}); \quad \boxed{P_0=120\text{ W}}$$

(Note that $P_0=2\bar{P}$.)

1.2.4 A.C circuit elements

- **Purely resistive load**

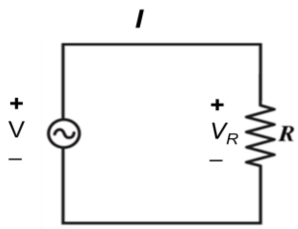


Figure 1.14(a)

In a purely resistive a.c. circuit, the current I_R and applied voltage V_R are in phase. i.e

$$V_R=V_o\sin\omega t \text{ and } I_R=I_o\sin\omega t$$

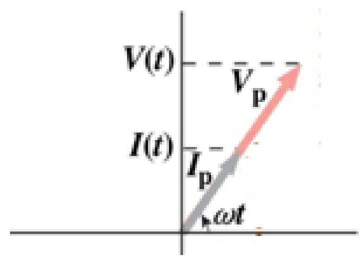


Figure 1.14(b)

Ohm's law is expressed as;

$$V_R=IR \text{ or } R=\frac{V}{I}\ (\Omega)$$

When $V(t)>0$

$$V(t)-V_R=0, \quad V_m\sin\omega t-IR=0$$

$$I=\frac{V(t)}{R}=\frac{V_p}{R}\sin\omega t$$

$$I=I_m\sin\omega t$$

$$I_{rms}=\frac{V_{rms}}{R}$$

$$I_m=\frac{V_m}{R}$$

- **Purely inductive load**

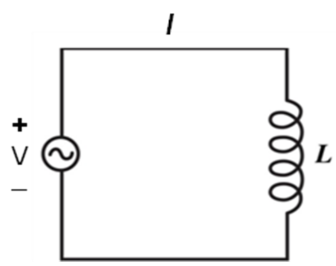


Figure 1.15(a)

When $V(t)>0$:

$$V(t)+\square_L=0$$

$$V_m\sin\omega t-L\frac{dI}{dt}=0$$

$$\frac{dI}{dt}=\frac{V_m}{L}\sin\omega t$$

$$I=\frac{1}{L}\int V(t)\,dt=\frac{V_m}{L}\int\sin\omega t\;dt=-\frac{V_m}{\omega L}\cos\omega t$$

$$I=\frac{V_m}{\omega L}\sin\left(\omega t-\frac{\pi}{2}\right)$$

$$I_m=\frac{V_m}{\omega L}\equiv\frac{V_m}{X_L}$$

$$X_L=\omega L\ (\Omega)$$

In a purely inductive a.c. circuit, the current I_L lags the applied voltage V_L by 90° (or $\pi/2$ rads)

i.e. $V_L = V_o \sin \omega t$ and $I_L = I_o \sin(\omega t - \pi/2)$

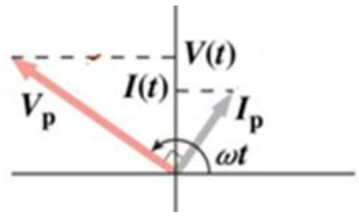


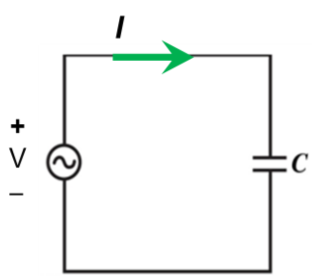
Figure 1.15(b)

In a purely inductive circuit the opposition to the flow of alternating current is called the **inductive reactance, X_L** i.e

$X_L \to 0$ as $\omega \to 0$ DC: short circuit

$X_L \to \infty$ as $\omega \to \infty$ .HF: open circuit

- **Purely capacitive load**



The current starts out at a large value and charges the plates of the capacitor. There is initially no resistance to hinder the flow of the current while the plates are not charged. As the charge on the plates increases, the voltage across the plates increases and the current flowing in the circuit decreases.

Figure 1.16(a)

When $V(t) > 0$,

$$V(t) - V_C = 0, \quad V_m \sin \omega t - \frac{q}{C} = 0$$

$$q = C\,V(t) = CV_m \sin \omega t$$

$$I = \frac{dq}{dt} = C\frac{dV}{dt} = C\,V_m\,\omega \cos \omega t$$

$$I = C\,V_p\,\omega \sin\left(\omega t + \frac{\pi}{2}\right)$$

So, I leads *V* by 90°

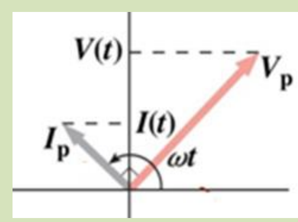


Figure 1.16(b)

In a purely capacitive a.c. circuit, the current I_C leads the applied voltage V_C by 90° (or $\pi/2$ rads) i.e.

$V_L = V_o \sin \omega t$ and $I_C = I_o \sin(\omega t + \pi/2)$). $I_m = C\,V_m\,\omega \equiv \dfrac{V_m}{X_C}$

$$X_C = \frac{1}{\omega C}\ (\Omega)$$

where C is the capacitance in farads. X_C is inversely proportional to frequency **ω**.

In a purely capacitive circuit the opposition to the flow of alternating current is called the **capacitive reactance, X_C**

$X_C \to \infty$ as $\omega \to 0$ Static: open circuit

$X_C \to 0$ as $\omega \to \infty$ HF: short circuit. Also, $V_{rms} = I_{rms} X_C$

Summary

Circuit Element	**Peak Current vs Voltage**	**Phase Relation**
Resistor	$I_o = \frac{V_o}{R}$	V & I in phase
Capacitor	$I_o = \frac{V_o}{X_C} = \frac{V_o}{1/\omega C}$	V lags I 90°
Inductor	$I_o = \frac{V_o}{X_L} = \frac{V_o}{\omega L}$	V leads I 90°

1.2.5 Some Applications

The Antenna

A current varying circuit produces a changing magnetic field and likewise a varying magnetic field produces accompanying varying electric field (that is electricity and magnetism co-exists). Energy stored in an LC circuit is continually transferred between the electric field of the capacitor and the magnetic field of the inductor. However, this energy transfer continues for prolonged periods of time only when the changes occur slowly. If the current alternates rapidly, the circuit loses some of its energy in the form of electromagnetic waves. As a fact, electromagnetic waves are radiated by any circuit carrying an alternating current. The fundamental mechanism responsible for this radiation is the acceleration of a charged particle. Whenever a charged particle accelerates it radiates energy. An alternating voltage applied to the wires of an antenna forces electric charges in the antenna to oscillate. This is a common technique for accelerating charged particles and is the source of the radio waves emitted by the broadcast antenna of a radio station.

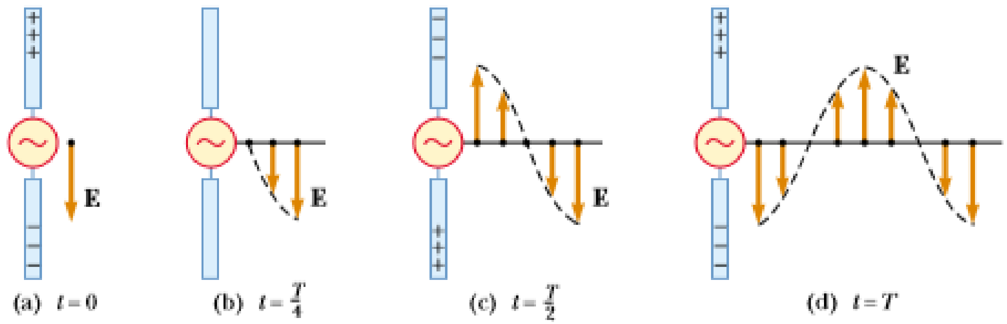


Figure 1.17

Production of electromagnetic wave by oscillating electric charges. The electric field is set up by the charges oscillating in an antenna and the field moves away from the antenna with the speed of light.

PART 2: MAGNETISM

2.1 Basic concepts of magnetism and electromagnetism

(i) When a charged particle is moving through a magnetic field, a magnetic force acts on it

Charged objects interact in terms of electric fields. An electric field surrounds any stationary or moving electric charge. In addition to an electric field, the region of space surrounding any moving electric charge also contains a magnetic field. A magnetic field also surrounds any magnetic substance.

A magnetic field is represented by the symbol $\boldsymbol{B}$. The direction of the magnetic field $\boldsymbol{B}$ at any location is the direction in which a compass needle points at that location. The following figure shows how the magnetic field of a bar magnet can be traced with the aid of a compass.

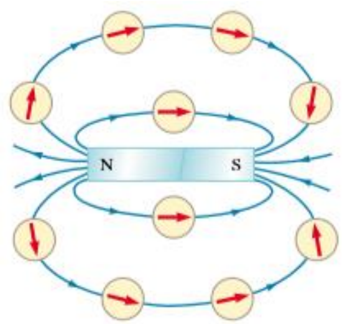


Figure 2.1

Notice that the magnetic field lines outside the magnet, point away from the north poles and toward the south poles. Magnetic field B at some point in space can be defined in terms of magnetic force $\boldsymbol{F_B}$ that the field exerts on a test object (i.e. charged particle moving with a velocity **v**). Assuming that no electric or gravitational fields are present at the location of the test object, experiments on various charged particles moving in a magnetic field results in the following:

- The magnitude $\boldsymbol{F_B}$ of the magnetic force exerted on the particle is proportional to the charge $\boldsymbol{q}$ and to the speed **v** of the particle.
- The magnitude and direction of $\boldsymbol{F_B}$ depend on the velocity of the particle and on the magnitude and direction of the magnetic field $\boldsymbol{B}$.
- When a charged particle moves parallel to the magnetic field vector, the magnetic force acting on the particle is zero.
- When the particle's velocity vector makes any angle with the magnetic field, the magnetic force acts in a direction perpendicular to both **v** and $\boldsymbol{B}$; that is, $\boldsymbol{F_B}$ is perpendicular to the plane formed by **v** and $\boldsymbol{B}$. (as depicted in the figure **a** below).
- The magnetic force exerted on a positive charge is in the direction opposite the direction of the magnetic force exerted on a negative charge moving in the same direction (figure **b** below).

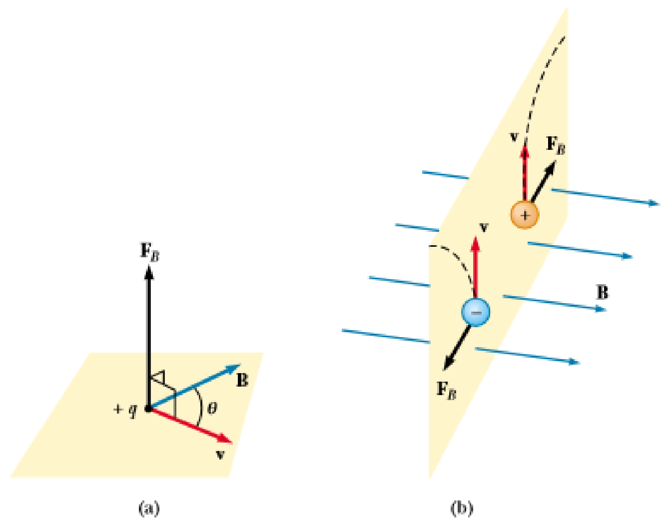


Figure 2.2

- The magnitude of the magnetic force exerted on the moving particle is proportional to $\sin\theta$, where θ is the angle the particle's velocity vector makes with the direction of ***B***. We can summarize these observations by writing the magnetic force in the form:

$$\mathbf{F_B} = q\mathbf{v} \times \mathbf{B}$$

where the direction of $\boldsymbol{F_B}$ is in the direction of $\mathbf{v} \times \mathbf{B}$ if $\boldsymbol{q}$ is positive, which is perpendicular to both **v** and **B** by definition of the cross product. This equation is an operational definition of the magnetic field at some point in space. That is, the magnetic field is defined in terms of the force acting on a moving charged particle.

Its magnitude is given by $F_B = qvB\sin\theta$

From the equation above, it can be observed that the force on a charged particle moving in a magnetic field has its maximum value when the particle's motion is perpendicular to the magnetic field, corresponding to θ = 90°, so that sin θ = 1. The magnitude of this maximum force has the value $F_B(\max) = qvB$

Also F is zero when is parallel to (corresponding to θ = 0° or 180°), so no magnetic force is exerted on a charged particle when it moves in the direction of the magnetic field or opposite the field.

Given that **F** is in newtons, ***q*** is in coulombs, and v is in meters per second, then the SI unit of magnetic field is in the weber (Wb) per square meter (1 T = 1 Wb/m^2, also called the tesla (T).

Experiments show that the direction of the magnetic force is always perpendicular to both **v** and **B**, as shown in the figures below for a positively and negatively charged particles. This is determined by the right-hand rule number 1:

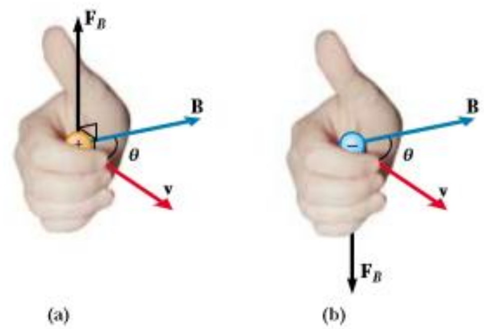


Figure 2.3

1. Point the fingers of your right hand in the direction of the velocity v.

2. Curl the fingers in the direction of the magnetic field, moving through the smallest angle.

3. Your thumb is now pointing in the direction of the magnetic force exerted on a positive charge.

(ii) Magnetic Force on a Current-Carrying Conductor

If a magnetic field exerts a force on a single charged particle when it moves through a magnetic field, the magnetic forces are exerted on a current-carrying wire, as well. This follows from the fact that the current is a collection of many charged particles in motion; hence, the resultant force on the wire is due to the sum of the individual forces on the charged particles. The force on the particles is transmitted to the bulk of the wire through collisions with the atoms making up the wire.

(iii)Motion of a Charged Particle in a Magnetic Field

Consider the case of a positively charged particle moving in a uniform magnetic field so that the direction of the particle's velocity is perpendicular to the field.

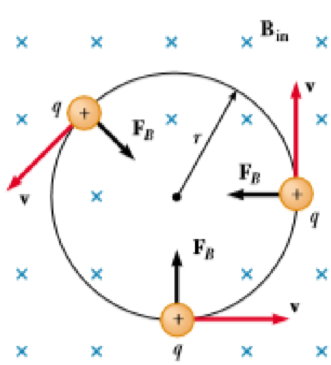


Figure 2.4

The crosses and the label $\mathbf{B_{in}}$ indicate that **B** is directed into the page. Application of the righthand rule at a point P shows that the direction of the magnetic force F at that location is upward.

The force causes the particle to change its direction of travel and follow a curved path.

If the right-hand rule is applied at any point, the magnetic force is seen to be always directed toward the center of the circular path. Therefore, the magnetic force causes a centripetal acceleration, which changes only the direction of v and not its magnitude. Since ***F*** produces the centripetal acceleration, we can equate its magnitude, qvB in this case, to the mass of the particle multiplied by the centripetal acceleration v 2/r. From Newton's second law, we find that;

$$F_B = qvB = \frac{mv^2}{r}$$

which gives;

$$r = \frac{mv}{qB}$$

This equation says that the radius of the path is proportional to the momentum mv of the particle and is inversely proportional to the charge and the magnetic field. The equation is often called the cyclotron equation, because it's used in the design of these instruments (popularly known as atom smashers).

