State Institute of Education, Kashmir

Bemina, Bypass, Srinagar, J&K

STUDY MATERIAL BASED ON JKBOSE SYLLABUS

CLASS: 12TH

SUBJECT: PHYSICS

Patron & Mentor:

Mr. Mohammad Younis Malik

Director, School Education, Kashmir.

Chief-Coordinator:

Ms. Suriya Bano,

Jt.Director, (Trgs.)/Principal, SIE, Kashmir.

Coordinator:

Mr. Fayaz Ahmad, SIE, Srinagar.

Delopement Team

- **Mr. Mohd Iqbal Shah (Sr.Lecturer Physics GHSS Handwara | Best Teacher Awardee 2016)**
- **Mr. Irshad A.Shah Pirzada (Principal, Govt. Higher secondary Institute),**
- **Dr. Sajad Ahmad Mir (Sr. Lecturer Physics, BHSS Pampore)**
- **Mr Tariq Ahmad Bhat (Sr.Lecturer Physics, HSS Kulgam),**
- **Mr Masroor Ahmad Ganaie (Sr. Lecturer Physics, BHSS Yaripora Kulgam)**
- **Mr. Shabir (Science teacher, HS Hawoorah Kulgam)**

Preface

Welcome to this study material which has been prepared and developed as part of the vision and under the mentorship of worthy Director, School Education, Kashmir, Mr. Mohammad Younis Malik. It is he who wanted to provide a quality study material to the students so that attending the coaching centres by the students of higher secondary level is minimised to a large extent. Besides, keeping in view the situation for the last few years wherein the Education sector has been badly hit, the initiative will prove to be of great significance. Accordingly, the worthy DSEK entrusted the said job to State Institute of Education, Kashmir. A two-day workshop was immediately conducted in this regard on 9th & 10th of March 2020 whereinthe best subject experts from the School Education Department were involved so that a proper strategy and plan of action would be adopted to accomplish the said task. It is expected that this study material enhances student excess to high quality learning materials, mantaining highest standards of accademic rigor at little to no cost. The study material discusses each topic of syllabus in a lucid way so that it can also prove a supporting material for the students for the annual examination. Each chapter has been supplemented with solved numericals followed by self-evaluation section which is comprised of multiple choice questions and unsolved numericals to check the level of understanding from the text matter. The answers to multiple choice questions and unsolved numiricals are given at the end of each chapter so that students can verify their answers. Every effort has been made to make this study material error-free yet in case there is any omission, typing/printing mistakes, or any other error, the same is requested to readers to send at principalsiekashmir@gmail.com/sajadphysics@gmail.com. We are thankful to the faculty members of SIE, DIETs and the Field subject experts especially the ones who were practically involved in getting this document set and wish all the best to all the stakeholders, especially the students of the valley.

Coordinator

Jt. Director (Trgs.)/Principal
SIE, Kashmir

Index

JKBOSE 12th class Syllabus in Physics

Unit 1. Electrostatics

Electric charges; conservation of charge, Coulomb's law-force between two point charges, forces between multiple charges, superposition principle and continuous charge distribution.

Electric field, electric field due to a point charge, electric field lines, and electric dipole, electric field due to a dipole, Torque on a dipole in uniform electric field.

Electric flux, statement of Gauss's theorem and its application to find field due to infinitely long straight wire, uniformly charged infinite plane sheet and uniformly charged thin spherical shell (field inside and outside).

Electric potential, potential difference, electric potential due to point charge, a dipole and system of charges. Equipotential surfaces, Electric potential energy of a system of two point charges and of electric dipole in an electrostatic field.

Conductor and insulators, free charges and bound charges inside a conductor. Dielectrics and electric polarization, capacitors and capacitance, combination of capacitors in series and in parallel, capacitance of parallel plate capacitor. Van de Graaff generator.

Unit 2. Current Electricity

Electric Current, flow of electric charges in a metallic conductor, drift velocity, mobility and their relation with electric current. Ohm's law, electric resistance, V_I cgaracteristics (Linear, non-leaner), Electric energy and Power, Electric resistivity and Conductivity, Carbon resistors, Colour code of carbon resistors; Temperature dependence of resistance.

Internal resistance of a cell, potential difference and emf of a cell, combination of cells in series and in parallel. Elementary idea of secondary cells. Krchoff's laws and their applications. Wheat stone bridge, Meter bridge.

Potentiometer-principle and its application to measure potential difference for comparing emf of two cells; measurement of internal resistance of a cell.

Unit 3. Magnetic Effects of Current and Magnetism

Concept of magnetic field, Oersted's experiment, Biot-Savart law and its application to current carrying circular loop. Ampere's law and its applications to infinite long straight wire, straight and toroidal solenoids.

Force on a moving charge in a uniform magnetic and electric fields. Cyclotron. Force on a current carrying conductor in a uniform magnetic field. Force between two parallel current carrying conductors-definition of ampere.

Torque experienced by a current loop in uniform magnetic field, Moving coil galvanometer-its current sensitivity and conversion to ammeter and voltmeter.

Current loop as a magnetic dipole and its magnetic dipole moment. Magnetic dipole moment of a revolving electron. Magnetic field intensity due to a magnetic dipole (bar magnet) along its axis and perpendicular to its axis. Torque on a magnetic dipole (bar magnet) in uniform magnetic filed, Bar magnet as an equivalent solenoid, magnetic field lines, Earth's magnetic field and magnetic elements. Para, dia, ferro-magnetic substances with examples. Electromagnets and factors affecting their strength, Permanent magnets.

Unit 4. Electro-magnetic Induction and Alternating Currents

Electromagbetic induction, Faraday's laws, Induced emf and current; Lenz's law, Eddy currents, Self and Mutual inductance.

Alternatimg currents, paek and rms value of alternatimg current/voltage. Reactance and impedence, LC oscillations (qualitative treatment only) and LCR circuits series, Resonance, power in A.C. circuits, wattless current, AC Generator and Transformer.

Unit 5. Electro-magnetic Waves

Need for displacement current, Electro-magnetic waves and their characteristics (qualitative idea only), Transverse nature of electromagnetic waves.

Electromagnetic spectrum (radio-waves, micro-waves, infra-red, visible, ultraviolet, X-rayss, gamma rays) including elementary facts about their uses.

Unit 6. Optics

Ray Optics- Reflection of light; Spherical mirrors; Mirror formula. Refraction of light-total internal reflection and its applications, Optical fibers, Refraction at spherical surfaces, Lenses, Thin lenses formula, Lens-makers formula, Newton's relation: displacement method to find position of images (conjugate points), Magnification, Power of lens, Combination of thin lenses. Microscopes and Astronomical telescopes (reflecting and refracting) and their Magnifying powers.

Wave Optics- wave front and Hugen's principle, reflection and refraction of plane wave at a plane surface using wavefronts. Proof of laws of reflection and refraction using Huygen's principle, Interference, Young's double slit experiment and expression for fringe width, Coherent sources and sustained interference of light.

Diffraction due to a single slit, width of central maximum. Resolving power of Microscopes and Astronomical telescopes, Polarization, Plane polarized light, Brewster's law, uses of plane polarized light and Polaroids.

Unit 7. Dual Nature of Matter and Radiation

Dual nature of radiation, Photoelectric effect, Hertz and Lenard's observation. Einstein's photoelectric equation-particle nature of light.

Matter waves, wave nature of particles, de-Broglie relation, Davisson-Germer experiment (experimental details should be omitted; only conclusion should be explained)

Unit 8. Atomic Nuclei

Alpha-particle scattering experiment, Rutherford's model of atom, Bohr's Model of atom; energy levels, Hydrogen spectrum, Continuous and characteristics of X-rays. Composition and size of nucleus; atomic masses, isotopes, isobars, isotones, Radioactivity (alpha, bets and gamma) particles/rays and their properties. Radioactive decay law, Mass-energy relation, Mass defect, Binding energy per nucleon (B.E/nucleon) and its variation with mass number, Nuclear fission and Nuclear fusion.

Unit 9. Electronic Devices

Energy bands in solids, conductors, insulators and semiconductors, Semiconductor diode, I-V characteristics in forward and reverse bias, Diode as rectifier; I-V characteristics of LED, Photo diode, Solar cell and Zener diode-Zener diode as voltage regulator, Junction transistors and its action; Characteristics of a transistor, Transistor as an amplifier (common emitter configuration) and Oscillator (common emitter). Logic gates (OR, AND, NOT), concept of NAND and NOR gates, Transistor as a switch

Unit 10. Communication System

Elements of communication system (block diagram only), Band width of signals (speech, T.V and digital data); bandwidth of transmission medium, Propagation of electromagnetic waves in the atmosphere, sky and space wave propagation.

Need for modulation; Production and detection of an amplitude modulated wave.

UNIT1

ELECTROSTATICS

1. Electric Charge

Electric charge is a fundamental property like mass, length etc associated with elementary particles for example electron, proton and many more.

- Electric charge is the property responsible for electric forces which acts between nucleus and electron to bind the atom together.

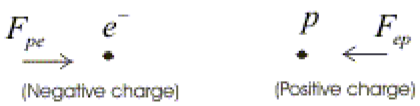
- Charges are of two kinds
  (i) negative charge
  (ii) positive charge

- Electrons are negatively charged particles and protons, of which nucleus is made of, are positively charged particles. Actually nucleus is made of protons and neutrons but neutrons are uncharged particles.

- electric force between two electrons is same as electric force between two protons kept at same distance apart i. e., both set repel each other but electric force between an electron and proton placed at same distance apart is not repulsive but attractive in nature.

  Conclusion
  (a) Like charges repel each other

  (b) Unlike charges attract each other



- Assignment of negative charge on electron and positive charge on proton is purely conventional , it does not mean that charge on electron is less than that on proton.

- Importance of electric forces is that it encompasses almost each and every field associated with our life; being it matter made up of atoms or molecules in which electric charges are exactly balanced or adhesive forces of glue associated with surface tension, all are electric in nature.

Unit

- Charge on a system can be measured by comparing it with the charge on a standard body.

- SI unit of charge is Coulomb written as C.

- 1 Coulomb is the charge flowing through the wire in 1 second if the electric current in it is 1A.

- Charge on electron is -1.602×10^{-19} C and charge on proton is positive of this value.

2. Basic properties of electric charge

(i) Additivity of charges

- Charges adds up like real numbers i. e., they are Scalars more clearly if any system has n number of charges q_1, q_2, q_3, q_n then total charge of the system is
$q = q_1 + q_2 + q_3 + \ldots\ldots\ldots\ldots\ldots q_n$

- Proper sign have to be used while adding the charges for example if
$q_1 = +1C$
$q_2 = -2C$
$q_3 = +4C$
then total charge of the system is
$q = q_1 + q_2 + q_3$
$q = (+1) + (-2) + (+4)$ C
$q = (+3)$ C

(ii) Charge is conserved

- Charge of an isolated system is conserved.

- Chage can not be created or destroyed but charged particles can be created or destroyed.

(iii) Quantization of charge

- All free charges are integral multiples of a unit of charge e, where $e = -1.602 \times 10^{-19}$ C i. e., charge on an electron or proton.

- Thus charge q on a body is always denoted by
$q = ne$
where n = any integer positive or negative

3. Frictional Electricity

- If we pass a comb through hairs, comb becomes electrically charged and can attract small pieces of paper.

- Many such solid materials are known which on rubbing attract light objects like light feather, bits of papers, straw etc.

- Explaination of appearance of electric charge on rubbing is simple.

- Material bodies consists of large number of electrons and protons in equal number and hence is in neutral in their normal state. But when the body is rubbed for example when a glass rod is rubbed with silk cloth, electrons are transferred from glass rod to silk cloth. The glass rod becomes positively charged and the silk cloth becomes negatively charged as it recieves extra electrons from the glass rod.

- In this case rod after rubbing, comb after passing through dry hairs becomes electrified and these are the example of frictional electricity.

4. Coulomb's law

- Coulomb's law is the law of forces between electric charges.
  Statement
  " It states that two stationary point charges q_1 and q_2 repel or attract each other with a force F which is directly proportional to the product of charges and inversly proportional to the square of distance between them."

  This dependence can be expressed by writing

$$F \propto \frac{q_1 q_2}{r^2} \qquad (1)$$

- These forces are attractive for unlike charges and repulsive for like charges .

- We now try to express Coulomb's law in vector form for more clearity of magnitude and direction of forces.

- Consider two point charges q_1 and q_2 at points with position vector $\mathbf{r_1}$ and $\mathbf{r_2}$ with respect to the origin

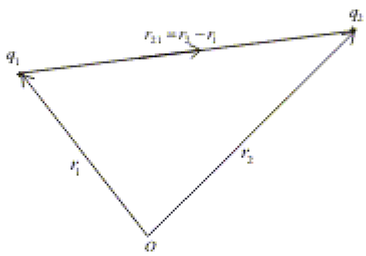


Figure 1.

vector $\mathbf{r_{21}} = \mathbf{r_2} - \mathbf{r_1}$ is the difference between $\mathbf{r_2}$ and $\mathbf{r_1}$ and the distance of separation r is the magnitude of vector $\mathbf{r_{21}}$.
pointwise it can be written as
$\mathbf{r_1}$ = position vector of charge q_1 with respect to origin
$\mathbf{r_2}$ = position vector of charge q_2 with respect to origin
$\mathbf{r_{21}}$ = vector from 1 to 2 ($r_2 - r_1$)
$\mathbf{r_{12}} = -\mathbf{r_{21}}$ = vector from 2 to 1 ($\mathbf{r_1} - \mathbf{r_2}$)
$\mathbf{r} = \mathbf{r_{12}} = \mathbf{r_{21}}$ = distance between 1 and 2.

q_1 and q_2 are of same charge type

q_1 and q_2 are of different charge type

Coulomb's law can then be expressed as
$\mathbf{F}_{21}$ = force on q_2 due to q_1

$$\mathbf{F}_{21} = \frac{kq_1q_2\mathbf{r}_{21}}{r^3} \quad (2a)$$

and, $\mathbf{F}_{12}$ = force on q_1 due to q_2

$$\mathbf{F}_{12} = -\mathbf{F}_{21} \frac{kq_1q_2\mathbf{r}_{12}}{r^3} \quad (2b)$$

Special Case

for simplicity we can choose q_1 being placed at origin
$\mathbf{r}_1 = 0$
and if we write $\mathbf{r}_2 = \mathbf{r}$ the position vector of q_2 then
$\mathbf{F}_{21}$ = force on q_2 due to q_1

$$\mathbf{F}_{21} = \frac{kq_1q_2\mathbf{r}}{r^3} \quad (3a)$$

$\mathbf{F}_{12}$ = force on q_1 due to q_2

$$\mathbf{F}_{12} = \frac{-kq_1q_2\mathbf{r}}{r^3} \quad (3b)$$

unit vector $\hat{r}_{21}$ and $\hat{r}_{12}$ can be defined as
$\hat{\mathbf{r}}_{21} = \mathbf{r}_{21}/r$ directed from q_1 to q_2
$\hat{\mathbf{r}}_{12} = \mathbf{r}_{12}/r$ directed from q_2 to q_1 (4)
$= -\mathbf{r}_{21}/r$
force can now be written in terms of unit vector given as follows

$$\mathbf{F}_{21} = \frac{kq_1q_2\hat{\mathbf{r}}_{21}}{r^2} \quad (5a)$$

$$\mathbf{F}_{12} = \frac{kq_1q_2\hat{\mathbf{r}}_{12}}{r^2} \quad (5b)$$

from this we can immidiately find factors giving magnitude and the directions

- in equation (2) we find a positive constant K and experimentally found value of k is

$K = 8.98755 \times 10^{9}\ Nm^2/C^2$
$K \cong 9 \times 10^{9}\ Nm^2/C^2$
sometimes K is written as $1/4\pi\varepsilon_0$ where ε_0 is the permittivity of the vaccum whose value is
$K = 1/4\pi\varepsilon_0$
$(\varepsilon_0 = 9 \times 10^{-12}\ C^2/Nm^2)$

5. Principle Of Superposition

- Coulomb's law gives the electric force acting between two electric charges.
- Principle of superposition gives the method to find force on a charge when system consists of large number of charges.
- According to this principle when a number of charges are interacting the total force on a given charge is vector sum of forces exerted on it by all other charges.
- This principle makes use of the fact that the forces with which two charges attract or repel one another are not affected by the presence of other charges.
- If a system of charges has n number of charges say q_1, q_2, ..., q_n, then total force on charge q_1 according to principle of superposition is
  $\mathbf{F} = \mathbf{F}_{12} + \mathbf{F}_{13} + \ldots\ldots\ldots\ldots \mathbf{F}_{1n}$
  Where $\mathbf{F}_{12}$ is force on q_1 due to q_2 and $\mathbf{F}_{13}$ is force on q_1 due to q_3 and so on.

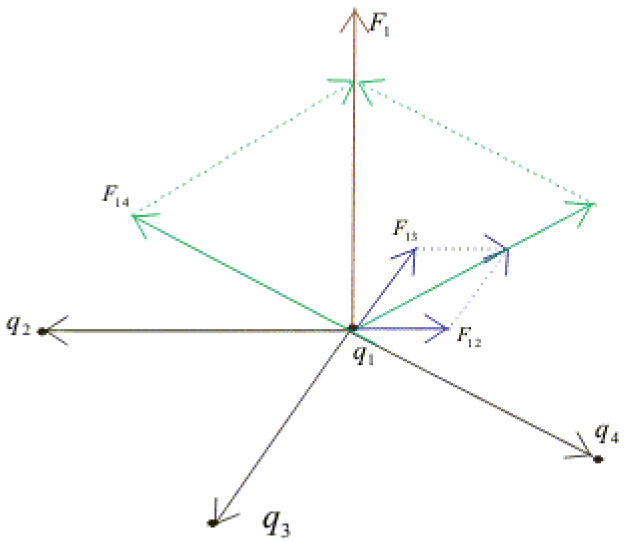


Figure 2. Force due to system of multiple charges

- $\mathbf{F_{12}}$, $\mathbf{F_{13}}$, ... $\mathbf{F_{1n}}$ can be calculated from Coulomb's law i. e.

$\mathbf{F_{12}}=\frac{kq_1q_2\hat{\mathbf{r}}_{12}}{4\pi\varepsilon_0(r_{12})^2}$

- to,

$\mathbf{F_{1n}}=\frac{kq_1q_2\hat{\mathbf{r}}_{1n}}{4\pi\varepsilon_0(r_{1n})^2}$

- The total force F_1 on the charge q_1 due to all other charges is the vector sum of the forces $\mathbf{F_{12}}$, $\mathbf{F_{13}}$, ... $\mathbf{F_{1n}}$.

$\mathbf{F_1} = \mathbf{F_{12}} + \mathbf{F_{13}} +$ ....

$$F_1 = \frac{1}{4\pi\varepsilon_0}\left[\frac{q_1q_2}{r_{12}^2}\hat{r}_{12} + \ldots\ldots\ldots\ldots \frac{q_1q_n}{r_{1n}^2}\hat{r}_{1n}\right]$$

$or,$

$$F_1 = \frac{q_1}{4\pi\varepsilon_0}\sum_{i=2}^{n}\frac{q_i}{r_{1i}^2}\hat{r}_{1i}$$

- The vector sum is obtained by parellogram law of addition of vector.
- Similarly force on any other charge due to remaining charges say on q_2, q_3 etc. can be found by adopting this method.

6. Electric Field

- Electrical interaction between charged particles can be reformulated using the concept of electric field.
- To understand the concept consider the mutual repulsion of two positive charged bodies as shown in fig (a)

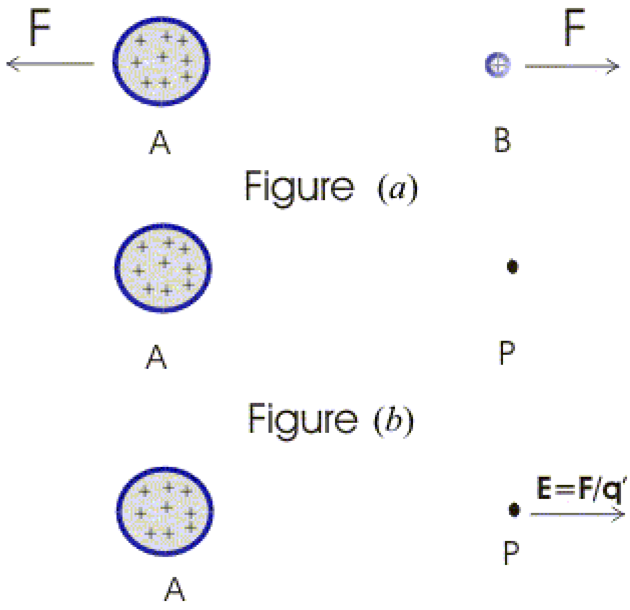


Figure (a)

Figure (b)

Figure (c)

- Now if remove the body B and label its position as point P as shown in fig (b), the charged body A is said to produce an electric field at that point (and at all other points in its vicinity)
- When a body B is placed at point P and experiences force F, we explain it by a point of view that force is exerted on B by the field not by body A itself.
- The body A sets up an electric field and the force on body B is exerted by the field due to A.
- An electric field is said to exists at a point if a force of electric origin is exerted on a stationary charged (test charge) placed at that point.
- If F is the force acting on test charge q placed at a point in an electric field then electric field at that point is
  $\mathbf{E} = \mathbf{F}/q$
  or $\mathbf{F} = q\mathbf{E}$
- Electric field is a vector quantity and since F = qE the direction of E is the direction of F.
- Unit of electric field is ($N.C^{-1}$)

Q. Find the dimensions of electric field
Ans. [$MLT^{-3}A^{-1}$]

7. Calculation of Electric Field

- In previous section we studied a method of measuring electric field in which we place a small test charge at the point, measure a force on it and take the ratio of force to the test charge.
- Electric field at any point can be calculated using Coulomb's law if both magnitude and positions of all charges contributing to the field are known.

- To find the magnitude of electric field at a point P, at a distance r from the point charge q, we imagine a test charge q'to be placed at P. Now we find force on charge q' due to q through Coulomb's law.

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

- electric field at P is

$$E=\frac{q}{4\pi\varepsilon_0 r^2}$$

- The direction of the field is away from the charge q if it is positive

- Electric field for either a positive or negative charge in terms of unit vector r directed along line from charge q to point P is

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

- r = distance from charge q to point P.
- When q is negative , direction of E is towards q, opposite to r.

Electric Field Due To Multiple Charges

- Consider the number of point charges q_1, q_2,........... which are at distance r_{1P}, r_{2P},.................. from point P as shown in fig

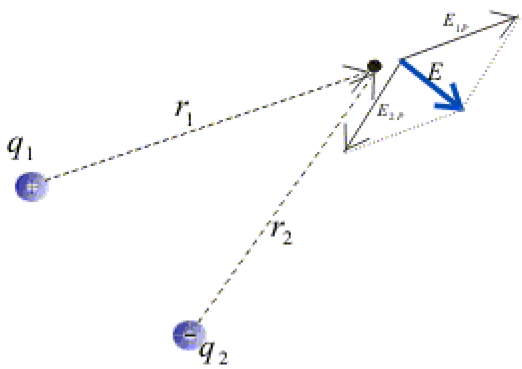


Figure:- figure shows the resultant electric field due to two point charges at point P

- The resultant electric field is the vector sum of individual electric fields as
E = E_{1P} + E_{2P} + ...

$$E=\frac{1}{4\pi\varepsilon_0}\left(\frac{q_1}{r_{1P}^2}\hat{r}_{1P}+\frac{q_2}{r_{2P}^2}\hat{r}_{2P}+\dots\dots\dots\right)$$

This is also a direct result of principle of superposition discussed earlier in case of electric force on a single charge due to system of multiple charges.

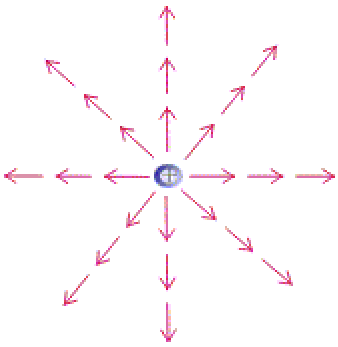
- E is a vector quantity that varies from one point in space to another point and is determined from the position of square charges.

8. Electric Field Lines

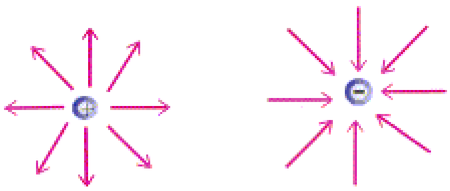
- For a single positive point charge q, electric field is

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

- now to get feel of this field one can sketch a few representative vectors as shown in fig below

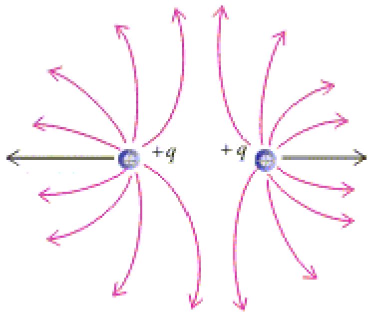


- Since electric field varies as inverse of square of the distance that points from the charge the vector gets shorter as you go away from the origin and they always points radially outwards.
- Connecting up these vectors to form a line is a nice way to represent this field .
- The magnitude of the field is indicated by the density of the field lines.
- Magnitude is strong near the center where the field lines are close togather, and weak farther out, where they are relatively apart.
- So, electric field line is an imaginary line drawn in such a way that it's direction at any point is same as the direction of field at that point.
- An electric field line is, in general a curve drawn in such a way that the tangent to it ateach point is the direction of net field at that point.
- Field lines of a single position charge points radially outwards while that of a negative charge are radially inwards as shown below in the figure.

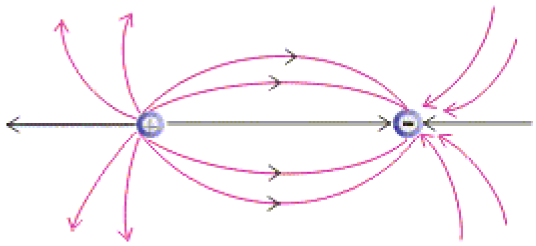


$q>0$ $\quad$ $q<0$

- Field lines around the system of two positive charges gives a different picture and describe the mutual repulsion between them.



- Field lines around a system of a positive and negative charge clearly shows the mutual attraction between them as shown below in the figure.

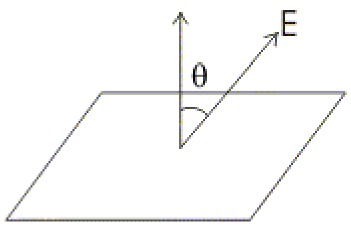


- Some important general properties of field lines are
  1.Field lines start from positive charge and end on a negative charge.
  2.Field lines never cross each other if they do so then at the point of intersection there will be two direction of electric field.
  3.Electric field lines do not pass through a conductor , this shows that electric field inside a conductor is always zero.
  4.Electric field lines are continuous curves in a charge free region.

9. Electric Flux

- Consider a plane surface of area ΔS in a uniform electric field E in the space.

- Draw a positive normal to the surface and θ be the angle between electric field E and the normal to the plane.



- Electric flux of the electric field through the choosen surface is then
  $\Delta\varphi = E\,\Delta S\cos\theta$

- Corresponding to area ΔS we can define an area vector ΔS of magnitude ΔS along the positive normal. With this definition one can write electric flux as
  $\Delta\varphi = \mathbf{E}\cdot\mathbf{\Delta S}$

- direction of area vector is always along normal to the surface being choosen.

- Thus electric flux is a measure of lines of forces passing through the surface held in the electric field.

Special Cases

- If E is perpendicular to the surface i. e., parallel to the area vector then θ = 0 and $\Delta\varphi = E\ \Delta S \cos 0$
- If θ = π i. e., electric field vector is in the direction opposite to area vector then $\Delta\varphi = -E\ \Delta S$
- If electric field and area vector are perpendicular to each other then θ = π/2 and Δφ = 0
- Flux is an scaler quantity and it can be added using rules of scaler addition.
- For calculating total flux through any given surface , divide the surface into small area elements. Calculate the flux at each area element and add them up.
- Thus total flux φ through a surface S is $\varphi \cong \Sigma\mathbf{E}.\mathbf{\Delta S}$
- This quantity is mathematically exact only when you take the limit ΔS→0 and the sum in equation 3 is written as integral $\varphi = \int\Sigma\mathbf{E}.\mathbf{dS}$

Electric Dipole

- Electric dipole is a pair of equal and opposite charges, +q and -q, separated by some distance 2a.
- Total charge of the dipole is zero but electric field of the dipole is not zero as charges q and -q are separated by some distance and electric field due to them when added is not zero.

(A)Field of an electric dipole at points in equitorial plane

- We now find the magnitude and direction of electric field due to dipole.

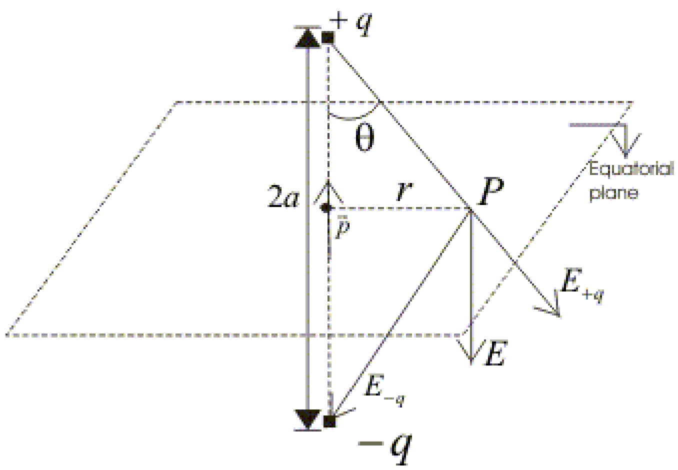


Figure:-Electric field of dipole at points in equatorial plane

- P point in the equitorial plane of the dipole at a distance r from the centre of the dipole. Then electric field due to +q and -q are

$$E_{-q} = \frac{-q\hat{P}}{4\pi\varepsilon_0(r^2 + a^2)} \qquad (1a)$$

$$E_{+q} = \frac{q\hat{P}}{4\pi\varepsilon_0(r^2 + a^2)} \qquad (1b)$$

and they are equal

$\hat{P}$ = unit vector along the dipole axis (from -q to +q)

- From fig we can see the direction of E_{+q} and E_{-q}. Their components normal to dipole cancel away and components along the dipole add up.
- Dipole moment vector points from negative charge to positive charge so in vector form.

E = -(**E**$_{+q}$ + **E**$_{-q}$) cos θ

$$E = -\frac{q}{4\pi\varepsilon_0}\left[\frac{1}{r^2+a^2} + \frac{1}{r^2+a^2}\right]\frac{a}{\sqrt{r^2+a^2}}$$

or,

$$E = -\frac{2qa}{4\pi\varepsilon_0(r^2+a^2)}\hat{p}$$

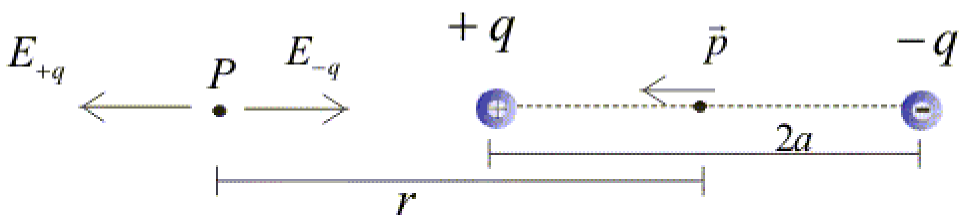
(2)

At large distances (r>>a), above equation becomes

$$E = \frac{-2qa\hat{P}}{4\pi\varepsilon_0(r^3)} \qquad (3)$$

and they are equal

(B) Field of an electric dipole for points on the axis



- Let P be the point at a distance r from the centre of the dipole on side of charge q. as shown in the fig.

$$E_{-q} = \frac{-q\hat{P}}{4\pi\varepsilon_0(r+a)^2} \qquad (4a)$$

$\hat{P}$ = unit vector along the dipole axis (from -q to +q)

also

$$E_{+q} = \frac{q\hat{P}}{4\pi\varepsilon_0(r-a)^2} \qquad (4b)$$

- Total field at P is

E = **E**$_{+q}$ + **E**$_{-q}$

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

or,

$$E = \frac{q}{4\pi\varepsilon_0}\left[\frac{4ar}{(r^2-a^2)}\right]\hat{p}$$

(5)

for r>>a

$$E = \frac{4qa\hat{P}}{4\pi\varepsilon_0 r^3} \quad (6)$$

- For equation (3) and (6) charge q and dipole separation 2a appear in combination qa. This leads us to define dipole moment vector **P** of electric dipole. Thus, electric dipole moment $\mathbf{P} = q \times 2a\,\hat{\mathbf{P}}$ (7)
- Unit of dipole moment is Coulomb's meter (Cm).
- In terms of electric dipole moment, field of a dipole at large distances becomes

  (i) At point on equitorial plane (r>>a)
  $E = -P/4\pi\varepsilon_0 r^3$

$$E = \frac{-\mathbf{P}}{4\pi\varepsilon_0 r^3}$$

- (ii) At point on dipole axis (r>>a)

$$E = \frac{2\mathbf{P}}{4\pi\varepsilon_0 r^3}$$

- **Note:-**
  (i) Dipole field at large distances falls off as $1/r^3$
  (ii) Both the direction and magnitude of dipole an angle between dipole moment vector **P** and position vector **r**

(C) Dipole in a uniform external field

- Consider a dipole in a uniform electric field **E** whose direction makes an angle θ with dipole axis (line joining two charges)

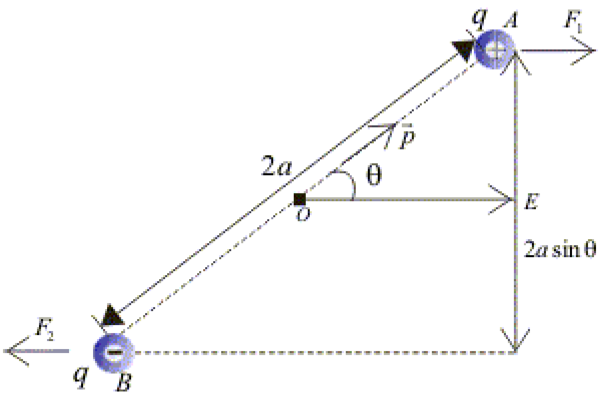


Figure a:- torque on dipole is τ=pEsinθ

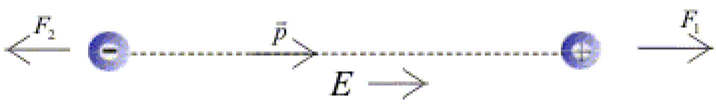


Figure b:- Dipole in equilibrium in uniform electric field.