(iv) Magnetic Field of a Long, Straight Wire and Ampère's Law

The direction of the magnetic field B in a long wire is found by right-hand rule number 2:

Point the thumb of your right hand along a wire in the direction of positive current, as in Figure. The fingers then naturally curl in the direction of the magnetic field B. When the current is reversed, B also reverses. These observations lead to a mathematical expression for the strength of the magnetic field due to the current I in a long, straight wire:

$$B = \frac{\mu_o I}{2\pi r}$$

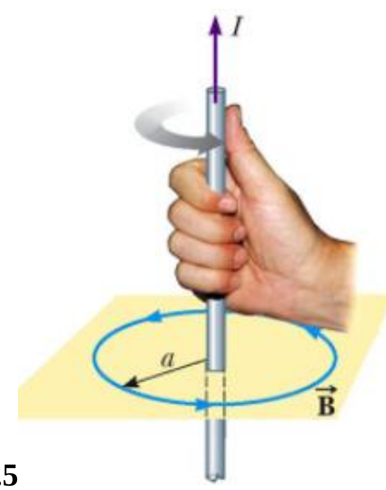


Figure 2.5

(v) Ampère's Law and a Long, Straight Wire

The above equation enables one to calculate the magnetic field due to a long, straight wire carrying a current. This statement, known as Ampère's circuital law, can be written

$$\sum B_E \Delta l = \mu_o I$$

where B_E is the component of parallel to the segment of length Δl and means that we take the sum over all the products $B_E \Delta l$ around the closed path. Ampère's law is the fundamental law describing how electric currents create magnetic fields in the surrounding empty space.

(vi)Magnetic Force Between Two Parallel Conductors

As shown earlier, a magnetic force acts on a current-carrying conductor when the conductor is placed in an external magnetic field. Because a conductor carrying a current creates a magnetic field around itself, it is easy to understand that two current-carrying wires placed close together exert magnetic forces on each other. Consider two long, straight, parallel wires separated by the distance d and carrying currents I1 and I2 in the same direction, as shown below

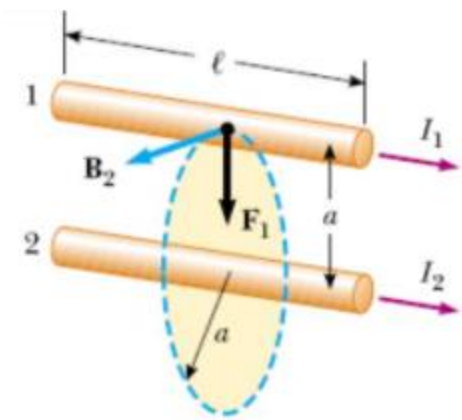


Figure 2.6: Two current carrying conductors

Wire *l* is directly above wire 2. What's the magnetic force on one wire due to a magnetic field set up by the other wire? In this calculation, we are finding the force on wire 1 due to the magnetic field of wire 2. The current I_2, sets up magnetic field B_2 at wire 1. The direction of B_2 is perpendicular to the wire, as shown in the figure.

From;

$$B = \frac{\mu_o l}{2\pi r},$$

is found that the magnitude of this magnetic field is

$$B_2 = \frac{\mu_o l_2}{2\pi d}$$

From $F_{max} = BIl$ the magnitude of the magnetic force on wire 1 in the presence of field due to I_2 is;

$$F_1 = B_2 I_1 l = \frac{\mu_o I_2}{2\pi d} I_1 l = \frac{\mu_o I_1 I_2 l}{2\pi d}$$

We can rewrite this relationship in terms of the force per unit length:

$$\frac{F_1}{l} = \frac{\mu_o I_1 I_2}{2\pi d}$$

The direction of F_1 is downward, toward wire 2, as indicated by right-hand rule number 1. This calculation is completely symmetric, which means that the force on wire 2 is equal to and opposite to F_1, as expected from Newton's third law of action–reaction. We have shown that parallel conductors carrying currents in the same direction attract each other If a conductor carries a steady current of 1A, then the quantity of charge that flows through any cross section in 1s is 1C.

There are several important differences between electric and magnetic forces:

- **I.** The electric force acts in the direction of the electric field, whereas the magnetic force acts perpendicular to the magnetic field.
- **II.** The electric force acts on a charged particle regardless of whether the particle is moving, whereas the magnetic force acts on a charged particle only when the particle is in motion.
- **III.** The electric force does work in displacing a charged particle, whereas the magnetic force associated with a steady magnetic field does no work when a particle is displaced.

From the last statement and on the basis of the work–kinetic energy theorem, we conclude that the kinetic energy of a charged particle moving through a magnetic field cannot be altered by the magnetic field alone. In other words,

2.2 Basic laws applications in computing systems

PART 3: WAVE THEORY

3.1 Periodic motions

3.2 Sinusoidal waves.

3.3 Wave interactions: Interference, reflection and refraction of waves.

3.4 Sound waves.

3.5 Light as a wave

3.6 The Hyugen's theorem

3.1 Periodic Motions

3.1.1 Definition of a wave

A wave is the transfer of energy from one point to another through a rhythmic disturbance in both space and time without transfer of matter. This occurs in a pulse or a continuous (periodic) wave motion. All the waves carry energy, but the amount of energy transmitted through a medium and the mechanism responsible for that that transport differs.

3.1.2 Types of waves

Waves can either be electromagnetic or mechanical waves. **Electromagnetic waves** do not require any medium of transmission and have the same speed.

A mechanical wave is a periodic disturbance that is created by a vibrating object and requires a material medium (solid, liquid or gas) for its propagation transporting energy as it moves.

The mechanism by which a mechanical wave propagates itself through a medium involves particle interaction; one particle applies a push or pull on its adjacent neighbor, causing a displacement of that neighbor from the equilibrium or rest position. These waves are also known as elastic waves because their propagation depends upon the elastic and inertial properties of the medium through which they pass.

Examples for mechanical waves are sound waves and water waves. In these waves, the particles of the medium just vibrate to and fro about their mean position creating a sine wave pattern which continues to move in an uninterrupted fashion until it encounters another wave along the medium or until it encounters a boundary with another medium. This type of wave pattern is referred to as a **traveling wave**.

The medium must possess the following properties for the propagation of the waves:

The medium should be able to return to its original condition after being disturbed, i.e., the medium must possess elasticity.

The medium must be capable of storing energy.

The frictional resistance must be negligible so as not to damp the oscillatory movement.

There are two basic types of wave motion for mechanical waves: **longitudinal** waves and **transverse** waves.

I. Longitudinal Waves

In a longitudinal wave the particle displacement is parallel to the direction of wave propagation. The particles do not move down and up with the wave; they simply oscillate back and forth about their individual equilibrium positions. E.g. Sound waves in air

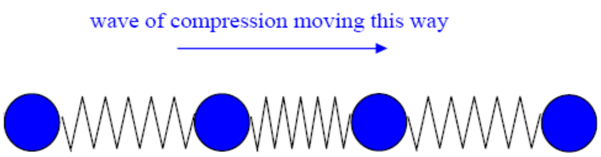


Figure 3.1

II. Transverse Waves

In a transverse wave the particle displacement is perpendicular to the direction of wave propagation. The particles do not move along with the wave; they simply oscillate up and down about their individual equilibrium positions as the wave passes by. E.g. A ripple on a pond and a wave on a string.

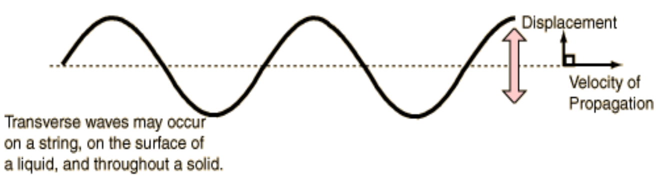


Figure 3.2

3.1.3 General wave equation

Consider a transverse wave traveling along an "ideal" string in the ***x*** direction. An "ideal" string is one in which the disturbance keeps its form as it travels. This means that friction and energy dissipation are ignored.

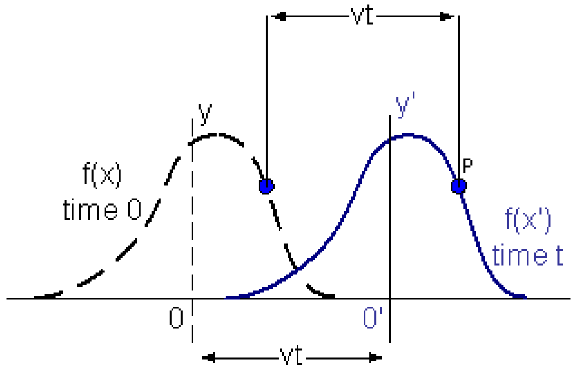


Figure 3.3

The coordinate *y* indicates the transverse displacement of a particular point on the string at position x and time *t*.

$$y(x,0) = f(x)$$

If *f* is a function that describes the shape of the wave at time 0, then at time *t* the waveform must still be described by a function *f*. i.e.

$$y(x,t) = f(x') = f(x - vt,)$$

This is because we assumed that the shape does not change as the wave travels. Moving in the positive x direction at speed *v* for some time *t*, the wave will travel a distance *vt*.

And for a pulse to the left the equation changes to

$$y(x,t) = f(x + vt)$$

If we follow the motion of point P, a particular part (or phase) of the wave, since the wave keeps its shape as it travels, the value y_p (transverse position of P) must not change. From the equation, this is only possible if the ***x*** coordinate of P increases as ***t*** increases in such as way as ***x-vt*** remains constant. This is true for any location on the waveform and for all time ***t***.

$$x - vt = constat$$

Now consider a sine wave traveling along an "ideal" string in the x direction. The coordinate y_{max} indicates the maximum displacement or amplitude of the sine curve and λ is the wavelength. The displacement *y* has the same value at any x as it does at x+λ, x+2λ, etc. Wavelength (in meters) equals velocity (in meters/second) times period T (in seconds).

The function *y* is called the wave function or waveform and it describes the actual geometric shape of the pulse at that time.

3.2 Sinusoidal Waves

This is a periodic wave which resembles a sine or a cosine wave.

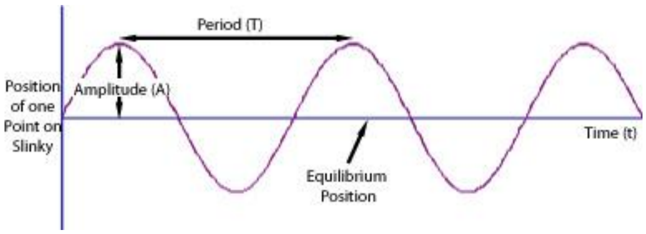


Figure 3.4

3.2.1 Characteristics a sinusoidal wave

(i) Crest and Trough: The *crest* of a wave is where the wave is at its maximum positive displacement from the equilibrium position, and the *trough* is where it is at its maximum negative displacement.

(ii) Amplitude: this is the maximum displacement of a particle from the mean position. The displacement at the crest is the wave's amplitude, while the displacement at the trough is the negative amplitude.

(iii) Wavelength: - the distance between two successive crests or two successive troughs or the distance between two successive particles that are in phase (at identical points on the wave form).

(iv) Period (T): This is the time (measured in seconds) taken to make one cycle by a wave.

(v) Frequency: is the number of wave cycles completed by one point along the wave in one second. The frequency of a wave is related to the period of a wave by the following equation:

$$f = \frac{1}{T}$$

The frequency of the wave is the same as the frequency of the simple harmonic oscillation of one element of the medium. The frequency is measured in cycles per second, or hertz (Hz).

(vi) Wave speed: Wave speed is a description of how fast a wave travels. The speed of a wave (v) is related to the frequency, period, and wavelength by the following simple equations:

$$v = \frac{\lambda}{T} \text{ or } v = \lambda f$$

Where v is the wave speed, λ is the wavelength, T is the period, and f is the frequency. Wave speed is commonly measured in units of meters per second (m/s).

A **sinusoidal wave Function** at an instant $t = 0$ can be expressed as

$$y(x,0) = y_{\max} \sin(ax)$$

Where $y_{\max}$ is the amplitude and a is a constant to be determined. At *x=0*, $y(0,0) = y_{\max} \sin(a.0) = 0$ the next value of *x* for which *y = 0* is $x = \lambda/2$.

Thus $y\left(\frac{\lambda}{2},0\right) = y_{\max} \sin a\left(\frac{\lambda}{2}\right) = 0$.

Therefore $a = 2\pi/\lambda$

The function describing the position of the element of the medium through which sinusoidal wave is travelling can be written as

$$y(x,0) = f(x) = y_{\max} \sin \frac{2\pi}{\lambda}(x)$$

If the wave moves to the right with a speed *v*, then the wave function at some later time *t* is

$$y(x,t) = f(x') = y_{\max} \sin \frac{2\pi}{\lambda}(x - vt)$$

But from definition of velocity $v = \frac{\lambda}{T}$ then the equation above becomes

$$y(x,t) = y_{\max} \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right)$$

$$y(x,t) = y_{\max} \sin\left(\frac{2\pi x}{\lambda} - \frac{2\pi t}{T}\right)$$

Hence $y(x,t) = y_{\max} \sin(kx - \omega t)$

Where $k = \frac{2\pi}{\lambda}$ is the *wave number* (radians per unit distance) and $\omega = \frac{2\pi}{T} = 2\pi f$ is the angular frequency (in radians per second).

The speed of a sinusoidal wave can be given as $v = \frac{\omega}{k}$

The equation $y(x,t) = y_{\max} \sin(kx - \omega t)$ assumes that the vertical position *y* of an element of the medium is zero at *x=0* and *t=0*. This need not be the case. Thus the general sinusoidal equation is expressed as

$$y = y_{\max} \sin(\kappa x - \omega t + \phi)$$

Where ϕ is the phase angle.

If the displacement y doesn't equal zero at position x = 0 at time t = 0 then we need to add a phase constant.

$$y(x,t) = y_{\max} \sin(\kappa x - \omega t - \phi)$$

Phase shift in space:

$$\begin{aligned} y(x,t) &= y_{\max} \sin(\kappa x - \phi - \omega t) \\ &= y_{\max} \sin\left[\kappa (x - \frac{\phi}{k}) - \omega t\right] \end{aligned}$$

Phase shift in time:

$$\begin{aligned} y(x,t) &= y_{\max} \sin(\kappa x - \omega t - \phi) \\ &= y_{\max} \sin\left[\kappa x - \omega (t + \frac{\phi}{\omega})\right] \end{aligned}$$

3.2.2 Sinusoidal wave on a string

The equation $y = A\sin(\kappa x - \omega t)$ can be used to describe the motion of any element of a string. An element of a string moves only vertically, and so its *x* coordinate remains constant. Therefore, the transverse speed v_y and the transverse acceleration a_y of elements of the string are

$$\begin{aligned} v_y(x,t) &= \frac{dy}{dt} = -\omega A\cos(kx - wt) \\ &= -\omega A\cos\frac{2\pi}{\lambda}(x - vt) \end{aligned}$$

$$\begin{aligned} a_y(x,t) &= \frac{d^2y}{dt^2} = -\omega^2 A\sin(kx - wt) \\ &= -\omega^2 y = -\omega A\sin\frac{2\pi}{\lambda}(x - vt) \end{aligned}$$

In these expressions we must use partial derivatives because y depends on both x and t.