- Force $\mathbf{F_1}$ of magnitude q**E**, acts on positive charge in direction of electric field and a force $\mathbf{F_2}$ of same magnitude acts on negative charge but it acts in direction opposite to F_1.

- Resultant force on dipole is zero, but since two forces do not have same line of action they constitute a couple.

- We now calculate torque (**r** × **F**) of these forces about zero.

  Torque of F_1 about O is
  $\boldsymbol{\tau_1} = \mathbf{OB} \times \mathbf{F_1}$

$= q(\mathbf{OB} \times \mathbf{E})$

Torque of F_2 about O is

$\boldsymbol{\tau}_2 = \mathbf{OA} \times \mathbf{F}_2$

$= -q(\mathbf{OA} \times \mathbf{E})$

$= q(\mathbf{AO} \times \mathbf{E})$

net torque acting on dipole is

$\boldsymbol{\tau} = \boldsymbol{\tau}_1 + \boldsymbol{\tau}_2$

$= q(\mathbf{OB} + \mathbf{AO}) \times \mathbf{E}$

$= q(\mathbf{AB} \times \mathbf{E})$

AB = 2a and $\mathbf{p} = 2q\mathbf{a}$ (dipole moment)

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

- Direction of torque is perpendicular to the plane containing dipole axis and electric field.
- Effect of torque is to rotate the dipole to a position in which dipole moment **p** is parallel to **E** the electric field vector is shown above in figure b and for uniform electric field dipole is in equilibrium in this position.
- magnitude of this torque is
  $\boldsymbol{\tau} = |\boldsymbol{\tau}| = pE\sin\theta$

Examples

Question 1

Two point charges q_1 and $q_2$are located with points having position vectors $\mathbf{r_1}$ and $\mathbf{r_2}$

(1) Find the position vector $\mathbf{r_3}$ where the third charge q_3 should be placed so that force acting on each of the three charges would be equal to zero.

(2) Find the amount of charge q_3

Question 2

Consider a thin wire ring of radius R and carrying uniform charge density λ per unit length.

(1) Find the magnitude of electric field strength on the axis of the ring as a function of distance x from its centre.

(2) What would be the form of electric field function for x>>R.

(3) Find the magnitude of maximum strength of electric field.

Question 3

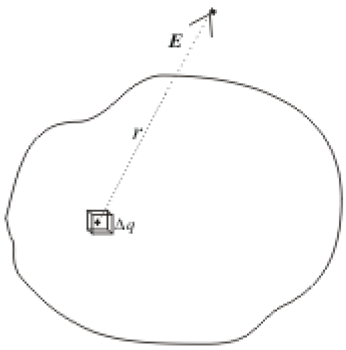
Two equally charged metal balls each of mass m Kg are suspended from the same point by two insulated threads of length l m long. At equilibrium, as a result of mutual separation between balls, balls are separated by x m. Determine the charge on each ball.

1. Introduction

- Gauss's law was suggested by Kark Fredrich Gauss(1777-1855) who was german scientist and mathematecian.
- Gauss's law is basically the relation between the charge distribution producing the electrostatic field to the behaviour of electrostatic field in space.
- Gauss's law is based on the fact that flux through any closed surface is a measure of total amount of charge inside that surface and any charge outside that surface would not contribute anything to the total flux.
- Before we look further to study Gauss's law in detail let's study electric field due to continous charge distributions.

2. Electric field due to continous charge distributions

- So far as in the previous chapter we have discussed force and field due to discrete charges.
- We now assume that charges on a surface are located very close togather so that such a system of charges can be assumed to have continous distribution of charges.
- In a system of closely spaced charges, total charge could be continously distributed along some line, over a surface or throughout a volume.
- First divide the continous charge distribution into small elements containing Δq amount of charge as shown in fig 1



- Electric field at point A due to element carrying charge Δq is

$$\Delta \boldsymbol{E} = \frac{1}{4\pi\varepsilon_0}\frac{\Delta q}{r^2}\hat{\boldsymbol{r}} \qquad (1)$$

where r is the distance of element under consideration from point A and **r**ˆis the unit vector in the direction from charge element towards point A.

- Total electric field at point A due to all such charge elements in charge distribution is

$$\boldsymbol{E} \cong \frac{1}{4\pi\varepsilon_0}\sum_i \frac{\Delta q_i}{r_i^2}\hat{r}_i \qquad (2)$$

where index i refers to the i_{th} charge element in the entire charge distribution.

- Since the charge is distributed continously over some region , the sum becomes integral. Hence total field at A within the limit Δq→0 is,

$$\boldsymbol{E} = \frac{1}{4\pi\varepsilon_0}\lim_{\Delta q_i \to 0}\sum_i \frac{\Delta q_i}{r_i^2}\hat{r}_i = \frac{1}{4\pi\varepsilon_0}\int \frac{dq}{r^2}\hat{\boldsymbol{r}} \qquad (3)$$

and integration is done over the entire charge distribution.

- If a charge q is uniformly distributed along a line of length L, the linear charge density λ is defined by

$$\lambda = \frac{q}{L} \qquad (4)$$

and the unit of λ is Coulumb/meter(C/m).

- For charge distributed non-uniformly over a line, linear charge density is

$$\lambda = \frac{dq}{dL} \qquad (5)$$

where dQ is the amount of charge in a small length element dL.

- For a charge Q uniformly distributed over a surfsce of area A, the surface charge density σ is

$$\sigma = \frac{q}{A} \qquad (6)$$

and unit of surface charge density is C/m^2. For non uniform charge distributed over a surface charge density is

$$\sigma = \frac{dq}{dA} \qquad (7)$$

where dA is a small area element of charge dQ.

- Similarly for uniform charge distributios volume charge density is

$$\rho = \frac{q}{V} \qquad (8)$$

and for non uniform distribution of charges

$$\rho = \frac{dq}{dV} \qquad (9)$$

and unit of volume charge distribution is C/m^3.

3. Gauss's Law

- We already know about electric field lines and electric flux. Electric flux through a closed surface S is

$$\phi = \int_S \boldsymbol{E} \cdot \boldsymbol{da} \qquad (10)$$

which is the number of field lines passing through surface S.

- **Statement of Gauss's Law**
“ELectric flux through any surface enclosing charge is equal to q/ε_0 , where q is the net charge enclosed by the surface”
mathematically,

$$\oint \boldsymbol{E} \cdot \boldsymbol{da} = \frac{q}{\varepsilon_0} \qquad (11)$$

where q_{enc} is the net charge enclosed by the surface and **E** is the total electric field at each point on the surface under consideration.

- It is the net charge enclosed in the surface that matters in Gauss's law but the total flux of electric field **E**depends also on the surface choosen not merely on the charge enclosed.
- So if you have information about distribution of electric charge inside the surface you can find electric flux through that surface using Gauss's Law.
- Again if you have information regarding electric fluxthrough any closed surface then total charge enclosed by that surface can also be easily calculated using Gauss's Law.
- Surface on which Gauss's Law is applied is known as Gaussian surface which need not be a real surface.
- Gaussian surface can be an imaginary geometrical surface which might be empty space or it could be partially or fully embedded in a solid body.

- Again consider equation 11

$$\oint E \cdot da = \frac{q}{\varepsilon_0}$$

In left hand side of above equation **E·da** is scalar product of two vectors namely electric field vector **E** and area vector **da**. Area vector **da** is defined as the vector of magnitude |da| whose direction is that of outward normal to area element da. So,**da=n**ˆda where **n**ˆ is unit vector along outward normal to da.

$$E \cdot da = |E||da|\cos\theta \qquad (12)$$

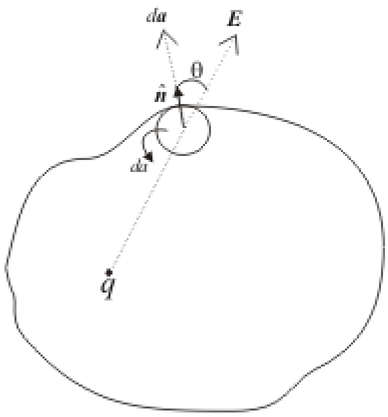


Figure 1

From above discussion we can conclude that,

(1) If both **E** and surface area da at each points are perpandicular to each other and has same magnitude at all points of the surface then vector **E** has same direction as that of area vector as shown below in the figure.

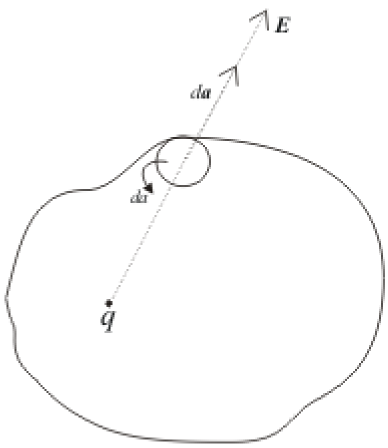


Figure 2

since **E** is perpandicular to the surface

$$\int E \cdot da = \int |E||da|\cos 0^0 = EA$$

(2) If E is parallel to the surface as shown below in the figure

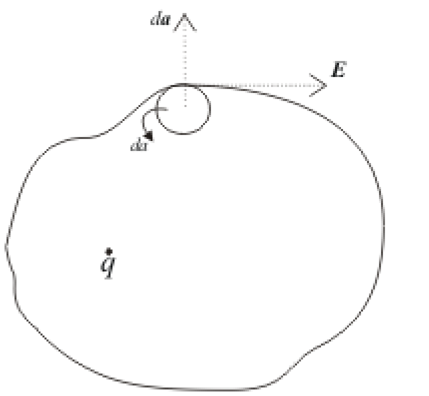


Figure 3

$\int \boldsymbol{E}\cdot \boldsymbol{da} = \int |\boldsymbol{E}||\boldsymbol{da}|\cos 90^0 = 0$

E at all points on the surface

4. Applications of Gauss's Law

(A) Derivation of Coulumb's Law

- Coulumb's law can be derived from Gauss's law.
- Consider electric field of a single isolated positive charge of magnitude q as shown below in the figure.

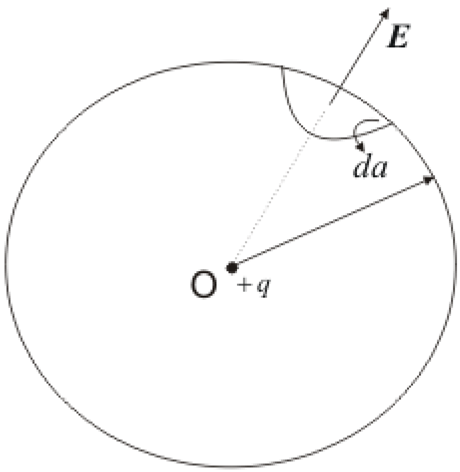


Figure 4

- Field of a positive charge is in radially outward direction everywhere and magnitude of electric field intensity is same for all points at a distance r from the charge.
- We can assume Gaussian surface to be a sphere of radius r enclosing the charge q.
- From Gauss's law

  $\oint \boldsymbol{E}\cdot \boldsymbol{da} = E\oint da = \frac{q_{enc}}{\varepsilon_0}$

  since **E** is constant at all points on the surface therefore,

$$EA = \frac{q}{\varepsilon_0}$$

$or,$

$$E = \frac{q}{\varepsilon_0 A}$$

surface area of the sphere is A=4πr^2
thus,

$$E = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}$$

- Now force acting on point charge q' at distance r from point charge q is

$$F = q'E$$

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

This is nothing but the mathematical statement of Coulumb's law.

(B) Electric field due to line charge

- Consider a long thin uniformly charged wire and we have to find the electric field intensity due to the wire at any point at perpandicular distance from the wire.
- If the wire is very long and we are at point far away from both its ends then field lines outside the wire are radial and would lie on a plane perpandicular to the wire.
- Electric field intensity have same magnitude at all points which are at same distance from the line charge.
- We can assume Gaussian surface to be a right circular cylinder of radius r and length l with its ends perpandicular to the wire as shown below in the figure.

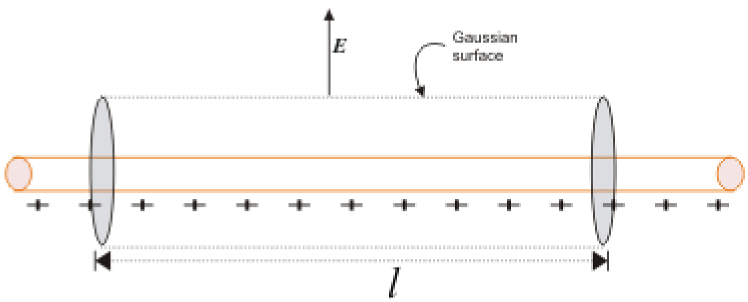


Figure 5. Cylindrical Gaussian surface for calculation of electric field due to line charge

- λ is the charge per unit length on the wire. Direction of **E** is perpandicular to the wire and components of **E** normal to end faces of cylinder makes no contribution to electric flux. Thus from Gauss's law

$$\oint \mathbf{E}\cdot d\mathbf{a} = \frac{q_{enc}}{\varepsilon_0}$$

- Now consider left hand side of Gauss's law

$$\oint \mathbf{E}\cdot d\mathbf{a} = E\oint da$$

Since at all points on the curved surface **E** is constant. Surface area of cylinder of radius r and length l is A=2πrl therefore,

$$\oint \mathbf{E}\cdot d\mathbf{a} = E(2\pi rl)$$

- Charge enclosed in cylinder is q=linear charge density x length l of cylinder,
  or, q=λl
  From Gauss's law

  $$\oint \mathbf{E}\cdot d\mathbf{a} = \frac{q}{\varepsilon_0}$$

  $$or,\ E(2\pi rl) = \frac{\lambda l}{\varepsilon_0}$$

  $$\Rightarrow E = \frac{\lambda}{2\pi r\varepsilon_0}$$

  $$\Rightarrow E \propto \frac{\lambda}{r}$$

  Thus electric field intensity of a long positively charged wire does not depends on length of the wire but on the radial distance r of points from the wire.

(C) Electric field due to charged solid sphere

- We'll now apply Gauss's law to find the field outside uniformly charged solid sphere of radius R and total charge q.
- In this case Gaussian surface would be a sphere of radius r>R concentric with the charged solid sphere shown below in the figure.From Gauss's law

  $$\oint \mathbf{E}\cdot d\mathbf{a} = \frac{q_{enc}}{\varepsilon_0}$$

  where q is the charge enclosed.
- Charge is distributed uniformly over the surface of the sphere. Symmetry allows us to extract **E** out of the integral sign as magnitude of electric field intensity is same for all points at distance r>R.
- Since electric field points radially outwards we have

  $$\oint \mathbf{E}\cdot d\mathbf{a} = E\oint da$$

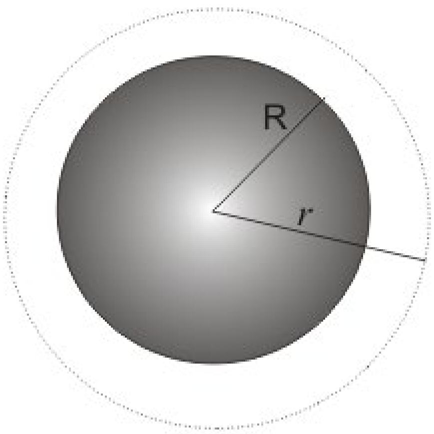


Figure 6

also as discussed magnitude of **E** is constant over Gaussian surface so,

$$E\oint da = E(4\pi r^2)$$

where $4\pi r^2$ is the surface area of the sphere.
Again from Gauss's law we have

$$E(4\pi r^2) = \frac{q}{\varepsilon_0}$$

$$\Rightarrow E = \frac{q}{4\pi\varepsilon_0 r^2}$$

THus we see that magnitude of field outside the sphere is exactly the same as it would have been as if all the charge were concentrated at its center.

(D) Electric field due to an infinite plane sheet of charge

- Consider a thin infinite plane sheet of charge having surface charge density σ(charge per unit area).
- We have to find the electric field intensity due to this sheet at ant point which is distance r away from the sheet.
- We can draw a rectangular gaussian pillbox extending equal distance above and below the plane as shown below in the figure.

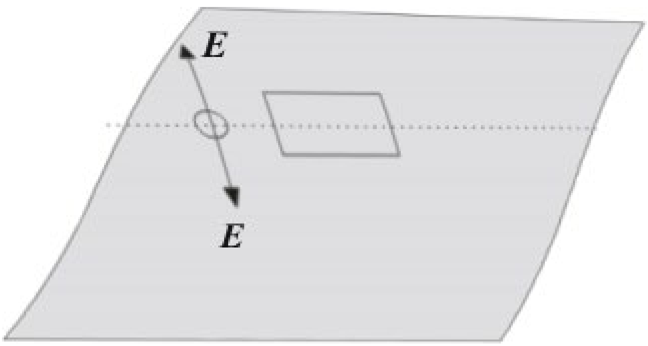


Figure 7

- By symmetry we find that **E**on either side of sheet must be perpandicular to the plane of the sheet, having same magnitude at all points equidistant from the sheet.
- No field lines crosses the side walls of the Gaussian pillbox i.e., component of **E** normal to walls of pillbox is zero.
- We now apply Gauss's law to this surface

  $$\oint \mathbf{E} \cdot d\mathbf{a} = \frac{q_{enc}}{\varepsilon_0}$$

  in this case charge enclosed is
  q=σA
  where A is the area of end face of Gaussian pillbox.
- **E** points in the direction away from the plne i.e., **E** points upwards for points above the plane and downwards for points below the plane. Thus for top and bottom surfaces,

  $$\oint \mathbf{E} \cdot d\mathbf{a} = 2A\,|\,\mathbf{E}\,|$$

  thus
  $2A|\mathbf{E}|=\sigma A/\varepsilon_0$

or,
$|\mathbf{E}|=\sigma/2\varepsilon_0$
Here one important thing to note is that magnitude of electric fied at any point is independent of the sheet and does not decrease inversly with the square of the distance. Thus electric field due to an infinite plane sheet of charge does not falls of at all.

1. Introduction

- We already have an introduction of work and energy while studying mechanics.
- We know that central forces are conservative in nature i.e., work done on any particle moving under the influence of consrevative forces does not depend on path taken by the particle but depends on initial and final positions of the particle.
- Electrostatic force given by Coulumb's law is also a central force like gravitational force and is conservative in nature.
- For conservative forces, work done on particle undergoing displacement can be expressed in terms of potential energy function.
- In this chapter we will apply work and energy considerations to the electric field and would develop the concept of electric potential energy and electric potential.

2. Electric potential energy

- Conside a system of two point charges in which positive test charge q' moves in the field produced by stationary point charge q shown below in the figure.

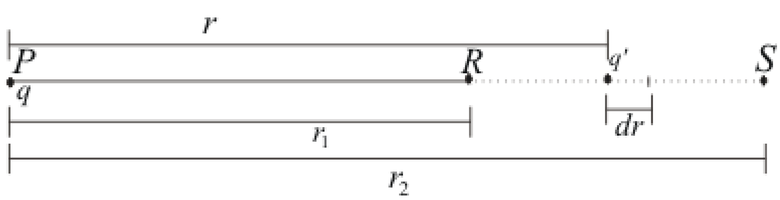


Figure 1

- Charge q is fixed at point P and is displaced from point R to S along a radial line PRS shown in the figure.
- Let r_1 be the distance between points P and R and r_2 be the distance between P and S.
- Magnitude of force on positive test charge as given by Coulumb's law is

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

- If q' moves towards S through a small displacement d**r** then work done by this force in making the small displacement d**r** is
  dW=**F**·d**r**

$$dW = \frac{1}{4\pi\varepsilon_0}\frac{qq'}{r^2}dr \qquad (2)$$

- Total work done by this force as test charge moves from point R to S i.e., from r_1 to r_2 is,

$$W = \int_{r_1}^{r_2} F dr = \int_{r_1}^{r_2} \frac{1}{4\pi\varepsilon_0}\frac{qq'}{r^2}dr$$

or

$$W = \frac{qq'}{4\pi\varepsilon_0}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad (3)$$

- Thus for this particular path work done on test charge q' depends on end points not on the path taken.
- Work done W in moving the test charge q' from point R to S is equal to the change in potential energy in moving the test charge q' from point R to S. Thus,
  $W=U(r_1)-U(r_2)$ (4)
  where

  $$U(r_1) = \frac{qq'}{4\pi\varepsilon_0 r_1}$$

  is the potential energy of test charge q' when it is at point R and

  $$U(r_2) = \frac{qq'}{4\pi\varepsilon_0 r_2}$$

  is the potential energy of test charge q' when it is at point S.
- Thus potential energy of test charge q' at any distance r from charge q is given by

$$U = \frac{qq'}{4\pi\varepsilon_0 r} \qquad (5)$$

  Equation 5 gives the electric potential energy of a pair of charges which depends on the separation between the charges not on the location of charged particles.
- If we bring the test charge q' from a very large distance such that $r_2=\infty$ to some distance $r_1=r$ then we must do work against electric forces which is equal to increase in potential energy as given by equation 5.
- Thus potential energy of a test charge at any point in the electric field is the work done against the electric forces to bring the charge from infinity to point under consideration.

3. Electric Potential

- We now move towards the electric potential which is potential energy per unit charge.
- Thus electrostatic potential at any point of an electric field is defined as potential energy per unit charge at that point.
- Electric potential is represented by letter V.
  V=U/q' or U=q'V (6)
- Electric potential is a scalar quantity since both charge and potential energy are scalar quantities.
- S.I. unit of electric potential is Volt which is equal to Joule per Coulumb. Thus,
  1 Volt = 1 JC^{-1}

- In equation 4 if we divide both sides by q' we have

$$\frac{W}{q'} = \frac{U(r_1)}{q'} - \frac{U(r_2)}{q'} = V(r_1) - V(r_2) \qquad (7)$$

where V(r_1) is the potential energy per unit charge at point R and V<r_2) is potential energy per unit charge at point S and are known as potential at points R and S respectively.

</r

- Again consider figure 1. If point S in figure 1 would be at infinity then from equation 7

$$V(r_1) - V(\infty) = \frac{W}{q'}$$

Since potential energy at infinity is zero therefore V(∞)=0. Therefore

$$V(r_1) = \frac{W}{q'}$$

hence electric potential at a point in an electric field is the ratio of work done in bringing test charge from infinity to that point to the magnitude of test charge.

- Dimensions of electric potential are $[ML^2T^{-3}A^{-1}]$ and can be calculated easily using the concepts of dimension analysis.

4. Electric potential due to a point charge

- Consider a positive test charge +q is placed at point O shown below in the figure.

Figure 2

- We have to find the electric potential at point P at a distance r from point O.
- If we move a positive test charge q' from infinity to point P then change in electric potential energy would be

$$U_P - U_\infty = \frac{qq'}{4\pi\varepsilon_0 r}$$

- Electric potential at point P is

$$V_P = \frac{U_P - U_\infty}{q'} = \frac{q}{4\pi\varepsilon_0 r} \qquad (8)$$

- Potential V at any point due to arbitrary collection of point charges is given by

$$V = \frac{1}{4\pi\varepsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_i} \qquad (9)$$

- here we see that like electric field potential at any point independent of test charge used to define it.
- For continous charge distributions summation in above expressin will be replaced by the integration

$$V = \frac{1}{4\pi\varepsilon_0}\int\frac{dq}{r} \qquad (10)$$

where dq is the differential element of charge distribution and r is its distance from the point at which V is to be calculate

5. Relation between electric fiels and electric potential

- Consider the electric field E due to a point charge +q at point O in a radially outward direction shown below in the figure.

Figure 3

- Suppose R and S are two points at a distance r and r+dr from point O where dr is vanishingly small distance and V is electric potential at point R.
- Now force on any test charge q' at point R in terms of electric field is
  $\mathbf{F}=q'\mathbf{E}$
- Work done by the force in displacing test charge from R to S in field of charge q is
  $dW = \mathbf{F}\cdot d\mathbf{r} = q'\mathbf{E}\cdot d\mathbf{r}$
  and, change in potential energy is
  $dU = -dW = -q'\mathbf{E}\cdot d\mathbf{r}$
  Change in electric potential would be
  $dV = dU/q$
  or $dV = -\mathbf{E}\cdot d\mathbf{r}$ (11)
- From equation 11 electric field is
  $E=-(dV/dr)$ (12)
  the quantity dV/dr is the rate of change of potential with the distance and is known as potential gradient. Negative sign in equation 12 indicates the decrease in electric potential in the direction of electric field.
- For cartesian coordinate system
  $E=E_xi+E_yj+E_zk$
  and,
  $dr=dxi+dyj+dzk$
  from equation 11
  $dV=-E\cdot dr$
  or, $dV=-(E_xdx+E_ydy+E_zdz)$ (13)

- Thus components of E are related to corresponding derivatives of V in the following manner
  $E_x=dV/dx$ (14a)
  $E_y=dV/dy$ (14b)
  $E_z=dV/dz$ (14c)
  In equation (14a) we see that V is differentiated with respect to coordinate x keeping other coordinates constant. Same is the case with equations (14b) and (14c) in case of y and z coordinates respectively.

6. Equipotential surfaces

- Surface over which the electric potential is same everywhere is called an equipotential surface.
- Equipotential surfaces are graphical way to represent potential distribution in an electric field.
- We can draw equipotential surfaces through a space having electric field.
- For a positive charge , electric field would be in radially outward direction and the equipotential surfaces would be concentric spheres with centers at the charge as shown below in the figure.

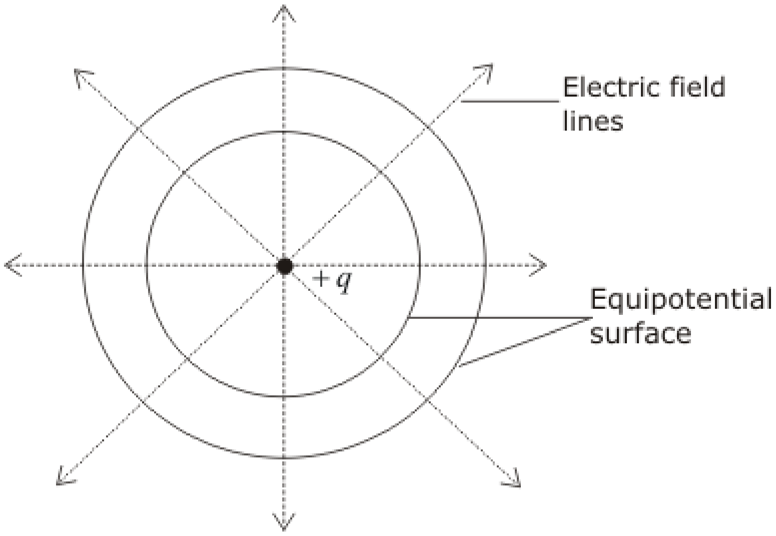


Figure 4

- Since electric potential remains same everywhere on an equipotential surface from this it follows that PE of a charged body is same at all points on this surface.This shows that work done in moving a charged body between two points on an equipotential surface would be zero.
- At every point on equipotential surface electric field lines are perpandicular to the surface. This is because potential gradient along any direction parallel to the surface is zero i.e., $E=-dV/dr=0$
  so component electric parallel to equipotential surface is zero

7. Potential due to an electric dipole

- We already know that electric dipole is an arrangement which consists of two equal and opposite charges +q and -q separated by a small distance 2a.
- Electric dipole moment is represented by a vector **p** of magnitude 2qa and this vector points in direction from -q to +q.
- To find electric potential due to a dipole consider charge -q is placed at point P and charge +q is placed at point Q as shown below in the figure.

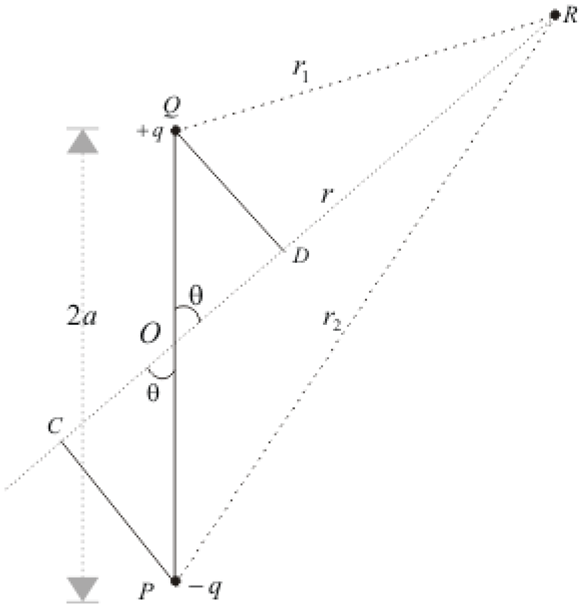


Figure 5

- Since electric potential obeys superposition principle so potential due to electric dipole as a whole would be sum of potential due to both the charges +q and -q. Thus

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r_1} - \frac{q}{r_2}\right) \qquad (15)$$

where r_1 and r_2 respectively are distance of charge +q and -q from point R.

- Now draw line PC perpandicular to RO and line QD perpandicular to RO as shown in figure.
From triangle POC
cosθ=OC/OP = OC/a
therefore OC=acosθ similarly OD=acosθ
Now ,
r_1 = QR≅RD = OR-OD = r-acosθ
r_2 = PR≅RC = OR+OC = r+acosθ

$$V = \frac{q}{4\pi\varepsilon_0}\left(\frac{1}{r - a\cos\theta} - \frac{1}{r + a\cos\theta}\right) = \frac{q}{4\pi\varepsilon_0}\left(\frac{2a\cos\theta}{r^2 - a^2\cos^2\theta}\right)$$

since magnitude of dipole is
|**p**| = 2qa

$$V = \frac{1}{4\pi\varepsilon_0}\left(\frac{p\cos\theta}{r^2 - a^2\cos^2\theta}\right) \qquad (16)$$

- If we consider the case where r>>a then

$$V = \frac{p\cos\theta}{4\pi\varepsilon_0 r^2} \qquad (17)$$

again since pcosθ= **p**·**r**ˆ where, **r**ˆ is the unit vector along the vector OR then electric potential of dipole is

$$V = \frac{p \cdot \hat{r}}{4\pi\varepsilon_0 r^2} \qquad (18)$$

for r>>a

- From above equation we can see that potential due to electric dipole is inversly proportional to r^2 not ad 1/r which is the case for potential due to single charge.
- Potential due to electric dipole does not only depends on r but also depends on angle between position vector **r** and dipole moment **p**.

8. Work done in rotating an electric dipole in an electric field

- Consider a dipole placed in a uniform electric field and it is in equilibrium position. If we rotate this dipole from its equllibrium position , work has to be done.
- Suppose electric dipole of moment **p** is rotated in uniform electric field **E** through an angle θ from its equilibrium position. Due to this rotation couple acting on dipole changes.
- If at any instant dipole makes an angle φ with uniform electric field then torque acting on dipole is
  Γ=pEsinφ (19)
  again work done in rotating this dipole through an infitesimaly small angle dφ is
  dW=torque x angular displacement
  =pEsinφdφ
- Total work done in rotating the dipole through an angle θfrom its equilibrium position is

$$W = \int_0^\theta pE\sin\varphi d\varphi = pE[-\cos\varphi]_0^\theta = pE(1-\cos\theta) \qquad (21)$$

This is the required formula for work done in rotating an electric dipole placed in uniform electric field through an angle θ from its equilibrium position.

9. Potential energy of dipole placed in uniform electric field

- Again consider equation 20 which gives the work done in rotating electric dipole through an infinetesimly small angle dφ is
  dW=pEsinφdφ
  which is equal to the change in potential energy of the system
  dW=dU=pEsinφdφ (22)
- If angle dφ is changed from 90^0 to θ then in potential energy would be

$$W = \int_0^\theta pE\sin\varphi d\varphi = pE(-\cos\varphi)_0^\theta = pE(1-\cos\theta)$$

- We have choosen the value of φgoing from π/2 to θ because at π/2 we can take potential energy to be zero (axis of dipole is perpandicular to the field). Thus $U(90^0)=0$ and above equation becomes

$$U(\theta) - U(90^0) = \int_{90^0}^\theta pE\sin\varphi d\varphi = pE(-\cos\varphi)_{90}^\theta = -pE\cos\theta = -\boldsymbol{p}\cdot\boldsymbol{E}$$

1. Introduction

- A capacitor (formerly known as condenser) is a device that can store electronic charge and energy.
- All capacitors consists of a combination of two conductors separated by an insulator.
- The insulator is called dielectric which could be oil, air or paper and many more such materials are there wich can act as a dielectric medium between conducting plates of a capacitor.
- Figure 1 below shows the symbol used to represent a capacitor.

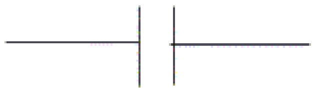


Figure 1

- Now plates of the capacitor are connected to the terminals of a battery, shown below in figure 2, in order to charge it's conducting plates.

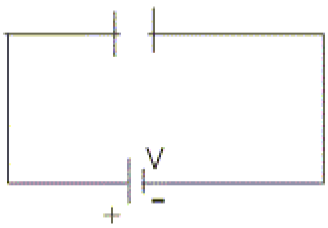


Figure 2

- As soon as capacitor is connected to the battery , charge is transferred from one conductor to another.
- Plate connected to positive terminal of the battery becomes positively charged with charge +Q in it and plate connected to negative terminal of the battery becomes negatively charged with cahrge -Q on it i.e. both plates have equal amount of opposite charge .
- Once the capacitor is fully charged potential difference between the conductors due to their equal and opposite charges becomes equal to the potential difference between the battery terminals.
- For a given capacitor $Q \propto V$ and the ratio Q/V is constant for a capacitor.
  Thus,
  $$Q=CV \qquad (1)$$
  where the proportionality constant C is called the capacitance of the capacitor.
- Capacitance of any capacitor depends on shape , size and geometrical arrangement of the conductors.
- When Q is in coulumbs (C) and V is in volts(V) then the S.I. unit of capacitance is in farads(F) where
  1F=1 coulumb/volt
- One farad is the capacitance of very large capacitor and it's submultiples such as microfarad($1\mu F=10^{-6}$) or picofarad($1pF=10^{-12}$) are generally used for practical applications.

2. Calculation of capacitance

- For calculating capacitance of a capacitor first we need to find the potential difference between it's two conducting plates having charge +Q and -Q.
- For simple arrangements of conductors like two equivalent parallel plates kept at distance d apart or two concentric conducting spheres etc., potential difference can be found first by calculating electric field from gauss's law or by Coulumb's law.
- After calculating electric field , potential difference can be found by integrating ellectric field using the relation

  $V_a - V_b = \int E.dr$

  where the limits of integration goes from a to b.
- Once we know the potential difference between two conductors of the capacitor , it's capacitance can be calculated from the relation

  $C=Q/V$ (2)
- Calculation of capacitance of some simple arrangements would be illustrated in following few articles.