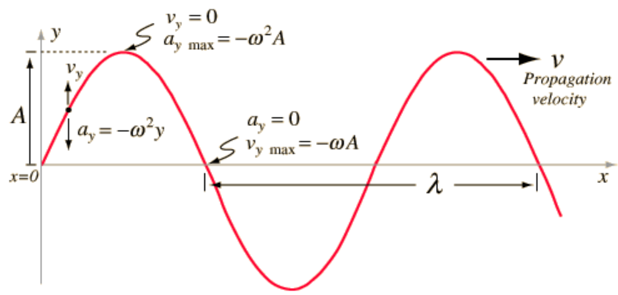


Figure 3.5

The maximum values of the transverse speed and transverse acceleration are absolute values of the coefficients of the cosine and sine functions. i.e.

$$v_{y\,\max} = -\omega A$$

and

$$a_{y\,\max} = -\omega^2 A$$

3.3 Wave interactions

Wave interference: is the phenomenon that occurs when two waves meet while traveling along the same medium. The interference of waves causes the medium to take on a shape that results from the net effect of the two individual waves upon the particles of the medium. Interference may be constructive or destructive

I. Constructive interference

Consider two pulses of the same amplitude traveling in different directions along the same medium. Suppose that each displaced upward 1 unit at its crest and has the shape of a sine wave. As the sine pulses move towards each other, they will eventually overlap completely. At that moment, the resulting shape of the medium would be an upward displaced sine pulse with an amplitude of 2 units as shown below.

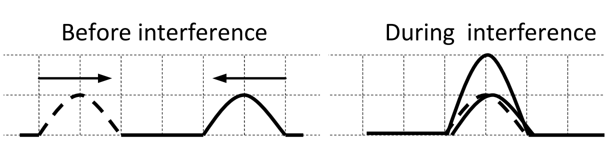


Figure 3.6

This type of interference is sometimes called *constructive interference* and occurs at any location along the medium where the two interfering waves have a displacement in the same direction. In this case, both waves have an upward displacement; consequently, the medium has an upward displacement that is greater than the displacement of the two interfering pulses. Constructive interference is observed at any location where the two interfering waves are displaced upward. It is also observed when both interfering waves are displaced downward as shown below.

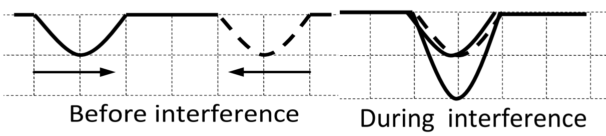


Figure 3.7

In this case, a sine pulse with a maximum displacement of -1 unit interferes with a sine pulse with a maximum displacement of -1 unit resulting to a sine pulse with a maximum displacement of -2 units.

II. Destructive interference

This is a type of interference that occurs at any location along the medium where the two interfering waves have a displacement in the opposite direction. For instance, when a sine pulse with a maximum displacement of +1 unit meets a sine pulse with a maximum displacement of -1 unit, destructive interference occurs. This is depicted in the diagram below.

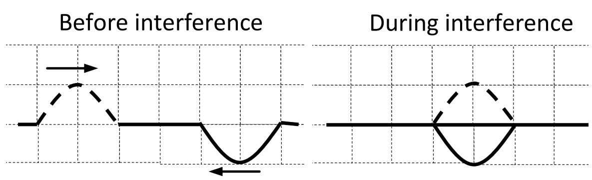


Figure 3.8

In the diagram above, the interfering pulses have the same maximum displacement but in opposite directions. The result is that the two pulses completely destroy each other when they are completely overlapped. At the instant of complete overlap, there is no resulting displacement of the particles of the medium. This "destruction" is not a permanent condition. When it is said that the two pulses *destroy each other*, it means that when overlapped, the effect of one of the pulses on the displacement of a given particle of the medium is *destroyed* or canceled by the effect of the other pulse. When two pulses with opposite displacements (i.e., one pulse displaced up and the other down) meet at a given location, the upward pull of one pulse is balanced (canceled or destroyed) by the downward pull of the other pulse. Once the two pulses pass through each other, there is still an upward displaced pulse and a downward displaced pulse heading in the same direction that they were heading before the interference. Destructive interference leads to only a momentary condition in which the medium's displacement is less than the displacement of the largest-amplitude wave.

The two interfering waves do not need to have equal amplitudes in opposite directions for destructive interference to occur. For example, a pulse with a maximum displacement of +1 unit could meet a pulse with a maximum displacement of -2 units. The resulting displacement of the medium during complete overlap is -1 unit.

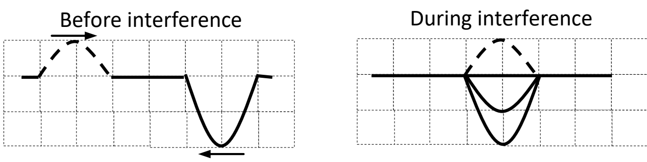


Figure 3.9

This is still destructive interference since the two interfering pulses have opposite displacements. However, the destructive nature of the interference does not lead to complete cancellation.

The meeting of two waves along a medium does not alter the individual waves or even deviate them from their path. Note that, two waves will meet, produce a net resulting shape of the medium, and then continue on doing what they were doing before the interference.

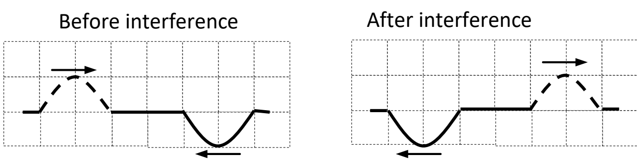


Figure 3.10

The task of determining the shape of the resultant demands that the principle of superposition to be applied.

- **Superposition of Waves**

Superposition is when two waves are superimposed on each other and add up. The phenomenon is described by the Principle of Superposition, which states:

When two waves are travelling in the same direction and speed, at any point on the combined wave the total displacement of any particle equals the vector sum of displacements of the waves.

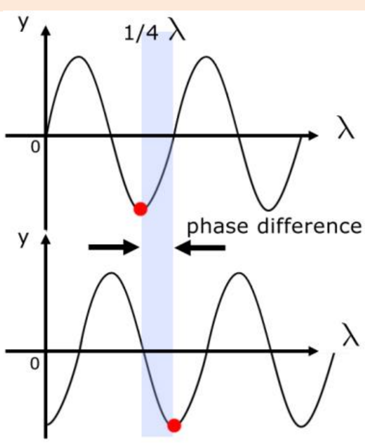


Figure 3.11

The task of determining the complete shape of the entire medium during interference demands that the principle of superposition be applied for every point (or nearly every point) along the medium.

Waves that obey this principle of superposition are called **linear waves** while those do not obey this principle are called **non linear waves**. One effect of the superposition is that two travelling waves can pass through each other without being destroyed or even altered.

- **Superposition of sinusoidal waves**

Let us apply this principle to two sinusoidal waves traveling in the same direction in a linear medium. If the two waves are traveling to the right and have the same frequency, wavelength, and amplitude but differ in phase, we can express their individual wave functions as

$$y_1 = A\sin(\kappa x - \omega t)$$

$$y_2 = A\sin(\kappa x - \omega t + \phi)$$

Where ϕ is the phase angle, $k = \frac{2\pi}{\lambda}$ and $\omega = 2\pi f$. Hence the resultant wave function y is

$$y = y_1 + y_2 = A[\sin(kx - \omega t) + \sin(kx - \omega t + \phi)]$$

To simplify this expression, we use the trigonometry identity

$$\sin a + \sin b = 2\cos\left(\frac{a-b}{2}\right)\sin\left(\frac{a+b}{2}\right)$$

If we let $a = kx - \omega t$ and $b = kx - \omega t + \phi$, we find the resultant wave function reduces to

$$y = 2A\cos\left(\frac{\phi}{2}\right)\sin\left(kx - \omega t + \frac{\phi}{2}\right)$$

The resultant wave function y also is sinusoidal and has the same frequency and wavelength as the individual waves, since the sine function incorporates the same values of k and ω that appear in the original wave functions. The amplitude of the resultant wave is $2A\cos\left(\frac{\phi}{2}\right)$ and its phase is $\frac{\phi}{2}$. If the phase constant ϕ equals 0, then $\cos\left(\frac{\phi}{2}\right) = 1$, and the amplitude of the resultant wave is 2A (twice the amplitude of either individual wave). In this case, the waves are said to be everywhere in phase and thus interfere constructively. This means, the crests and troughs of the individual wave's y_1 and y_2 occur at the same positions and combine to form wave y of amplitude 2A.

In general, constructive interference occurs when $\cos\left(\frac{\phi}{2}\right) = \pm 1$.

- **Conditions for interference**

The waves from light sources must be coherent with each other. This means that they must be of the same frequency, with a constant phase difference between them.

The amplitude (maximum displacement) of interfering waves must have the same magnitude. Slight variations produce lack of contrast in the interference pattern.

Problem

Suppose the two waves of the same frequency are travelling in the same direction, say to the right. Then they may be represented by equations;

$y_1 = A_2 \sin(\kappa x - \omega t)$ and $y_2 = A_2 \sin(\kappa x - \omega t + \phi)$.

Here ϕ is the phase difference between the two waves. By the principle of superposition, show that the resultant wave can be shown to have the form

$$y = y_1 + y_2 = A\sin(\kappa x - \omega t + \varphi) \ ,$$

where

$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos\phi}$ is the amplitude of the resulting wave, and

$\tan\varphi = \dfrac{A_2 \sin\phi}{A_1 + A_2 \cos\phi}$ is the phase difference between the first wave and the resulting wave

Problem

A transverse sinusoidal wave on a string has a period T = 25.0 ms and travels in the negative x direction with a speed of 30m/s. At t = 0, a particle on the string at x = 0 has a transverse position of 2cm and is traveling downward with a speed of 2m/s.

a) What is the amplitude of the wave?
b) What is the initial phase angle?
c) What is the maximum transverse speed of the string?
d) Write the wave function for the wave.

3.2.3 Standing (Stationary) Waves

Stationary or Standing waves have become very important in physics since has given insights into sound and many other important topics e.g. AC circuit theory, quantum mechanics, nanotechnology.

- **Formation of stationary waves**

The conditions for standing waves are:

- two waves travelling in *opposite directions* along the *same line of travel* and in the *same plane*

- the waves have the *same speed*
- the waves have the *same frequency*
- the waves have the same approximate amplitude

As a result of **superposition** (waves adding/subtracting), a resultant wave is produced. Now, depending on the phase difference between the waves, this resultant wave appears to move slowly to the right or to the left or disappear completely. It is only when the phase difference is exactly zero, that is when the two waves are exactly in phase, that 'standing/stationary waves' occur.

1. Two waves having the same amplitudes approach each other from opposite directions.
2. The two waves are 180^o out of phase with each other and therefore cancel out(black horizontal line).
3. The phase difference between the two waves narrows. The resultant grows but is not in phase with either of the two waves.
4. The phase difference between the two waves is narrower still. The resultant is larger but is still out of phase with the two waves.
5. The phase difference between the two waves is now zero. The resultant has its maximum value and is in phase with the two waves.

These 'in phase' waves produce an amplitude that is the sum of the individual amplitudes, and is called an **antinode**. Between two antinodes is a region where the superposition is zero. This is called a **node**.

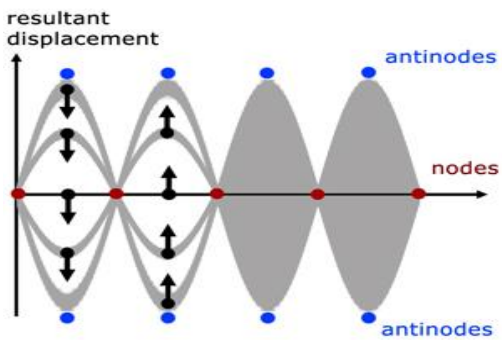


Figure 3.12

S*tanding wave pattern*s are only created within the medium at specific frequencies of vibration. These frequencies are known as *harmonic frequencies*, or merely harmonics. At any frequency other than a harmonic frequency, the interference of reflected and incident waves leads to a resulting disturbance of the medium that is irregular and non-repeating

Let consider wave functions for two transverse sinusoidal waves having the same amplitude, frequency, and wavelength but traveling in opposite directions in the same medium: where y_1 represents a wave traveling to the right and y_2 represents one traveling to the left.

$$y_1 = A\sin(\kappa x - \omega t)$$

$$y_2 = A\sin(\kappa x + \omega t)$$

Adding these two functions gives the resultant wave function y:

$$y = y_1 + y_2 = A[\sin(kx - \omega t) + \sin(kx + \omega t)] \text{....A}$$

Where $k = \frac{2\pi}{\lambda}$, $\omega = 2\pi f$. When we use the trigonometric identity

$$\sin(a \pm b) = \sin(a)\cos(b) \pm \cos(a)\sin(b)$$

Equation A reduces to

$$y = (2A\sin kx)\cos \omega t \text{ ....B}$$

This equation represents the wave function of a standing wave. Notice that Equation B does not contain a function of $(\kappa x - \omega t)$ thus, it is not an expression for a traveling wave.

Every particle of the medium oscillates in simple harmonic motion with the same frequency $\omega = 2\pi f$ (according to the $\cos \omega t$ factor in the equation). However, the vertical displacement (amplitude) of the simple harmonic motion of a given particle (given by the factor *2A sinkx*, the coefficient of the cosine function) depends on the lateral position x of the particle in the medium.

Considering the magnitude *2A sinkx*, at different longitudinal positions (***x***), the maximum displacement of a particle of the medium has a minimum value of zero when x satisfies the condition *sin(kx) =0* that is, when $kx = \pi, 2\pi, 3\pi....$

Because $k = 2\pi/\lambda$, these values for kx give

$$x = \frac{\lambda}{2}, \lambda, \frac{3\lambda}{2}, ... = \frac{n\lambda}{2}, n = 0,1,2,3....... \quad \text{C}$$

These points of zero displacement are called *nodes*

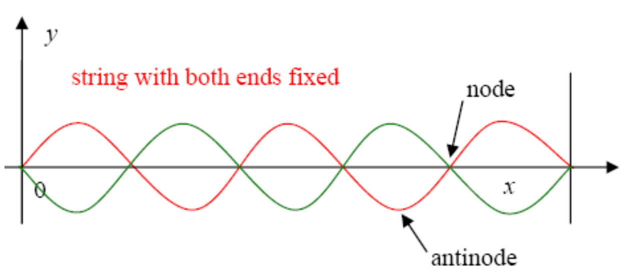


Figure 3.13

The particle with the greatest possible displacement from equilibrium has amplitude of 2*A*, and we define this as the amplitude of the standing wave. The positions in the medium at which this maximum displacement occurs are called *antinodes*. The antinodes are located at positions for which the coordinate x satisfies the condition $\sin kx = 0$, that is, when

$$kx = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{3\pi}{2} \ldots.$$

Thus, the position of the antinodes are given by

$$x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4} \ldots\ldots = \frac{n\lambda}{4}, n = 1,3,5 \ldots\ldots$$

D

- **Properties of stationary waves**

The above diagram shows how a standing wave moves up and down over time.

- separation of adjacent nodes is half a wavelength (λ/2)
- separation of adjacent antinodes is also λ/2
- hence separation of adjacent nodes and antinodes is λ/4
- the maximum amplitude is 2a (twice that of a single wave)
- a standing wave does not transfer energy(its two components however, do transfer energy in their respective directions)

Example

Two waves traveling in opposite directions produce a standing wave. The individual wave functions are

$$y_1 = (4.0cm)\sin(3.0x - 2.0t)$$

$$y_2 = (4.0cm)\sin(3.0x + 2.0t)$$

Where x and y are measured in centimeters.

Find the amplitude of the simple harmonic motion of the particle of the medium located at $x = 2.3cm$.

Find the positions of the nodes and antinodes if one end of the spring is at $x=0$.

Example

What is the amplitude of the simple harmonic motion of a particle located at an antinode?

Solution

Use the equation of a standing wave $y = (2A\sin kx)\cos \omega t$. A=4.0cm, k=3.0rad/cm and w=2.0rad/s. thus

$$y = (2A\sin kx)\cos \omega t = [(8.0cm\sin 3.0x)]\cos 2.0t$$

Amplitude is the coefficient of the cosine function

$$A = [(8.0cm \sin 3.0 \times 2.3)] = 8.0 \sin(6.9rad) = 4.6cm$$

With $k = 2\pi/\lambda$ =3.0 rad/cm, we get the wavelength as $\lambda = 2\pi/3.0$. Therefore nodes are located at

$$x = \frac{n\lambda}{2} = n\left(\frac{\pi}{3}\right)cm, n = 0,1,2,3.......$$

And the antinodes are located at $x = \frac{n\lambda}{4} = n\left(\frac{\pi}{6}\right)cm, n = 1,3,5.......$

The maximum displacement of a particle at an antinode is the amplitude of the standing wave, which is twice the amplitude of the individual traveling waves

$y_{max} = [(8.0cm \sin 3.0x)] = 8.0cm(\pm 1) = \pm 8.0cm$ where we have used the maximum value of $\sin kx = \pm 1$

Let check this result by evaluating the coefficient of our standing-wave function at the positions we found for the antinodes:

$$y_{max} = (8.0)\sin 3.0x\,|_{x=n(\pi/6)}$$

$$= (8.0)\sin\left[3.0n\left(\frac{\pi}{6}\right)rad\right] = (8.0)\sin\left[n\left(\frac{\pi}{2}\right)rad\right] = \pm 8.0cm$$

We have used the fact that n is an odd integer; thus, the sine function is equal to ± 1 depending on the value of n.

Example

Two waves in a long string are described by the equations

$y_1 = (0.0150m)\sin\left(\frac{x}{2} - 40t\right)$ and $y_2 = (0.0150m)\sin\left(\frac{x}{2} + 40t\right)$

where y_1 , y_2 , and x are in meters and t is in seconds.

- Determine the positions of the nodes of the resulting standing wave**.**

Solution

The equation of a standing wave is given as

$$y = (2A \sin kx)\cos \omega t = \left(0.030 \sin \frac{x}{2}\right)\cos 40t$$

Node positions are $x=\frac{\lambda}{2},\lambda,\frac{3\lambda}{2},......=\frac{n\lambda}{2}$

But $\lambda=\frac{2\pi}{k}=\frac{2\pi}{2}=\pi$

Therefore $x=\frac{\pi}{2},\pi,\frac{3\pi}{2},......=\frac{n\pi}{2},n=1,2,3$

- What is the maximum transverse position of element of the string at the position $x=0.04m$ in terms of time t?

$$y=(2A\sin kx)\cos\omega t=\left(0.030\sin\frac{0.04}{2}\right)\cos(40t)$$

$$=1.728Cos40t\times10^{-5}m$$

3.2.4 Interaction of surface waves

Surface waves occur at the boundary between media (i.e. water and air); the particles move in circles - simultaneously perpendicular and parallel to the wave direction (e.g. ocean waves).

Wave fronts move in one of three ways:

- straight lines like waves on the ocean
- spreading concentric circles like ripples in a pond
- spreading concentric spheres like light or sound waves

There are also three ways in which waves interact with objects or other waves:

(i) Reflection

- Bouncing back of a wave when it meets a surface or a boundary.
- At a free boundary (i.e. the end of a rope loosely tied to a post), the reflected wave is identical to the original because the end of the rope is free to move up and down the post, just as if someone were shaking it.
- At a fixed boundary (i.e. the end of a rope tied in place to a post), the reflected wave is identical to the original, but upside down or inverted.

(ii) Diffraction

- Bending of waves around an edge

- Wave fronts travel in a straight line until they hit an object.
- The end of the wave nearest the object bends and begins to spread out from that point as if it were a new wave

(iii)Refraction

- Bending of waves when they pass from one medium to another.
- Because waves travel at different speeds through different media, waves bend when they change from one medium to another.
- Short-wavelength/high-frequency waves bend more than long-wavelength/low-frequency waves, which is why violet (wavelength = 400nm) is always on the inside of a rainbow and red (wavelength = 650nm) is always on the outside.

When sound enters a new medium, it is reflected, transmitted, or absorbed as seen above. Sound can be absorbed by certain materials: rubber, cork, and acoustic tiles, for example. Sound absorbing materials have high absorption coefficients. The absorption coefficient is the fraction of the original sound intensity absorbed by the material at some given frequency.

3.4 Sound Waves

3.4.1 Classification of sound

Sound waves are longitudinal waves that travel through any material medium with a speed that depends on the properties of the medium. As the waves travel, the particles in the medium vibrate to produce changes in density and pressure along the direction of motion of the wave. These changes result in a series of high-pressure and low-pressure regions. If the source of the sound waves vibrates sinusoidally, the pressure variations are also sinusoidal.

Sound waves are divided into three categories that cover different frequency ranges.

- *Audible waves*- are waves that lie within the range of sensitivity of the human ear. They can be generated in a variety of ways, such as by musical instruments, human vocal cords, and loudspeakers.
- *Infrasonic waves* -are waves having frequencies below the audible range. Elephants can use infrasonic waves to communicate with each other, even when separated by many kilometers.
- *Ultrasonic waves*- are waves having frequencies above the audible range. Ultrasonic waves are also used in medical imaging.

3.4.2 The Doppler Effect

[https://phys.libretexts.org/Bookshelves/University_Physics/Book%3A_University_Physics_(OpenStax)/Book%3A_University_Physics_I_-_Mechanics_Sound_Oscillations_and_Waves_(OpenStax)/17%3A_Sound/17.08%3A_The_Doppler_Effect]

3.5 Light as a Wave

Light is an electromagnetic radiation with wavelengths capable of causing the sensation of vision. These wavelengths range from approximately 4000 Å (Angstroms) to 7700 Å. i.e. extreme blue to extreme red. NB: 1Å = 1 x 10^{-10} m.

Maxwell showed that light is an electromagnetic wave – meaning that it is made up of perpendicular electric, **E**, and magnetic, **B**, fields oscillating together. A single ray can be represented by the figure below.

The magnetic field however interacts much more weakly with matter than the electric field and can often be ignored. Light can then be modeled as if it were a *single oscillating transverse wave*.

- **What is it that travels?**
  - Light is a form of energy that is always moving.
  - Light also carries momentum – a powerful laser can support a small ball.
  - Light can convey information from one place to another.
  - Light travel in vacuum (with speed $c = 3.0 \times 10^8$ m s^{-1}) since it is an electromagnetic wave

- **Electromagnetic properties of light**

Light can be thought as an electromagnetic waves meaning it is composed of electric and magnetic fields and has properties common to all waves.

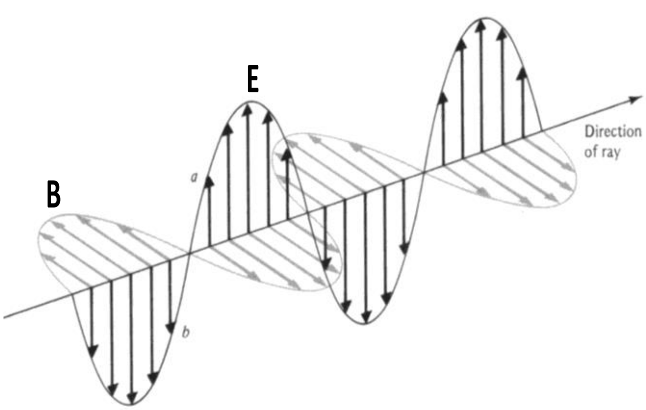


Figure 3.14

Finite speed of light, c

All electromagnetic radiation travels with same velocity in free space (vacuum) which is given the symbol, **c** where $c = f\lambda$**,** from which $f = \dfrac{c}{\lambda}$

NB: The velocity of light depends on the media in which the light is travelling and generally, **v < c** for other media. f is not affected by the medium of transmission.

There are two types of light

- Polychromatic light
- Monochromatic light

The normal everyday light we see is known as a white light. Using a prism it can be shown to be composed of seven colors i.e ROYGBIV.

As each of these colors is composed of different frequencies and hence wavelengths, white light is said to be **polychromatic**. Light composed of a single wavelength i.e. appearing to be of one color is said to be a **monochromatic** light.

❑ The rectilinear propagation of light

All light is an energy and hence is produced by a source. We can reduce all sources of light to idealized sources which we call **point sources**. This is a source which emits light from a point. It radiates light in all directions spherically. This can be represented by a ray of light i.e. lines of light pointing in the direction of propagation of the light. The lines are straight and thus we say light undergoes **rectilinear propagation**.

NB: The exception to this condition occurs when light passes through an opening (aperture) whose width, a is the same or less than the wavelength λ of inclined light. This is referred to as *diffraction*, which is spreading out of light.

3.6 Huygens Wave Theory

The Huygens principle (1690) is a simple scattering theory that can be used to understand many optical phenomena. It claims that, light is propagated as spherical waves and from a source of light, light can be represented by a spherical wave front.

It states that:

Any point on a wavefront behaves as a new (secondary) source, from which secondary waves *(wavelets)* spread out in a *forward* direction only with the same frequency and speed as the primary wave. The neglect of backwaves is reasonable since points from which wavelets arise are ***not*** independent sources, but are set in motion as a result of a wavefront from the original (primary) source.

The surface which touches, or envelops, all of these wavelets at a particular instant forms a new (secondary) wavefront.

Huygen's theory assumes that the effect of the wavelets is limited to that part which touches the new wavefront. This assumption was later shown to be a reasonable consequence of *interference* between wavelets (Fresnel 1800's) – they cancel each other out. This theory can only explain reflection and refraction but not other effects like diffraction.

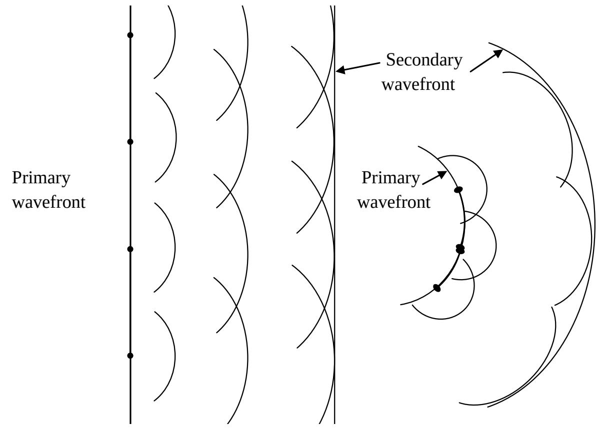


Figure 3.15

3.7 Electromagnetic Waves

3.7.1 Properties of Electromagnetic Waves

1. *Electromagnetic waves travel at the speed of light.*
2. *Electromagnetic waves are transverse waves, because the electric and magnetic fields are perpendicular to the direction of propagation of the wave and to each other.*
3. *The ratio of the electric field to the magnetic field in an electromagnetic wave equals the speed of light.*
4. *Electromagnetic waves carry both energy and momentum, which can be delivered to a surface.*

3.7.2 The Electromagnetic Spectrum

The various types of electromagnetic waves are as listed in the following figure which shows the electromagnetic spectrum. Note the wide ranges of frequencies and wavelengths. No sharp dividing point exists between one type of wave and the next. Remember that all forms of the various types of radiation are produced by the same phenomenon—accelerating charges. The names given to the types of waves are simply for convenience in describing the region of the spectrum in which they lie. Radio waves are the result of charges accelerating through conducting wires.

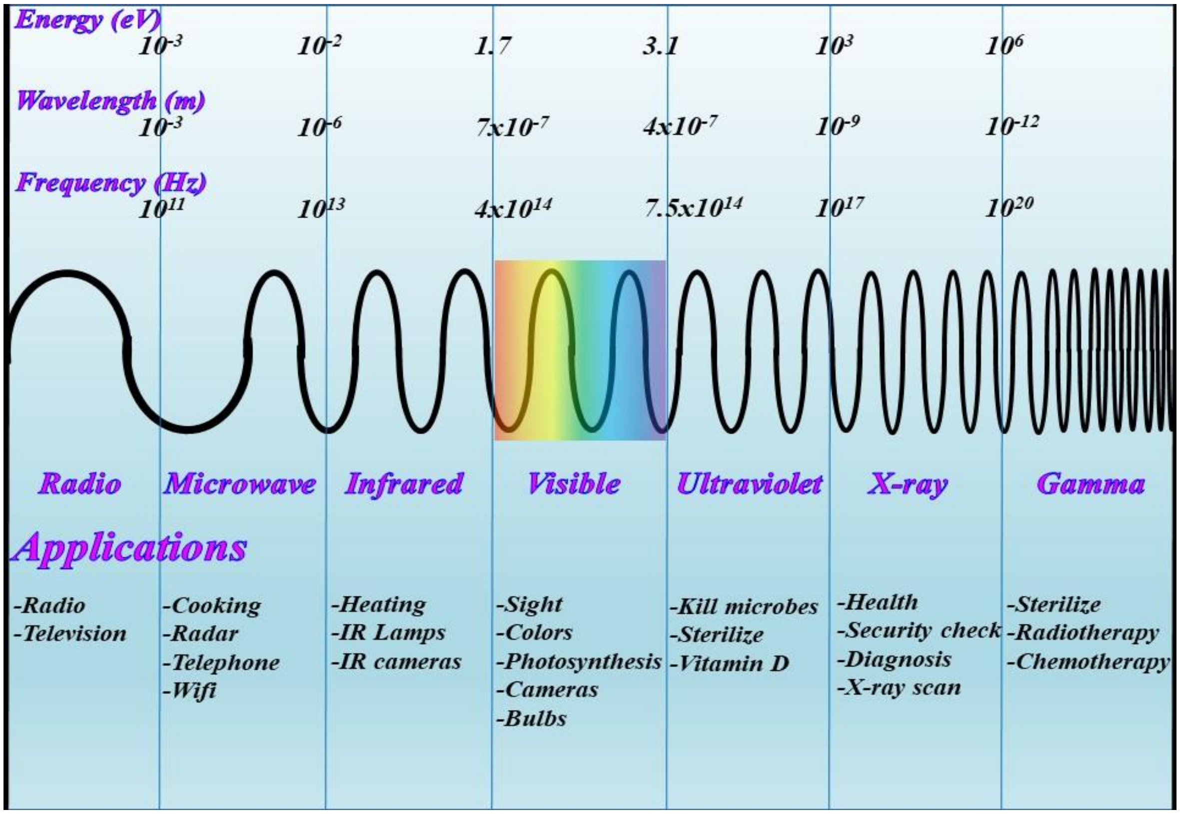


Figure 3.16

Radio waves: Results from charges accelerating through conducting wires. Ranging from more than 10^4 m to about 0.1 m in wavelength, they are generated by such electronic devices as LC oscillators and are used in radio and television communication systems.