3. Parallel plate capacitor

- A parallel plate capacitor consists of two large plane parallel conducting plates separated by a small distance shown below in the figure 3.

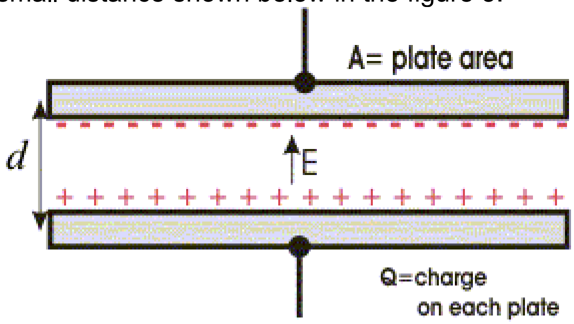


Figure 3

- Suppose two plates of the capacitor has equal and opposite charge Q on them. If A is the area of each plate then surface charge density on each plate is

  $\sigma=Q/A$
- We have already calculated field between two oppositely charged plates using gauss's law which is

  $E=\sigma/\varepsilon_0=Q/\varepsilon_0 A$

  and in this result effects near the edges of the plates have been neglected.
- Since electric field between the plates is uniform the potential difference between the plates is

  $V=Ed=Qd/\varepsilon_0 A$

  where , d is the separation between the plates.
- Thus, capacitance of parallel plate capacitor in vacuum is

  $C=Q/V=\varepsilon_0 A/d$ (3)
- From equation 3 we see that quantities on which capacitance of parallel plate capacitor depends i.e.,ε_0 , A and d are all constants for a capacitor.
- Thus we see that in this case capacitance is independent of charge on the capacitor but depends on area of it's plates and separation distance between the plates.

4.Cylinderical capacitor

- A cylinderical capacitor is made up of a conducting cylinder or wire of radius a surrounded by another concentric cylinderical shel of radius b (b>a).

- Let L be the length of both the cylinders and charge on inner cylender is +Q and charge on outer cylinder is -Q.
- For calculate electric field between the conductors using Gauss's law consider a gaussian consider a gaussian surface of radius r and length L^1 as shown in figure 4.

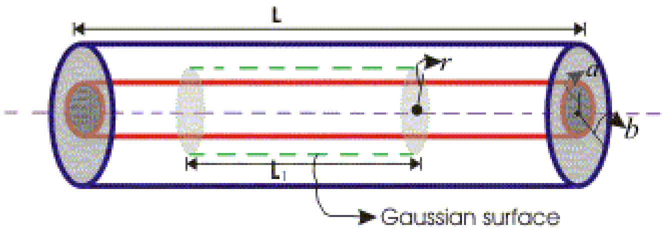


Figure 4

- According to Gauss's law flux through this surface is q/ε_0 where q is net charge inside this surface.
- We know that electric flux is given by

$$\varphi = E.A$$
$$= EA\cos\theta$$
$$= EA$$

since electric field is constant in magnitude on the gaussian surface and is perpandicular to this surface. Thus,

$$\varphi = E(2\pi rL)$$

since $\varphi = q/\varepsilon_0$

=> $E(2\pi rL) = (\lambda L)/\varepsilon_0$

where λ = Q/L = charge per unit length

=>

$E = \frac{\lambda}{2\pi\varepsilon_0 r}$

- or,

$E = \frac{\lambda}{2\pi\varepsilon_0 r}$

- (4)
- If potential at inner cylinder is V_a and V_b is potential of outer cylinder then potential difference between both the cylinders is

$$V = V_a \text{ and } V_b = \int E dr$$

where limits of integration goes from a to b.

- Potential of inner conductor is greater then that of outer conductor because inner cylinder carries positive charge. Thus potential difference is

$V = \frac{Q\ln(b/a)}{2\pi\varepsilon_0 L}$

- Thus capacitance of cylinderical capacitor is

C=Q/V

or,

$$C = \frac{2\pi\varepsilon_0 L\, Q ln(b/a)}{ln(b/a)} \qquad (5)$$

- From equation 5 it can easily be concluded that capacitance of a cylinderical capacitor depends on length of cylinders.
- More is the length of cylinders , more charge could be stored on the capacitor for a given potential difference.

5. Spherical capacitor

- A spherical capacitor consists of a solid or hollow spherical conductor of radius a , surrounded by another hollow concentric spherical of radius b shown below in figure 5

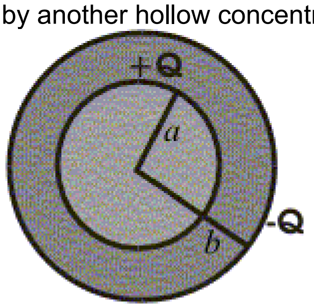


Figure 5

- Let +Q be the charge given to the inner sphere and -Q be the charge given to the outer sphere.
- The field at any point between conductors is same as that of point charge Q at the origin and charge on outer shell does not contribute to the field inside it.
- Thus electric field between conductors is

$$E = \frac{Q}{4\pi\varepsilon_0 r^2}$$

- Potential difference between two conductors is

$V=V_a-V_b$

$=-\int E.dr$

where limits of integration goes from a to b.

On integrating we get potential difference between to conductors as

$$V = \frac{Q(b-a)}{4\pi\varepsilon_0 ab}$$

- Now , capacitance of spherical conductor is

$C=Q/V$

or,

$$C = \frac{4\pi\varepsilon_0 ab}{(b-a)} \qquad (6)$$

- again if radius of outer conductor aproaches to infinity then from equation 6 we have

$C=4\pi\varepsilon_0 a$ (7)

- Equation 7 gives the capacitance of single isolated sphere of radius a.
- Thus capacitance of isolated spherical conductor is proportional to its radius.

6. Capacitors in series and parallel combinations

For prectical applications , two or more capacitors are often used in combination and their total capacitance C must be known.To find total capacitance of the arrangement of capacitor we would use equation

$$Q=CV$$

(i) Parallel combination of capacitors

- Figure below shows two capacitors connected in parallel between two points A and B

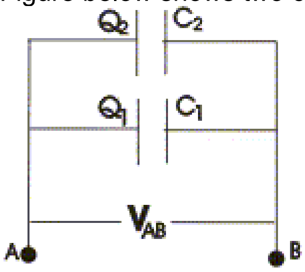


Figure 6

- Right hand side plate of capacitors would be at same common potential V_A. Similarly left hand side plates of capacitors would also be at same common potential V_B.
- Thus in this case potential difference $V_{AB}=V_A-V_B$ would be same for both the capacitors, and charges Q_1 and Q_2 on both the capacitors are not necessarily equal. So,

$$Q_1=C_1V \text{ and } Q_2=C_2V$$

- Thus charge stored is divided amongst both the capacitors in direct proportion to their capacitance.
- Total charge on both the capacitors is,

$$Q=Q_1+Q_2$$
$$=V(C_1+C_2)$$

and

$$Q/V=C_1+C_2 \qquad (8)$$

So system is equivalent to a single capacitor of capacitance

$$C=Q/V$$

where,

- When capacitors are connected in parallel their resultant capacitance C is the sum of their individual capacitances.
- The value of equivalent capacitance of system is greater then the greatest individual one.
- If there are number of capacitors connected in parallel then their equivalent capacitance would be

$$C=C_1+C_2+ C_3.......... \qquad (10)$$

(ii) Series combination of capacitors

- Figure 7 below shows two capacitors connected in series combination between points A and B.

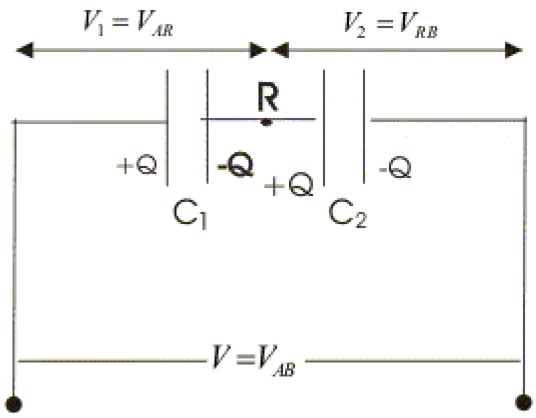


Figure 7

- Both the points A and B are maintained at constant potential difference V_{AB}.
- In series combination of capacitors right hand plate of first capacitor is connected to left hand plate of next capacitor and combination may be extended foe any number of capacitors.
- In series combination of capacitors all the capacitors would have same charge.
- Now potential difference across individual capacitors are given by

  $V_{AR}=Q/C_1$

  and,

  $V_{RB}=Q/C_2$
- Sum of V_{AR} and V_{RB} would be equal to applied potential difference V so,

  $V=V_{AB}=V_{AR}+V_{RB}$

  $=Q(1/C_1 + 1/C_2)$

  or,

  $$\frac{V}{Q}=\frac{1}{C_1}+\frac{1}{C_2}=\frac{Q}{C}$$

  where

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

  i.e., resultant capacitance of series combination C=Q/V, is the ratio of charge to total potential difference across the two capacitors connected in series.
- So, from equation 12 we say that to find resultant capacitance of capacitors connected in series, we need to add reciprocals of their individual capacitances and C is always less then the smallest individual capacitance.
- Result in equation 12 can be summarized for any number of capacitors i.e.,

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

7. Energy stored in a capacitor

- Consider a capacitor of capacitance C, completely uncharged in the begning.
- Charhing process of capacitor requires expanditure of energy because while charging a capacitor charge is transferred from plate at lower potential to plate at higher potential.

- Now if we start charging capacitor by transporting a charge dQ from negative plate ti the positive plate then work is done against the potential difference across the plate.
- If q is the amount of charge on the capacitor at any stage of charging process and φ is the potential difference across the plates of capacitor then magnitude of potential difference is $\varphi=q/C$.
- Now work dW required to transfer dq is
  $dW=\varphi dq=qdq/C$
- To charge the capacitor starting from the uncharged state to some final charge Q work required is
  Integrating from 0 to Q
  $W=(1/C)\int qdq$
  $=(Q^2)/2C$ (14a)
  $=(CV^2)/2$
  $=QV/2$

  Which is the energy stored in the capacitor and can also be written as
  $U=(CV^2)/2$ ---(15)

- From equation 14c,we see that the total work done is equal to the average potential V/2 during the charging process ,multiplied by the total charge transferred
- If C is measured in Farads ,Q in coulumbs and V in volts the energy stored would in Joules
- A parallel plate capacitor of area A and seperation d has capacitance

  $C=\varepsilon_0 A/d$

- electric field in the space between the plates is
  $E=V/d$ or $V=Ed$

  Putting above values of V and C in equation 14b we find
  $W=U=(1/2)(\varepsilon_0 A/d)(Ed)^2$
  $=(1/2)\varepsilon_0 E^2(Ad)$
  $=(1/2)\varepsilon_0 E^2.V$ ---(16)

- If u denotes the energy per unit volume or energy density then
  $u=(1/2)\varepsilon_0 E^2$ x volume
- The result for above equation is generally valid even for electrostatic field that is not constant in space.

8. Effect of Dielectric

- Dielectric are non conducting materials for ex- Glass,mica,wood etc.
- What happened when space between the two plates of the capacitor is filled by a dielectric was first discovered by faraday.
- Faraday discovered that if the space between conductors of the capacitor is occupied by the dielectric,the capacitance of capacitor is increased.
- If the dielectric completely fills the space between the conductors of the capacitor ,the capacitance is increased by an factor K which is characterstics of the dielectric and This factor is known as the dielectric constan.

- Dielectric constant of vaccum is unity.
- Consider a capacitor of capacitance C_0 is being charged by the connecting it to a battery.
- If Q_0 is the amount of charged on the capacitor at the end of the charging and V_0 is potental diffrence across the plates of the capacitor then
  $C_0=Q_0 /V_0$ ----(17)

  Thus charge being placed on the capacitor is
  $Q_0=C_0V_0$

- If the battery is diconnected and space between the capacitor is filled by a dielectric the P.D decrease to a new value
  $V=V_0/K$.
- Since the original charge is still on the capacitor,the new capacitance will be
  $C=Q_0/V=KQ_0/V_0=KC_0$----(19)
- From equation 19 it follows that C is greater then C_0.
- Again if the dielectric is inserted while the battery is still connected then battery would have to supply some amount of charge to maintain the P.D between the plates and then total charge on the plates would be $Q=KQ_0$.
- In either of the cases ,capacitance of the capacitor is increase by the amount K.
- For a parallel plate capacitor with dielectric of dielectric constant K between its plates its capacitance becomes

  $C=\varepsilon A/D$ ----(20)
  where $\varepsilon=K\varepsilon_0$

- When a sufficiently strong electric field is applied to any dielectric material it becomes a conductor and this phenomenon is known as dielectric breakdown.
- The maximun electric field a material can withstand without the occurence of breakdown is called dielectric strength of that material.
- Thus field across the capacitor should never exceed breakdown limits in order to store charge on capacitor without leaking.

UNIT 2nd

CURRENT ELECTRICITY

(1) Introduction

- In our previous few chapters of electrostatics,we have discussed various terms and characterstics related to charge at rest
- Now in this chapter we will study about the moving charges,phenomenom related to them and various effects related to charge in motion
- Consider two metallic conducting balls charged at different potential are hanged using a non conducting insulating wires .Since air is an insulator ,no charge transfer takes place
- Now if we join both the metallic wire using a conducting metallic wire then charge will flow from metallic ball at higher potential to the one at lower potential.
- This flow of charge will stop when the two balls would be at the same potentials.
- If somehow we could maintian the potential between the metallic balls,we will get constant flow of the charge in metallic wire,connecting the two conducting balls
- This flow of charge in metallic wire due to the potential dicference between two conducters used is called electric current about which we would be dicussing in this chapter.

(2)Electric current and Current density

Electric Current

- We already had a brief idea about the electric current which wed defined as the state of motion of the electric charge .Now we are going to study about the electric current in details
- Quantitatively electric current is defined as the time rate of flow of the net charge of the area of crosssection of the conducter i.e
  Electric current = Total charge flowing / time taken
- if q is the amount of charge flowing through the conducter in t sec,The current through the conducter is given by
  $I=q/t$ (1)
- SI unit of the current is Ampere(A) named so in the honour of french scientist Andee marie Ampere(1775-1836).Now,
  1 Ampere= 1 Coulumb/ 1 sec=1 Cs^{-1}
- Thus current through any conducter is said to be 1 ampere,if 1 C of charge is flowing through the conducter in 1 sec
- Small amount of currents are accordingly expressed in milliamperes (1mA=10^{-3} A) or in micro ampere (1 mA=10^{-6} A)
- Direction of electric current is in the direction of the flow of positive charged carriers and this current is known as conventional current.
- Direction of the flow of electron in conductor gives the direction of electronic current. Direction of conventional current is opposite to that of electronic current
- Electric current is a scalar quantity .Although electric current represent the direction of the flow of positive charged carrier in the conductor,still current is treated as scalar quantity as current in wires in a circuit does not follows the laws of vector addition

Current density

- The current density at a point in the conductor is defined as the current per unit cross-section area.Thus if the charge is flowing per unit time uniformaly over the area of crosss-section A of the conductor,then current density J at any point on that area is defined as
  J=I/A -(2)
- It is the characterstic property of point inside the conductor nor of the conductor as a whole
- Direction of current density is same as the direction of conventional current
- Note that current density is a vector quantity unlike electric current
- Unit of current density is Ampere/meter^2 (Am^{-2})

(3) Drift Velocity

- Metallic conductors have large numbers of elctrons free to move about.These elctrons which are free to move are called conduction electrons
- Thus valence electrons of atom becomes the conduction electrons of the metals
- At room temperature,these conduction electrons moves randomly inside the conductor more or less like a gas molecule
- During motion,these conduction electrons collide with ions(remaining positive charged atom after the valence electrons move away) again and again and there direction of motion changes after each and every collision.
- As a results of these collisions atoms moves in a zig-zag path
- Since in a conductor there are large number of elctrons moving randomly inside the conductor.Hence they have not net motion in any particular direction.Since the number of electrons crossing an imaginary area ΔA from left to right inside the conductor very nearly equals the number of electron crossing the same area element from right to left in a given interval of time leaving flow of electric current through that area nearly equals to zero
- Now when we applied some P.D using a battery across the two ends of the conductor,then an electric field sets up inside the conductor
- As a result of this electric field setup inside the conductor,conduction electron expeirence a force in direction opposite to electric field and this force accelerates the motions of the electrons
- As a result of this accelerated motion electrons drifts slowly along the length of the conductor towards the end at higher potential
- Due to this acceleration velocity of electron's increases only for short interval of time as each acclerated electrons suffers frequent collision with positive ions and looses their Kinectic energy
- After each collision electrons starts fresh in random direction ,again get accelerated and loose their gained Kinetic energy in another collision
- This extra velocity gained by the electrons is lost in subsequent collison and the processes continued till the electron reach positive end of the conductor
- Under the effect of electric field inside the conductor ,free electrons have random thermal velocities due to the room tmeperature and small velocities with which they drift towards the positive end of the conductor.
- if τ is the average time between two successive collisions and E is the strength of applied electric field then force on electron due to applied electric field is
  F=eE
  Where e is the amount of charge on electron
- if m is the mass of electron ,then acceleration produced is given by
  a=eE/m
- Since electron is acclerated for an average time interval τ,additional velocity acquired by the electron is
  $v_d=a\tau$
  or $v_d=(eE/m)\tau$ (3)

- This small velocity imposed on the random motion of electrons in a conductor on the application of electric field is known as drift velocity
- This drift velocity is defined as the velocity with which free electrons gets drifted towards the positive end of the conductor under the influence of externally applied electric field

(4) Relation between drift velocity and electric current

- Consider a conducting wire of lenghth L and having uniform cross-section area A in which electric field is present

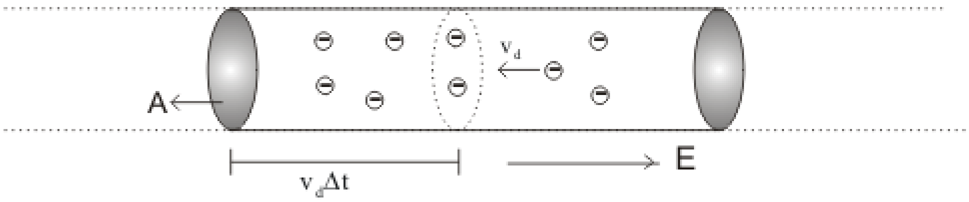


Figure 1:- Electrons moving opposite to electric field in a conductor

- Consider in the wire that there are n free electrons per unit volume moving with the drift velocity v_d
- In the time interval Δt each electron advances by a distance $v_d\Delta t$ and volume of this portion is $Av_d\Delta t$ and no of free electron in this portion is $nAv_d\Delta t$ and all these electrons crosses the area A in time Δt
- Hence charge crossing the area in time Δt is
  $\Delta Q = neAv_d\Delta t$
  or
  $I = \Delta Q/\Delta t = neAv_d$ (4)
  This is the relation between the electric current and drift velocity
- If the moving charge carriers are positive rather than negative then electric field force on charge carriers would be in a direction of electric fields direction and drift velocity would be in left to right direction opposite to what shown in fig-1
- In terms of drift velocity current density is given as
  $j = I/A = nev_d$ (5)

(5) Ohm's Law and Resistance

- Ohm's law is the relation between the potential difference applied to the ends of the conductor and current flowing through the conductor.This law was expressed by George Simon Ohm in 1826
- Statement of Ohm's Law
  'if the physical state of the conductor (Temperature and mechanical strain etc) remains unchanged ,then current flowing through a conductor is always ditectly proportional to the potential difference across the two ends of the conductor
  Mathematically
  $V \propto I$
  or
  $V = IR$ (6)

Where constant of proportionallity R is called the electric resistance or simply resistance of the conductor

- Value of resistance depends upon the nature ,dimension and physically dimensions of the conductor
- Ohm's Law can be deducted using drift velocity relation as given in equation -3 .Thus from the equation
  $v_d=(eE/m)\tau$
  but Now $E=V/l$
  Therfore
  $v_d=(eV/ml)\tau$
  Also $I=neAv_d$
  Substituting the value of v_d in I relation
  $I=(ne^2A\tau/ml)\ V$ (7)
  or $V/I=(ml/ne^2A\tau)=R$ a constant for a given conductor
  Thus
  $V=IR$
  Mathematical expression of Ohm's Law
  From Ohm's Law
  $V=IR$ or $R=V/I$ (8)
  Thus electric resitance is the ratio of potential difference across the two ends of conductor and amount of current flowing through the conductor
- electric resistance of a conductor is the obstraction offered by the conductor to the flow of the current through it.
- SI unit of resistance is ohm (Ω) where
  1 Ohm=1 volt/1 Ampere
  or $1\Omega=1VA^{-1}$
- Dimension of resistance is $[ML^2T^{-3}A^{-2}]$

(6) Resistivity and conductivity

- In terms of drift velocity ,electric current flowing through a conducting wire of length L and uniform area of cross-section A
  is
  $I=dQ/dt=neAv_d=(ne^2A\tau/ml)\ V$
  The above can be rearranged to give the ohm's law i.e,
  $V=IR$
  where $R=(ml/ne^2A\tau)$ Now $R=\rho l/A$ (9)
  Where ρ is called the specific resistance or resistivity of the conductor
  And $\rho=m/ne^2\tau$ (10)
- From equation (9) ,we can see that resistance of the wire is proportional to its length and inversly proportional to its cross-sectional area.
- Thus resistance of a long and thin wire will greater then the resistance of short and thick wire of the same material
- Now from equation (9)
  $R=\rho l/A$ (11)
  And from ohm law $R=V/I$
  Therefore
  $\rho=(V/I)(A/L)$

$=(V/L) / (I/A)$

$=E/J$ (12)

Where $E=V/L$ is the electric field at any point inside the wire and $J=I/A$ is current density at any point in the wire. Unit of resistivity is ohm-meter.

- Thus from equation (12) ,electric resistivity can also be defined as the ratio of electric field intensity at any point in the conductor and the current density at that point.
- The greater the resistivity of the material ,greater would be the field needed to establish a given current densisty
- Perfect conductor have zero resistivities and for perfect insulators resistivity would be infinite
- Metals and alloys have lowest resistivities and insulators have high resistivities and exceeds those of metals by a factor of 10^{22}
- The reciprocal of resistivity is called conductivity and is represented by σ
- Unit of conductivity is $ohm^{-1}meter^{-1}(\Omega^{-1}m^{-1})$ and
  σ is defined as
  $\sigma=1/\rho$
  Since $\rho=E/J$
  or $\sigma=J/E$
  or $J=\sigma E$ (13a)
- The above relation can also be written in vector form as both J and E are vector quatities where vector **J** being directed towards **E**
  $\mathbf{J}=\sigma\mathbf{E}$ (13b)

(7) variation of resistivity with temperature

- Resistance and hence resistivity of conductor depends on mubers of factors
- One of the most important factors is dependence of resistance of metals on temperature
- Resistivity of the metallic conductor increases with increase on temperature
- when we increase the temperature of the metallic conductor,its constituent atoms vibrate with greater amplitudes then usual.This results
  to the more frequent collison between ions and electrons
- As a result average time between the two successive collision decreases resulting the decrease in drift velocity
- Thus increase collison with the increase in tempearture results in increase resistivity
- For small temperature variations ,resistivity of the most of the metals varies according to the following relations
  $\rho(T)=\rho(T_0)[1 + \alpha(T-T_0)]$ (14)
  Where $\rho(T)$ and $\rho(T_0)$ are the resistivies of the material at temperature T and T_0 respectively and α is the constant for given materail and is known as coefficient of resistivity.
- Since resistance of a given conductors depends on the length and crosssectional area of the conductor through the relation
  $R=\rho l/A$
  Hence temperature variation of the resistance can be given as
  $R=R(T_0)[1 + \alpha(T-T_0)]$ (15)
- Resistivity of alloys also increase with temperature but this increase is much small as compared to metals
- Resistivies of the non-metals decreases with increase in temperature .This is because at high temperature more electrons becomes avialable for conduction as they set themselves loose from atoms and hence temperature coefficient of resistivity is negative for non-metals
- A similar behavior occurs in case of semi-conductors .temperature coefficient of resistivity is negative for semi-conductors and its value is often large for a semi-conductor materials

(8) Current Voltage relations

- We know that current through any electrical device such as resistors depends on potential defference between the terminals
- Devices obeying ohm's law follow a linear relationship between current following and potential applied where current is directly proportional to voltage applied .Graphical relation between V and I is shown below in figure

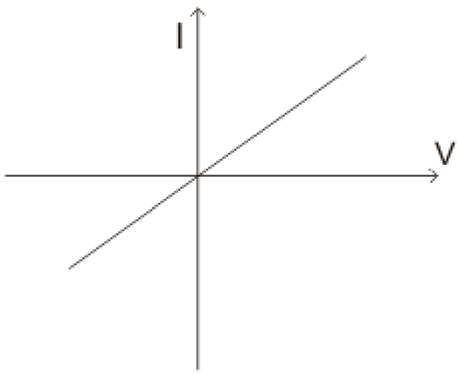


Figure 2:- IV relation for resistors obeying Ohm's Law

- Graph for a resistor obeying ohm's law is a straight line through the origin having some finite slope
- There are many electrical devices that does not obey the ohm's law and current may depends on voltage in more complicated ways.Such devices are called non-ohmic devices for examples vaccum tubes,semiconductor diodes ,transistors etc
- Consider the case of a semi conductor junction diode which are used to convert alternating current to direct current and are used to perform variety of logic functions is a non=ohmic device
- Graphical voltage relation for a diode is shown below in the figure

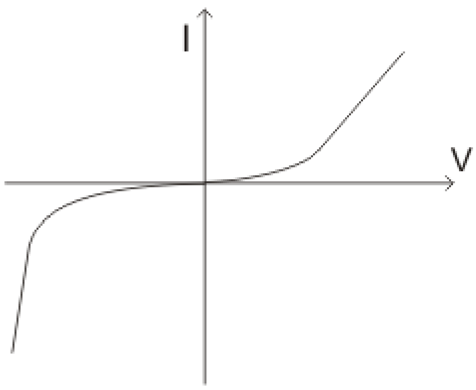


Figure 3:- IV relation for semiconducting diode

- Figure clearly shows a non linear depeence of current on voltage and diode clearly does not follow the ohm's w
- When a device does not follow obey ohm's law,it has non linear voltage -current relation and the quantity V/I is no longer a constant however ratio is still known as resistance which now varies with current

- In such cases we define a quantity dV/dI known as dynamic resistance which expresses the relation between smaal change in current and resulting change in voltage
- Thus for non-ohmic electrical devices resistance is not constant for different values of V and I

(9) Colour code of carbon resistors

- **Commercially resistors of different type and values are avilable in the market but in electronic circuits carbon resistors are more frequently used**
- **In carbon resistors value of resistance is indicated by four coloured bands marked on its surface as shown below in figure**

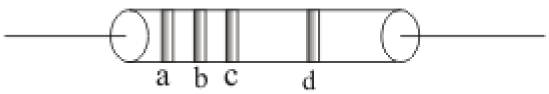


Figure 4:- Carbon resistor with strips

- **The first three bands a,b.c determine the value of the resistance and fourth band d gives the tolerance of the resistance**
- **The colour of the first and second band respectively gives the first and second significant figure of the resistance and third band c gives the power of the ten by which two significant digits are multiplied for obtainng the value of the resistance**
- **value of different colurs for making bands in carbon resistors are given below in the table**

Colur	Figure(first and second band)	Multiplier(for third band)	tolerance
Black	0	1	-
Brown	1	10	-
Red	2	10^2	-
Orange	3	10^3	-
Yellow	4	10^4	-
Green	5	10^5	-
Blue	6	10^6	-
Violet	7	10^7	-
Gray	8	10^8	-
white	9	10^9	-
Gold	-	10^{-1}	5%
Silver	-	10^{-2}	10%
no Colour	-	-	20%

-
- **For example in a given resistor let first strip be brown ,second strip be red and third be orange and fourth be gold then resitance of the resitor would be**
  12X 10^3 +/- 5%

(10)Combination of Resistors

- We have earlier studied that sevral capacitors can be connected in series or parallel combination to form a network. In same way sevral resistor may be combined to form a network.
- Just like capacitors resistors can be grouped in series and parallel.
- Equivalent resistance of the combination of any number of resistors is a single resistance which draw same current as the combination of different resistances draw when the same potential difference is applied across it.

(A) Resistors in Series

- Resisors are said to be connected in series combinaton. If same current flows through each resistor when same potential difference is applied across the combination.
- Consider the figure given below

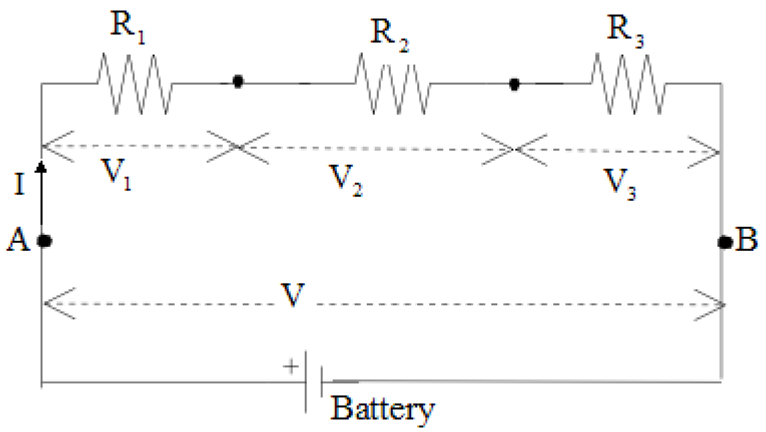


Figure 5:- Series combination of three resistors connected to a battery

- In figure given above three resistors if resistance R_1, R_2 and R_3 are connected ibn series combination.
- If battery is connected across the series combination so as to mintain potential difference V between points A and B, the current I would pass through each resistor.
- If V_1, V_2 andV_3 is the potential difference across each resistor R_1, R_2 and R_3 respectively, then according to Ohm's Law,
  $V_1=IR_1$
  $V_2=IR_2$
  $V_3=IR_3$
  Since in series combination current remains same but potential is divieded so,
  $V=V_1+V_2+V_3$
  or, $V=I(R_1+R_2+R_3)$
  If $R_{eq}$is the resistance equivalent to the series combination of R_1, R_2 and R_3 then ,
  $V=IR_{eq}$
  where, $R_{eq}=R_1+R_2+R_3$
- Thus when the resistors are connected in series, equivalent resistance of the series combination is equal to the sum of individual resistances.
- Value of esistance of the series combination is always greater then the value of largest individual resisnces.
- For n numbers of resistors connected in series equivalent resistance would be
  $R_{eq}=R_1+R_2+R_3+$..........................$+R_n$

(B) Resistors in parallel

- Resistors are said to be connected in parallel combination if potential difference across each resistors is same.
- Thus , in parallel combination of resistors potential remains the same but current is divided.
- Consider the figure given below

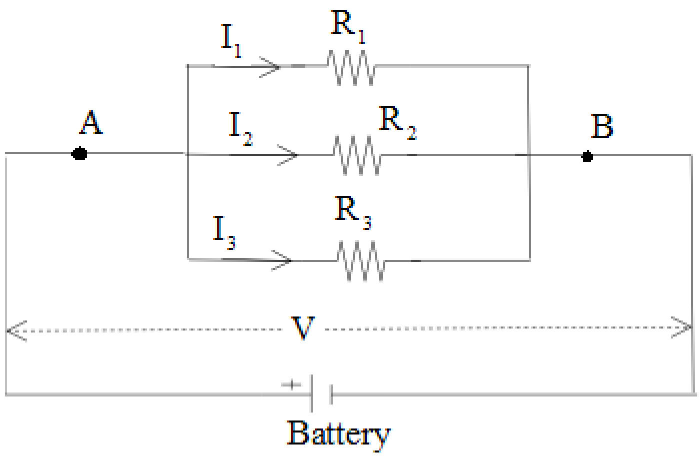


Figure 6:- Parallel combination of three resistors connected to a battery

- <u>Battery</u> B is connected across parallel combination of resistors so as to maintain potential difference V across each resistors.Then total current in the circuit would be
  $I=I_1+I_2+I_3$ (16)
- Since potential difference across each resistors is V. Therefore, on applying Ohm's Law
  $V=I_1R_1=I_2R_2=I_3R_3$
  or,

$$I_1=\frac{V}{R_1}, I_2=\frac{V}{R_2}, I_3=\frac{V}{R_3}$$

From equation (16)

$$I=\frac{V}{R_1}+\frac{V}{R_2}+\frac{V}{R_3}=V\left(\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}\right)$$

- If R<sub.eq< sub=""> is the equivalent resistance of parallel combination of three resistors heaving resistances R_1, R_2 and R_3 then from Ohm's Law
  $V=IR_{eq}$
  or,

$$I=\frac{V}{R_{eq}} \qquad (17)$$

Comparing equation (16) and (17) we get

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

</sub.eq<>

- For resistors connected in parallel combination reciprocal of equivalent resistance is equal to the sum of reciprocal of individual resistances.

- Value of equivalent resistances for capacitors connected in parallel combination is always less then the value of the smallest resistance in circuit.
- If there are n number of resistances connected in parallel combination, then quivalent resistance would be reciprocal of

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

(1) Introduction

- In previous chapter we have already studied about electric current and resistance.
- We know that a force must be applied on free charges of a conductor in order to maintain a continous current in the conductor.
- Here a question arises how can we maintain this force in order to maintain a continous flow of current. You will find answer to this question while studying this chapter.
- In this chapter we will learn about ElectroMotive Force(emf) and sources of emf (responsible for driving charge round the closed circuit). We'll also learn about electric circuits and measurements.

(2) ElectroMotive Force(emf)

- Consider a conductor lying in presence of electric field as shown below in the figure such that an electric field exists inside the conductor.

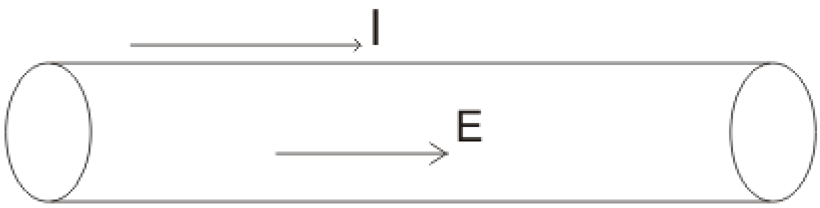


Figure 1. Current and Field E in a conductor

- We know that when electric field exists in a conductur electric current begins to flow inside the conductor. Now a question arises what happens to the
charge carriers when they reach the ends of the conductor and would this current remains constant with the passage of time.
- We can easily conclude that for an open ended conductor as shown in the figure , charges would accumulate at the ends of the conductor resulting a change in electric field with the passage of time. Due to this electric current would not remain constant and would flow only for a very short interval of time , digrametically shown below in the figure.