Microwaves (short-wavelength radio waves): Have wavelengths ranging between about 1 mm and 30 cm and are generated by electronic devices. Their short wavelengths make them well suited for the radar systems used in aircraft navigation and for the study of atomic and molecular properties of matter. Microwave ovens are an interesting domestic application of these waves.

Infrared waves: Produced by hot objects and molecules, have wavelengths ranging from about 1 mm to the longest wavelength of visible light, $7x10^{-7}$ m. They are readily absorbed by most materials. The infrared energy absorbed by a substance causes it to get warmer because the energy agitates the atoms of the object, increasing their vibrational or translational motion. The result is a rise in temperature. Infrared radiation has many practical and scientific applications, including physical therapy, infrared photography, and the study of the vibrations of atoms.

Visible light: Is the most familiar form of electromagnetic waves, may be defined as the part of the spectrum that is detected by the human eye. Light is produced by the rearrangement of electrons in atoms and molecules. The wavelengths of visible light are classified as colors ranging from violet $\lambda \sim 4 x 10^{-7}$ m) to red ($\lambda \sim 7 x 10^{-7}$ m). The eye's sensitivity is a function of wavelength and is greatest at a wavelength of about $5.6 x 10^{-7}$ m (yellow green).

Ultraviolet (UV) light: Covers wavelengths ranging from about $4 x 10^{-7}$ m (400 nm) down to $6 x 10^{-10}$ m (0.6 nm). The Sun is an important source of ultraviolet light (which is the main cause of suntans).

X-rays: Are electromagnetic waves with wavelengths from about 10^{-8} m (10 nm) down to 10^{-13} m (10^{-4} nm). The most common source of x-rays is the acceleration of high-energy electrons bombarding a metal target. X-rays are used as a diagnostic tool in medicine and as a treatment for certain forms of cancer.

Gamma rays: Are electromagnetic waves emitted by radioactive nuclei—have wavelengths ranging from about 10^{-10} m to less than 10^{-14} m. They are highly penetrating and cause serious damage when absorbed by living tissues. Accordingly, those working near such radiation must be protected by garments containing heavily absorbing materials, such as layers of lead. When astronomers observe the same celestial object using detectors sensitive to different regions of the electromagnetic spectrum, striking variations in the object's features can be seen.

Dangers associated with electromagnetic waves

Some specific properties of electromagnetic waves that can make them harmful to human health include:

Ionizing radiation: Some EM waves may possess enough energy to capable of ionizing atoms and molecules causing serious damage to biological tissues. Such includes X-rays, gamma rays, and ultraviolet (UV) radiation.

High frequency: High frequencies EM waves can penetrate deep into tissues causing damage. Microwaves for example, which are used in many communication devices can burn or cause other tissue damage under prolonged exposure.

Prolonged exposure: Non-ionizing electromagnetic waves, e.g. from cell phones and Wi-Fi routers, can be harmful under prolonged exposure. Studies have shown that prolonged exposure to these waves can lead to DNA damage, cell death, and an increased risk of cancer.

Electromagnetic interference: EM waves can interfere with medical devices, such as pacemakers and insulin pumps, leading to malfunction and potential harm to the patient.

PART 4: ELECTRONICS

4.1 Review of semiconductor properties

4.2 The P-N junction and its V-I characteristics; diode resistance.

4.3 Conventional Photo and light emitting diodes.

4.4 Transistors: *The bipolar junction transistor*; common base, common emitter, common collector configurations and their characteristics, transistor as a switch. *Field effect transistor* and their characteristics.

4.5 Applications in basic logic circuits and in computer storage devices.

4.1 Review of semiconductor properties

A semiconductor material is one whose electrical properties lie in between those of insulators and good conductors. Examples are: germanium and silicon. In terms of energy bands, semiconductors can be defined as those materials which have almost an empty conduction band and almost filled valence band with a very narrow energy gap (of the order of 1 eV) separating the two as shown in figure 4.1 below.

At 0ºK, there are no electrons in the conduction band and the valence band is completely filled. However, with increase in temperature, width of the forbidden energy bands is decreased so that some of the electrons are liberated into the conduction band. In other words, conductivity of semiconductors increases with temperature. Moreover, such departing electrons leave behind positive holes in the valence band. Hence, semiconductor current is the sum of electron and hole currents flowing in opposite directions.

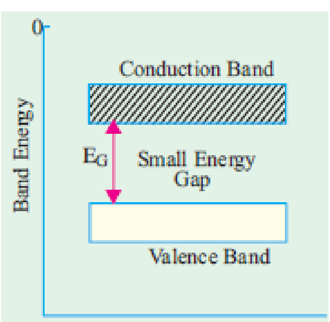


Figure 4.1(a)

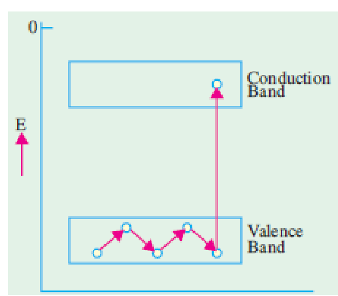


Figure 4.1(b)

❑ Atomic Binding in Semiconductors

Semiconductors like germanium and silicon, have crystalline structure. Their atoms are arranged in an ordered array known as crystal lattice. Both these materials are tetravalent i.e. each has four valence electrons in its outermost shell.

The neighbouring atoms form covalent bonds by sharing four electrons with each other so as to achieve inert gas structure (i.e. 8 electrons in the outermost orbit). A two dimensional view of the germanium crystal lattice is shown below in which circles represent atom cores consisting of the nuclei and inner 28 electrons. Each pair of lines represents a covalent bond. The dots represent the valence electrons. It is seen that each atom has 8 electrons under its influence.

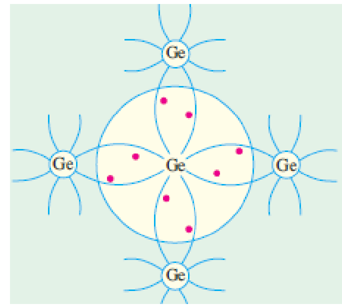


Figure 4.2

❑ Majority and Minority Carriers

In a piece of pure germanium or silicon, no free charge carriers are available at 0ºK. However, as its temperature is raised to room temperature, some of the covalent bonds are broken by heat energy and as a result; electron-hole pairs are produced. These are called thermally-generated charge carriers. They are also known as intrinsically available charge carriers. Ordinarily, their number is quite small. An intrinsic of pure germanium can be converted into a P-type semiconductor by the addition of an acceptor impurity which adds a large number of holes to it. Hence, a P-type material contains following charge carriers:

(a) large number of positive holes—most of them being the added impurity holes with only a very small number of thermally generated ones;

(b) a very small number of thermally-generated electrons (the companions of the thermally generated holes mentioned above).

Obviously, in a P-type material, the number of holes (both added and thermally-generated) is much more than that of electrons. Hence, in such a material, holes constitute majority carriers and electrons form minority carriers as shown in Fig. 4.3 (a)

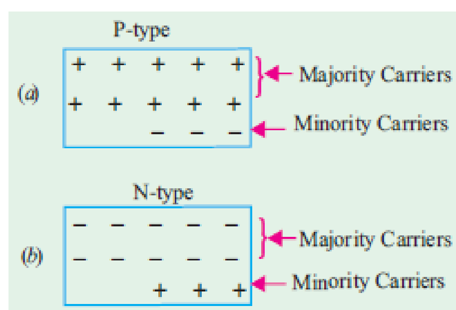


Figure 4.3

Similarly, in an N-type material, the number of electrons (both added and thermally-generated) is much larger than the number of thermally-generated holes. Hence, in such a material, electrons are majority carriers whereas holes are minority carriers as shown in Fig. 4.3 (b).

4.2 The P-N Junction

It is possible to manufacture a single piece of a semiconductor material half of which is doped by P-type impurity and the other half by *N*-type impurity as shown in Fig. 4.4.

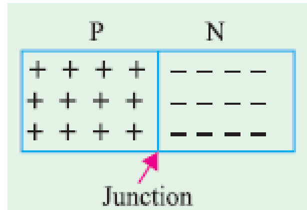


Figure 4.4

The plane dividing the two zones is called ***junction.*** Theoretically, junction plane is assumed to lie where the density of donors and acceptors is equal. The *P-N* junction is fundamental to the operation of diodes, transistors and other solid-state devices.

Let us see if anything unusual happens at the junction. It is found that following three phenomena take place:

1. A thin ***depletion layer*** or region (also called space-charge region or transition region) is established on both sides of the junction and is so called because it is depleted of ***free charge carriers***. Its thickness is about 10^{-6} m.

2. A barrier potential or junction potential is developed across the junction.

3. The presence of depletion layer gives rise to junction and diffusion capacitances

Formation of Depletion Layer

Suppose that a junction has just been formed (Fig. 4.5);

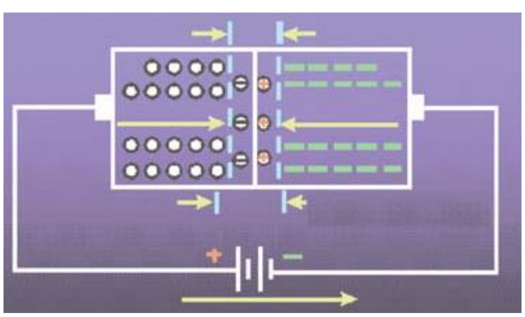


Figure 4.5

At that instant, holes are still in the *P*-region while the electrons are in the *N*-region. However, there is greater concentration of holes in *P*-region than in *N*-region (where they exist as minority carriers). Similarly, concentration of electrons is greater in *N*-region than in *P*-region (where they exist as minority carriers). This concentration differences establishes density gradient across the junction resulting in carrier diffusion. Holes diffuse from *P* to *N*-region and electrons from *N*-to *P*-region and terminate their existence by recombination [Fig. 4.6 (*a*)].

This recombination of free and mobile electrons and holes produces the narrow region at the junction called depletion layer. It is so named because this region is devoid of (or depleted of) ***free and mobile***

charge carriers like electrons and holes—there being present only positive ions which are not free to move.

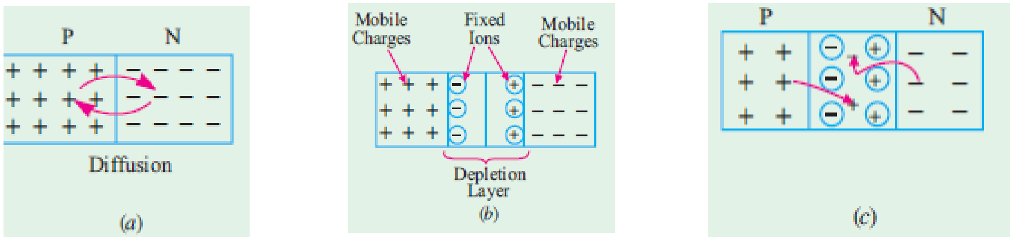


Figure 4.6

It might seem from above that eventually all the holes from the *P*-side would diffuse to the *N*-side and all the electrons from the *N*-side would diffuse to the *P*-side but this does not occur due to the formation of ions on the two sides of the junction. The impurity atoms which provide these migratory electrons and holes are left behind in an ionized state bearing a charge which is opposite to that of the departed carrier. Also, these impurity ions, just like germanium atoms, are fixed in their positions in the crystal lattice in the *P*- and *N*- regions of the diode. Hence, as shown in Fig. 4.6 (*b*), they form parallel rows or 'plates' of opposite charges facing each other across the depletion layer. Obviously, row of ***fixed*** positive ions in the *N*-region is produced by the migration of electrons from the *N*- to *P*- region. Similarly, the row of ***fixed*** negative ions in the *P*-region is produced by the migration of holes from the *P*- to *N*-region. If a majority carrier (either an electron or a hole) tries to cross into depletion layer, it can meet either of the following two facts:

(i) either it can be trapped or captured by the row of fixed impurity ions of opposite sign which guard its own region. For example, a hole trying to approach the depletion layer may be neutralized by the row of fixed negative ions situated in the P-region itself at the edge of the depletion layer. So will be the case with the electron trying to approach the depletion layer from N-region [Fig.4.7 (c)]

(ii) it may succeed in entering the depletion layer where it will be repelled by the row of similarly charged impurity ions guarding the other region. But its life will be cut short by recombination with a majority carrier of opposite sign which has similarly entered the depletion layer from the other half of the diode.

Ultimately, an equilibrium condition is reached when depletion layer has widened to such an extent that no electrons or holes can cross the P-N junction.

Junction or Barrier Voltage

Even though depletion layer is cleared of charge carriers, it has oppositely-charged fixed rows of ions on its two sides. Because of this charge separation, an electric potential difference V_B is established across the junction even when the ***junction is externally isolated*** (Fig. 4.7).

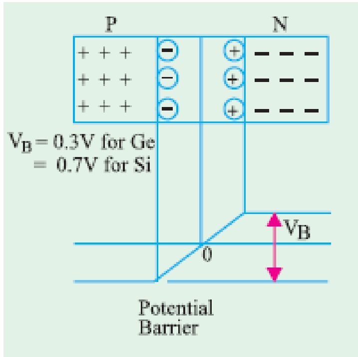


Figure 4.7

It is known as ***junction or barrier potential***. It stops further flow of carriers across the junction unless supplied by energy from an external source. At room temperature of 300ºK, V_B is about 0.3 V for Ge and 0.7 V for Si. The value of barrier voltage is given by;

$$V_B = V_T log_e \ N_a N_d / n_i^2$$

where $N_a N_d / n_i^2$ and V_T have the meanings explained.

The value of V_T at room temperature of 300ºK is given by;

$$V_T = V_{300} = \frac{kT}{e} = \frac{1.38 \times 10^{-23} \times 300}{1.6 \times 10^{-19}} = 26 \text{ mV}$$

$$V_B = 26 \log_e (N_a N_d / n_i^2) \text{ mV}$$

Barrier voltage depends on doping density, electronic charge and temperature. For a given junction, the first two factors are constant, thus making V_B dependent on temperature. With increase in temperature, more minority charge carriers are produced, leading to their increased drift across the junction. As a result, equilibrium occurs at a lower barrier potential. It is found that both for Ge and Si, V_B decreases by about 2 mV / ºC.

$\therefore \Delta V_B = -0.002 \ \Delta t$ where Δt is the rise in temperature in ºC.

The strong field set up by V_B causes drift of carriers through depletion layer. As seen from Fig. 4.8 (*b*), under the influence of this field, ***holes drift from N-to P-region and electrons from P- to -N region***. This drift current must be equal and opposite to the diffusion current [Fig. 4.8 (*a*)] because under condition of equilibrium and with no external supply, net current through the crystal is zero.

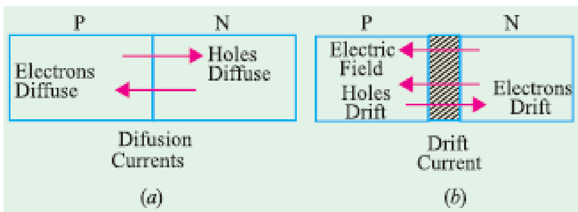


Figure 4.8

To summarize the main points here, we have:

1. As soon as P-N junction is formed, free electrons and holes start diffusing across the junction and recombining.
2. Their recombination leads to the appearance of a depletion layer across the junction which contains no mobile carriers but only immobile ions.
3. These immobile ions set up a barrier potential and hence an electric field which sets up drift current that is equal and opposite to the diffusion current when final equilibrium is reached.

P-N Junction Biasing

Forward Biased P-N Junction

Suppose, positive battery terminal is connected to *P*-region of a semiconductor and the negative battery terminal to the *N*-region as shown in Fig. 4.9 (*a*).

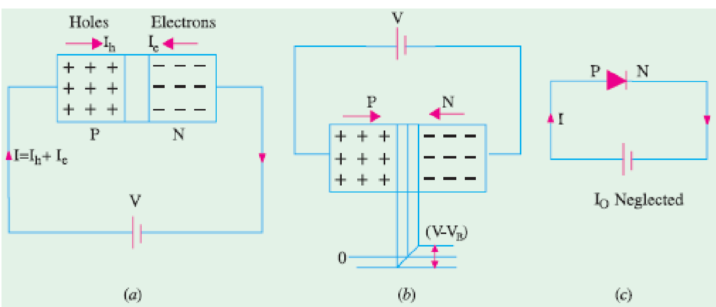


Figure 4.9

In that case the junction is said to be biased in the ***forward direction*** because it permits easy flow of current across the junction.

As soon as battery connection is made, holes are ***repelled*** by the positive battery terminal and electrons by the negative battery terminal with the result that both the electrons and the holes are driven ***towards*** the junction where they recombine. This ***en masse*** movement of electrons to the left and that of holes to the right of the junction constitutes a large current flow through the semiconductor. Obviously, the junction offers ***low resistance*** in the forward direction.

Forward I-V Characteristics

A typical *V/I* characteristic for a forward biased *P-N* junction is shown in Fig. 4.10. It is seen that forward current rises exponentially with the applied forward voltage. However, at ordinary room temperature, a p.d. of about 0.3 V is required before a reasonable amount of forward current starts flowing in a germanium junction. This voltage is known as ***threshold voltage* (V^{th}) or *cut-in voltage*** or

knee voltage V_K. It is practically the same as barrier voltage V_B. Its value for silicon junction is about 0.7 volt. For $V < V^{th}$, current flow is negligible.

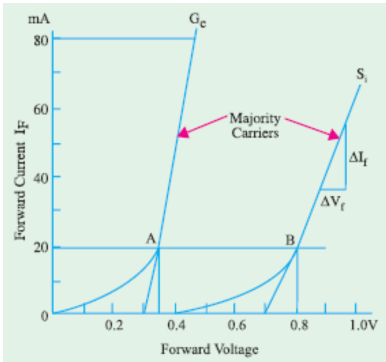


Figure 4.10

For point B in Fig. 4.10, the forward resistance for Si is $R_F = 0.8V/20\ mA = 40\ \Omega$

Similarly, for point A on the Ge curve,

$R_F = 0.36\ V/20\ mA = 18\ \Omega$

Reverse Biased P-N Junction

When battery connections to the semiconductor are made as shown in Fig. 4.11 (*a*), the junction is said to ***reverse-biased.*** In this case, holes are attracted by the negative battery terminal and electrons by the positive terminal so that both holes and electrons move ***away*** from the junction and ***away*** from each other. Since there is no electron-hole combination, no current flows and the junction offer high resistance. Another way of looking at the process is that in this case, the applied voltage increases the barrier potential to ($V + V_B$), thereby blocking the flow of majority carriers.

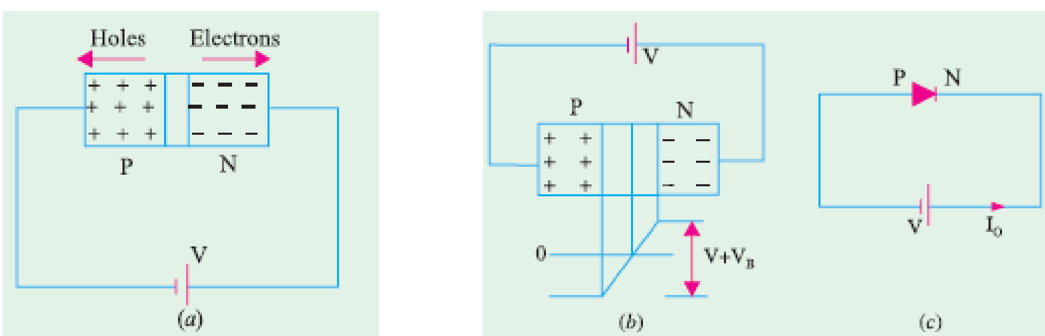


Figure 4.11

Reverse V/I Characteristic

As said earlier, the reverse saturation current is also referred to as ***leakage current*** of the *P-N* junction. Fig. 4.12 shows *V/I* characteristics of a reverse-biased *P-N* junction. It is seen that as reverse voltage is increased from zero, the reverse current quickly rises to its maximum or saturation value. Keeping temperature constant as the reverse voltage is increased, *Io* is found to increase only slightly. This slight increase is due to the impurities on the surface of the semiconductor which behaves as a resistor and hence obeys Ohm's law. This gives rise to a very small current called ***surface leakage current***.

Unlike the main leakage (or saturation) current, this surface leakage current is independent of temperature but depends on the magnitude of the reverse voltage. A reverse-biased junction can be represented by a very large resistance. As seen from Fig. 4.13, in the case of Si, for a reverse voltage of about 15 V, I_0= 10 μA. Hence, reverse resistance is RR = 15 V/10 μA = 1.5 MΩ

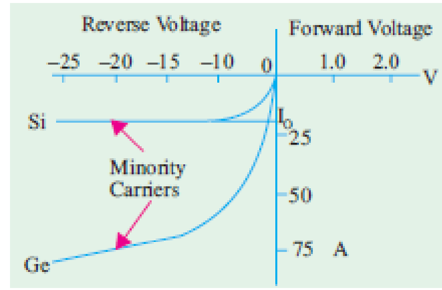


Figure 4.12

4.3 Transistors

A transistor is a semiconductor device that controls current between two terminals based on the current or voltage at a third terminal. It is used for *amplification* or *switching* of electrical signals. Transistors gained widespread applications in the early 1960s, replacing vacuum tubes in electronic circuits. The first transistor was invented at Bell Laboratories, USA in 1947. All of the complex electronics devices and systems developed since then are based on the developments in semiconductor transistors. It has been greatly miniaturized over the years. Transistors are now responsible for making components much smaller leading to the information age with the integrated circuit replacing discrete components like the transistor.

Because integrated circuits are designed around discrete components like transistors, it is essential to have an understanding of transistor operations since they are foundational to understanding integrated circuits. Today, integrated circuits have millions of transistors etched on with a piece of silicon chip. To understand how this works, you need to have at least a grasp of how the transistor works.

Types of transistors.

There are two basic types of transistors:

- the bipolar junction transistor (BJT) and
- the field-effect transistor (FET).

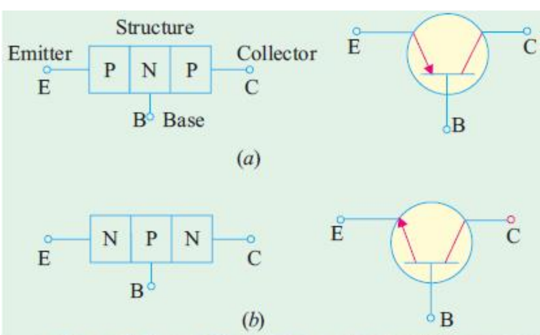
The bipolar junction transistor: This is used in two broad areas of electronics: as a linear amplifier to boost an electrical signal or as an electronic switch. Basically, the BJT consists of two back-to-back P-N junctions manufactured in a single piece of a semiconductor crystal. These two junctions give rise to three regions called *emitter*, *base* and *collector*.

The field-effect transistor: The acronym 'FET' stands for field effect transistor. It is a three-terminal unipolar solid-state device in which current is controlled by an electric field as is done in vacuum tubes. Broadly speaking, there are two types of FETs:

4.3.1 Bipolar Junction Transistor

The *diodes* as discussed earlier are bipolar devices as their operation depends upon holes and electrons as current carriers. Two additional bipolar devices are the NPN and the PNP bipolar transistors. The basic structure of the bipolar junction transistor (BJT) determines its operating characteristics. DC bias is important to the operation of transistors in terms of setting up proper currents and voltages in a transistor circuit. Two important parameters are α_{DC} and β_{DC}.

As shown in Fig. 4.13 below, junction transistor is simply a sandwich of one type of semiconductor material between two layers of the other type. Fig. 4.13 (a) shows a layer of N-type material sandwiched between two layers of P-type material. It is described as a PNP transistor. Fig. 4.13(b) shown an NPN-transistor consisting of a layer of P-type material sandwiched between two layers of N-type material.



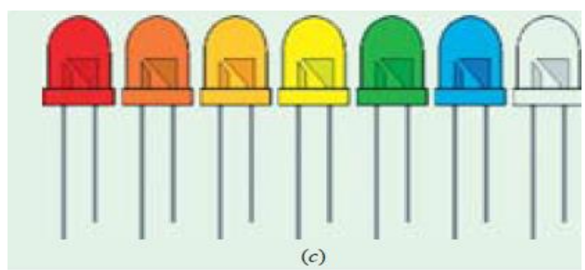


Figure 4.13

The emitter, base and collector are provided with terminals which are labeled as *E*, *B* and *C*. The two junctions are: emitter-base (*E*/*B*) junction and collector-base (*C*/*B*) junction. The symbols employed for *PNP* and *NPN* transistors are also shown in Fig. 4.13. The arrowhead is always at the emitter (not at the collector) and in each case, its direction indicates the ***conventional*** direction of current flow. For a *PNP* transistor, arrowhead points from emitter to base meaning that emitter is positive with respect to base (and also with respect to collector). In a transistor, for normal operation, collector and base have the same polarity with respect to the emitter. For *NPN* transistor, it points from base to emitter meaning that base (and collector as well)* is positive with respect to the emitter.

(i). Emitter: It is more heavily doped than any of the other regions because its main function is to supply majority charge carries (either electrons or holes) to the base.

(ii). Base: It forms the middle section of the transistor. It is very thin (10^{-6} m) as compared to either the emitter or collector and is very ***lightly-doped.***

(iii). Collector: Its main function (as indicated by its name) is to collect majority charge carriers coming from the emitter and passing through the base. In most transistors, collector region is made physically larger than the emitter region because it has to dissipate much greater power. Because of this difference, there is no possibility of inverting the transistor *i.e.* making its collector the emitter and its emitter the collector. Fig. 4.13 (*c*), shows the picture of C1815 (front and the back view) transistor.

Transistor Biasing (Common emitter, Common base and Common collector).

For proper working of a transistor, it is essential to apply voltages of correct polarity across its two junctions.

- The base-emitter junction behaves like any other PN junction when viewed alone. If the base-emitter junction is forward biased, the transistor is on. If it is reverse biased, the transistor is off, just like a diode.
- The base-emitter junction in a transistor essentially turns the transistor on or off. Now, the base-collector junction will not have that same power, but the base-emitter junction will determine whether the transistor is turned on or off. The base-collector junction also behaves as a PN junction, but it will not have the ability to make it turn on or off.

Biasing proper

Considering the different types of biasing, the effects of simultaneously biasing both of the junctions in a transistor are important to understand. There are four possible combinations to bias the two junctions, but only three play key roles:

Case 1: Reverse biasing base-emitter (BE) and reverse biasing collector-base (CB) junctions.
This voltage application at both junctions will cause them to turn off, a condition known as cutoff. This is essential for digital where a transistor only operate as a switch (on or off). The off condition is the cut

off while the on condition is called saturation. This case is not used in linear operations like amplifiers since the entire spectrum between on and off is used.

Case 2: Forward biasing base-emitter (BE) junction and reverse biasing collector base (CB) junction. This allows for maximum current flow between the emitter and the collector. This is almost the same current that is going through the emitter that will be in the collector, as well to add up. In this case, the base current is very small, so the current that will have a large flow of current through the current but only a very small amount will actually go out of the base connection here. Current going from the emitter to base is very heavy. The emitter is heavily doped with free electrons so we have lots of free electrons here. The base is very lightly doped with just a few holes. If electrons recombined with holes, they can exit through the base.

It is worthwhile to remember that for normal operation;

1. base-emitter junction is always forward biased and

2. collector-base junction is always reverse-biased.

This type of biasing is known as *FR* biasing. In Fig. 4.14b, two batteries respectively provide the dc emitter supply voltage V_{EE} and collector supply voltage V_{CC} for properly biasing the two junctions of the transistor. In Fig. 4.14(*a*), **P**ositive terminal of V_{EE} is connected to **P**-type emitter in order to repel or **P**ush holes into the base. The negative terminal of V_{CC} is connected to the collector so that it may ***attract*** or ***pull*** holes through the base.

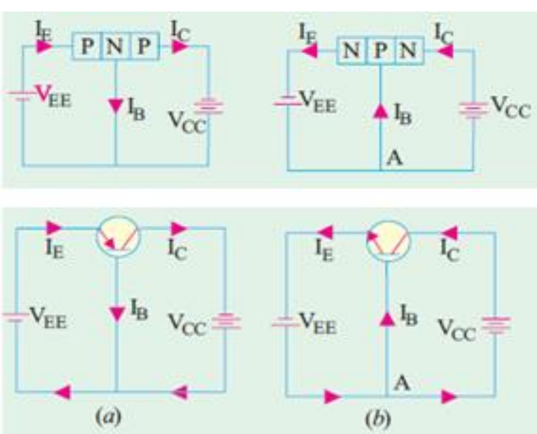


Figure 4.14

Similar considerations apply to the *NPN* transistor of Fig. 4.14 (*b*). It must be remembered that a transistor will never conduct any current if its emitter-base junction is not forward-biased. This is because there would be no current due to majority charge carriers (However, there would be an extremely small current due to minority charge carriers which is called leakage current of the transistor).

Important Biasing Rule

For a *PNP* transistor, both collector and base are negative with respect to the emitter (the letter *N* of Negative being the same as the middle letter of *PNP*). Of course, collector is ***more negative*** than base [Fig. 4.15 (*a*)]. Similarly, for *NPN* transistor, both collector and base are positive with respect to the emitter (the letter **P** of **P**ositive being the same as the middle letter of *N***P***N*). Again, collector is ***more positive*** than the base as shown in Fig. 4.15 (*b*).