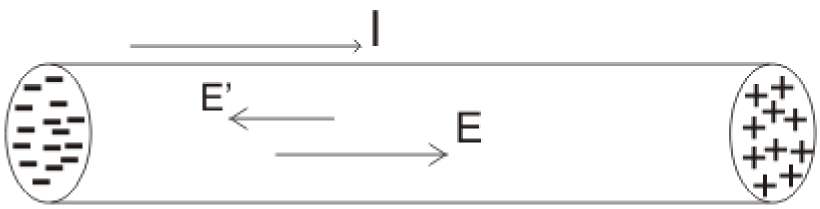


Figure 2a. Charges accumulates at the ends of the conductor and develops field E' opposite to E

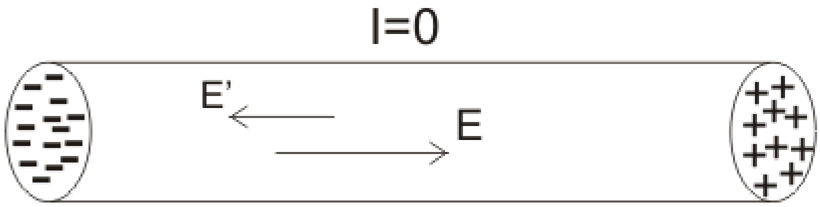


Figure 2b. Finally total field E_{tot}= E+E'=0 and current flowing in conductor becomes zero

- Thus, in order to maintain a steady current throughout a conducting path the path must be in the form of a closed loop forming a complete circuit. Even this condition is not sufficient to maintain a steady current in the circuit.
- This is because charge always moves in the direction of decreasing potential and electric field always does a positive work on the charge.
- Now after travelling through a complete circuit when charge returns to a point where it has started, potential at that point must be same as the potential at that point in the begning of the journey but flow of current always involves loss of potential energy.
- Hence we need some external source in the circuit in which maintains a potential difference at its terminals by increasing the potential energyof the electric charge.
- Such a source make charge travel from lower potential to higher potential energy in direction opposit to the electrostatic force trying to push charge from higher potential to lower potential.
- This force that makes charge move from lower potential to higher potential is called electro mative force (EMF).
- The source or device which provide emf in a complete circuit is known as source of EMF and examples of such devices are generator, batteries, thermocuples etc.
- The source of EMF are basically energy converters that convert mechanical, thermal, chemical or any other form of energy into electrical potential energy and transform it into the circuit to which the source of emf is connected.
- Now we know that a source of emf or battery maintains a potential difference between its two terminals as shown in below figure

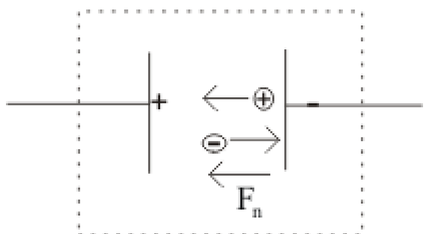


Figure 3. Internal structure of the battery

- Generally a battery consist of two terminals one positive and other is negative.
- Some internal force F_n generally non electric in nature is exerted on the charges of the material of the battery. This non-electric force depends on the nature of source of EMF.
- These Force(F_n) drives the positive charges of the material towards P and nigative charges of the material towards Q. This battery force
  F_n is directed Q to P.
- Positive charge accumulate on plate Pand negative charge accumulate on plate Q and a potential developes between plates P and Q. Thus an electric field would set up inside the battery from P to Q which excert an electric force on the charge of the material.
- When a steady state is reached, the the electric force and battery force F_n would become equal and opposit
  mathematically,
  $qE=F_n$ (1)
  and after a steady state is reached no further accumulation of charhe takes place.
- Workdone by battery force F_n in taking poisitive charge from terminal Q to terminal P would be

$W=F_n d$
where d is distance between plates P and Q.

- Workdone by force F_n per unit charge is
  $EMF=W/q=F_n d/q$ (2)
  where the quantity E is known as E.M.F. of the battery.
- For steady state
  $ENF=qEd/q=ED=V$ (3)
  where $V_{eq}=Ed$ is the potential difference across the terminals of the battery when nothing is connected externally between P and (i.e. when circuit is open)

(3)Internal Resistance of Battery (or cell)

- The resistance offered by medium in between plates of battery (electrolytes and electrodes of the cell) to the flow of current within the battery is called internal resistance of the battery.
- Internal resistance of a battery usually d branch containing batteryenoted by r and in electric circuit its representation is shown below in the figure

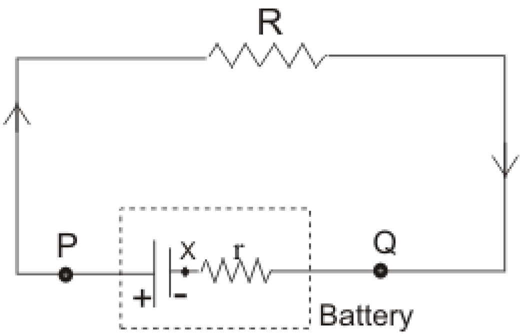


Figure 4. r Represents internal resistance of the battery

- Internal resistance of a battery depends on factors like seperation between plates, plate area, nature of material of plate etc. For an ideal cell r=0 , but real batteries or sourcesof emf always has same finite internal resistance.
- If P and Q are two terminals of the battery shown below in the figure

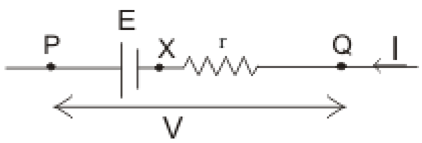


Figure 5. Terminal voltage of a cell

then potential difference between terminals P and Q is
$V_P=(V_P-V_x) - (V_Q - V_x) = E-Ir$
let $V_P-V_Q=V$
$V=E-IR$
now for I=0 and V=EMF
and this potential difference V is called the terminal difference of the cell or battery and defined as the emf of the battery when no current drawn from it.

- For real battery equation(4) which gives $V=E-Ir$ whereI is the current in the branch containing battery.
- From figure(4) potential difference across the external resistance R of the circuit would be equal to terminal potential difference of the cell. Thus
  $V=IR$ also $V=E-Ir$
  or, $IR=E-Ir$
  which gives
  $I=E/(R+r)$ =Net EMF/Net resistance

- From equation(4) we can calculate that when current is drawn from the battery terminal potential difference is less than the EMF of the battery.

(4)Electric Energy and Power

- To understand the process of energy transfer in a simple circuit consider a simple circuit as shown in the figure given below

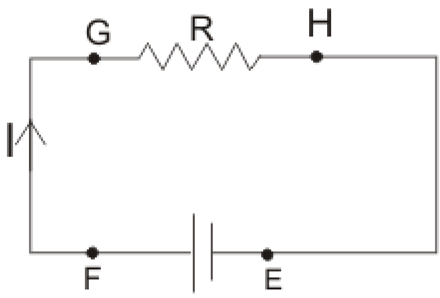


Figure 6. Circuit consisting of a resistance R

- Positive terminal of the battery as we all know is always at higher potential.
- Let ΔQ amount of charge begin to flow in the circuit from point E through the battery and resistor and then back to point E.
- When Charge ΔQ moves from point E to point F through the battery electric potential potential energy of the system increased by the amount
  $\Delta U=\Delta Q\ V$ -(6) and the electric energy of the battery decreased by the same amount.
- When charge ΔQ moves from point G to S through resistoe R, there comes a decrease in electric potential energy.
- This loss in potential energy appears as the increased in thermal energy of the resistor.
- Thermal energy of the resistor increases because when the charge moves through the resistor they loss there electrical potential energy by colliding with the atom in the resistor. This way electrical energy is transformed internal energy crossesponding to increase in vibrational motion of the atom of the resistor and this cause increase in temprature of the resistor.
- The connecting wires are assumed to have negligible resistance and no energy transfer occur for the path FG and HE.
- In time Δt charge ΔQ moves through the resistor i.e. from G to H. The rate at which it loss potential energy
  $\Delta U/\Delta t=(\Delta Q/\Delta t)\Delta V=I\Delta V$ where I is thecurrent in the resistor and ΔV is the potential difference across it.
- This charge ΔQ regain its energy when it passes through the battery at the cost of conservation of chemical energy of electrolyte to the electrical energy.
- This loss of potential energy as stated earlier appears as increased thermal energy of the resistor. If P represents the rate at which energy is delivered to the resistor then
  $P=I\Delta V$
- We know that ΔV =IR for a resistor hence alternative forms of equation(8) are
  $P=I^2R=\Delta V^2/R$ where I is expressed in amperes, ΔV iv volts and resustance R in ohm(Ω)
- SI unites of power is watt such that
  1watt=1volt*1ampare
  Bigger unit of electric power are Kilowatt(KW) and Megawatt(MW).

(5) Kirchoff's Rules

- We have already analyzed simple circuit using ohm's laws and reducing these circuit to series and parallel combination of resistors
- But we also come across circuits containing sources of EMF and grouping of resistors can be far more complex and can not be easily reduced to a single equivalent resistors
- Such complex circuits can be anaylzed using two kirchoff's rules

(A) The junction Rule (or point rule)

- This law states that "The algebric sum of all the currents entering juntion or any point in a circuit must be equal to the sum of currents leaving the junction"
- Alternatively this rule can also be stated as " Algebric sum of the currents meeting at a point in a electric circuit is always zero i.e
  $\Sigma I=0$ at any point in a circuit
- This law is based on the law of conservation of charge
- Consider a point P in an electric circuit at which current I_1,I_2,I_3 and I_4 are flowing through conductors in
  the direction shown below in the figure below

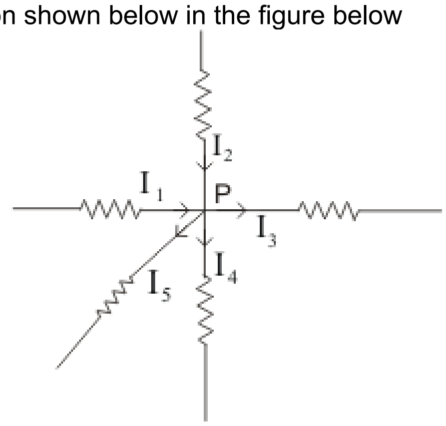


Figure 7. Current at a junction is zero

- If we take current flowing towards the junction as positive and current away from the junction as negative,then from kirchoff's law
  $I_1+I_2+(-I_3)+(-I_4)=0$
  or,
  $I_1+ I_2=I_3+ I_4$
- From this law ,we conclude that netcharge coming towards a point must be equal to the net charge going away from this point in the same interval of time

(B) The Loop Rule (or Kirchoff's Voltage Law)

- The rule states that " the sum of potential difference across all the circuit elements along a <u>closed loop</u> in a circuit is zero
  $\Sigma V=0$ in a <u>closed loop</u>

- Kirchoff's loop rule is based on the law of conservation of energy becuase total amount of energy gained and losed by a charge round a trip in a closed loop is zero
- when applying this kirchoff's loop rule in any DC circuit,we first choose a closed loop in a circuit that we are analyzing
- Next thing we have to decide is that whether we will traverse the loop in a clockwise direction or in anticlockwise direction and the answer is that ,the choice of direction of travel is arbitrary to reach the same point again
- When traversing the loop ,we will be following convention to note down drop or rise in the voltage across the resistors or battery
  i) If the resistor is being traversed in the direction of the current then change in PD across it is negative i.e -IR
  ii)If the resistor is being traversed in the direction opposite to the current then change in PD across it is negative i.e IR
  iii) If a source of EMF is traversed in the direction from -ve terminal to its positive terminal then change in electric potential is positive i.e E
  iv)If a source of EMF is traversed in the direction from +ve terminal to its negative terminal then change in electric potential is negative i.e -E
- We would now demonstrate the use of kirchoff's loop law in finding equations in simple circuit
- Consider the circuit as shown below

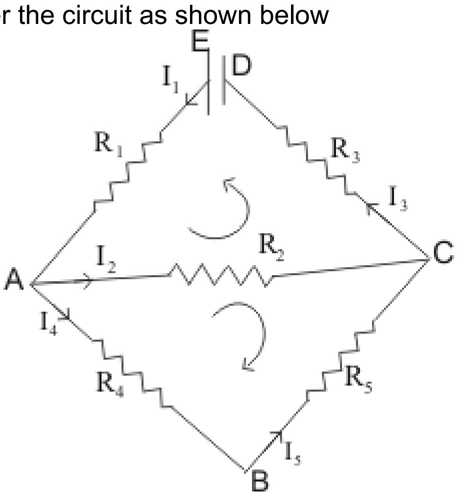


Figure 8. Circuit for explaining Kirchoff's loop rule

- First consider loop ABDA.Lets traverse loop in anticlock wise direction.From kirchoff's loop law
  $\Sigma V=0$
- Neglecting internal resistance of the cell and using sign conventions stated previously we find
  $-I_3R_3+E-I_1R_1-I_2R_2=0$
  or
  $I_1R_1+I_2R_2+I_3R_3=E$
  And similarly if we traverse the loop ABCA in clock wise direction
  $-I_2R_2+I_5R_5+I_4R_4=0$
  or,
  $I_5R_5+I_4R_4-I_2R_2=0$

(6)Grouping of the cell's

- A limited ammount of current can be drawn from a single cell or battery
- There are situations where single cell fails to meet the current requirement in a circuits
- To overcome the problem cells can be grouped in series and in parallel combinations or mixed grouping of cells is done in order to obtain a large value
  of electric current

(A) Series combination

- Figure below shows the two cells of emf's E_1 and E_2 and internal resistance r_1 and r_2 respectively connected in series combination through external resistance

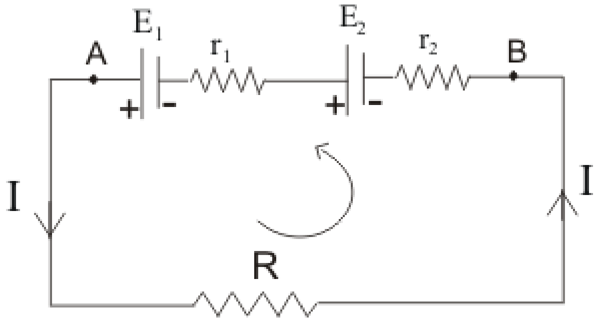


Figure 9a. Cells grouped in series

- Points A and B in the circuit acts as two terminals of the combination
- Applying kirchoff's loop rule to above closed circuit
  $-Ir_2-Ir_1-IR+E_1+E_2=0$
  or
  $I=E_1+E_2/R+(r_1+r_2)$
  Where I is the current flowing through the external resistance R
- Let total internal resistance of the combination by $r=r_1+r_2$ and also let $E=E_1+E_2$ is the total EMF of the two cells
- Thus this combination of two cells acts as a cell of emf $E=E_1+E_2$ having total internal resistance $r=r_1+r_2$ as shown above in the figure

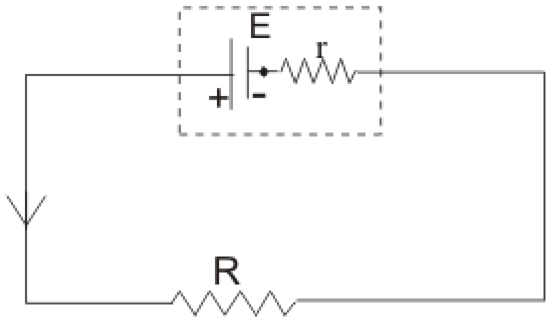


Figure 9b. Equivalent cell of emf E and equivalent internal resistance r

(B) Parallel combinations of cells

- Figure below shows the two cells of emf E_1 and E_2 and internal resistance r_1 and r_2 respectively connected in parallel combination through external resistance

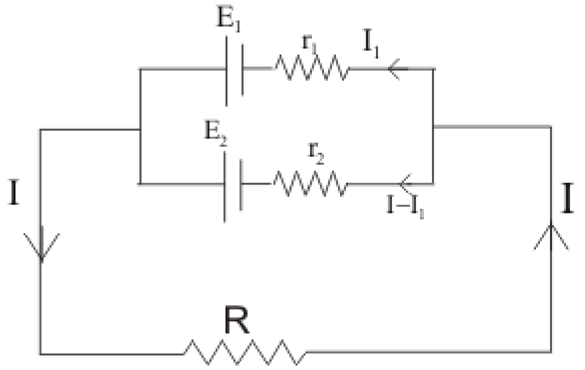


Figure 10. Cells grouped in parallel combination

- Applying kirchoff's loop rule in loop containing E_1 ,r_1 and R,we find
  E_1-IR-I_1r_1=0 ------------------------(1)
  Similarly applying kirchoff's loop rule in loop containing E_2 ,r_2 and R,we find
  E_2-IR-(I-I_1)r_2=0 ------------------------(2)
- Now we have to solve equation 1 and 2 for the value of I,So multiplying 1 by r_2 and 2 by r_1 and then adding these equations results in following equation
  $IR(r_1+r_2)+r_2r_1I-E_1r_2-E_2r_1=0$
  which gives

$$I=\frac{E_1r_2+E_2r_1}{R(r_1+r_2)+r_2r_1}$$

We can rewrite this as

$$I=\frac{\dfrac{E_1r_2+E_2r_1}{r_1+r_2}}{R+\dfrac{r_1r_2}{r_1+r_2}}=\frac{E}{R+r} \qquad (11)$$

where

$$E=\frac{E_1r_2+E_2r_1}{r_1+r_2}$$

and

$$r=\frac{r_1r_2}{r_1+r_2}$$

E is the resulting EMF due to parallel combination of cells and r is resulting internal resistance.

(7) Wheat stone bridge

- Wheat stone bridge was designede by british physicist sir Charles F wheatstone in 1833
- It is a arrangement of four resistors used to determine resistance of one resistors in terms of other three resistors
- Consider the figure given below which is an arrangement of resistors and is knowns as wheat stone bridge

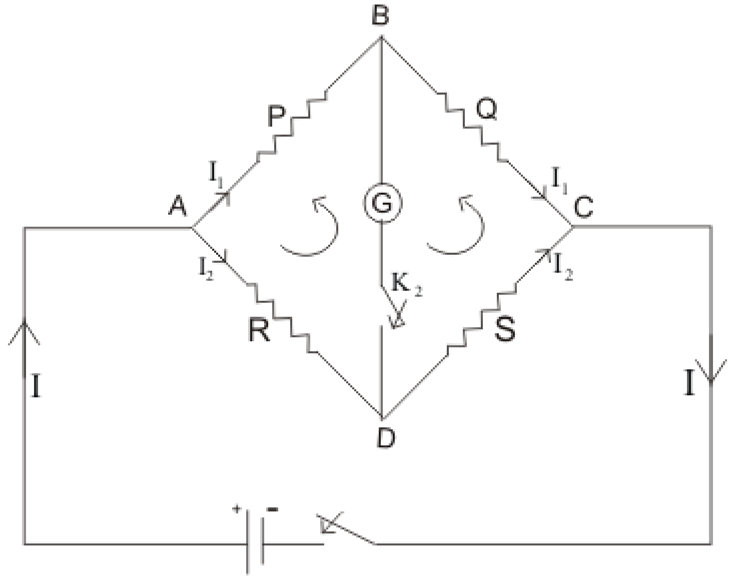


Figure 11. Wheatstone bridge circuit arrangement

- Wheatstone bridge consists of four resistance P,Q,R and S with a battery of EMF E.Two keys K_1 and K_2 are connected across terminals A and C and B and D respectively
- ON pressing key K_1 fisrt and then pressing K_2 next if galvanometer does not show any deflection then wheatstone bridge is said to be balanced
- Galavanometer is not showing any deflection this means that no current is flowing through the galvanameter and terminal B and D are at the same
  potential .THus for a balanced bridge
  $V_B=V_D$
- Now we have to find the condition for the balanced wheatstone bridge .For this applying kirchoff's loop rule to the loop ABDA ,we find the relation
  $-I_2R+I_1P=0$
  or $I_1P=I_2R$ --(a)
  Again applying kirchoff's rule to the loop BCDB
  $I_1Q-I_2S=0$
  or $I_1Q=I_2S$ --(b)
  From equation a and b we get
  $I_1/I_2=R/P=S/Q$
  or
  $P/Q=R/S$ (12)
- equation 12 gives the condition for the balanced wheatstone bridge
- Thus if the ratio of the resistance R is known then unknown resistance S can easily be calculated
- One important thing to note is that when bridge is balanced positions of cell and galvanometer can be exchanged without having any effect on the balance of the bridge
- Sensitivity of the bridge depends on the relative magnitudes of the resistance in the four arm of the bridge is maximum for same order of four resistance.

(8) Meterbridge (slide wire bridge)

- Meter bridge is based on the principle of wheatstone bridge and it is used to find the resistance of an unknown conductor or to compare two unknown
  resistance

- Figure below shows a schematic diagram of a meter bridge

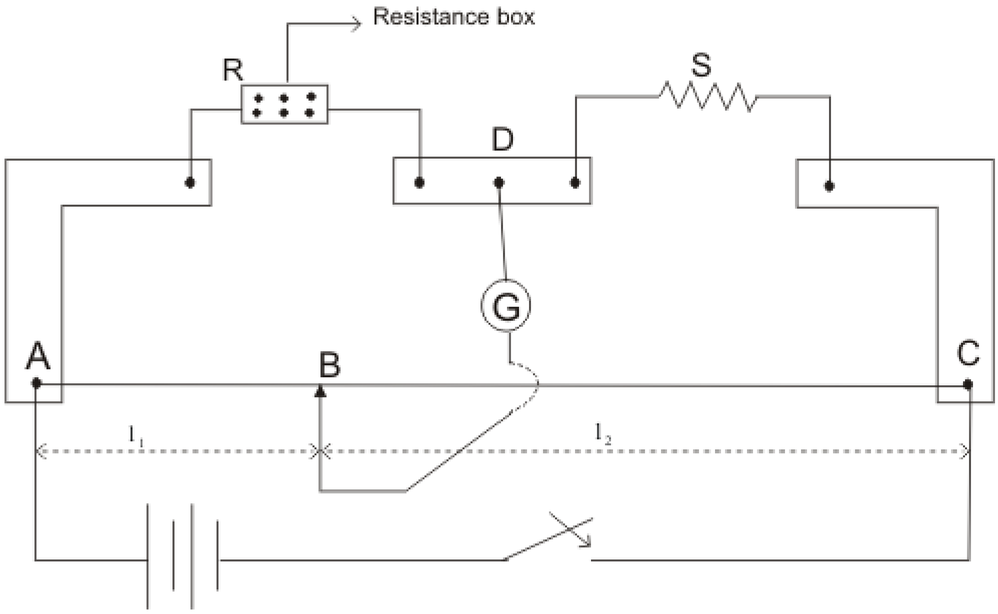


Figure 12. A meter bridge

- In above figure AC is a 1m long wire made of maganin or constanan having uniform area of cross-section
- This wire is stretched along a scale one a wooden base
- Ends A and C of the wire are screwed to two L shaped copper strips as shown in figure
- A resistance box R and an unknown resistance S are connected as shown in figure
- One terminal of galvanometer is connected to point D and another terminal is joined to a jockey that can be slided on a bridge wire
- when we adjust the suitable resistance of value R in the resistance box and slide this jockey along the wire then a balance point is obtained sat at point B
- Since the circuit now is the same as that of wheatstone bridge ,so from the condition of balanced wheatstone bridge we have
  $P/Q=R/S$
  Here resistance P equals
  $P=\rho l_1/A$
  And $Q=\rho l_2/A$
  where ρ is the resistivity of the material of the wire and A is the area of cross-section of wire
  Now $P/Q=(\rho l_1/A)(A/\rho l_2)=l_1/l_2$

(9) Potentiometer

- Potentiometer is an accurate instruments used to compare emf's of a cells,Potential difference between two points of the electric wire
- Potentiometer is based on the principle that potential drop across any portion of th wire of uniform crossection is proportional to the length of that portion of thw wire when a constant current flows through the wire
- Figure below shows the construction of a potentiometer which consists of a number of segments of wire of uniform area of cross-section stretched on a wooden board between two copper strips .Meter scale is fixed parallel to the lenght of the wire

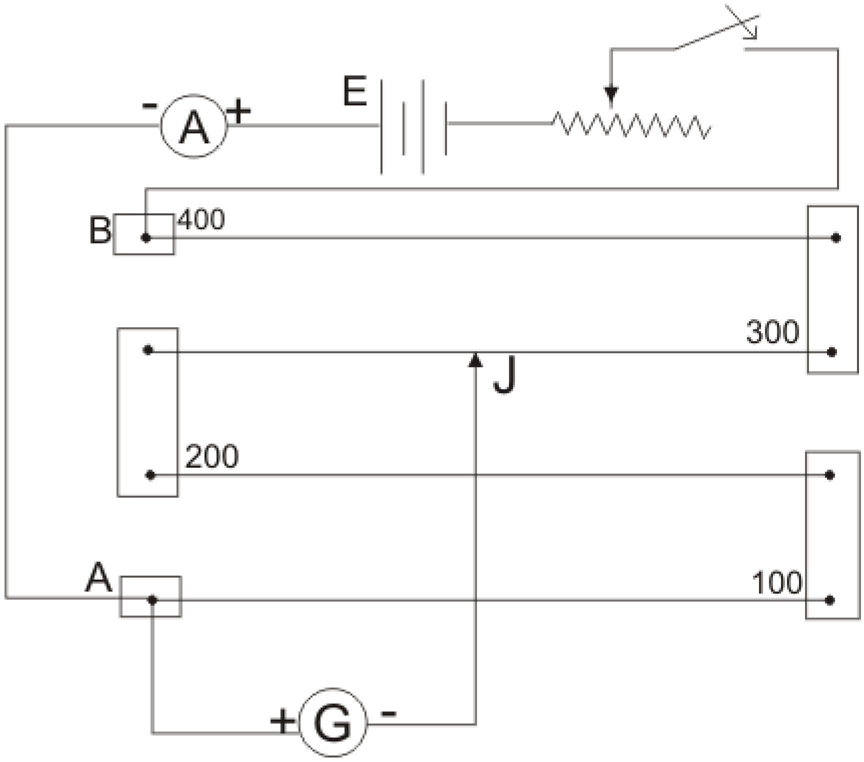


Figure 13. Potentiometer

- A battery is connected across terminals A and B through a rehestat so that a constant currents flows through the wire
- Potentiometer is provided with a jockey J with the help of which contact can be made at any point on the wire
- Suppose A and ρ are the area of cross-section and resistivity of the material of the wire the resitance
  $R=\rho l/A$ ----------------------------(i)
  where l is the lenght of the wire
- If I is the current flowing through the wire then from Ohm's Law,
  $V=IR$ ------------------------------(ii)
  Where V is the potential differene across the position of the wire of length l
  Thus ,from (i) and (ii)
  $V=IR=I(\rho l/A)=kl$
  where $K=\rho I/A$
  => V is proportional to l when current I is constant
- $K=V/l$ is also known as potential gradient which is the fall of potential per unit length of wire
- Senstivity of a potentiometer depends on its potential gradient .If the potential gradient of a potentiometer is small then the potentiometer is more sensitive and hence more accurate

(A) Comparison of EMF's of two cells using potentiometer

- Consider the circuit arrangement of potentiometer given below used for comparison of emf's of two cells

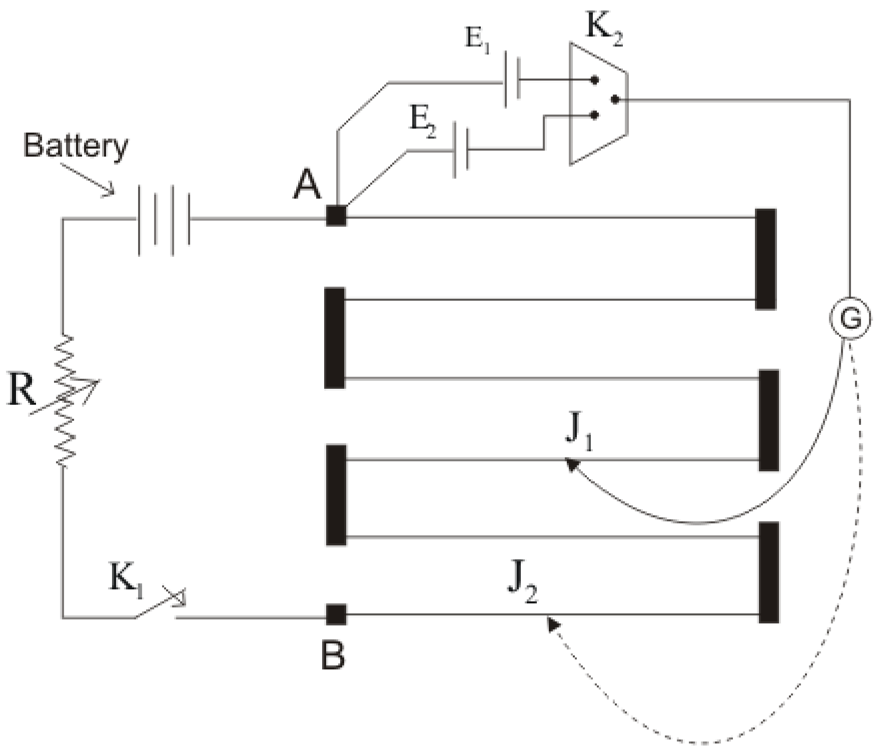


Figure 14. Potentiometer arrangement for comparison of emf's of two cells

- Positive terminals of two cells of emf's E_1 and E_2(whose emf are to be compared) are connected to the terminals A and negative terminals are connected to jockey through a two way key K_2 and a galvanometer
- Now first key K_1 is closed to establish a potential difference between the terminals A and B then by closing key Ksub>2 introduce cell of EMF E_1 in the circuit and null point junction J_1 is dtermined with the help of jockey.If the null point on wire is at length
  l_1 from A then
  $E_1=Kl_1$
  Where K -> Potential gradient along the length of wire
- Similarly cell having emf E_2 is introduced in the circuit and again null point J_2 is determined .If length of this null point from
  A is l_2 then
  $E_2=Kl_2$
  Therefore
  $E_1/E_2=l_1/l_2$
  This simple relation allows us to find the ratio of E_1/E_2
- if the EMF of one cell is known then the EMF of other cell can be known easily

(B) Determination of internal resistance of the cell

* Potentiometer can also be used to determine the internal resistance of a cell

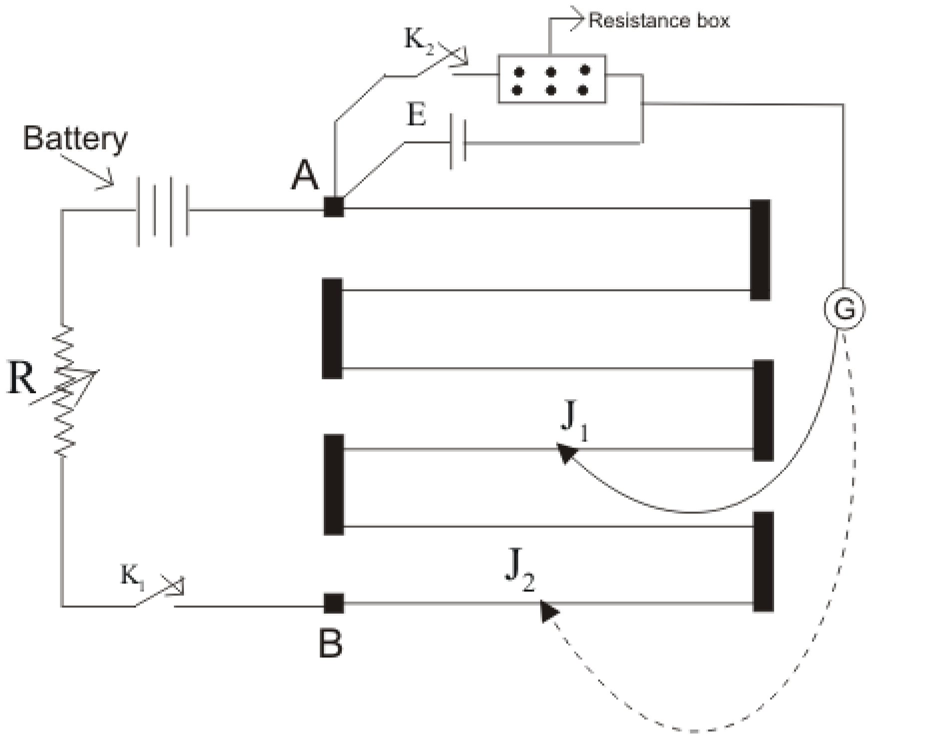


Figure 15. Potentiometer arrangement for determining internal resistance of a cell

* For this a cell whose internal resistance is to be determined is connected to terminal A of the potentiometer across a resistance box through a key K_2
* First close the key K_1 and obtain the null point .Let l_1 be the length of this null point from terminal A then
  $E=Kl_1$
* When key K_2 is closed ,the cell sends current through resistance Box (R).If E_2 is the terminal Potential difference and null point is obtained at length $l_2(AJ_2)$ then
  $V=Kl_2$
  Thus
  $E/V=l_1/l_2$
  But $E=I(R+ r)$ and $V=IR$
  This gives
  $E/V=(r+R)/R$
  So $(r+R)/R=l_1/l_2$
  giving
  $r=R(l_1/l_2-1)$
* Using above equation we can find internal resistance of any given cell

(1) Introduction

- We have already learned the electric current and the physics behind it in the previous chapter
- We have also discussed the mechanism of flow of current in a conducter but not the physical consquences of flow of elctric current are related to other forms of energy
- In this chapter we will study about causes and consequences of electric current
- In the nutshell ,what we will study in this chapter is the connection between the electricity and thermal energy.

(2) Heating effect of current

- In previous chapter while discussing electric energy and power ,we learned that IΔV amount of energy is lost per second when a current I flows through a potential ΔV and this energy appears in the form of heat energy
- Due to the conversion of electric energy into heat energy the conductor becomes hot .This effect is known as Joule's Heating and this heating is thermodynamics irreversible.

Cause Behind Joule's Heating:-

- Explanation behind the Joule's heating is that when a potential difference ΔV is maintained between the ends of a conductor,the free electrons in the conductor are accelerated towards the higher potential end of the conductor
- In their way electrons frequently collided withe the positive ions of the conductor due to which their velocity decreased
- This the energy electrons gained on account of acceleration is transferred to the positive lattice ions or atoms and electrons then returns to their equilibrium distribution of velocities
- Thus ,lattice ions receives energy randomly at the average rate of IΔV per unit time
- Ions spends this energy by vibrating about their mean positions resulting in rise in the temperature of the conductor
- This way Joule's heating nothing but the conversion of electrical energy into heat energy

(3) Thermoelectricity

- We know that currents flows in a conductor whenever there is a electric potential difference betweens the ends of the conductor
- If there is a temperature difference between the ends of the conductor then thermal energy flows from hotter end to the colder ends
- Thermal energy flows may also be carries by the electrons in the conductor and hence resulting the presence of electric current
- At the hotter end of the conductor electrons have slightly higher kinetic energy and hence they move faster
- So there is net flow of current towards the end of the conductor with lower temperature.Thus an electric current exists in the conductor due to the difference in the temperature of two ends of the conductor

This phenonmenon due to which electricity is produced when two ends of the conductor are kept at different temperature is known as thermoelectricity

4)Seeback effect

- Seeback effect was first discovered by Thomas John seaback
- It stated that when two different conductor are joined to form a circuit and the two junctions are held at the different temperature then an emf is developed which results in the flow of the electric current through the circuit.Arrangement is shown as below in figure

- Maginitude of thermo-electric emf depends upon the nature of the two metals and on the temperature difference between terminals

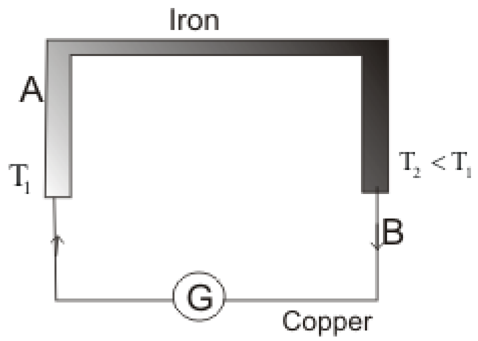


Figure 1. Thermocouple of iron and copper and direction of current is from Cu to Fe through hot junction

- Seaback effect is reversible i.e, if the hot and cold junctions are reversed the direction of thermoelectric current is alse reversed
- Seaback investigated thermo-electric properties of a large number of metals and arranged them in a series known as thermo-electric series or seaback series and is given as follows

  Bi,Ni,Co,Pt,Cu,Mn,Hg,Pb,Sn,Au,Ag,Zn,Cd,Fe,As,Sb,Te

- When any two of these metals in the series is used to form a thermocouple ,the thermo emf is greater when two metals used are farther apart in the circuit
- Figure 1 shows the thermocouple of Cu and Fe.The current in this couple flows from Cu to Fe through the hot junction
- The thermo emf of this couple is only 1.3 milivolt for a temperature difference of 100 C between the hot and cold junction

(5) variation of thermo-emf with temperature

- To study the effect of difference of temperature of the two junctions consider a thermo-couple of two dis-similar metals A and B
- Now consider that cold junction is at temperature 0 C and the temperature of the hot junction is raised gradually
- It is found in experiment that thermo emf varies with the temperature of the hot junction

- Figure below shows the graph of variation of thermo emf in the circuit with the variation of temperature of the hot junction.

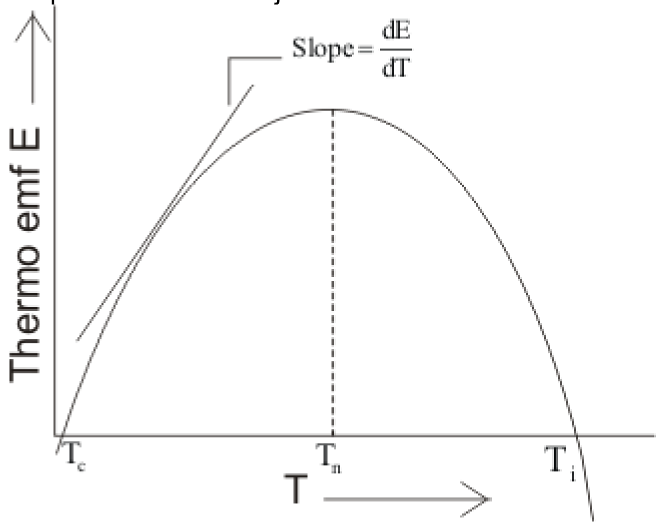


Figure 2. Variation of thermo emf with temperature of hot junction

- From graph it is clear that thermoemf is zerp when both the junctions are at the same temperature 0 C and gradually increases as the temperature of the hot junction increases
- The temperature of the hot junction at which thermoemf in the thermocouple becomes maximum is called neutral temperature (T_n) for that thermocouple
- For a given thermo-couple of two metals neutral temperature has a fixed value
- On further increasing the temperature of the hot junction,after T_n has reached,The thermo-emf decreases and becomes equal at a particular temperature called temperature of inversion T_i.
- Beyond T_i ,if the temperature of hot junction is still increased ,the thermo-emf again started to increase but in reverse direction
- Temperature of inversion T_i is as much above the neutral temperature as neutral temperature is above the temperature of the cold junction.Thus mathimatically
  $T_n-T_c=T_i-T_n$
  or $T_n=(T_i+T_c)/2$
  Hence neutral temperature is the mean of the temperature of inversion and temperature of the cold junction
- Thermo-emf is the property of each material and can be easily measured for a junction of two dissimilar metals at different temperatures
- Thermo-emf of number of thermocouples is given by the simple relation
  $E=\alpha T+\beta T^2$
  where T is the temperature difference between the two junction and α and β are the parameters of the materoal used
- Also the rate of change of thermo-emf with temperature i.e dE/dT is called thermo-power or seaback coefficent S .Mathematically
  $S=dE/dT$

6) Peltier Effect

- Peltier effect is named after his discover Jean Peltier who in 1934 discovered a thermo-electric effect which is converse of Seaback effect
- Peltier discovered that "when an electric current is passed through two disimilar conductor connected to form a thermo-couple ,heat is evolved at one junction and absorbed at the other end.The absorption and evolution of heat depends on the direction of flow of current

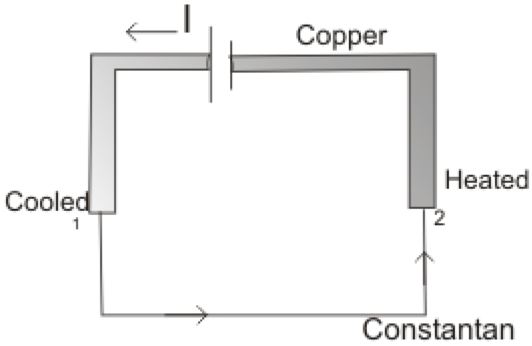


Figure 3. In copper constantan thermocouple if an electric current flows as shown then junction n1 gets cooled and 2 get heated

- Peltier effect is entirely reversible in nature

Peltier Coefficent

- Peltier coefficent is defined as the amount of heat energy absorbed or evolved due to peltier effect at the junction of two dissimilar metals when one coulumb of charge passes through the junction
- Peltier coefficent is denoted by π
- Value of Peltier Coefficent is different for different thermo-couple .Its value also depends upon the temperature of the junction
- If q amount of charge passes through the junction then
  Energy absorbed or evolved=πq
  if V is the contact Potential difference
  Then workdone=qV
  Now heat absorbed=Workdone
  So π=V
- Hence Peltier coefficent (in J/C) at a junction is numerically equal to the contact v in(Volts)

(7) Thomson effect

- Thomson effect is related to the emf that develops between two parts of the single metal when they are at different temperature
- Thus thomson effect is the absorption or evolution of heat along a conductor when current passes through it when one end of the conductor is hot and another is cold
- If two parts of the metal are at small temperature difference dT ,themn the electric potential difference is proportional to dT $dV \propto dT$
  or
  $dV=\sigma dT$
  where σ is the constant of proportinality and is known as thomson coefficent
- Peltier coefficent and thomson coeficent are related to thermopower according to following relations
  $\pi=Ts=T(dE/dT)$
  and $\sigma=-T(ds/dT)=-T(d^2E/dT^2)$
- We have seen that all the three effects are defined in terms of three coefficent namely seaback,peltier and thomson coefficent but the basic quantity is thermo-power which is the rate of change of thermo-emf with temperature

(8)Applications of thermoelectricity or thermo-electric effect are

- To measure temperature using thermo electric thermometer
- To detect heat radiation using thermopiles
- Thermoelectric refregerator or generator

UNIT3

Magnetic effect of current and magnetism

- We have allready studied about thermal effects of current and now in the present chapter we are studied about magnetic effect of current.
- Earlier it was believe that there is no connection between electric and magnetic force and both of them are completely different.
- But in 1820 Oersted showed that the electric current through a wire deflect the magnetic needle placed near the wire and the direction of deflection of needle is reversed if we reverse the direction of current in the wire.
- So, Oersted's experiments establishes that a magnetic field is assoiated with current carrying wire.
- Again if we a magnetic needlle near a bar magnet it gets deflectid and rests in some other direction.
- This needle experiences the tourque which turn the needle to a definite direction.
- Thus, the reagion near the bar magnet or current carrying where magnetic needle experience and suffer deflection is called magnetic field.

(2) The Magnetic Field

- We all ready know that a stationery charges gets up a electric field E in the space surrounding it and this electric field exerts a force **F**=q_0**E** on the test charge q_0 placed in magnetic field.
- Similarly we can describe the intraction of moving charges that, a moving charge excert a magnetic field in the space surrounding it and this magnetic field exert a force on the moving charge.
- Like electric field, magntic field is also a vector quantity and is represented by symbol **B**
- Like electric field force which depend on the magnitude of charge and electric field, magnetic force is propotional to the magnitude of charge and the strength of magnetic field.
- Apart from its dependence on magnitude of charge and magnetic field strength magnetic force also depends on velocity of the particle.
- The magnitude magnetic force increase with increase in speed of charged particle.
- Direction of magnetic force depends on direction of magnetc field B and velocity **v** of the chared particle.
- The direction of magnetic force is not alonge the direction of magnetic field but direction of force is always perpendicular to direction of both magnetic field **B** and velocity **v**

- Test charge of magnitude q_0 is moving with velocity **v** through a point P in magnetic field **B** experience a deflecting force **F** defined by a equation
  F=q**v** X **B**
- As mentioned earlier this force on charged particle is perpendicular to the plane formed by **v** and **B** and its direction is determined right hand thumb rule.

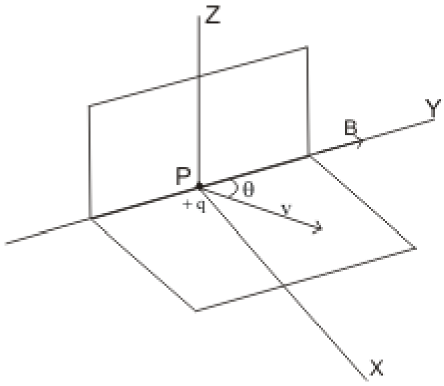


Figure 1. Particle having charge +q and velocity v through a point P in magnetic field

- When moving charge is positive the direction of force **F** is the direction of advance of hand screw whose axis is perpendicular to the plane formed by **v** and **B**.

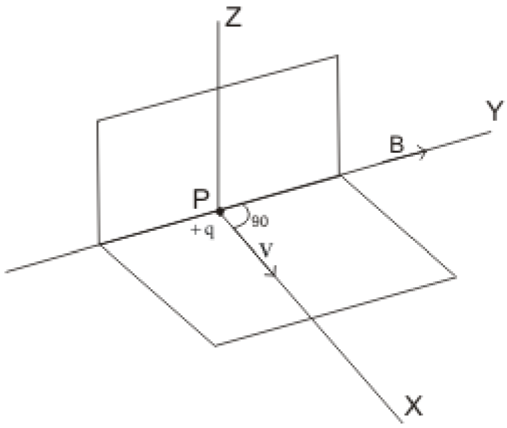


Figure 2. Particle having charge +q and velocity v through a point P in magnetic field with B perpendicular to v

- Direction of force would be opposit to the direction of advance screw for negative charge moving in same direction.
- Magnitude of force on charged particle is
  $F=q_0vB\sin\theta$
  where θ is the angle between v and B.
- If **v** and **B** are at right angle to each other i.e. θ=90 then force acting on the particle would be maximum and is given by
  $F_{max}=q_0vB$ ----(3)
- When θ=180 or θ=0 i.e. v is parallel or antiparallel to B then froce acting on the particle would be zero.
- Again from equation 2 if the velocity of the palticle in the magnetic field is zero i.e., particle is stationery in magnetic field then it does not experience any force.
- SI unit of strength of magnetic field is tesla (T). It can be defined as follows
  $B=F/qv\sin\theta$
  for F=1N,q=1C and v=1m/s and θ=90
  $1T=1NA^{-1}m^{-1}$

Thus if a charge of 1C when moving with velocity of 1m/s along the direction perpendicular to the magnetic field experiences a force of 1N then magnitude of field at that point is equal to 1 tesla (1T).

Another SI unit of magnetic field is weber/m^2 Thus
1 $Wb\text{-}m^{-2}$=1T=1$NA^{-1}m^{-1}$
In CGS system, the magnetic field is expressed in 'gauss'. And 1T= 10^4 gauss. Dimention formula of magnetic field (B) is [$MT^{-2}A^{-1}$]

3) Lorentz Force

- We know that force acting on any charge of magnitude q moving with velocity **v** inside the magnetic field B is given by
  F=q(**v** X **B**)
  and this is the magnetic force on charge q due to its motion inside magnetic field.
- If both electric field E and magnetic field B are present i.e., when a charged particle moves through a reagion of space where both electric field and magnetic field are present both field exert a force on the particle and the total force on the particle is equal to the vector sum of the electric field and magnetic field force.
  F=q**E**+q(**v** X **B**) (4)
- This force in equation(4) is known as Lorentz Force.
- Where important point to note is that magnetic field is not doing any work on the charged particle as it always act in perpendicular direction to te motion of the charge.

(4)Motion of Charged Particle in The Magnetic Field

- As we have mentioned earlier magnetic force F=(vXB) does not do any work on the particle as it is perpendicular to the velocity.
- Hence magnetic force does not cause any change in kinetic energy or speed of the particle.
- Let us consider there is a uniform magnetic field **B** perpendicular to the plane of paper and directed in downward direction and is indicated by the symbol C in figure shown below.

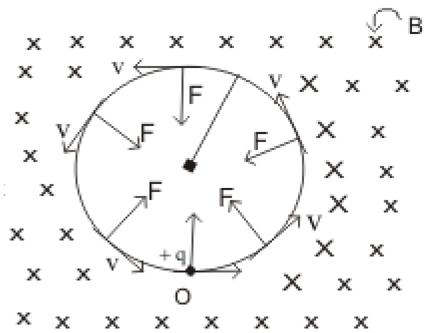


Figure 3. Motion of charged particle in uniform magnetic field

- Now a charge particle +q is projected with a velocity v to the magnetic field at point O with velocity v directed perpendicular to the magnetic field.
- Magnetic force acting on the particle is
  F=q(**v** X **B**) = qvBsinθ
  Since v is perpendicular to B i.e., angle between v and B is θ=90 Thus charged particle at point O is acted upon by the force of magnitude

$F=qvB$
and the direction of force would be perpendicular to both **v** and **B**

- Since the force f is perpendicular to the velocity, it would not change the magnitude of the velocity and the peffect of this force is only to change the direction of the velocity.
- Thus under the action of the magnetic force of the particle will more along the circle perpendicular to the field.
- Therefore the charged particle describe an anticlockwise circular path with constant speed v and here magnetic force work as centripetal force. Thus
$F=qvB=mv^2/r$
where radius of the circular path traversed by the particle in the magnetic in field B is given as
$r=mv/qB$ ---(5)
thus radius of the path is proportional to the momentum mv pof the charged particle.
- $2\pi r$ is the distance traveled by the particle in one revolution and the period T of the complete revolution is
$T=2\pi r/v$
From equation(5)
$r/v=m/qB$
time period T is
$T=2\pi m/qB$ (6)
and the frequency of the particle is $f=1/T=qB/2\pi m$ (7)
- From equation (6) and (7) we see that both time period and frequency does not dependent on the velocity of the moving charged particle.
- Increasing the speed of the charged particle would result in the increace in the radius of the circle. So that time taken to complete one revolution would remains same.
- If the moving charged particle exerts the magnetic field in such a that velocity **v** of particle makes an angle θ with the magnetic field then we can resolve the velocity in two components

$v_{parallel}$: Compenents of the velocity parallel to field
$v_{perpendicular}$:component of velocity perpendicular to magnetic field B

- The component v_{par} would remain unchanged as magnetic force is perpendicular to it.
- In the plane perpendicular to the field the particle travels in a helical path. Radius of the circular path of the helex is

$r=mv_{perpendicular}/qB=mv\sin\theta/qB$

(5) Cyclotron

- Cyclotron is a machine for producing high energy particles ,first developed by E.O.Lawrence and M.S.Livingston in 1931.Figure below shows the path of a charged particle in a cyclotron

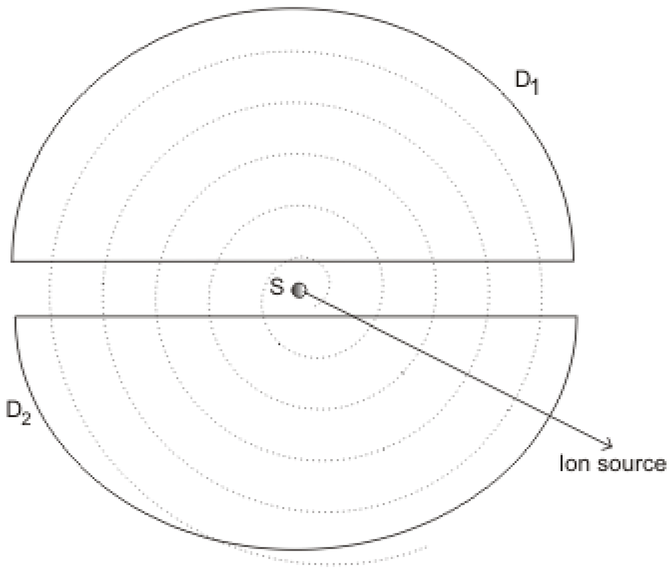


Figure 4. Path of a charged particle in a cyclotron

Construction

- cyclotron consists of two horizontal D-shaped hollow metal segments D_1 and D_1 with a small gap between them
- These D'ees are placed in between the poles of a large electromagnet so that that magnetic field is Perpendicular to the plane of the D'ees
- The whole space inside the D'ees is evacuated to pressure of about 10^{-6} mm of Hg
- An ion source S is kept at the center between the D'ees
- The two D'ees are connected to the terminals of high freqency oscillating A.C circuit.This changes the charge of each D'ees several million time per sec

Theory and working

- Suppose that any instant ,alternating potential is in the direction which makes D_1 positive and D_2 negative
- A positive ion starting from source S will be attracted by the Dee D_2
- Since Uniform magnetic field B acts at right angles to the plane of the Dees ,the positively charged ion of the charge q and mass m will move in a circular motion of radius
  $r=mv/qB$
  where v is the speed of the particle and it is constant
- After traversing half a cycle the ion comes to the edge of D_2.If we adjust the frequency of the oscillator in such a way that by the time ,ion comes to the edge of D_2,potential difference changes direction so as to make D_1 negative and D_2 positive.
- The ion will then get then attracted to D_1 and its speed will increasae due to acceleration
- Once inside D_1 ,the ion is now in electric field free zone and again it will move in a circular path with constant speed which is higher then the previous constant speed in D_2.Radius of the path in D_1 will be larger then D_2
- After traversing the semi -circular path in D_1 ,the ion will come to the egde of D_1 where if the direction of electric field changes ,it will receive additional energy
- This way the ion will continue travelleing in semi circles of increasing radii every time it goes from D_2 to D_1 and from D_1 to D_2
- Time taken by the ion to traverse the semi-circular path in the Dee is given by
  $t=\pi r/v$
- Thus by adjusting the magnetic field B,t can be made the same as that required to change the potential of the D_1 and D_2,so that positive charge ion always crosses the alternating electric field across the gap in correct phase

- Ions gain tremendous amount of energy after traversing through reversal rotation.when they come near the circumference of the Dees,an auxillary electric field is used to deflect them from the circular path to eventually reach a target
- Frequency F of charged particle moving in a cyclotron is
  f=ω/2π =υ/2πr=Bυ/2πm -- (10)

  where υ=1/2t

- If f and B are adjusted to keep charged ion always in phase each time ,the ion crosses the gap .it receive additional energy and at the same time it describes a flat spiral of increasing radius
- KE of ion emerging from the cyclotron if R is radius of the D'e is

$$KE = \frac{1}{2}M\left(\frac{BqR}{M}\right)^2$$
$$= \frac{q^2B^2R^2}{2M} = 2\pi^2R^2f^2M$$

- Above relation shows that the maximum energy attained by the ion is limited by the radius R of the Dees ,magnetic field B or the frequency f
- Maximum energy acquired by the charged particle in a particular cyclotron is independent of the alternating potential i.e when the voltage is small the ion makes a large number of the turns before reaching the periphery and for the large voltage number of turns is small.Total energy remians the same in both the cases,provided both B and R are unchanged
- These days cyclotron are not in wide use but others based on principle of cylotron are used

6) Magnetic force on a current carrying wire

- We know that current flowing in a conductor is nothing but the drift of free electron's from lower potential end of the conductor to the higher potential end
- when a current carrying conductor is placed in a magnetic field ,magnetic forces are exerted on the moving charges with in the conductor
- Equation -1 which gives force on a moving charge in a magnetic field can also be used for calculating the magnetic force exerted by magnetic field on a current carrying conductor (or wire)
- Let us consider a straight conducting wire carrying current I is placed in a magnetic field B(x).Consider a small element dl of the qire as shown below in the figure

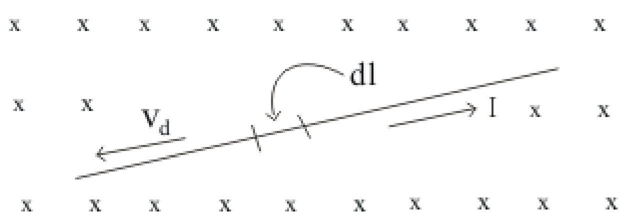


Figure 5.

- Drift velocity of electrons in a conductor and current **I** flowing in the conductor is given by
  $I=neAv_d$
  Where A is the area of cross-section of the wire and n is the number of free electrons per unit volume
- Magnetic force experienced by each electron in presence of magnetic field is
  $\mathbf{F}=e(\mathbf{v_d} \times \mathbf{B})$
  where e is the amount of charge on an electron
- Total number of electron in length dl of the wire
  N=nAdl
- Thus magnetic force on wire of length dl is
  $\mathbf{dF}=(nAdl)(e\mathbf{v_d} \times \mathbf{B})$
  if we denote length dl along the direction of current by the vector **dl** the above equation becomes
  $\mathbf{dF}=(nAev_d)(\mathbf{dl} \times \mathbf{B})$
  or $\mathbf{dF}=I(\mathbf{dl} \times \mathbf{B})$ -- (12)
  where the quantity **IdL** is known as current element
- If a <u>straight</u> wire of length l carrying current **I** is placed in a uniform magnetic field then force on wire would be equal to
  $\mathbf{dF}=I(\mathbf{L} \times \mathbf{B})$ -- (13)

Direction of force

- Direction of force is always perpendicular to the plane containing the current element **IdL** and magnetic field **B**

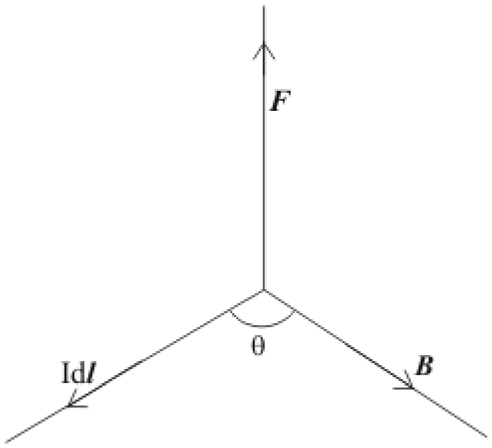


Figure 6.

- Direction of force when current element **IdL** and **B** are perpendicular to each other can also be find using either of the following rules

  i) Fleming'e left hand rule:-
  If fore finger ,the middle finger and thumb of the left hand are stretched in such a way that the all are mutually perpendicular to each other then,if the fore finger points in the direction of the field (B) and middle finger points in the direction of current **I** ,the thumb will point in the direction of the force
  ii) Right hand palm Rule:
  Stretch the finger and thumb of the right hand so that they are perpendicular to each other .If the fingers point in the direction of current **I** and the palm in the direction of field B then the thumb will point in the direction of force

7) Torque on a current carrying rectangular loop in a magnetic field

- Consider a rectangular loop ABCD being suspended in a uniform magnetic field B and direction of B is paralle to the plane of the coil as shown below in the figure

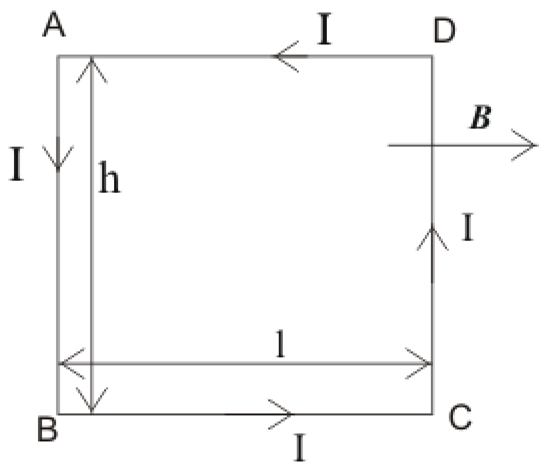


Figure 7.

- Magnitude of force on side AM according to the equation(13) is
  F_{AB}=IhB (angle between I and B is 90^0)
  And direction of force as calculated from the right hand palm rule would be normal to the paper in the upwards direction
- Similarly magnitude of force on CD is
  F_{CD}=ihB
  and direction of F_{CD} is normal to the page but in the downwards direction going into the page
- The forces F_{AB} and F_{CD} are equal in magnitude and opposite in direction and hence they constitute a couple
- Torque τexerted by this couple on rectangular loop is
  τ=IhlB
  Since torque = one of the force * perpendicular distance between them
- No force acts on the side BC since current element makes an angle θ=0 with B due to which the product (**IL**X**B**) becomes equal to zero
- Similary on the side DA ,no magnetic force acts since current element makes an angle θ=180^0 with **B**
- Thus total torque on rectangular current loop is
  τ=IhlB
  =IAB --(15)
  Where A=hl is the area of the loop
- If the coil having N rectangular loop is placed in magnetic field then torque is given by
  τ=NIAB ----(16)
- Again if the normal to the plane of coil makes an angle θ with the uniform magnetic field as shown below in the figure then

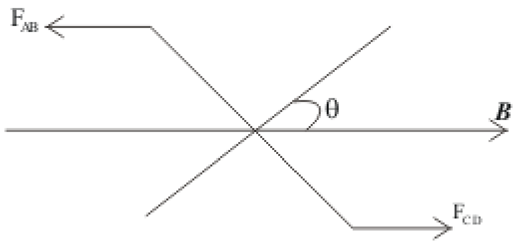


Figure 8.

τ=NIABsinθ

- We know that when an electric dipole is placed in external electric field then torque experienced by the dipole is
  τ=**P** X **E**=PEsinθ
  Where **P** is the electric dipole moment
- comparing expression for torque experienced by electric dipole with the expression for torque on a current loop i.e ,
  τ=(NIA)Bsinθ
  if we take NIA as magnetic dipole moment (m) analogus to electric dipole moment (p),we have
  m=NIA -- (18)
  then
  τ=**m** X **B** -- (19)
- The coil thus behaves as a magnetic dipole
- The direction of magnetic dipole moment lies along the axis of the loop
- This torque tends to rotate the coil about its own axis .Its value changes with angle between the plane of the coil and the direction of the magnetic field
- Unit of magnetic moment is Ampere.meter^2 (Am^2)
- Equation (18) and (19) are obtained by comsidering a rectangular loop but thes equations are valid for plane loops of any shape
-

1) Introduction

- In the previous chapter we have defined concept of magnetic field represented by vector **B**
- We defined magnetic field B in terms of force it exerts on moving charges and on current carrying conductors
- We also know that magnetic field is produced by the motion of the electric charges or electric current
- In this chapter we would study the magnetic field produced by the steady current
- we would study about how various factors affect the magnitude and direction of the magnetic field
- We will also learn to calculate the equation for magnetic field B if the current configuration is known using Biot-savart's law and ampere circuital law

(2)Biot Savart Law

- We know that electric current or moving charges are source of magnetic field
- A Small current carrying conductor of length **dl** (length element) carrying current **I** is a elementary source of magnetic field .The force on another similar conductor can be expressed conveniently in terms of magnetic field **dB** due to the first
- The dependence of magnetic field **dB** on current **I** ,on size and orientation of the length element **dl** and on distance **r** was first guessed by Biot and savart
- The magnitude of the magnetic field **dB** at a distance **r** from a current element **dl** carrying current **I** is found to be proportional to **I** ,to the length dl and inversely proportional to the square of the distance |**r**|
- The direction of the magnetic Field is perpendicular to the line element **dl** as well as radius **r**
- Mathematically, Field **dB** is written as

$$d\boldsymbol{B} = \left(\frac{\mu_0}{4\pi}\right) I \frac{d\boldsymbol{l} \times \hat{\boldsymbol{r}}}{r^2}$$

or,

$$d\boldsymbol{B} = \left(\frac{\mu_0}{4\pi}\right) I \frac{d\boldsymbol{l} \times \boldsymbol{r}^3}{r^2} \qquad (1)$$

Here ($\mu_0/4\pi$) is the proportionality constant such that
$\mu_0/4\pi = 10^{-7}$ Tesla Meter/Ampere(Tm/A)

- Figure below illustrates the relation between magnetic field and current element

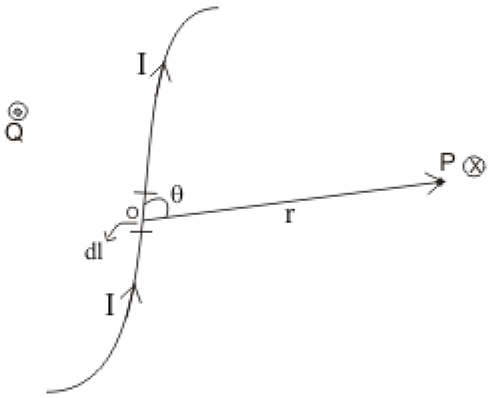


Figure 1. Field at point P is perpendicular to the plane of paper pointing into it

- if in figure, Consider that line element **dl** and radius vector **r** connecting line element mid point to the field point P at which field is to be found are in the plane of the paper
- From equation (1) ,we expect magnetic field to be perpendicular to both **dl** and **r**.Thus direction of **dB** is the direction of advance of right hand screw whose axis is perpendicular to the plane formed by **dl** and **r** and which is rotated from **dl** to **r** (right hand screw rule of vector product)
- Thus in figure ,**dB** at point P is perpendicular directed downwards represented by the symbol (x) and point Q field is directed in upward direction represented by the symbol (•)
- The magnitude of magnetic field is

$$dB = \left(\frac{\mu_0}{4\pi}\right)\frac{I\,|\,\mathbf{dl}\,|\sin\theta}{r^2} \qquad (2)$$

  where θ is the angle between the line element dl and radius vector **r**
- The resultant field at point P due to whole conductor can be found by integrating equation (1) over the length of the conductor i.e.
  B=∫d**B**

  Relation between permeability (μ_0 and permittivity (ε_0) of the free space
- We know that
  $\mu_0/4\pi=10^{-7}$ N/A^2 ----(a)
  and
  $1/4\pi\varepsilon_0=9*10^9$ N-m^2/C^2 ----(b)
  Dividing equation a by b we get
  $\mu_0\varepsilon_0 =1/(9*10^{16})$ (C/Am)^2
  we know that
  1C=1A-s

  So $\mu_0\varepsilon_0 =1/(3*10^8$ m/s$)^2$
  And $3*10^8$ m/s is the speed of the light in free space
  So $\mu_0\varepsilon_0 =1/c^2$
  or $c=1/\sqrt{(\mu_0\varepsilon_0)}$

3) Comparison between Coulomb's laws and Biot Savart laws

- Both the electric and magnetic field depends inversely on square of distance between the source and field point .Both of them are long range forces
- Charge element dq producing electric field is a scalar whereas the current element **Idl** is a vector quantity having direction same as that of flow of current

- According to coulomb's law ,the magnitude of electric field at any point P depends only on the distance of the charge element from any point P .According to Biot savart law ,the direction of magnetic field is perpendicular to the current element as well as to the line joining the current element to the point P
- Both electric field and magnetic field are proportional to the source strength namely charge and current element respectively. This linearity makes it simple to find the field due to more complicated distribution of charge and current by superposing those due to elementary changes and current elements

(4) Applications of Biot Savart law

In this section we will now apply Biot-Savart law as studied in previous section to calculate field B in some important cases

i) Magnetic Field due to steady current in an infinitely long straight wire

- Consider a straight infinitely long wire carrying a steady current I
- We want to calculate magnetic field at a point P at a distance R from the wire as shown below in figure

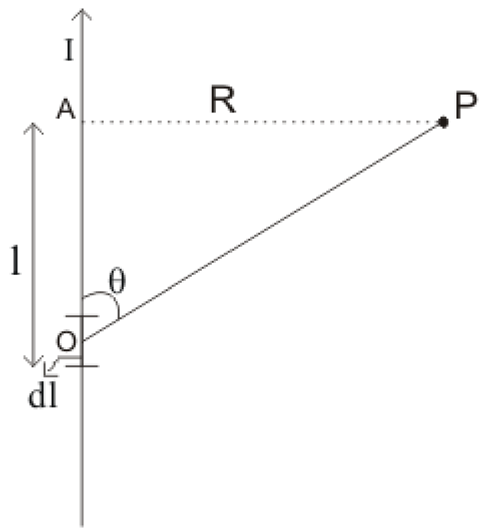


Figure 2. Infinitely long wire carrying current I

- From Biot,-Savart law ,magnetic field d**B** due to small current element of the wire at point O at a distance |**r**|=r from point P is

$$d\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right) I \frac{d\mathbf{l} \times \mathbf{r}^3}{r^2} \qquad (4)$$

- since current element Id**l** and vector **r** makes an angle θ with each other ,the magnitude of the product d**l**X**r** is dlrsinθ and is directed perpendicular to both dl and r vector as shown in the figure