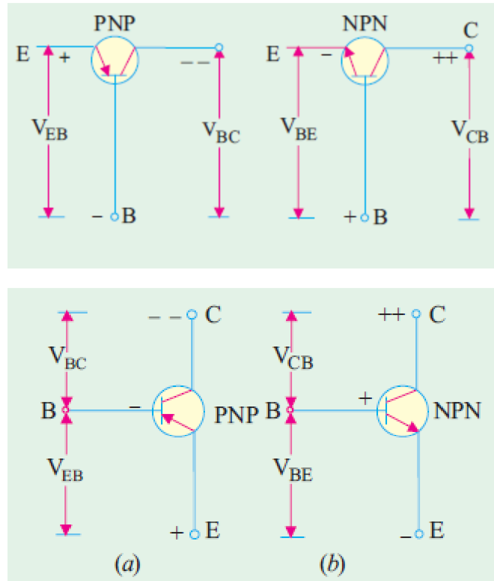


Figure 4.15

It may be noted that different potentials have been designated by double subscripts. The first subscript always represents the point or terminal which is more positive (or less negative) than the point or terminal represented by the second subscript. For example, in Fig. 4.15 (*a*), the potential difference between emitter and base is written as V_{EB} (and not V_{BE}) because ***emitter is positive with respect to base***. Now, between the base and collector themselves, collector is more negative than base. Hence, their potential difference is written as V_{BC} and not as V_{CB}. Same is the case with voltages marked in Fig. 4.15.

Transistor Currents

The three primary currents which flow in a properly-biased transistor are I_E, I_B and I_C. In Fig. 4.16 (*a*) are shown the directions of flow as well as relative magnitudes of these currents for a *PNP* transistor connected in the common-base mode. It is seen that again,

$$I_E = I_B + I_C$$

It means that a small part (about 1-2%) of emitter current goes to supply base current and the remaining major part (98-99%) goes to supply collector current.

Moreover, I_E flows into the transistor whereas both I_B and I_C flow out of it. Fig. 4.16 (*b*) shows the flow of currents in the same transistor when connected in the common-emitter mode. It is seen that again, $I_E = I_B + I_C$.

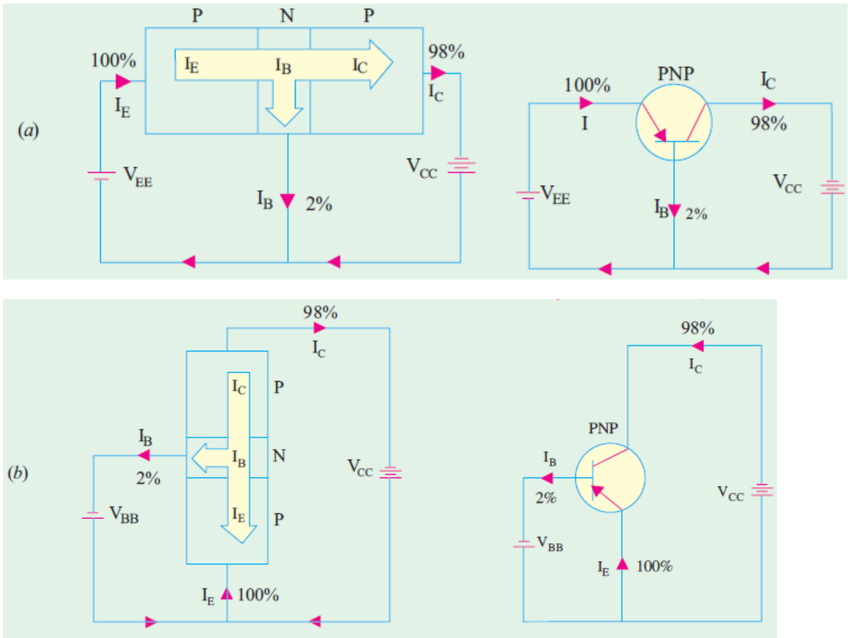


Figure 4.16

By normal convention, currents flowing ***into*** a transistor are taken as positive whereas those flowing ***out*** of it are taken as negative. Hence, I_E is positive whereas both I_B and I_C are negative.

Applying Kirchhoff's Current Law, we have

$$I_E + (-I_B) + (-I_C) = 0 \text{ or}$$

$$I_E - I_B - I_C = 0 \text{ or } I_E = I_B + I_C$$

This statement is true ***regardless of transistor type or transistor configuration.***

Note.

- Here, the leakage currents which exist in a transistor is not taken into account.
- The four basic guide posts about all transistor circuits are:

  a) conventional current flows along the arrow whereas electrons flow against it;

  b) *E/B* junction is always forward-biased;

  c) *C/B* junction is always reverse-biased;

  d) $I_E = I_B + I_C$

❑ Transistor Circuit Configurations

Basically, there are three types of circuit connections (called configurations) for operating a transistor.

1. common-base (*CB*),

2. common-emitter (*CE*),
3. common-collector (*CC*).

The term 'common' is used to denote the electrode that is common to the input and output circuits. Because the common electrode is generally grounded, these modes of operation are frequently referred to as grounded-base, grounded-emitter and grounded-collector configurations as shown in Fig. 4.17 for a *PNP* – transistor.
Since a transistor is a 3-terminal (and not a 4-terminal) device, one of its terminals has to be common to the input and output circuits.

CB Configuration

In this configuration, emitter current I_E is the input current and collector current I_C is the output current. The input signal is applied between the emitter and base whereas output is taken out from the collector and base as shown in Fig. 4.17 (*a*).The ratio of the collector current to the emitter current is called dc alpha (α_{dc}) of a transistor.

$$\therefore \qquad \alpha_{dc} = \frac{-I_C}{I_E}$$

The negative sign is due to the fact that current IE flows into the transistor whereas I_C flows out of it. Hence, I_E is taken as positive and I_C as negative.

$$\therefore \qquad I_C = -\alpha_{dc} I_E$$

If we write α_{dc} simply as α, then $\alpha = I_E / I_C$. It is also called forward current transfer ratio ($-h_{FB}$). In h_{FB}, subscript *F* stands for forward and *B* for common-base. The subscript d.c. on a signifies that this ratio is defined from dc values of I_C and I_E.

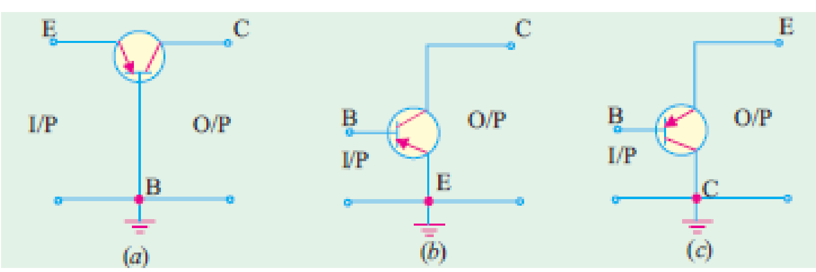


Figure 4.17

The α of a transistor is a measure of the quality of a transistor; higher the value of α, better the transistor in the sense that collector current more closely equals the emitter current. Its value ranges from 0.95 to 0.999. Obviously, it applies only to CB configuration of a transistor. As seen from above and Fig. 4.18.

A more accurate expression is $\alpha_{dc} = \dfrac{I_C - I_{CBO}}{I_E}$

Negative sign has been omitted, since we are here concerned with only magnitudes of the currents involved.

$$I_C = \alpha I_E \text{ and } I_B = I_E - \alpha I_E = (1-\alpha) I_E$$

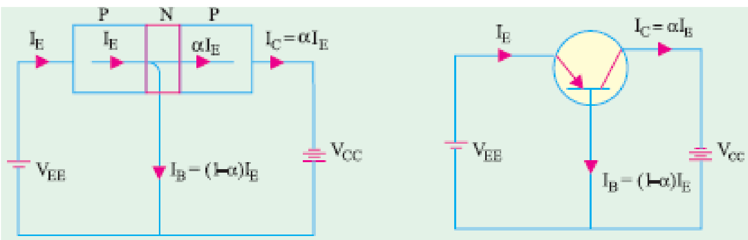


Figure 4.18

Incidentally, there is also an a.c. α for a transistor. It refers to the ratio of ***change*** in collector current to the ***change*** in emitter current.

$$\therefore \qquad \alpha_{ac} = \frac{-\Delta I_C}{\Delta I_E}$$

It is also, known as short-circuit gain of a transistor and is written as – h_{fb}. It may be noted that upper case subscript '*FB*' indicates dc value whereas lower case subscript '*fb*' indicates ac value. For all practical purposes, $\alpha_{dc} = \alpha_{ac} = \alpha$.

Example: *Following current readings are obtained in a transistor connected in CB configuration : I_E = 2 mA and IB = 20 mA. Compute the values of* α *and I_C.*

Solution. $I_C = I_E - I_B$

= 2 × 10–3 – 20 × 10–6 = **1.98 mA**

$\alpha = I_C / I_E$ = 1.98/2 = **0.99**

CE Configuration

Here, input signal is applied between the base and emitter and output signal is taken out from the collector and emitter circuit.

As seen from Fig. 4.19 (*b*), I_B is the input current and I_C is the output current. The ratio of the d.c. collector current to dc base current is called dc beta (βdc) or just β of the transistor.

∴ $\beta = -I_C / -I_B = I_C / I_B$ or $I_C = \beta I_B$ — Fig. 4.19 (*a*)

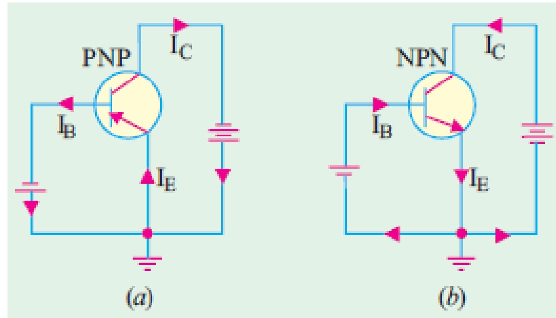


Figure 4.19

It is also called common-emitter d.c. ***forward transfer ratio*** and is written as h_{FE}. It is possible for β to have as high a value as 500. While analyzing ac operation of a transistor, we use ac β which is given by $\beta_{ac} = \Delta I_C / \Delta I_B$. It is also written as h_{fe}. The flow of various currents in a *CE* configuration both for *PNP* and *NPN* transistor is shown in Fig. 4.24b. As seen $I_E = I_B + I_C = I_B + \beta I_B = (1 + \beta) I_B$

- Relation Between α and β

$$\beta = \frac{I_C}{I_B} \text{ and } \alpha = \frac{I_C}{I_E}$$

$$\therefore \quad \frac{\beta}{\alpha} = \frac{I_E}{I_B}$$

Now $I_B = I_E - I_C$

$$\therefore \quad \beta = \frac{I_C}{I_E - I_C} = \frac{I_C / I_E}{I_E / I_E - I_C / I_E} \quad \text{or}$$

$$\beta = \frac{\alpha}{1-\alpha}$$

Cross-multiplying the above equation and simplifying it, we get;

$\beta(1-\alpha) = \alpha$ or $\beta = \alpha(1+\beta)$ or

$\alpha = \beta/(1+\beta)$

It shows that $(1-\alpha) = 1/(1+\beta)$

CC Configuration

In this case, input signal is applied between base and collector and output signal is taken out from emitter-collector circuit. Conventionally speaking, here I_B is the input current and I_E is the output current as shown in Fig. 4.19.

$$I_E = I_B + I_C = I_B + \beta I_B = (1+\beta) I_B$$

$\therefore$ output current = $(1+\beta) \times$ input current.

The current gain of the circuit is

$$\frac{I_E}{I_B} = \frac{I_E}{I_C} \cdot \frac{I_C}{I_B} = \frac{\beta}{\alpha} = \frac{\beta}{\beta/(1+\beta)} = (1+\beta)$$

The flow paths of various currents in a *CC* configuration are shown in Fig. 4.20.

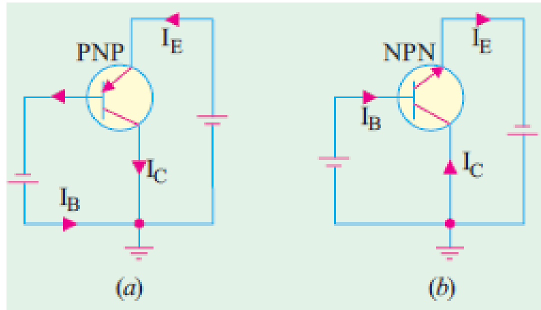


Figure 4.20

- **Relation Between Transistor Currents**

While deriving various equations, following definitions should be kept in mind.

$$\alpha = \frac{I_C}{I_E}, \quad \beta = \frac{I_C}{I_B}, \quad \alpha = \frac{\beta}{(1+\beta)} \text{ and } \beta = \frac{\alpha}{(1-\alpha)}$$

(i) $I_C = \beta I_B = \alpha I_E = \dfrac{\beta}{(1+\beta)} I_E$

(ii) $I_B = \dfrac{I_C}{\beta} = \dfrac{I_E}{(1+\beta)} = (1-\alpha) I_E$

(iii) $I_E = \dfrac{I_C}{\alpha} = \dfrac{1+\beta}{\beta} I_C = (1+\beta) I_B = \dfrac{I_B}{(1-\alpha)}$

(iv) *The three transistor d.c currents always bear the following ratio*

$$I_E : I_B : I_C \quad :: \quad 1 : (1-\alpha) : \alpha$$

This relates to the power distribution relationship in an induction motor.

Incidentally, it may be noted that for ac currents, small letters i_e, i_b and i_c are used.

Leakage Currents in a Transistor

(*a*) CB Circuit

Consider the *CB* transistor circuit shown in below figure. The emitter current (due to majority carriers) initiated by the forward-biased emitter base junction is split into two parts:

i. $(1-\alpha)\, I_E$ which becomes base current I_B in the external circuit and

ii. αI_E which becomes collector current I_C in the external circuit

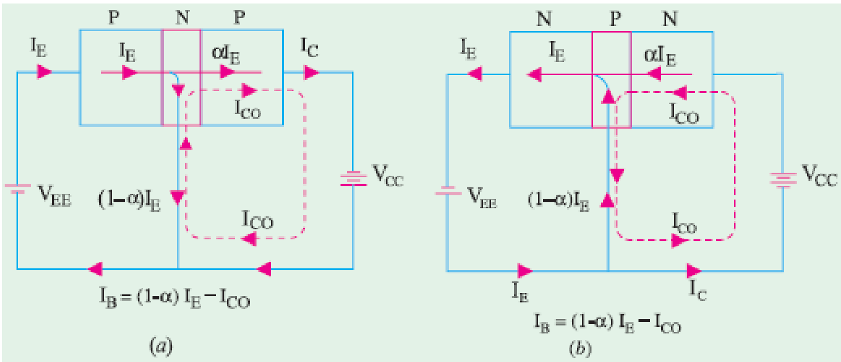


Figure 4.21

NB: Though *C/B* junction is reverse-biased for majority charge carriers (*i.e.* holes in this case), it is forward-biased so far as thermally-generated minority charge carriers (*i.e.* electrons in this case) are concerned. This current flows even when emitter is disconnected from its dc supply as shown in Fig. 4.21 (*a*) where switch, *S*1 is open. It flows in the ***same*** direction as the collector current of majority carriers (Actually, electrons, which form minority charge carriers in collector flow from negative terminal of collector battery, to collector, then to base through *C*/*B* junction and finally, to positive terminal of V_{CC}. However, conventional current flows in the opposite direction as shown by dotted line in Fig. 4.22 (*a*)). It is called leakage current I_{CBO}. The subscripts *CBO* stand for 'Collector to Base with emitter Open.' Very often, it is simply written as I_{CO}.