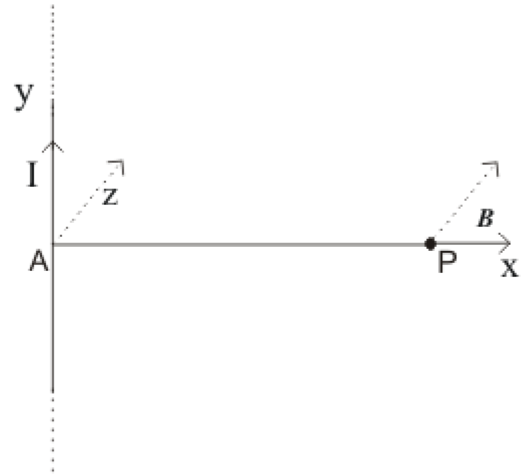


Figure 3. Direction of B is along z-axis perpendicular to the x-y plane in which current element and position vector lies

- Since from our choice of co-ordinate, we found out that field B lies along z-axis therefore we can write

$$d\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\frac{I\,|\,\mathbf{dl}\,|\sin\theta}{r^2}\mathbf{k} \qquad (5)$$

where **k** is the unit vector along z-axis

- we will now express sinθ and r in terms of R which is fixed distance for any point in space and l which describes the position of current element on the infinitely long wire .From figure 1 we have

$$\sin\theta = \frac{R}{(R^2 + l^2)^{1/2}}$$

and $r=(R^2 + l^2)^{1/2}$
Putting these values in the equation (5) we find

$$d\mathbf{B} = \left(\frac{\mu_0}{4\pi}\right)\frac{IdlR}{(R^2 + l^2)^{3/2}}\mathbf{k} \qquad (6)$$

- To find the field due to entire straight wire carrying wire ,we would have to integrate equation (6) **B**=∫d**B**

$$\mathbf{B} = \frac{\mu_0 IR\mathbf{k}}{4\pi}\int_{-\infty}^{\infty}\frac{dl}{(R^2 + l^2)^{3/2}}$$

- To evaluate the integral on the RHS substitute
l=RtanΦ and dl=Rsec^{2}Φ dΦ
Therefore

$$\mathbf{B} = \frac{\mu_0 I}{4\pi R}\mathbf{k}\int_{-\pi/2}^{\pi/2}\cos\phi\, d\phi$$

$$\mathbf{B} = \frac{\mu_0 I}{2\pi R}\mathbf{k} \qquad (7)$$

- From equation (7) ,we noticed that
  i) Magnetic field is proportional to the current **I**
  ii) It is inversely proportional to the distance R
  iii)Magnetic field is in the direction perpendicular to the straight wire and vector **AP**=**R**
- The magnetic line of force near a linear current carrying wire are concentric circles around the conductor in a plane perpendicular to the wire
- Hence the direction of field Bat point P at a distance R from wire, will be along the tangent drawn on a circle of radius R around the conductor as shown below in figure

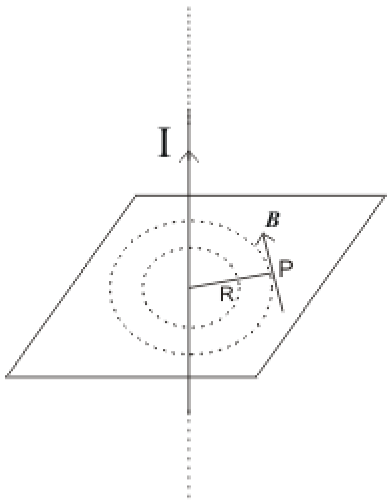


Figure 4. Direction of magnetic field at point P

Direction of B

- Direction of B can be found by right hand thumb rule i.e. grasp the wire with right hand ,the thumb pointing in direction of current ,the finger will curl around the wire in the direction of **B**
- The magnetic field lines are circular closed curve around the wire

Application of Biot Savart law

ii) Force between two long and parallel current carrying conductor

- It is experimentally established fact that two current carrying conductors attract each other when the current is in same direction and repel each other when the current are in opposite direction
- Figure below shows two long parallel wires separated by distance d and carrying currents I_1 and I_2

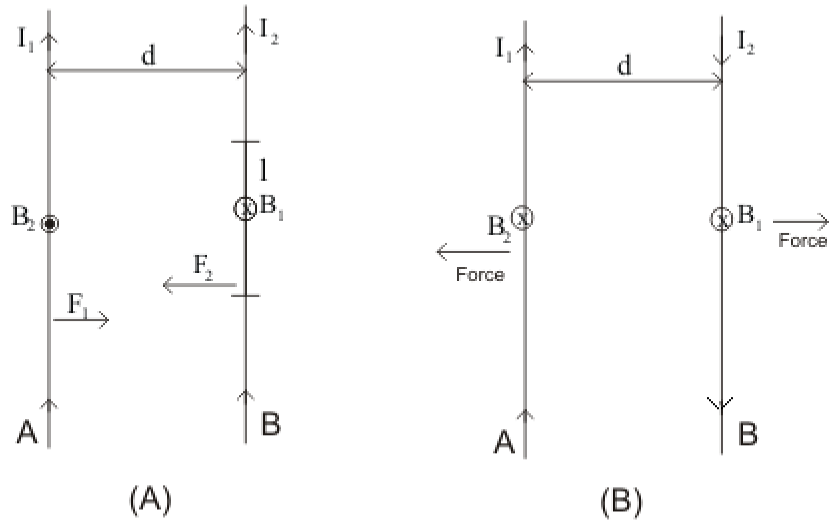


Figure 5. Current carrying wires exerts force on each other

- Consider fig 5(a) wire A will produce a field B_1 at all near by points .The magnitude of B_1 due to current I_1 at a distance d i.e. on wire b is
  $B_1=\mu_0 I_1/2\pi d$ ----(8)
- According to the right hand rule the direction of $\mathbf{B_1}$ is in downward as shown in figure (5a)
- Consider length l of wire B and the force experienced by it will be (I_2**l**X**B**) whose magnitude is

$$F_2 = I_2 lB = \frac{\mu_0}{2\pi}\frac{lI_1I_2}{d} \qquad (9)$$

- Direction of F_2 can be determined using vector rule .F_2 Lies in the plane of the wires and points to the left
- From figure (5) we see that direction of force is towards A if I_2 is in same direction as I_1 fig(5a) and is away from A if I_2 is flowing opposite to I_1 (fig 5b)
- Force per unit length of wire B is

$$\frac{F_2}{l} = \frac{\mu_0}{2\pi}\frac{I_1I_2}{d}$$

- Similarly force per unit length of A due to current in B is

$$\frac{F_1}{l} = \frac{\mu_0}{2\pi}\frac{I_1I_2}{d}$$

  and is directed opposite to the force on B due to A. Thus the force on either conductor is proportional to the product of the current
- We can now make a conclusion that the conductors attract each other if the currents are in the same direction and repel each other if currents are in opposite direction

iii) Magnetic Field along axis of a circular current carrying coil

- Let there be a circular coil of radius R and carrying current I. Let P be any point on the axis of a coil at a distance x from the center and which we have to find the field
- To calculate the field consider a current element I**dl** at the top of the coil pointing perpendicular towards the reader
- Current element I**dl** and **r** is the vector joining current element and point P as shown below in the figure

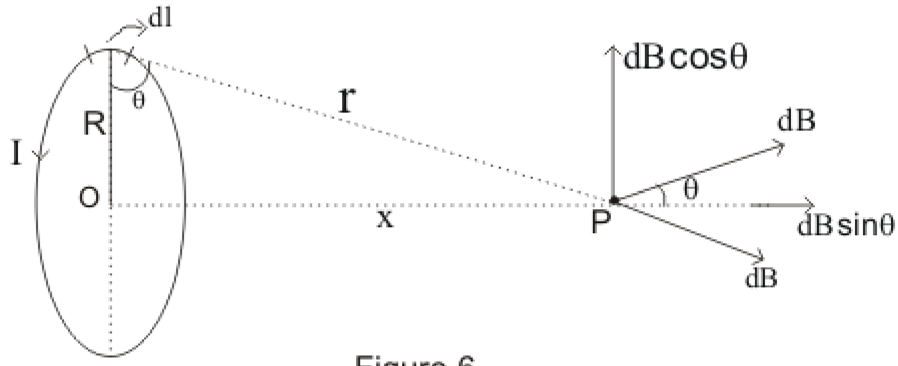


Figure 6.

- From Biot Savart law, the magnitude of the magnetic field due to this current element at P is

$$dB = \frac{\mu_0}{4\pi}\frac{Idl\sin\phi}{r^2} \qquad \text{---(10)}$$

where Φ is the angle between the length element dl and r
- Since I**dl** and **r** are perpendicular to each other so Φ=90.Therefore

$$dB = \frac{\mu_0}{4\pi}\frac{Idl}{r^2} \qquad \text{---------(11)}$$

- Resolving dB into two components we have dBsinθ along the axis of the loop and another one is dBcosθ at right angles to the x-axis
- Since coil is symmetrical about x-axis the contribution d**B** due to the element on opposite side (along -y axis) will be equal in magnitude but opposite in direction and cancel out. Thus we only have dBsinθ component
- The resultant B for the complete loop is given by,
B=∫dB

$$B = \int \frac{\mu_0}{4\pi}\frac{\mathrm{Idl}\sin\theta}{\mathrm{r}^2}$$

Now from figure 6
sinθ=R/r =R/√(R^2 + x^2) So
eq

$$B = \frac{\mu_0}{4\pi}\frac{IR}{(R^2+x^2)^{3/2}}\int_0^{2\pi R} dl$$

or,

$$B = \frac{\mu_0}{4\pi}\frac{IR(2\pi R)}{(R^2+x^2)^{3/2}}$$

or,

$$B = \frac{\mu_0}{2}\frac{IR^2}{(R^2+x^2)^{3/2}} \qquad \text{---(12)}$$

- If the coil has N number of turns then

$$B = \frac{\mu_0}{2}\frac{NIR^2}{(R^2+x^2)^{3/2}} \qquad \text{------(13)}$$

Direction of B

- Direction of magnetic field at a point on the axis of circular coil is along the axis and its orientation can be obtained by using right hand thumb rule .If the fingers are curled along the current, the stretched thumb will point towards the magnetic field
- Magnetic field will be out of the page for anti-clockwise current and into the page for clockwise direction

Field at center of the coil

- At the center of the coil x=0
  so

$$B_{center} = \frac{\mu_0 I R^2}{2R^2} = \frac{\mu_0 I}{2R} \qquad \text{----(14)}$$

Field at point far away from the center x>>>R

- In this case R in the denominator can be neglected hence

$$B = \frac{\mu_0 I R^2}{2x^3} \qquad \text{------(15)}$$

- For coil having N number of turns

$$B = \frac{\mu_0 N I R^2}{2x^3} \qquad \text{---------------(16)}$$

- If the area of the coil is πR^2 then

$$B = \frac{\mu_0 N I A}{2\pi x^3} \qquad \text{--------(17)}$$

- m=NIA represents the magnetic moment of the current coil. Thus from equation (17) we have

$$B = \frac{\mu_0 m}{2\pi x^3} \qquad \text{--------(18)}$$

iv) Magnetic Field at the center of a current carying arc

- Consider an arc of radius R carrying current I as shown below in the figure

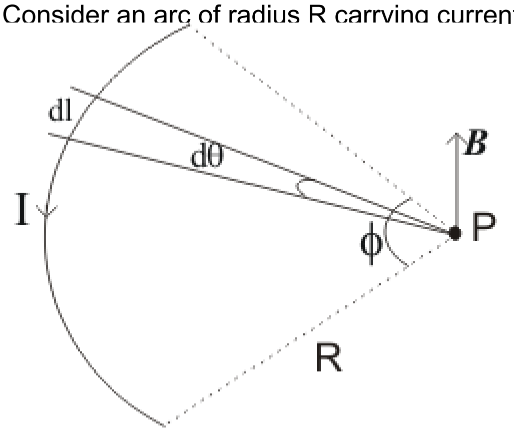


Figure 7.

- According to the Biot Savart law the magnetic field at any point P is given by

$$B = \frac{\mu_0}{4\pi}\int_0^{\phi}\frac{Idl}{R^2}$$

Here dl=RdΦ
So

$$B = \frac{\mu_0}{4\pi}\frac{I\phi}{R} \qquad \text{-------(19)}$$

- If l is the length of the arc then
l=RΦ so that

$$B = \frac{\mu_0}{4\pi}\frac{Il}{R^2} \qquad \text{-------(20)}$$

- Equation 19 and 20 gives us magnetic field only at the center of curvature of a circular arc of current
- For semi circular loop put Φ=π in equation 19 and for full circle Φ=2π in equation 19 and calculate to find the result
- If the circular current loop lies on the plane of the paper then magnetic field will be out of the page for anticlockwise current and into the page for clockwise current as shown below in the figure

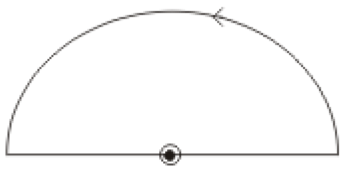


Figure 8a. Current in anticlockwise direction B comes out of the paper

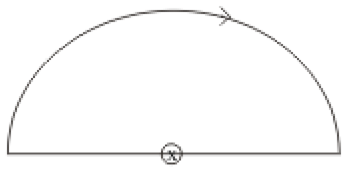


Figure 8b. Current in clockwise direction B goes into the paper

4) Ampere's circuital law

- Ampere's circuital law in magnetism is analogous to gauss's law in electrostatics
- This law is also used to calculate the magnetic field due to any given current distribution
- This law states that
  " The line integral of resultant magnetic field along a closed plane curve is equal to μ_0 time the total current crossing the area bounded by the closed curve provided the electric field inside the loop remains constant" Thus

  $$\oint \mathbf{B}.\mathbf{dl} = \mu_0 I_{enc} \quad \text{-----(21)}$$

  where μ_0 is the permeability of free space and I_{enc} is the net current enclosed by the loop as shown below in the figure

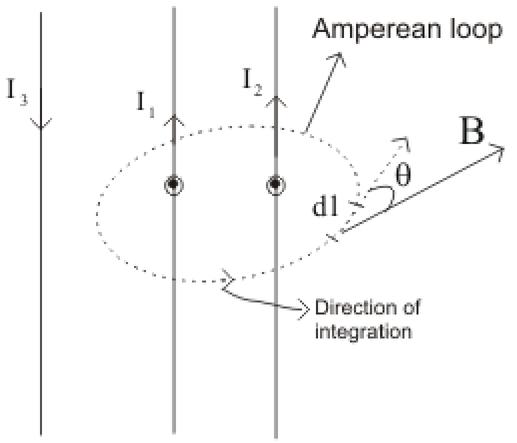


Figure 9. Ampere's law applied to a loop containing two long straight wires.

- The circular sign in equation (21) means that scalar product **B**.**dl** is to be integrated around the closed loop known as Amperian loop whose beginning and end point are same
- Anticlockwise direction of integration as chosen in figure 9 is an arbitrary one we can also use clockwise direction of integration for our calculation depending on our convenience
- To apply the ampere's law we divide the loop into infinitesimal segments **dl** and for each segment, we then calculate the scalar product of **B** and **dl**
- **B** in general varies from point to point so we must use **B** at each location of **dl**
- Amperion Loop is usually an imaginary loop or curve ,which is constructed to permit the application of ampere's law to a specific situation

Proof Of Ampere's Law

- Consider a long straight conductor carrying current I perpendicular to the page in upward direction as shown below in the figure

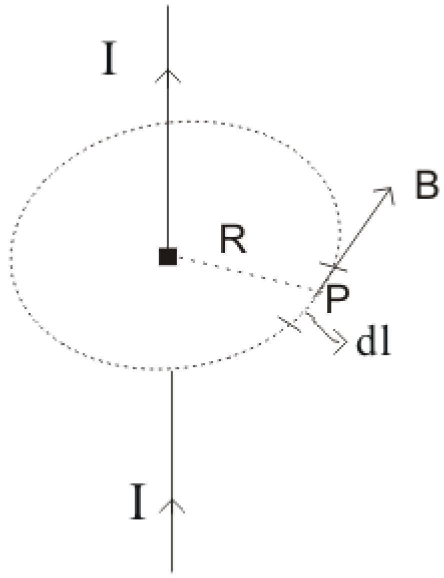


Figure 10. B is the magnetic field due to current carrying conductor at point P

- From Biot Savart law, the magnetic field at any point P which is at a distance R from the conductor is given by

$$B = \frac{\mu_0 I}{2\pi R}$$

- Direction of magnetic Field at point P is along the tangent to the circle of radius R withTh conductor at the center of the circle
- For every point on the circle magnetic field has same magnitude as given by

  $$B = \frac{\mu_0 I}{2\pi R}$$

  And field is tangent to the circle at each point
- The line integral of B around the circle is

$$\oint \mathbf{B.dl} = \oint \frac{\mu_0 I}{2\pi R} dl \quad = \frac{\mu_0 I}{2\pi R} \oint dl$$

since ∫dl=2πR ie, circumference of the circle so,

$$\oint \mathbf{B.dl} = \mu_0 I$$

This is the same result as stated by Ampere law

- This ampere's law is true for any assembly of currents and for any closed curve though we have proved the result using a circular Amperean loop
- If the wire lies outside the amperion loop, the line integral of the field of that wire will be zero

$$\oint \mathbf{B.dl} = 0$$

but does not necessarily mean that **B**=0 everywhere along the path ,but only that no current is linked by the path

- while choosing the path for integration ,we must keep in mind that point at which field is to be determined must lie on the path and the path must have enough symmetry so that the integral can be evaluated

5) Magnetic field of a solenoid

- A solenoid is a long wire wound in a close-packed helix carrying a current **I** and the length of the solenoid is much greater then its diameter
- Figure below shows a section of a stretched out solenoid in xy and yz plane

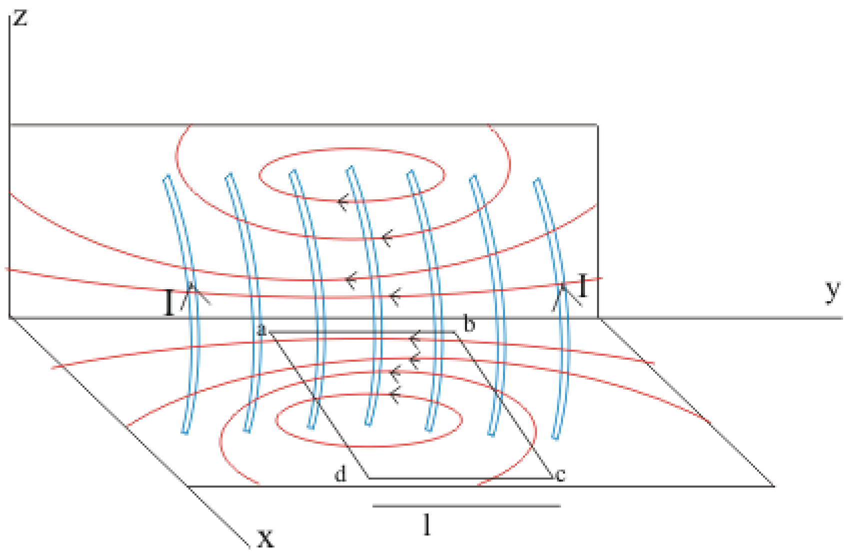


Figure 11. Magnetic field lines surrounding a solenoid

- The solenoid magnetic field is the vector sum of the field produced by the individual turns that make up the solenoid
- Magnetic field B is nearly uniform and parallel to the axis of the solenoid at interior points near its center and external field near the center is very small
- Consider a dashed closed path abcd as shown in figure .Let l be the length of side ab of the loop which is parallel to the is of the solenoid
- Let us also consider that sides bc and da of the loop are very-very long so that side cd is very much far away from the solenoid and magnetic field at this side is negligibly small and for simplicity we consider its equal to 0
- At side a magnetic field **B** is approximately parallel and constant. So for this side $\int B.dl=Bl$
- Magnetic field B is perpendicular to sides bc and da ,hence these portions of the loop does not make any contributions to the line integral as **B.dl**=0 for the side bc and da
- Side cd lies at external points solenoid where **B.dl**=0 as B=0 or negligibly small outside the solenoid
- Hence sum around the entire closed path reduces to Bl
- If N are number of turns per unit length in a solenoid then number of turns in length l is nl.The total current through the rectangle abcd is NIl and from ampere 's law

$Bl=\mu_0 NIl$

or $B=\mu_0 NI$ (22)

- we have obtained this relation for infinitely long solenoids considering the field at external points of the solenoid equal to zero.
- However for real solenoids external field is relatively weak rather then equal to zero
- Thus for actual solenoids relation 22 holds for internal points near the center of the solenoid
- Field at internal points of the solenoid does not depend on length and diameter of the solenoid and is uniform over the cross-section of a solenoid

(6) Magnetic Field of a toriod

- We will now apply Ampere circuital law to calculate magnetic field of a toriod
- A toriodal solenoid is a hollow circular ring with a large number of turns of a wire carrying current wound around the ring
- Suppose we have to find the magnetic field B at a point P inside the toriod as shown below in figure

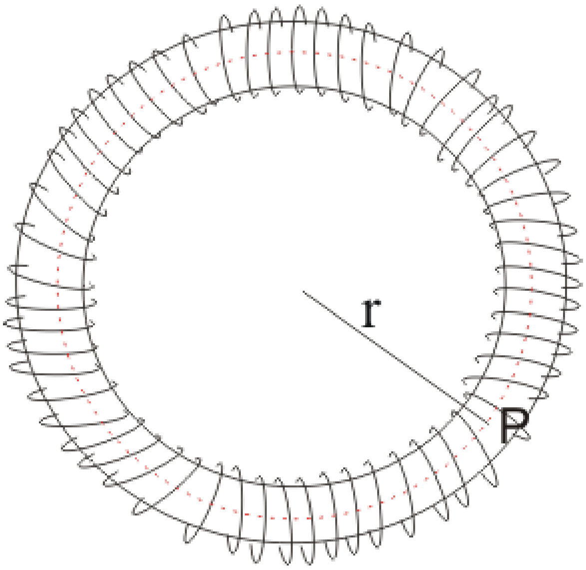


Figure 12. Toroidal solenoid

- In this case amperion loop would be a circle through point P and concentric inside the toriod
- By symmetry field will have equal magnitude at all points of this circle and this field is tangential to every point in the circle
  Thus

$$\oint \boldsymbol{B.dl} = \boldsymbol{B}\int \mathrm{dl} = 2\pi r\mathrm{B}$$

- If there are total N number of turns ,net current crossing the area bounded by the circle is NI where I is the current in the toriod
- using Ampere law

$$\oint \boldsymbol{B.dl} = \mu_0 \mathrm{NI}$$

or,

$$2\pi r\mathrm{B} = \mu_0 \mathrm{NI}$$

or,

$$\mathrm{B} = \frac{\mu_0 \mathrm{NI}}{2\pi r}$$

Thus we see that field B varies with r i.e. field B is not uniform over the cross-section of the core because the path $l=2\pi r$ is longer at the outer side of the section then at the inner side

- Imagine a concentric circle through point P' outside the toriod
- The net current passing through this circular disc is zero ,since the current NI passes in and same current passes out. Thus using Ampere's circuital law, the field B=0 outside the torriod

Unit10

(1) Introduction

- All substances possess magnetic properties and most general definition of magnetism defines it as a particular form of interactions originating between moving electrically charged particles.
- Magnetic interaction relates spatially separate material objects and it is transmitted by means of magnetic field about which we have already studied .This magnetic field is important characteristics of EM form of matter.
- We already know that source of magnetic field is a moving electric charge i.e. an electric current. On atomic scale, there are two types of macroscopic current associated with electrons.
  a) Orbital current is which electron in an atom moves about the nucleus in closed paths constituting electric currents loops
  b) Spin currents related to the internal degrees of freedom of the motion of electrons and this can only be understood through quantum mechanics.
- Like electrons in an atom, atomic nucleus may also have magnetic properties like magnetic moment but it is fairly smaller then that of electrons.
- Magnetic moment **m** is nothing but the quantitative measure of the magnetism of a particle.
- For an elementary closed loop with a current i in it, the magnitude |**m**| of a magnetic moment vector equals the current times the loop area S i.e.
  |**m**|=iS and direction of **m** can be determined using right hand rule.
- All micro structural elements of matter electrons, protons and neutrons are elementary carriers of magnetic moment and combination of these can be principle sources of magnetism
- Thus magnetic properties are inherent to all the substances i.e. they are all magnets
- An external magnetic field has an influence on these atomic orbital and spin currents and two basic effects of an external field are observed

  i) First is **diamagnetic effect** which is consequences of faraday's law of induction. According to the Lenz law's, a magnetic field always sets up an induced current with its magnetic field direction opposite to an initial field .Therefore diamagnetic moment created by the external field is always negative related to this field
  ii) Second effect occurs if there is a resultant non zero magnetic moment in the atom i.e. there is a spin magnetic moment and orbital magnetic moment .In this case external field will attempt to orient the intrinsic atomic magnetic moment in its own direction .As a result a positive moment parallel to the field is created and this is called **paramagnetic moment**.

- Because of the universality of the diamagnetic effect, all substances possess diamagnetic.
- However, diamagnetism is by no means actually observed in all matter. This is because in many instances the diamagnetic effect is masked by the more powerful paramagnetic effect.
- Thus in paramagnetic substances we actually always observe a difference effect produced by the prominent Para magnetism and weaker diamagnetism.

(2) Important terms used in magnetism

(a) Intensity of Magnetization:

- Intensity of magnetization is denoted by the letter I.
- It represents the extent to which the material is magnetized.

- When we place a material in the magnetic field, atomic dipoles of the material tends to align fully or partially in the direction of the field.
- So net magnetic moment is developed in the direction of the field in any small volume of the material.
- Intensity of magnetism is defined as the magnetic moment per unit volume of the magnetized material so,
  $\mathbf{I}=\mathbf{M}/V$ ----(1)
  where **M** is the total magnetic moment within volume due to the magnetizing field.
- Unit of I is Am^{-1}.

(b) Magnetic Field strength

- When a substance is placed in external magnetic field ,the material gets magnetized
- The actual magnetic field inside the material is the sum of external field and the field due to magnetization
- Now we can define a new vector **H** where
  $\mathbf{H}=\mathbf{B}/\mu_0 - \mathbf{I}$ ----(2)
  where **B** is the magnetic field induction inside the substance and **I** is the intensity of magnetization
- Unit of **H** is same as that of **I** i.e Am^{-1}
- CGS unit of **H** is oersted
- In the absence of any material I=0 so
  $\mathbf{H}=\mathbf{B}/\mu_0$ ----(3)

(c) Magnetic Susceptibility

- Magnetic Susceptibility is a measure of how easily a substance is magnetized in a magnetic field
- For paramagnetic and diamagnetic substances ,the intensity of magnetization **I** is directly proportional to the magnetic intensity .Thus
  $\mathbf{I}=\chi\mathbf{H}$ ----(4)
  where proportionality constant χ is known as Magnetic Susceptibility of the material
- Since H and **I** have unit unit so χ is a dimensionless constant and it is a pure number
- value of χ is zero in vacuum as there can no magnetization in vacuum

(d) Magnetic permeability

- Magnetic intensity is given by

  $\mathbf{H}=\mathbf{B}/\mu_0 - \mathbf{I}$
  or
  $\mathbf{B}=\mu_0(\mathbf{H}+\mathbf{I})$
  $=\mu_0(\mathbf{H}+\chi\mathbf{H})$
  $=\mu_0\mathbf{H}(1+\chi)$
  we can also write this as
  $\mathbf{B}=\mu\mathbf{H}$
  where $\mu=\mu_0(1+\chi)$ is a constant called permeability of the material
- μ_0 is the permeability of vacuum as χ=0 for vacuum
- The constant
  $\mu_r=\mu/\mu_0=1+\chi$
  is called the relative permeability of the material

3) Classification of magnetic material

(A) Phenomenological classification

- Such type of classification is based on <u>sign</u> and magnitude of magnetic susceptibility χ
- According to this type of classification there are three type of magnetic material
  i) Diamagnetic materials -> $\chi < 0$
  i.e magnetic susceptibility is negative
  ii) Paramagnetic material -> $\chi > 0$
  i.e magnetic susceptibility is positive and less then unity **iii)** Ferromagnetic material -> $\chi >> 0$
  i.e magnetic susceptibility is positive and is very high
- This approach ignores the nature of microscopic carriers of magnetism and does not consider their interaction
- Through this approach magnetic states like anti-ferromagnetic ,ferromagnetic cannot be recognized

(B) Main Effects of external Field

- Main effects related to the actions of external field on magnetic moments of atomic carriers are
  i) Diamagnetic effects
  ii) Paramagnetic effects
- It was first proposed by the Ampere that the magnetic properties of a material arises due to large number of tiny current loops within the material
- These tiny microscopic current loops are associated with the motion of electrons within the atoms and each current loop has a magnetic moment associated with it
- In addition to the orbital motion of electron around the nucleus electron also spin or rotate about their own axis
- Thus internal magnetic field in a material is produced by electron orbiting around the nucleus and by the spin of the electrons as shown below in the figure .This is how internal magnetism is produced in the material

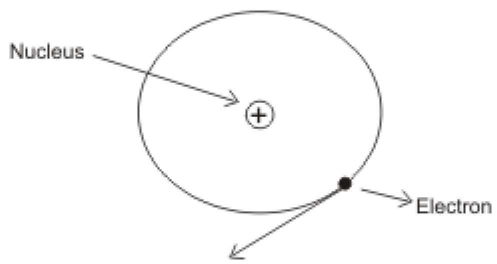


Figure 1(a) Electron orbiting round the nucleus

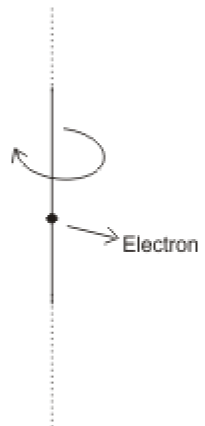


Figure 1 (b) Spin of electron about its own axis

(i) Diamagnetic effects

- Diamagnetic effects occurs in materials where magnetic field due to electronic motions i.e orbiting and spinning completely cancels each other

- Thus for diamagnetic materials intrinsic magnetic moments of all the atoms is zero and such materials are weakly affected by the magnetic field.
- The diamagnetic effects in material is a result of inductive action of the externally applied field on the molecular currents
- To explain the occurrence of this effect ,we first consider the Lenz law accordingly to which, whenever there is a change in a flux in a circuit, an induced current is setup to oppose the change in flux linked by the circuit
- Here the circuit under consideration is orbiting electrons in an atom, ions or molecules constituting the material under consideration
- We know that moving electron are equivalent to current and when there is a current ,there is a flux
- On application of external field ,the current changes to oppose the change in flux and this appear as a change in the frequency of the revolution
- The change in frequency gives rise to magnetization as a result of which each atom will get additional magnetic moment ,aligned opposite to the external field causing it.
- It is this additional magnetic moment which gives diamagnetic susceptibility a negative sign which is order of 10^{-5} for most diamagnetic material (e g. bismith,lead,copper,silicon,diamond etc).
- All substances are diamagnetic ,although diamagnetism may vary frequently be masked by a stronger positive paramagnetic effect on the part of external magnetic field and as a result of internal interactions
- Diamagnetic susceptibility is independent of temperature as effect of thermal motion is very less on electron orbits as long as it deform them

(ii) paramagnetic effect

- Materials having non zero permanent magnetic moment may either be paramagnetic or ferromagnetic but in this section we will only discuss paramagnetic effects
- Para magnetism occurs in material where the magnetic field produced by orbital and spin motion of the electron do not cancel each other completely
- Materials showing paramagnetic effects in the presence of external magnetic field have permanent magnetic moment of the atoms
- In the absence of external field paramagnetic material have magnetic moments but they are oriented randomly.
- These moments (both due to spin and orbital motion of electron) experience an orienting effect in the presence of externally applied magnetic field
- Due to this orienting effect material gets magnetized parallel to the external applied field resulting in positive paramagnetic susceptibility
- This alignment of atomic magnetic moments in paramagnetic substance is opposed by the thermal motion of the atoms ,so alignment increases with the decrease in temperature and increase in strength of applied magnetic field
- Thus there is a sharp dependence of paramagnetic susceptibility on temperature
- For paramagnetic substance magnetic susceptibility is of the order of 10^{-5} to 10^{-3} and it is temperature dependent

4) concept of ferromagnetism

- Ferromagnetism is the existence of the spontaneous magnetization ,even in the absence of an external magnetic field
- Internal magnetic field in ferromagnetism may be hundred or thousand times greater than that of diamagnetic and paramagnetic material
- Relation between I and H magnetization intensity and magnetic field is not linear. I and H are no longer have direct proportionality in case of ferromagnetic materials .hence magnetic susceptibility is very large but no longer constant
- Even in the absence of external field some ferromagnetic material exhibits large magnetization and can become permanent magnetized

- Some of the elements exhibiting ferromagnetic properties at room temperature are iron ,nickle,cobalt and gadolinium
- Because of complicated relation ship between I and H in case of ferromagnetic material, it is not possible to express I as a function of H
- So when a piece of unmagnetized iron is brought near a magnet or is subjected to the magnetic field of an electric current, the magnetization induced in iron by the field is described by a magnetization curve obtained by plotting the intensity of magnetization I against the field strength H
- Ferromagnetism can occur only in paramagnetic material i.e molecule and atoms of a ferromagnetic material also has unpaired electrons and hence non -zero permanent magnetic moment
- all ferromagnetic materials are composed of many small magnets or domains ,each of which consists of many atoms within a domain. Size of a domain is usually microscopic
- Within the domain, all magnetic moments are aligned ,but the alignment of magnetic moments varies from domain to domain which result in zero net magnetic moment of the macroscopic piece of material as a whole shown below in fig 2(a)

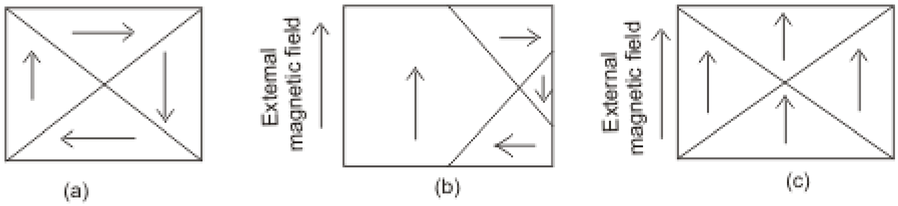


Figure 2

- when the substance is placed in an external field ,magnetism of substance can increase in two different ways
  i) By the displacement of the boundaries of the domain where domains oriented favorably with respect to the external field increase in and those oriented opposite to the external field are reduced in size as shown in fig 2(b)
  ii) By the rotation of domain that is the domain rotate until their magnetic moments are aligned more or less in the direction of the externally applied magnetic field
- In presence of week magnetic field material is magnetized mostly by the displacement of the domains and in presence of strong fields magnetization takes place mostly by the rotation of the domains.
- In case of ferromagnetic materials on removal of the external magnetic field ,material is not completely demagnetized and some residual magnetization remains in it
- Every ferromagnetic material has a critical temperature known as curie temperature (T_c) above which material becomes paramagnetic and this transition of material from ferromagnetic to paramagnetic is a phase change or phase transition analogous to those between solid, liquid and gaseous phases of the matter

(5) Hysteresis

- We have already mentioned that in case of ferromagnetic materials ,the relation between I and H is not linear
- This relation can even depend on the history of the sample i.e whether it has been previously magnetized or not
- when we place a ferromagnetic material in the magnetic field it gets magnetized by induction
- If the field strength is first increased from a zero to high value and then decreased again, it is observed that the original curve is not retraced, the induction lags behind and follows a characteristics curve. This phenomenon is known as Hysteresis and characteristics curve is known as hysteresis loop as shown below in the curve

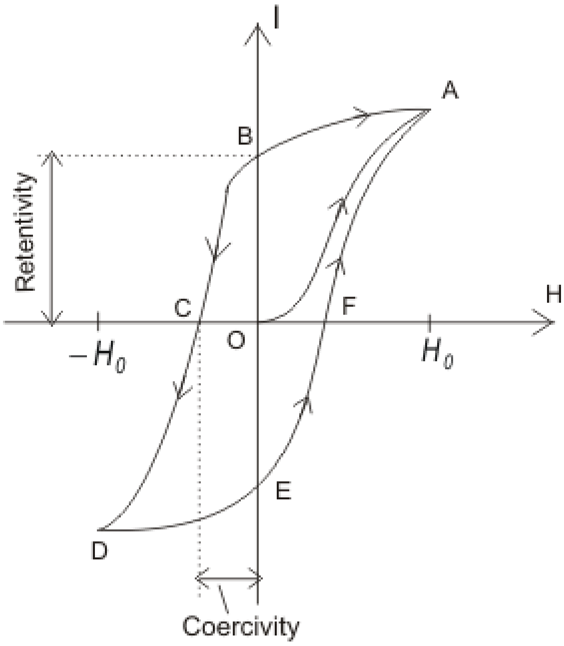


Figure 3:- Variation of I with H

- Figure 3 shows the variation of I and H.In the beginning I= 0 and H=0 as represented by the point O in the figure. At this instant the sample is in unmagnetized state
- As value of H is increased, I also increases non uniformly .If we increase H indefinitely the intensity of magnetization of ferromagnetic material approaches finite limit known as saturation
- Thus at $H=H_0$,the magnetization becomes nearly saturated and magnetization and magnetization varies along path OA
- Now if we begin to decrease the value of magnetic field ,the magnetization I of the substance also begin to decrease but this time not following the path AO but following a new path AB
- when H becomes equal to zero, I still have value equal to OB,This magnetization remaining in substance when magnetizing field becomes equal to zero is called the residual magnetism and the remaining value of I at point B is known as retentively of the material
- To reduce I to zero, we will increase field H in reverse direction and the magnetization I decreases following curve BC where at point C,I becomes equal to zero where H=OC
- The value OC of the magnetizing field is called coercivity of the substance
- the coercivity of the substance is a measure of the reverse magnetizing field required to bring magnetization I equal to zero
- if we further increase H beyond OC the sample begins to get magnetized in reverse direction ,again getting saturated at D at $H=-H_0$
- while taking back H from its negative value through zero to its original maximum positive value H_0,we symmetrical curve DEFA
- Thus we see that if the field strength is first increased from zero to saturated and then decreased again, it is observed that original curve is not retraced, the induction lags behind the field and follows a characteristic curve .This phenomenon is known as hysteresis and the characteristic curve (Here ABCDEFA) is known as Hysteresis loop

UNIT4

ELECTRO-MAGNETIC INDUCTION AND ALTERNATING CURRENT
Introduction

- While studying magnetism we learned that electricty and magnetism are interrelated and infact moving charges or electric currents produces magnetic field that can deflect magnetic compass needle
- A question arises,can moving magnetic field produce electricty .Answer is yes and it was first showed by a british scientist Michael faraday in 1931 who after performing various experiments found that moving magnetic field can give rise to the EMF
- Independently the effect was discovered by Joseph Henry in USA at about the same time
- In this chapter we will discuss about the electric and magnetic field changing with time
- More precisely we will consider the phenomenon related to time changing current or time changing magnetic fields

(2) Faraday's experiment

- faraday in 1831 first discovered that whenever the number it of magnetic lines of forces in a circuit changes ,a emf is produced in the circuit and is known as induced emf and this phenomenon is known as Electro Magnetic Induction
- If the circuit is closed one then a current flows through it which is known an induced current
- This induced emf and current lasts only for the time while magnetic flux is chnaging
- We now illustrate two examples of the sort that faraday and henry performed

(i) experiment I

- Figure below shows a closed circuit containing coil of insulated wire.
- Also note that circuit does not contain any source of emf so there is no deflection in the galvanometer

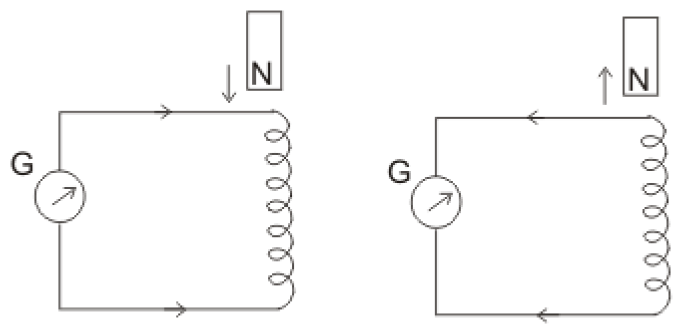


Figure 1. North pole of the magnet moving towards and away from closed circuit containing galvanometer in Faraday's experiment

- If we move bar magnet towards the coil keeping the coil stationary with north pole of the magnet facing the coil (say) then we notice deflection in needle of the galvanometer indicating the pressence of the current in the circuit
- This deflection observed is only for the time interval during which the magnet is in motion
- Now if we begin to move the magnet in the opposite direction then the galvanometer needle is now deflected in the opposite direction
- again if we move the magnet towards the coil ,with its south pole facing the coil,the deflection is now in opposite direction,again indicating that the current now setup in the coil is in reverse direction to that when the north pole faces the wire

- A deflection is also oberved in galvanometer when the magnet is held stationary and circuit is moved away from the magnet.
- It is further observed that faster is the motion of magnet ,larger is the deflection in the galavanometer needle.
- From this experiment faraday convinced that magnet moving towards the coil one way has the same effect moving coil towards the magnet the other way.

(ii) experiment -2

- Figure-2 given below shows a primary coil P connected to the battery and a secondary coild connected to the galvanometer

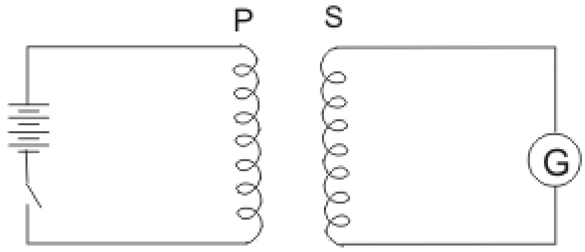


Figure 2. This experiment shows that current flowing in primary coil generated current in secondary coil

- Now we have replaced magnet of the previous experiment with a current carrying coil and expect to observe similar effect as current carrying coil produces magnetic field.
- The motion of either of the coils shows deflection in the galvanometer.
- Also galvanemter shows a sudden deflection in one direction when current was started in primary coil and in the opposite direction when the current was stopped.

3) Magnetic Flux

- The flux of magnetic field through a surface is defined in a similar manner as we defined flux in the electric field
- if dA be an area element on any arbitrary surface and **n** be the unit vector perpendicular to the area element then magnetic flux is defined as
  $\phi = \int \mathbf{B.n}dA$ ---(1)
- If the surface under considerations is a plane of area A and magnetic field is constant in both the magnitude and direction over the surface and θ is the angle between magnetic field **B** and unit normal to the surface then magnetic flux is given by
  $\Phi = BA\cos\theta$ --(2)
- Unit of flux is Wb(weber)
  $1Wb = 1T.m^2$

(4) Faradays law of electromagnetic induction

- From experiments on EM induction, faraday came to conclude then an emf was being induced in the coil when the magnetic flux linked to it was being changed either by
  (i) Moving magnet close or away from the coil(experiment 1)
  (ii) Chnaging amouny of current in the primary coil or motion of either of the coils relative to each other (experiment 2)
- From these observations faraday enunciated an important law. The magitude of the induced emf produced in the coil is given by
  $|\xi| = \frac{d\phi}{dt}$ ---(3)
  where Φ is the magnetic flux as given by equation (1) .ξ is expressed in volts

- If the circuit is a tightly wound coil of N turns then total flux linked by N turns coil is NΦ.So induced EMF is whole coil is
  $$|\xi| = \frac{Nd\phi}{dt} \qquad ---(4)$$
  Above equation only give the magnitude of the induced EMF and it does not give its direction.
- we know that unit of magnetic flux is 1 weber=1 T.m^2
- If magnetic flux linking a single turn changes at a rate of 1weber per second then Induced emf=Magnetic flux/time=1 weber/sec Now since weber=Nm/Amp
  Hence 1 weber/sec= 1 N.m/A.s =1 Joule/coulumb
- Hence the unit of emf is volt.

(5) Direction of Induced EMF: Lenz's Law

- The direction of induced current and emf is given by lenz's law
- Due to change in magnetic flux through a closed loop an induced current is established in the loop
- Lenz's law states that
  'The induced current due to the induced emf always flow in such a direction as to oppose the change causing it'
- Now we can combine faraday's law as given in equation (4) to find the direction of emf
- Thus we can say that **"The emf induced in a coil is equal to the negative rate of the change of the magnetic flux linked with it "**
  $$\xi = -\frac{Nd\phi}{dt} \qquad ---(5)$$

Explaination of lenz's law

>

- To explain this law again consider faradays experiment 1 in which north pole of the magnet moves towards a closed coil
- This movement of north pole of magnetic induces current in the coil in such a direction so that end of the coil ,facing and approaching north pole becomes a magnetic north pole
- The repulsion between two poles opposes the motion of the magnet towards the coil
- Thus work has to be done to push the magnet against the coil
- It is this mechanical work which causes the current to flow in the coil against its resistance R and supply the energy for the heat loss
- The mechanical workdone is converted to electrical energy which produces the heat energy
- If the direction of the induced current were such as not to oppsose the motion ,then we would be obtaining electrical energy continously without doing any work ,which is impossible
- So ,every things seems to be all right if we accept lenz's law otherwise the principle of conservation of the energy would be voilated
- Direction of the induced current can be found using Fleming right hand rule.
  " If we stretch thumb ,index and middle finger perpendicular to one another then index fingers points in direction of the magnetic field ,middle finger in direction of induced current and thum points in direction of the motion of the conducter"

6) Motional EMF

- We know that an emf is produced in the loop when the amount of magnetic flux linked with the circuit os changed.
- The flux Φ linked with the loop can be changed by
  (i) Keeping the loop at rest and changing the magnetic field i.e , there is no physical movemnt

of either the source of emf or the loop(or coil) through which the magnetic flux is linked but the magnetic field changes with time and this may caused by changing the electric current producing the field
(ii) Keeping the magnetic field constant and moving the loop or source of the magnetic field partly or wholey i.e the change is produced by the relative motion of the source of the magnetic field and the lopp (or coil) through which the magnetic field passes.

- In both the cases of producing the emf,the induced emf is given by the same law i.e it is equal to the time rate of change of the magnetic flux.
- In the later case emf induced due to the relative motion of source of magnetic field and coil is called motional emf.
- This phenomemon of motionla emf can be understood easily in lorentz force on moving charges
- Consider a thin conducting rod AB length l moving in the magnetic field **B** with constant velocity **v** as shown in the below in the figure

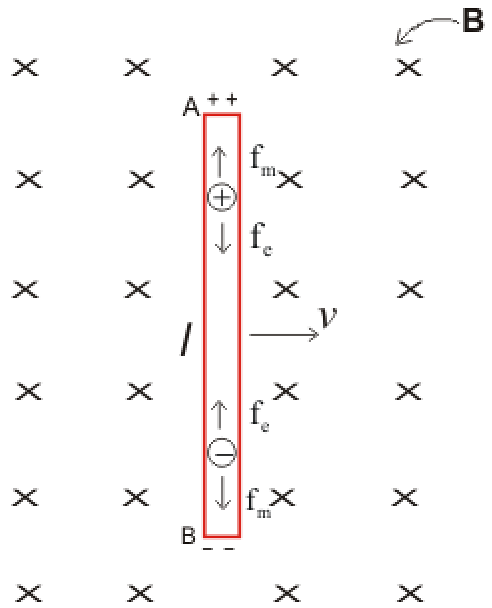
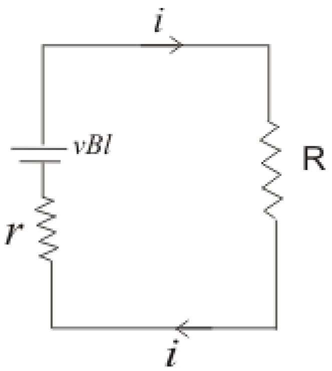


Figure 3. Conducting rod in uniform magnetic field

- This uniform magnetic field B is perpendicular to the plane of the diagram ,directed away from the reader
- Rod is moving in the magnetic field in such a way that its velocity is perpendicular to the magnetic field **B** and its own axis
- when we move conducting rod AB with velocity v,free electrons in it gain velocity in the direction of the motion and in the presence of magnetic field ,these free electrons experiences lorentz force perpendicular to both **B** and **v**
- The electrons under this force accumulate at end end ,providing it negative polarity and the other end deprived of the electrons becomes positively charged
- Magnetic force or lorentz force F_m acting on these moving electrons is $\mathbf{F_m}=q\mathbf{v} \times \mathbf{B}$
  where $q=-1.6*10^{-19}$ C ,charge on each electron
- According to fleming left hand rule force on negative charges is towards B hence negative charge accumulates at B and positive charge appears at A
- As charges accumulates at the ends of the rod,an electric field E is produced in the rod from A to B.This field E is of non electrostatics origins and is produced by the changing magnetic fields
- This electric field in turns produces a force on electron in the conducting rod which is opposite to the lorentz force so
  $\mathbf{F_e}=q\mathbf{E}$
- when enough charge accumulates a situation comes when this electric force cancels out lorentz force and then free electrons do not drift anymore.In this situation
  $F_m = F_e$ or $|q\mathbf{v} \times \mathbf{B}|=|q\mathbf{E}|$ or $vB=E$

- In this situation there is no force acting on the free electrons of the rod AB. Potential difference between the ends A and B of the rod would be
  V=El=vBl This is the emf induced in the rod AB due to its motion in the magnetic field
- Thus motional emf induced in the rod moving in magnetic field is ξ=vBl ----(6)
- If the velocity v of the rod makes an angle θ with the direction of the magnetic field ,the potential difference induced between the ends of the conductor will be
  ξ=vBlsinθ
  as vsinθ is the component of v perpendicular to B
- If the rod moves parallel to the field i.e, θ =0 no potential difference will be induced
- The emf associated with the moving rod in Figure 3 is analogous to that of a battery with the positive terminal at A and negative terminal at B
- If ends of A and B are connected by an external resistor and if r is the internal resistance of the rod then



  an electric field is produced in this resitor due to the potential differebce and a current is established in the circuit with direction from A to B in the external circuit
- Since magnitude of current of induced emf is
  ξ=vBl
  So current is,
  $$i = \frac{vBl}{r + R} \qquad ----(7)$$
  and direction of current can be found using lenz's law
- We now know the current flowing in the circuit from this we can calculate the power loss and force F connected with this motion Thus
  $$P = I^2 R = \frac{B^2 l^2 v^2}{(r + R)^2} R$$
  if r <<< R then it can be neglected so
  $$P = \frac{B^2 l^2 v^2}{R} \qquad ---(8)$$
  and Force
  $$F = \frac{P}{v} = \frac{B^2 l^2 v}{R} \qquad ---(9)$$

7) Induced Electric Fields

- In the earliar section we have studied that when a conductor moves in a magnetic field induced current is generated

- Now consider a situation in which conductor is fixed in a time varying magnetic field .In this situation magnetic flux through the conducting loop changes with time and an induced current is generated
- Figure below shows a soleniod encirceled by a conducting loop with a small galvanometer

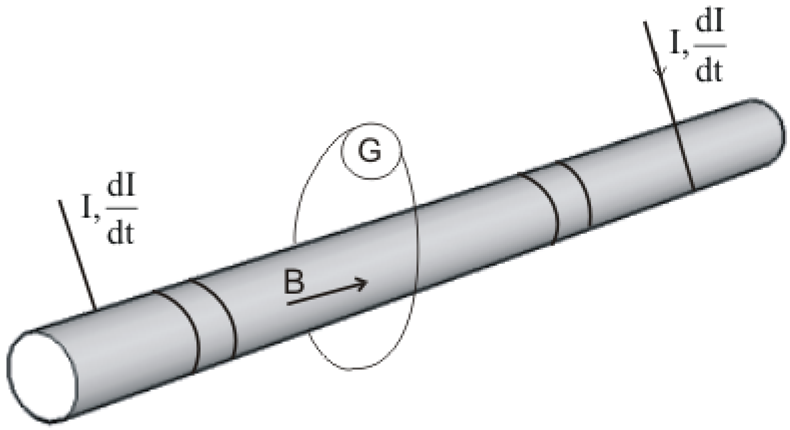


Figure 4. Solenoid with I flowing through its windings and current is changing at a rate dI/dt

- Current **I** through the soleniod sets up a magnetic field B along its axis and a magnetic flux Φ passes through the surface bounded by the loop
- Now when the current **I** through the soleniod changes ,the galvanometer deflects for the time during which the flux is changing .This indicates that an emf is induced in the conductor.
- From Faraday's law this emf is given by the relation eq Here as we earliar stated that the conductor is sationary and the flux through the loop is chnaging due to the magnetic field varying on time
- Since charges are at rest (v=0) so magnetic forces $\mathbf{F_m}=q(\mathbf{v} \times \mathbf{B})$ cannot set the charges to motion.Hence induced current in the loop appears becuase of the presence of an electric field **E** in the loop
- It is this electric field **E** which is responsible for the induced emf and hence for the current flowing in a fixed loop placed in a magnetic field varying with time
- This electric field produced here is purely a field of nonaelectrostatic origin i.e it originated due to the magnetic field varying with time, and induced emf may be defined as the line integral of this non-electrostatic field .Thus,

$$\xi = \oint \mathbf{E}.d\mathbf{l}$$

- Using faraday law

$$\xi = -\frac{d\phi}{dt}$$

So,

$$\oint \mathbf{E}.d\mathbf{l} = -\frac{d\phi}{dt} \qquad \text{---(10)}$$

- From equation (10) we see that line intregal of electric field induced by varying magnetic field differs from zero.This means we can not define a electrostatic potential corresponding to this field.
- Hence this electric field produced by changing electric field is non-electrostatis and non-conservative in nature.
- We call such a field as induced electric field.

(1) Introduction

- Before defining inductance first of all, we will define an inductor.
- Like capacitor, inductor is another component commonly in electronic circuits.
- An inductor consists of a coil wound on a core or former of a suitable material like solid or laminated iron. core or ferrites which are highly ferromagnetic substances.
- when a current through an inductor changes am emf is induced in it which opposes this change of current in the inductor.
- This property of inductor or coil due to which it opposes change of current through it called the inductance denoted by letter L.
- Unit of inductance is henry(H).

(2) Self Inductance

- Consider the figure given below

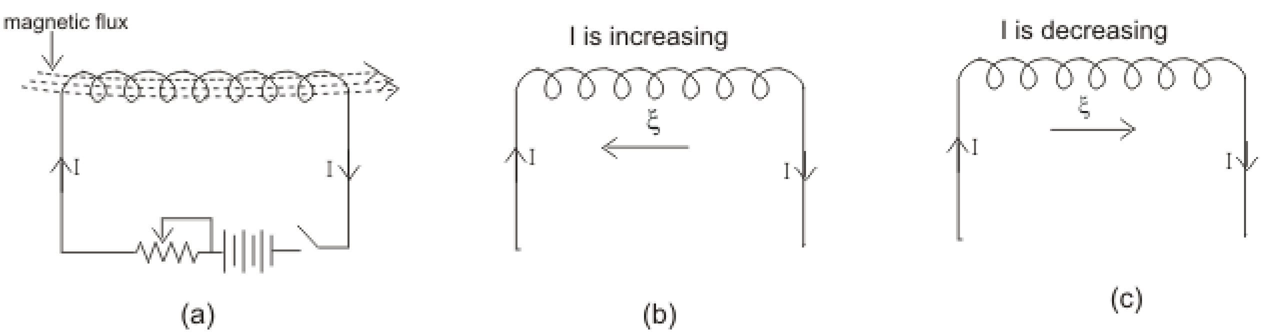


Figure 1. When current increases direction of induced emf is opposite to direction of current (b) and in case of decreasing current direction of induced emf is same as direction of current

- When we establish a current through an inductor or coil, it generates a magnetic field and this result in a magnetic flux passing through the coil as shown in figure 1(a).
- If we vary the amount of current flowing in the coil with time, the magnetic flux associated with the coil also changes and an emf ξ is induced in the coil.
- According to the Lenz's law, the direction of induced emf is such that it opposes its cause i.e. it opposes the change in current or magnetic flux.
- This phenomenon of production of opposing induced emf in inductor or coil itself due to time varying current in the coil is known as self induction.
- If I is the amount of current flowing in the coil at any instant then emf induced in the coil is directly proportional to the change in current i.e.

$$\xi \propto \left(\frac{-dI}{dt}\right)$$

or

$$\xi = -L\frac{dI}{dt} \qquad \text{--(1)}$$

where L is a constant known as coefficient of self induction.

- If (-dI/dt)=1 then ξ=L
  Hence the coefficient of self induction of a inductor or coil is numerically equal to the emf induced in the coil when rate of change of current in the coil is unity.
- Now from the faraday's and Lenz's laws induced emf is

$$\xi = -\frac{d\phi}{dt} \qquad \text{---(2)}$$

comparing equation 1 and 2 we have,

$$L\frac{dI}{dt} = \frac{d\phi}{dt}$$

or Φ=LI

- Again for I=1, Φ=L
  hence the coefficient of self induction of coil is also numerically equal to the magnetic flux linked with the inductor carrying a current of one ampere
- If the coil has N number of turn's then total flux through the coil is
  Φ_{tot}=NΦ
  where Φ is the flux through single turn of the coil .So we have,
  Φ_{tot}=LI
  or L=NΦ/I
  for a coil of N turns
- In the figure given below consider the inductor to be the part of a circuit and current flowing in the inductor from left to right

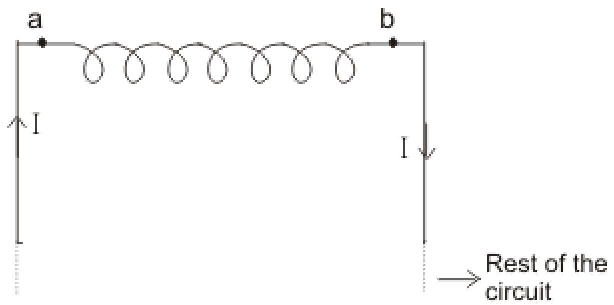


Figure 2. Inductor as the part of a circuit

- Now when a inductor is used in a circuit, we can use Kirchhoff's loop rule and this emf(Self induced emf) can be treated as if it is a potential drop with point A at higher potential and B at lower potential when current flows from a to b as shown in the figure
- We thus have
  V_{ab}=LdI/dt

(3) Self induction of a long solenoid

- Consider a long solenoid of length l, area of cross-section A and having N closely wound turns.
- If I is the amount of current flowing through the solenoid them magnetic field **B** inside the solenoid is given by,

$$B = \frac{\mu_0 NI}{l}$$

- Magnetic flux through each turn of the solenoid is,

$$\phi = BA = \frac{\mu_0 N^2 AI}{l}$$

$$\text{but}, \phi = LI$$

$$\text{So}, LI = \frac{\mu_0 N^2 AI}{l}$$

or, coefficent of self induction

$$L = \frac{\mu_0 N^2 A}{l} \qquad (3)$$

(4) Energy in an inductor

- Changing current in an inductor gives rise to self induced emf which opposes changes in the current flowing through the inductor.
- This self inductance thus plays the role the inertia and it is electromagnetic analogue of mass in mechanics.
- So a certain amount of work is required to be done against this self induced emf for establishing the current in the circuit.
- In order to do so, the source supplying current in a circuit must maintain Potential difference between its terminals which is done by supplying energy to the inductor.
- Power supplied to the inductor is given by relation
  P=ξI ---(4)
  where

  $$\xi = -L\frac{dI}{dt}$$

  L is Self inductance and
  dI/dt is rate of change of current I in the circuit.
- Energy dW supplied in time dt would be
  dW=Pdt
  =LI(dI/dt) dt
  =LIdI
  and total energy supplied while current I increases from o to a final value I is

  $$W = L\int_0^I IdI = \frac{1}{2}LI^2 \qquad \text{----(5)}$$

- Once the current reaches its final value and becomes steady ,the power input becomes zero.
- The energy so far supplied to the inductor is stored in it as a form of potential energy as long as current is maintained.

- When current in circuit becomes zero, the energy is returned to the circuit which supplies it.

1) Introduction

- We have already discussed about direct current (DC) which is produced by the voltage source whose pole does not change their polarity with time
- Hence direction of flow of direct current does not changes with time
- Alternating current on the other hand is produced by voltage source whose terminal polarity keeps alternating with time i.e. terminal which was positive at one instant of time becomes negative some time later and vice -versa
- Due to changing polarity of voltage source direction of flow of current also keep changing
- In this chapter we would learn how voltage and current changing with time are related to each other in various circuits with components namely resistors, capacitor and inductor.

(2) Alternating current and Alternating EMF

- An alternating current is one whose magnitude changes sinusoidal with time .Thus alternating current is given by

  $$i = i_0 \sin(\omega t + \phi) \qquad \text{-----(1)}$$

  Where
  i_0=current amplitude or peak value of alternating current
  If T is the time period of alternating current and f is the frequency, then

  $$\omega = \frac{2\pi}{T} = 2\pi f \qquad \text{-----(2)}$$

  Where ω is called angular frequency of A.C and φ is known as phase constant
- Instead of sine function AC can also be represented by cosine function and both representation leads to same results. We will discuss circuits with sine representation of A.C
- Figure below shows the variation of A.C with time

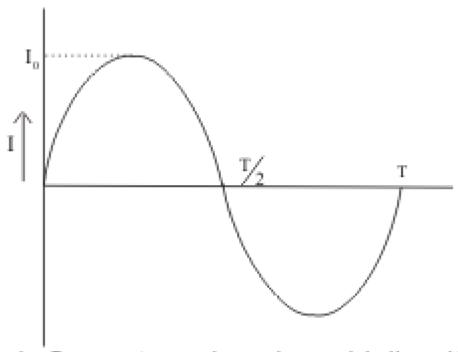


Figure 1. Current varying sinusoidally with time

- Complete set of variations of the current in one time period T is called cycle

- The emf or voltage whose magnitude changes sinusoidal with time is known as alternating emf and is represented by

  $$V = V_0 \sin(\omega t + \phi) \qquad \text{----(3)}$$

  where V_0 is the peak value of alternating current.

(3) Average or mean current

- When an alternating current passed through a moving coil galvanometer it shows no deflection ,this is because for one complete cycle mean value of alternating current is zero as AC flows in one direction during one half cycle and in opposite direction during another half cycle.
- But mean value of A.C is finite over half cycle.
- So, mean or average value of AC is defined either for positive half cycle or for negative half cycle
- So,

$$i_{avg(T/2)} = \frac{\int_0^{T/2} i dt}{\int_0^{T/2} dt} = \frac{\int_0^{T/2} i_0 \sin(\omega t + \phi)}{\int_0^{T/2} dt} = \frac{2i_0}{\pi} \cong .636 i_0 \qquad \text{----(4)}$$

- From equation (4),we see that the average value of A.C during the half cycle is .636 times or 63.6% of its peak value
- Similarly we can show that

$$V_{avg(T/2)} = \frac{2V_0}{\pi} \cong .636 V_0$$

- During next half cycle mean value of ac will be equal in magnitude but opposite in direction.
- Always remember that mean value of AC over a complete cycle is zero and is defined over a half cycle of AC.

4) Root Mean square value of AC

- We know that time average value of AC over one cycle is zero and it can be proved easily
- Instantaneous current I and time average of AC over half cycle could be positive for one half cycle and negative for another half cycle but quantity i^2 would always remain positive
- So time average of quantity i^2 is

$$\bar{i}^2 = \frac{\int_0^T i^2 dt}{\int_0^T dt}$$

$$= \frac{1}{T}\int_0^T i_0^2 \sin^2(\omega t + \phi)dt$$

$$= \frac{i_0^2}{2T}\int_0^T [1 - \cos 2(\omega t + \phi)]dt$$

$$= \frac{i_0^2}{2T}\left[t - \frac{\sin^2(\omega t + \phi)}{2\omega}\right]_0^T$$

$$= \frac{i_0^2}{2T}\left[T - \frac{\sin(4\pi + 2\phi) - \sin 2\phi}{2\omega}\right]$$

$$= \frac{i_0^2}{2} \qquad \text{----(5)}$$

This is known as the mean square current

- The square root of mean square current is called root mean square current or rms current. Thus,

$$i_{rms} = \sqrt{\bar{i}^2} = \frac{i_0}{\sqrt{2}} = 0.707 i_0 \qquad \text{----(6)}$$

thus ,the rms value of AC is .707i_0 of the peak value of alternating current

- Similarly rms value of alternating voltage or emf is

$$V_{rms} = \frac{V_0}{\sqrt{2}} \qquad \text{----(7)}$$

- If we allow the AC current represented by $i=i_0\sin(\omega t+\varphi)$ to pass through a resistor of resistance R,the power dissipated due to flow of current would be
  $P=i^2R$
- Since magnitude of current changes with time ,the power dissipation in circuit also changes
- The average Power dissipated over one complete current cycle would be

$$\overline{P} = \bar{i}^2 R$$

$$= (i_{rms})^2 R$$

If we pass direct current of magnitude i_{rms} through the resistor ,the power dissipate or rate of production of heat in this case would be
$P=(i_{rms})^2R$

- Thus rms value of AC is that value of steady current which would dissipate the same amount of power in a given resistance in a given tine as would gave been dissipated by alternating current
- This is why rms value of AC is also known as virtual value of current

(5) Phasor diagram

- Phasor diagrams are diagram representing alternating current and voltage of same frequency as vectors or phasors with the phase angle between them
- Phasors are the arrows rotating in the anti-clockwise direction i.e. they are rotating vectors but they represents scalar quantities
- Thus a sinusoidal alternating current and voltage can be represented by anticlockwise rotating vectors if they satisfy following conditions
- Length of the vector must be equal to the peak value of alternating voltage or current
- Vector representing alternating current and voltage would be at horizontal position at the instant when alternating quantity is zero
- In certain circuits when current reaches its maximum value after emf becomes maximum then current is said to lag behind emf
- When current reaches its maximum value before emf reaches its maximum then current is said to lead the emf
- Figure below shows the current lagging behind the emf by 90^0

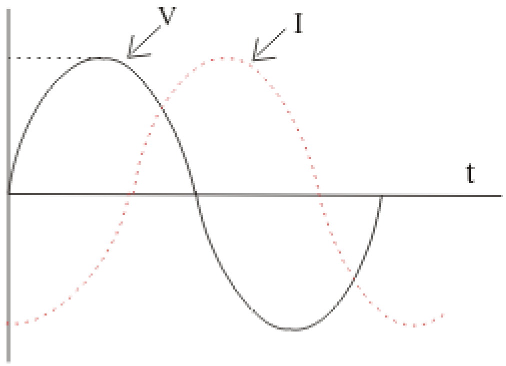


Figure 2 (a). Sinusoidal representation

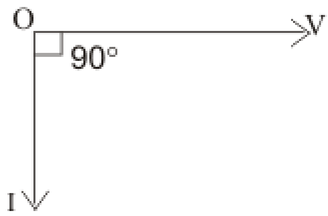


Figure 2 (b). Phaser diagram representation where ve
OV is peak value of voltage and OI is peak value of c

(6) A.C through pure resistor

- Figure below shows the circuit containing alternating voltage source $V=V_0\sin\omega$ connected to a resistor of resistance R

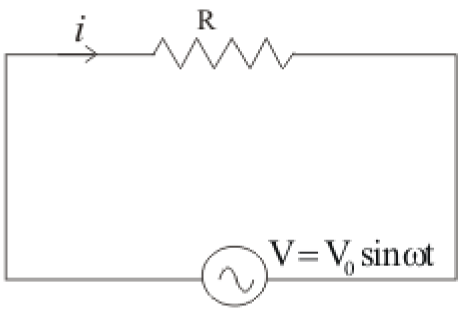


Figure 3. A.C. Circuit containing only resistor

- Let at any instant of time ,i is the current in the circuit ,then from Kirchhoff's loop rule
$V_0\sin\omega t Ri$
or
$i=(V_0/R)\sin\omega t$
$=i_0\sin\omega t$ ----(8)

Where,
$i_0=V_0/R$ ----(9)

- From instantaneous values of alternating voltage and current ,we can conclude that in pure resistor ,the current is always in phase with applied voltage
- Their relationship is graphically represented as

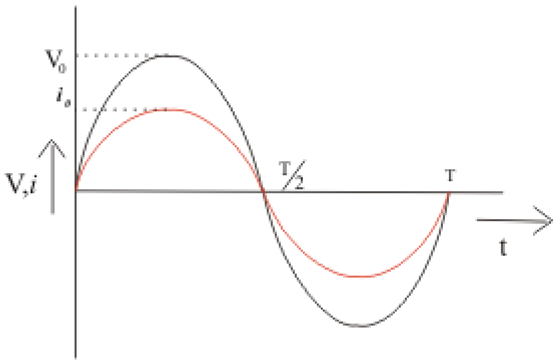


Figure 4(a). Sinusoidal representation

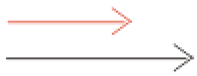


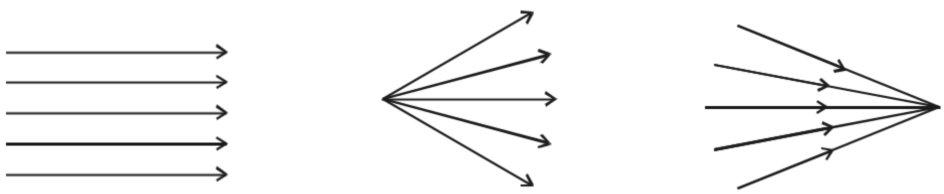
Figure 4(b). Phaser diagram representation of current and voltage through pure resistor

UNIT 5

Ray Optics

Light rays and beams

A ray of light is the direction along which the light energy travels. In practice a ray has a finite width and is represented in diagrams as straight lines. A beam of light is a collection of rays. A search light emits a parallel beam of light (Fig. 9.1a). Light from a lamp travels in all directions which is a divergent beam. (Fig. 9.1b). A convex lens produces a convergent beam of light, when a parallel beam falls on it (Fig. 9.1c).