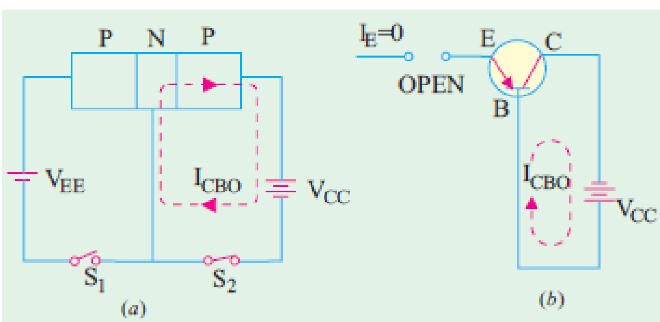


Figure 4.22

It should be noted that;

(i) I_{CBO} is exactly like the reverse saturation current I_S or I_0 of a reverse-biased diode discussed.

(ii) I_{CBO} is extremely temperature-dependent because it is made up of thermally-generated minority carriers.

As mentioned earlier, I_{CBO} doubles for every 10°C rise in temperature for *Ge* and 6°C for *Si*. If we take into account the leakage current, the current distribution in a *CB* transistor circuit becomes as shown in Fig. 4.22 both for *PNP* and *NPN* type transistors.

It is seen that total collector current is actually the sum of two components:

(i) current produced by normal transistor action *i.e.* component controlled by emitter current. Its value is a I_E and is due to majority carriers.

(ii) temperature-dependent leakage current *ICO* due to minority carriers.

$$\therefore \quad I_C = \alpha I_E + I_{CO} \text{.............} i$$

$$hence \quad \alpha = \frac{I_C - I_{CO}}{I_E}$$

Since $I_{CO} << I_C$, $\quad hence\ \alpha \equiv I_C / I_E$

(*iii*) Substituting the value of $I_E = (I_C + I_B)$ in Eq. (*i*) above, we get

Since $I_E = I_C + I_B$,

Then equation *i becomes;*

$$I_C = \alpha(I_C + I_B) + I_{CO}\ or$$

$$I_C(1-\alpha) = \alpha I_B + I_{CO} \therefore \quad I_C = \frac{\alpha I_B}{(1-\alpha)} + \frac{I_{CO}}{(1-\alpha)}$$

(*iv*) Eliminating I_C from Eq. (i) above, we get

$$(I_E - I_B) = \alpha I_E + I_{CO} \text{ or } \qquad I_B = (1-\alpha)I_E - I_{CO}$$

Transistor characteristics

These are curves which represents relationship between different d.c. currents and voltages of a transistor. They are helpful in studying the operation of a transistor when connected in a circuit. The three important characteristics of a transistor are:

1. *Input characteristic,*
2. *Output characteristic,*
3. *Constant-current transfer characteristic*

(*a*) Input Characteristic

It shows how I_B varies with changes in V_{BE} when V_{CE} is held constant at a particular value.

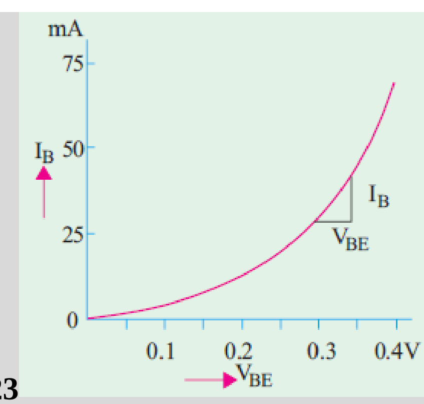


Figure 4.23

(b) Output or Collector Characteristic

It indicates the way in which I_C varies with changes in V_{CE} when I_B is held constant.

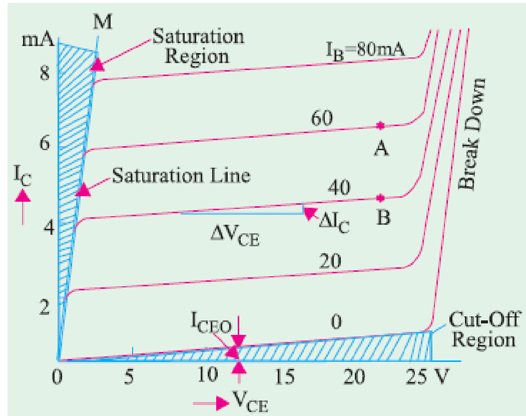


Figure 4.24

(c) Current Transfer Characteristic

It indicates how I_C varies with changes in I_B when V_{CE} is held constant at a given value.

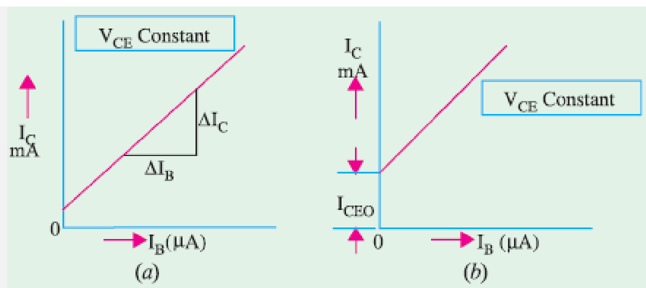


Figure 4.25

Importance of V_{CE}

The voltage V_{CE} is very important in checking whether the transistor is: (a) defective, (b) working in cut-off, (c) in saturation or well into saturation.

- When $V_{CE} = V_{CC}$, the transistor is in cut-off *i.e.* it is turned OFF.
- When $V_{CE} = 0$, the transistor is in saturation *i.e.* it is turned fully ON.
- When V_{CE} is less than zero *i.e.* negative, the transistor is said to be well into saturation. In practice, both these conditions are avoided.
- For amplifier operation, $V_{CE} = 1/2\ V_{CC}$ *i.e.* transistor is operated at approximately 1/2 ON. In this way, variations in I_B in either direction will control I_C in both directions. In other words, when I_B increases or decreases, I_C also increases or decreases.

However, if I_B is OFF, I_C is also OFF. On the other hand, if collector has been turned fully ON, maximum I_C flows. Hence, no further increase in

I_B can be reflected in I_C.

BJT Operating Regions

Transistors are three terminal active devices acting as either an insulator or a conductor by the application of a small signal voltage. The transistor's ability to change between these two states enables it to have two basic functions: "switching" (digital electronics) or "amplification" (analogue electronics).

A BJT has two junctions *i.e*. base-emitter (BE) and base-collector (BC) junctions either of which could be forward-biased or reverse-biased. With two junctions, there are four possible combinations of bias condition.

(i) both junctions reverse-biased,

(ii) both junctions forward-biased,

(iii) BE junction forward-biased, *BC* junction reverse-biased.

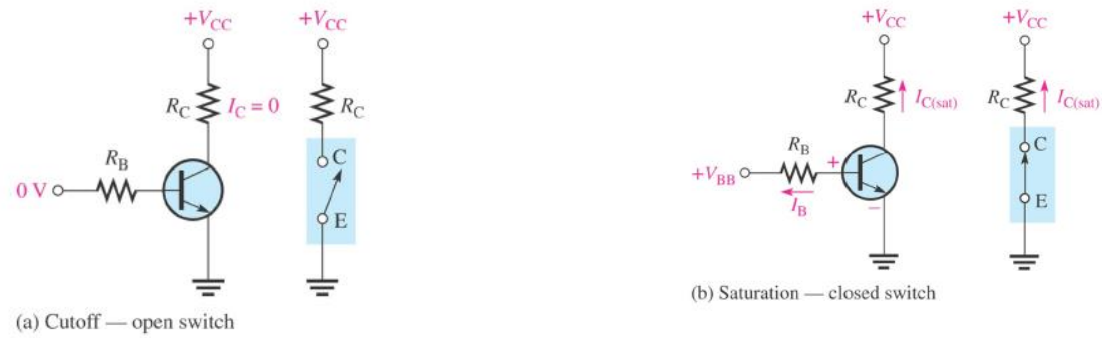
(iv) *BE* junction reverse-biased, *BC* junction forward-biased. This condition is generally not used, hence does not form the transistor operating regions.

Based on the above combinations, the bipolar transistors have the ability to operate within three different regions:

1. **Active Region** – The EB junction is forward-biased while CB junction is reverse-biased. For a CB and CE configurations, the voltages V_{BE} and V_{CE} are applied to the base B and collector C of the transistor and related by: $V_{CE} = V_{CB} + V_{BE}$. The transistor operates as an amplifier and $I_C = \beta I_B$
2. **Saturation region** – Both junctions are Forward-biased. The Base current (I_B) is high enough to give a Collector-Emitter voltage of 0 V (V_{CE} = 0) resulting in maximum Collector current I_C = I (saturation) flowing. The device is switched "fully-ON" operating as a switch.
3. **Cut-off region** – Both junctions are Reverse-biased. The Base current is zero or very small resulting in zero Collector current (I_C = 0) flowing, with V_{CE} = V_{CC}. The device is switched "fully-OFF" and operating as a switch.

BJT as a Switch

When used as an electronic switch, a transistor normally is operated alternating in cutoff and saturation.



(a) Cutoff — open switch

(b) Saturation — closed switch

$V_{CE} = V_{CC}$ and $I_C = 0$ **Figure 4.26** $V_{CE} = 0\,V$ and $I_C = V_{CC} / R_C$

4.3.2 Field Effect Transistors

- **What is a FET?**

The acronym 'FET' stands for *field effect transistor*. It is a three-terminal unipolar solid-state device in which current is *controlled by an electric field*. There are two main types of FETs;

(a) *Junction field effect transistor (JFET)*

(b) *Metal-oxide semiconductor FET (MOSFET).* It is also called insulated-gate FET (IGFET). It may be further subdivided into:

(i) depletion-enhancement MOSFET i.e. DEMOSFET.

(ii) enhancement-only MOSFET i.e. E-only MOSFET.

Both of these can be either P-channel or N-channel devices. The FET family tree is shown below;

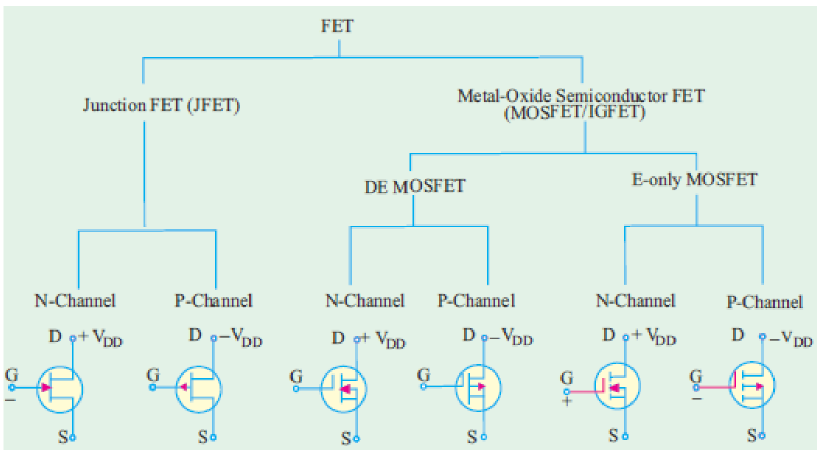


Figure 4.27

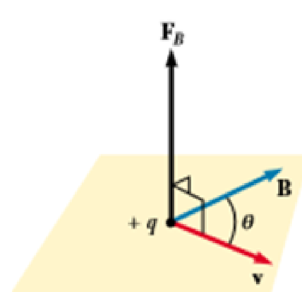
Reference Text Books

1. Theraja, B. L., Theraja, A. K., Patel, U., Uppal, S., Panchal, J., Oza, B., ... & Patel, R. (2005). A textbook of electrical technology Vol. II. *S. Chand publishers,*
2. Yariv, "Optical Electronics in Modern Communication", 5th edition, Oxford University Press, NY, 1997.
3. Sedra, A. S., Sedra, D. E. A. S., Smith, K. C., & Smith, K. C. (1998). *Microelectronic circuits*. New York: Oxford University Press.
4. Senior, J. M., & Jamro, M. Y. (2009). *Optical fiber communications: principles and practice*. Pearson Education.
5. Rosencher, E., & Vinter, B. (2002). *Optoelectronics*. Cambridge University Press.

Sample Question

1. Name two direct current (DC) sources of electromotive force (emf).
2. What are charge carriers in semiconductors?
3. Name two applications of transistors.
4. Name two properties of electromagnetic waves (EM)
5. List the following EM waves in their order of increasing energy.

   Radio, Infrared, Gamma, Visible light
6. State the Kirchoff's current law giving its mathematical expression.
7. Define electric field and determine the electric force on a proton placed in an electric field of 2.0×10^4 N/C directed along the positive x axis. (Take /e/ = 1.6×10-19 C)
8. State the Ampere's law giving its mathematical expression.
9. In. As an ambulance travels east down a highway at a speed of 35 m/s, its siren emits sound at a frequency of 400 Hz. What frequency is heard by a person in a car traveling west at 24 m/s as the car
   - **(i)** approaches the ambulance?
   - **(ii)** moves away from the ambulance?
10. In a transistor connected in CB configuration, the following current readings were recorded: $I_E = 2\ mA$ and $I_B = 20\ mA$. Compute the values of I_C and α .
11. When a charged particle, **+q** moves through a magnetic field **B**, with a velocity, **v**, it experiences a magnetic force, $\mathbf{F_B}$ given by; $\mathbf{F_B} = q\mathbf{v}\times\mathbf{B}$ as shown in the following diagram. Show that the maximum magnetic force it can experience is given by $F_B = qvB$.



12. Explain three properties of electric charges.
13. Define a semiconductor? Explain the effect of temperature increase on conductivity of

semiconductor materials.

14. A general sinusoidal equation is expressed as $y = B\sin(\kappa x - \omega t + \phi)$. What do the symbols $y, \quad B, \quad k, \quad \omega \quad and \quad \phi$ represent?

15. In an a.c circuit, the average power is given by $\overline{P} = I_{rms}V_{rms}$. Show that $2\overline{P} = P_o$, where P_o is the peak power.

16. Show that for a purely resistive circuit, both voltage and current are in phase.

17. Using a diagram, show how a BJT in common base (CB) configuration is biased.

18. Define resonance in RLC circuits. Show that, the resonance ω_o is given by $\omega_o = \dfrac{1}{\sqrt{LC}}$.

19. Show that for resistors arranged in series, the effective resistance is given by: $R_{eff} = R_1 + R_2 + R_3 + ...$

20. Give three differences between electric force and magnetic force

21. Using relevant examples, differentiate between intrinsic and extrinsic semiconductors

22. What is a P-N junction? Describe the formation of the depletion layer in a P-N junction device.

23. Show that for a positively charged particle moving in a uniform magnetic field, the particle follows a circular path with radius $r = \dfrac{mv}{qB}$, where the symbols have their usual meaning.

24. For red light of wavelength $750\, nm$, emitted by excited lithium atoms,

- **(i)** express this wavelength in metres
- **(ii)** calculate the light frequency (v)
- **(iii)** What is its wave number (k)
- **(iv)** calculate the energy (in Joules) associated the light

25. Describe the harmful effects of EM radiations