(a) Parallel beam *(b) Divergent Beam* *(c) Convergent Beam*

Fig. 9.1 Beam of light

9.1 Reflection of light

Highly polished metal surfaces reflect about 80% to 90% of the light incident on them. Mirrors in everyday use are therefore usually made of depositing silver on the backside of the glass. The largest reflector in the world is a curved mirror nearly 5 metres across, whose front surface is coated with aluminium. It is the hale Telescope on the top of Mount Palomar, California, U.S.A. Glass by itself, will also reflect light, but the percentage is small when compared with the case of silvered surface. It is about 5% for an air-glass surface.

9.1.1 Laws of reflection

Consider a ray of light, AO, incident on a plane mirror XY at O. It is reflected along OB. Let the normal ON is drawn at the point of incidence. The angle AON between the incident ray and the normal is called angle of incidence, *i* (Fig. 9.2) the angle BON between the reflected ray and the normal is called angle of reflection, *r*. Experiments show that : (i) *The incident ray, the reflected ray and the normal drawn to the reflecting surface at the point of incidence, all lie in the same plane.*

The angle of incidence is equal to the angle of reflection. (i.e) $i = r$.

These are called the laws of reflection.

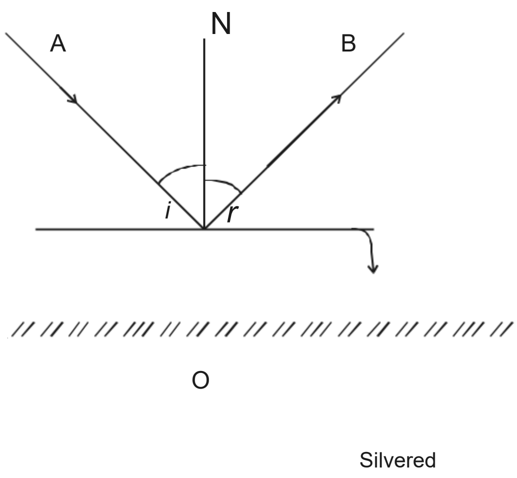


Fig. 9.2 Reflection at a plane mirror

9.1.2 Deviation of light by plane mirror

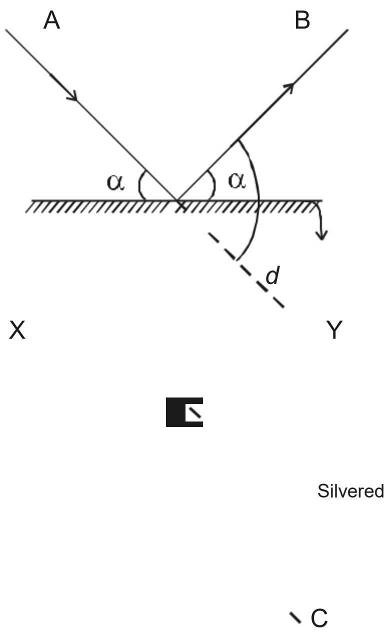


Fig. 9.3 Deviation of light by a plane mirror

angle of deviation, d = 2α

Consider a ray of light, AO, incident on a plane mirror XY (Fig. 9.3) at O. It is reflected along OB. The angle AOX made by AO with XY is known as the glancing angle α with the mirror. Since the angle of reflection is equal to the angle of incidence, the glancing angle BOY made by the reflected ray OB with the mirror is also equal to α.

The light has been deviated from a direction AO to a direction OB. Since angle COY = angle AOX, it follows that

So, in general, *the angle of deviation of a ray by a plane mirror or a plane surface is twice the glancing angle.*

9.1.3 Deviation of light due to rotation of a mirror

Let us consider a ray of light AO incident on a plane mirror XY at O. It is reflected along OB. Let α be the glancing angle with XY (Fig. 9.4). We know that the angle of deviation COB = 2α.

Suppose the mirror is rotated through an angle θ to a position X′Y′.

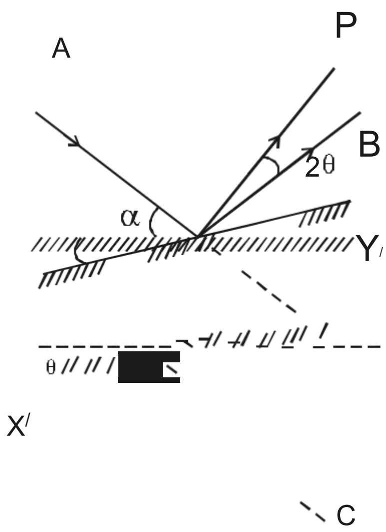


Fig. 9.4 Deviation of light due

to rotation of a mirror

The same incident ray AO is now reflected along OP. Here the glancing angle with X′Y′ is (α + θ). Hence the new angle of deviation COP = 2 (α + θ). The reflected ray has thus been rotated through an angle BOP when the mirror is rotated through an angle θ.

$$\angle BOP = \angle COP - \angle COB$$

$$\angle BOP = 2(\alpha+\theta) - 2\alpha = 2\theta$$

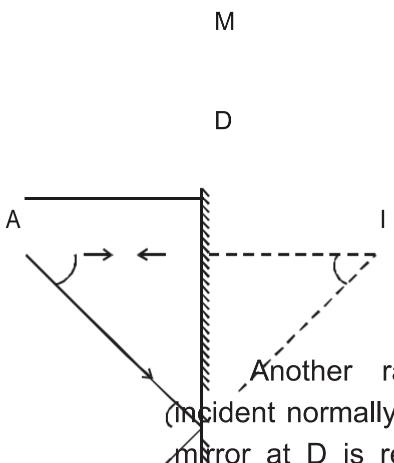
For the same incident ray, when the mirror is rotated through an angle, the reflected ray is rotated through twice the angle.

9.2 Image in a plane mirror

Let us consider a point object A placed in front of a plane mirror M as shown in the Fig. 9.5. Consider a

ray of light AO from the point object incident on the mirror and reflected along OB. Draw the normal ON to the mirror at O.

The angle of incidence AON = angle of reflection BON



Another ray AD incident normally on the mirror at D is reflected back along DA. When BO and AD are produced backwards, they meet at I. Thus the rays reflected from M appear to come from a point I behind the mirror.

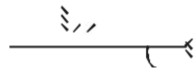


Fig. 9.5 Image in a plane mirror

From the figure

$\angle AON = \angle DAO$, alternate angles and $\angle BON = \angle DIO$, corresponding angles it follows that $\angle DAO = \angle DIO$.

The triangles ODA and ODI are congruent

AD=ID

For a given position of the object, A and D are fixed points. Since AD = ID, the point I is also fixed. It should be noted that AO = OI. So the object and its image in a plane mirror are at equal perpendicular distances from the mirror.

9.2.1 Virtual and real images

An object placed in front of a plane mirror has an image behind the mirror. The rays reflected from the mirror do not actually meet through I, but only appear to meet and the image cannot be received on the screen, because the image is behind the mirror. This type of image is called an unreal or virtual image (Fig. 9.6a).

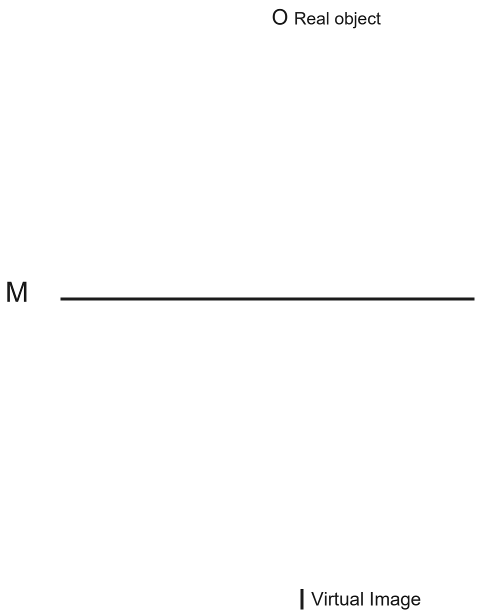
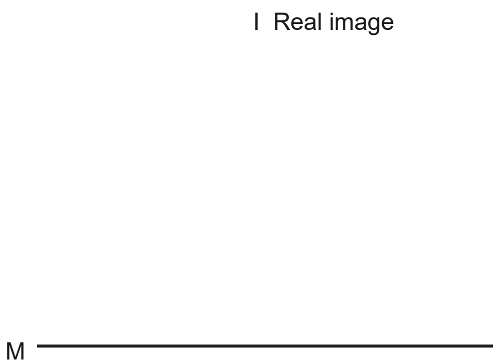


Fig. 9.6a Virtual image in a plane mirror



If a convergent beam is incident on a plane mirror, the reflected rays pass through a point I in front of M, as shown in the Fig. 9.6b. In the Fig. 9.6a, a real object (divergent beam) gives rise to a virtual image. In

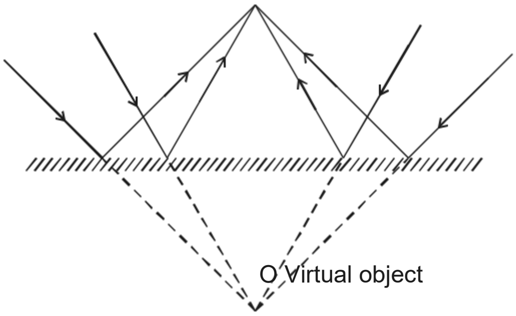


Fig. 9.6b Real image in a plane mirror

the Fig. 9.6b, a virtual object (convergent beam) gives a real image. Hence plane mirrors not only produce virtual images for real objects but also produce real images for virtual objects.

9.2.2 Characteristics of the image formed by a plane mirror

Image formed by a plane mirror is as far behind the mirror as the object is in front of it and it is always virtual.

The image produced is laterally inverted.

The minimum size of the mirror required to see the complete image of the object is half the size of the object.

If the mirror turns by an angle θ, the reflected ray turns through an angle 2θ.

If an object is placed between two plane mirrors inclined at an angle θ, then the number of images formed is $n = \frac{360}{\theta^{\circ}} - 1$

9.3 Reflection at curved surfaces

In optics we are mainly concerned with curved mirrors which are the part of a hollow sphere (Fig. 9.7). One surface of the mirror is silvered. Reflection takes place at the other surface. If the reflection takes place at the concave surface, (which is towards the centre of the sphere) it is called concave mirror. If the reflection takes place at the convex surface, (which is away from the centre of the sphere) it is called convex mirror. The laws of reflection at a plane mirror are equally true for spherical mirrors also.

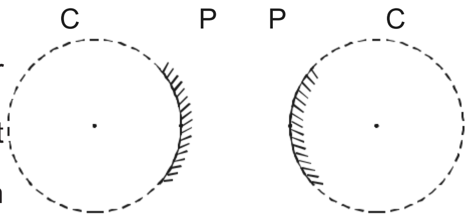


Fig.9.7 Concave and convex mirror

The centre of the sphere, of which the mirror is a part is called the *centre of curvature* (C).

The geometrical centre of the mirror is called its *pole* (P).

The line joining the pole of the mirror and its centre of curvature is called the *principal axis*.

The distance between the pole and the centre of curvature of the spherical mirror is called the *radius of curvature* of the mirror and is also equal to the radius of the sphere of which the mirror forms a part.

When a parallel beam of light is incident on a spherical mirror, the point where the reflected rays converge (concave mirror) or appear to

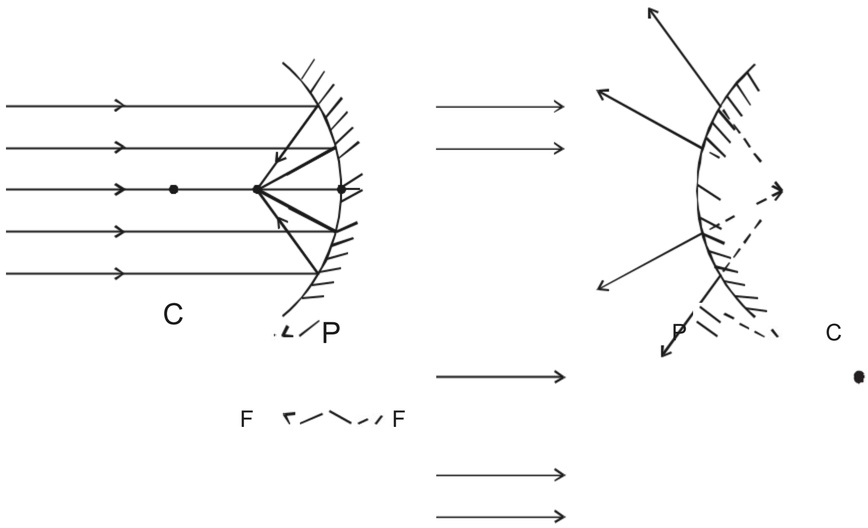


Fig. 9.8 Principal focus

diverge from the point (convex mirror) on the principal axis is called the *principal focus* (F) of the mirror. The distance between the pole and the principal focus is called the *focal length* (f) of the mirror (Fig. 9.8).

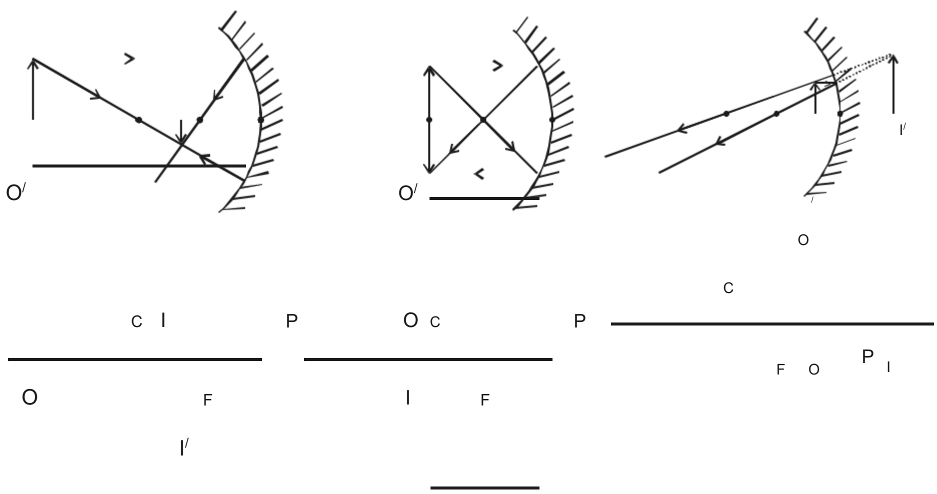
9.3.1 Images formed by a spherical mirror

The images produced by spherical mirrors may be either real or virtual and may be either larger or smaller than the object. The image can be located by graphical construction as shown in Fig. 9.9 by adopting any two of the following rules.

A ray parallel to the principal axis after reflection by a concave mirror passes through the principal focus of the concave mirror and appear to come from the principal focus in a convex mirror.

A ray passing through the centre of curvature retraces its path after reflection.

A ray passing through the principal focus, after reflection is rendered parallel to the principal axis.



I/

Fig. 9.9 Formation of images in concave mirror

A ray striking the pole at an angle of incidence *i* is reflected at the same angle *i* to the axis.

9.3.2 Image formed by a convex mirror

In a convex mirror irrespective of the position of the object, the image formed is always virtual, erect but diminished in size. The image lies between the pole and the focus (Fig. 9.10).

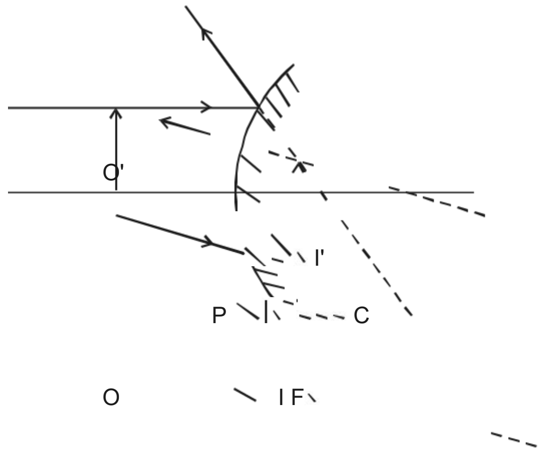


Fig. 9.10 Image formed by

convex mirror

In general, real images are located in front of a mirror while virtual images behind the mirror.

9.3.3 Cartesian sign convention

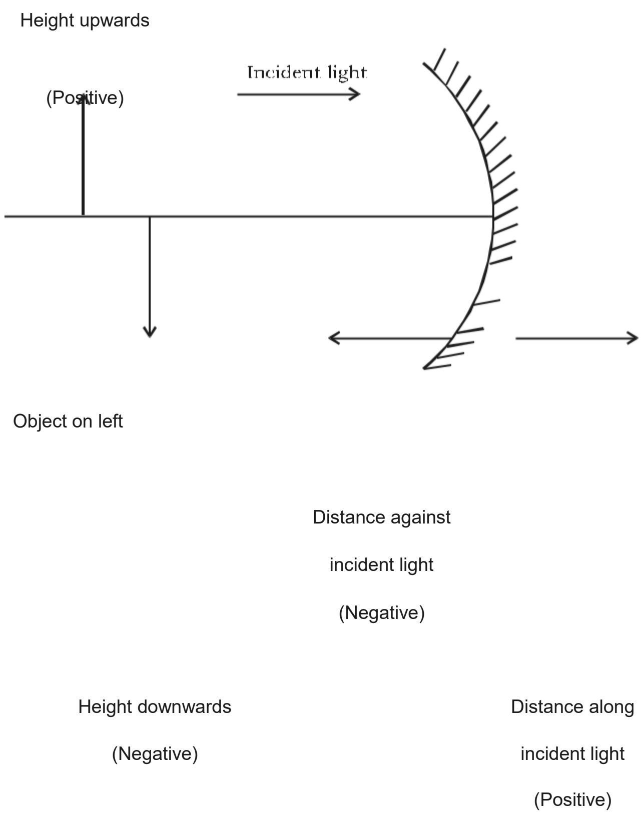


Fig. 9.11 Sign convention

The following sign conventions are used.

All distances are measured from the pole of the mirror (in the case of lens from the optic centre).

The distances measured in the same direction as the incident light, are taken as positive.

The distances measured in the direction opposite to the direction of incident light are taken as negative.

Heights measured perpendicular to the principal axis, in the upward direction are taken as positive.

Heights measured perpendicular to the principal axis, in the downward direction are taken as negative.

The size of the object is always taken as positive, but image size is positive for erect image and negative for an inverted image.

The magnification is positive for erect (and virtual) image, and negative for an inverted (and real) image.

9.3.4 Relation between u, v and f for spherical mirrors

A mathematical relation between object distance *u*, the image distance *v* and the focal length *f* of a spherical mirror is known as mirror formula.

(i) Concave mirror - real image

Let us consider an object OO′ on the principal axis of a concave mirror beyond C. The incident and the reflected rays are shown in the Fig 9.12. A ray O′A parallel to principal axis is incident on the concave mirror at A, close to P. After reflections the ray passes through the focus F. Another ray O′C passing

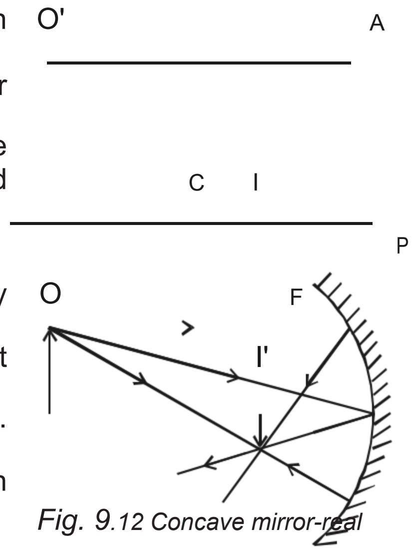


Fig. 9.12 Concave mirror-real image

through centre of curvature C, falls

normally on the mirror and reflected back along the same path. A third ray O′P incident at the pole P is reflected along PI′. The three reflected rays intersect at the point I′. Draw perpendicular I′I to the principal axis. II′ is the real, inverted image of the object OO′.

Right angled triangles, *II ′P* and *OO′P* are similar.

$$\therefore \frac{II'}{OO'} = \frac{PI}{PO} \qquad \text{... (1)}$$

Right angled triangles *II′F* and *APF* are also similar (A is close to

P ; hence AP is a vertical line)

$$\frac{II'}{AP} = \frac{IF}{PF}$$

$AP = OO'$. Therefore the above equation becomes,

$$\frac{\ '}{OO'} = \frac{IF}{PF} \qquad \text{... (2)}$$

Comparing the equations (1) and (2)

$$\frac{PI}{PO} = \frac{IF}{PF} \qquad \text{... (3)}$$

But, $IF = PI – PF$

Therefore equation (3) becomes,

$$\frac{PI}{PO} = \frac{PI - PF}{PF} \qquad \text{... (4)}$$

Using sign conventions, we have $PO = –u$,

PI = -v and PF = -f

Substituting the values in the above equation, we get

$$\frac{-v}{-u} = \frac{-v-(-f)}{-f} \quad \text{(or)}$$

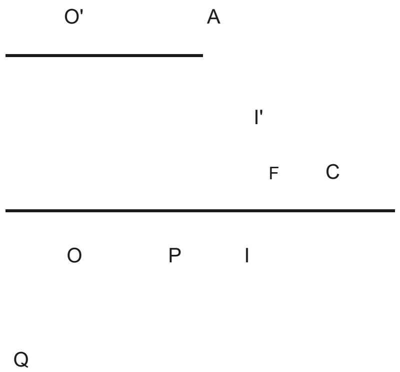
$$= \frac{v-f}{f} = \frac{v}{f} - 1$$

Dividing by v and rearranging, $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

This is called *mirror equation*. The same equation can be obtained for virtual image also.

(ii) Convex mirror - virtual image

Let us consider an object OO′ anywhere on the principal axis of a convex mirror. The incident and the reflected rays are shown in the Fig. 9.13. A ray O′A parallel to the principal axis incident on the convex mirror at A close to P. After reflection



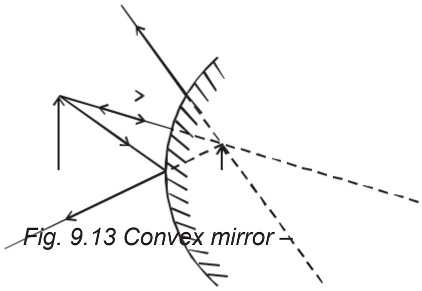


Fig. 9.13 Convex mirror – Virtual image

the ray appears to diverge from the focus F. Another ray O′C passing through centre of curvature C, falls normally on the mirror and is reflected back along the same path. A third ray *O ′P* incident at the pole *P* is reflected along *PQ*. The three reflected rays when produced appear to meet at the point *I ′*. Draw perpendicular *II′* to the principal axis. II′ is the virtual image of the object OO′.

Right angled triangles, *II ′P* and *OO ′P* are similar.

$$\therefore \frac{II'}{OO'} = \frac{PI}{PO} \quad \text{... (1)}$$

Right angled triangles *II ′F* and *APF* are also similar (A is close to P; hence AP is a vertical line)

$$\frac{II'}{AP} = \frac{IF}{PF}$$

$AP = OO'$. Therefore the above equation becomes,

$$\frac{II'}{OO'} = \frac{IF}{PF} \qquad \ldots (2)$$

Comparing the equations (1) and (2)

$$\frac{PI}{PO} = \frac{IF}{PF} \qquad \ldots (3)$$

But, $IF = PF – PI$. Therefore equation (3) becomes,

$$\frac{PI}{PO} = \frac{PF - PI}{PF}$$

Using sign conventions, we have $PO = -u$, $PI = +v$ and $PF = +f$.

Substituting the values in the above equation, we get

$$\frac{+v}{-u} = \frac{+f-(+v)}{+f} \quad \text{(or)} \quad -\frac{v}{u} = \frac{f-v}{f} = 1 - \frac{v}{f}$$

Dividing by v and rearranging we get, $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

This is called *mirror equation for convex mirror producing virtual image*.

9.3.5 Magnification

The linear or transverse magnification is defined as the ratio of the size of the image to that of the object.

Magnification = $\frac{\textit{size of the image}}{\textit{size of the object}} = \frac{h_2}{h_1}$

where h_1 and h_2 represent the size of the object and image respectively.

From Fig. 9.12 it is known that $\frac{II'}{OO'} = \frac{PI}{PO}$

Applying the sign conventions,

$II' = -h_2$ (height of the image measured downwards)

$OO' = +h_1$ (height of the object measured upwards) $PI = -v$ (image distance against the incident light)

PO = –*u* (object distance against the incident light)

Substituting the values in the above equation, we get

$$\text{magnification } m = \frac{-h_2}{+h_1} = \frac{-v}{-u} \quad \text{(or) } m = \frac{h_2}{h_1} = \frac{-v}{u}$$

For an erect image *m* is positive and for an inverted image *m* is negative. This can be checked by substituting values for convex mirror also.

Using mirror formula, the equation for magnification can also be obtained as

$$m = \frac{h_2}{h_1} = \frac{-v}{u} = \frac{f-v}{f} = \frac{f}{f-u}$$

This equation is valid for both convex and concave mirrors.

9.4 Total internal reflection

When a ray of light *AO* passes from an optically denser medium to a rarer medium, at the interface *XY*, it is partly reflected back into the same medium along *OB* and partly refracted into the rarer medium along *OC* (Fig. 9.14).

If the angle of incidence is gradually increased, the angle of refraction *r* will also gradually increase and at a certain stage *r* becomes

90^o. Now the refracted ray *OC* is bent so much away from the normal and it grazes the surface of separation of two media. *The angle of incidence in the denser medium at which the refracted ray just grazes the surface of separation is called the critical angle c of the denser medium.*

If *i* is increased further, refraction is not possible and the incident

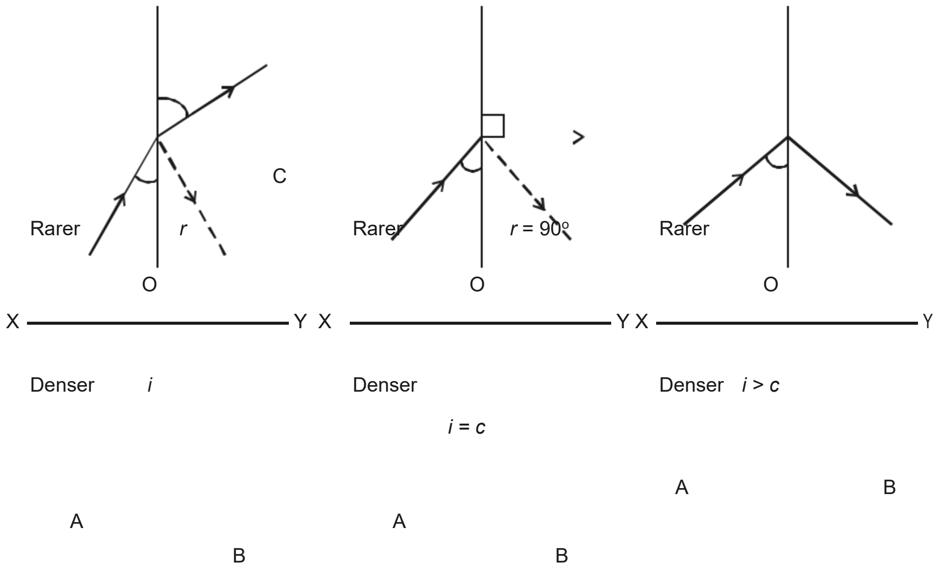


Fig. 9.14 Total internal reflection

ray is totally reflected into the same medium itself. This is called *total internal reflection.*

If μ_d is the refractive index of the denser medium then, from Snell's Law, the refractive index of air with respect to the denser medium is given by,

$\sin i$

μ =

$$\frac{\mu_a}{\mu_d} = \frac{\sin i}{\sin r}$$

$$\frac{1}{\mu_d} = \frac{\sin i}{\sin r} \quad (\because \mu_a = 1 \text{ for air})$$

If $r = 90^o$, $i = c$

$$\frac{\sin c}{\sin 90^o} = \frac{1}{\mu_d} \quad \text{(or)} \quad \sin c = \frac{1}{\mu_d} \text{ or } c = \sin^{-1}\frac{1}{\mu_d}$$

If the denser medium is glass, $c = \sin^{-1}\frac{1}{\mu_g}$

Hence for total internal reflection to take place (i) light must travel from a denser medium to a rarer medium and (ii) the angle of incidence inside the denser medium must be greater than the critical angle i.e. $i > c$.

Table 9.1 Critical angle for some media

(NOT FOR EXAMINATION)

Medium	Refractive index	Critical angle
Water	1.33	48.75^{o}
Crown glass	1.52	41.14^{o}
Dense flint glass	1.62	37.31^{o}
Diamond	2.42	24.41^{o}

9.4.1 Applications

(i) Diamond

Total internal reflection is the main cause of the brilliance of diamonds. The refractive index of diamond with respect to air is 2.42. Its critical angle is 24.41^{o}. When light enters diamond from any face at an angle greater than 24.41^{o} it undergoes total internal reflection. By cutting the diamond suitably, multiple internal reflections can be made to occur.

(ii) Optical fibres

The total internal reflection is the basic principle of optical fibre. An optical fibre is a very thin fibre made of glass or quartz having radius of the order of micrometer (10^{-6}m). A bundle, of such thin fibres forms a 'light pipe' (Fig. 9.15a).

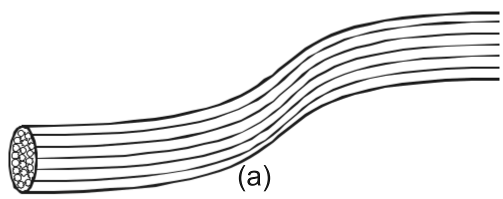
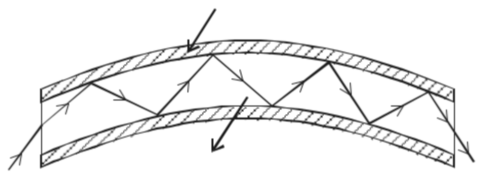

(a)

Fig. 9.15b shows the principle of light transmission inside an optical fibre. The refractive index of the material of the core is higher than that of the cladding. When the light is incident at one end of the fibre at a small angle, the light passes

Cladding (μ = 1.5)



Fibre (μ = 1.7)

(b) *Fig.9.15 An optical fibre*

inside, undergoes repeated total internal reflections along the fibre and finally comes out. The angle of incidence is always larger than the critical angle of the core material with respect to its cladding. Even if the fibre is bent or twisted, the light can easily travel through the fibre.

Light pipes are used in medical and optical examination. They are also used to transmit communication signals.

9.5 Michelson's method

Michelson, an American physicist, spent many years of his life in measuring the velocity of light and he devised a method in the year 1926 which is considered as accurate.

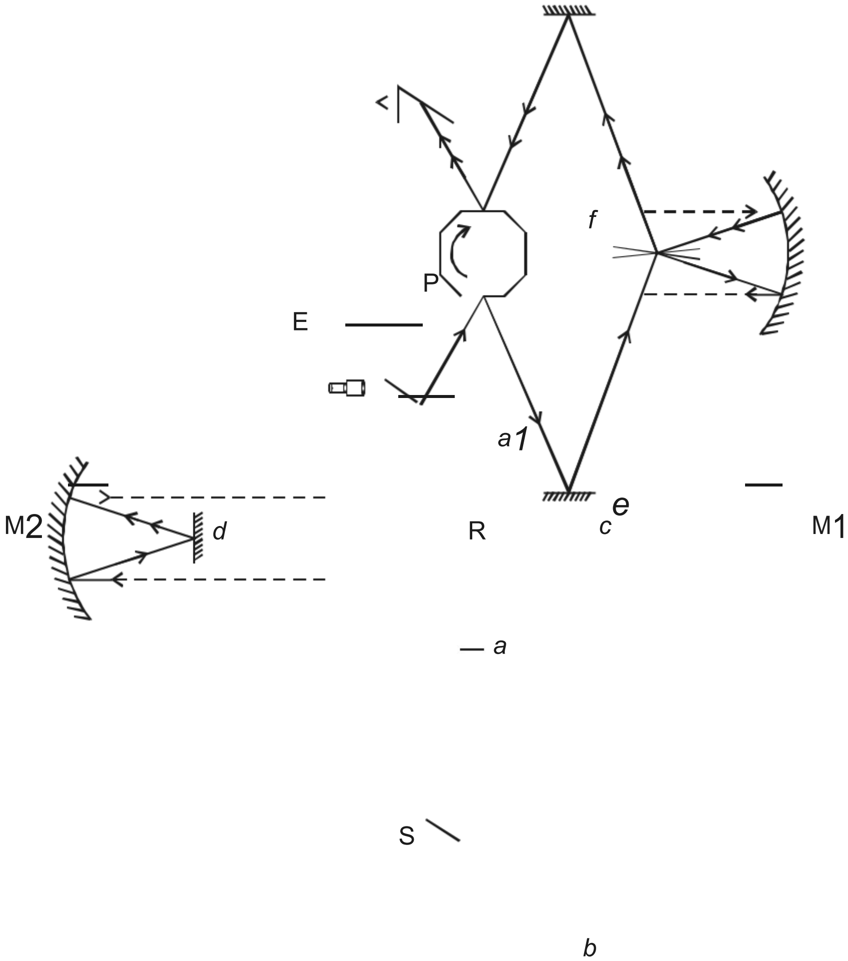


Fig. 9.16 Michelson's method

The experimental set up is shown in Fig. 9.16. Light from an arc source after passing through a narrow slit S is reflected from one face *a* of an octagonal mirror R. The ray after reflections at small fixed mirrors *b* and *c* is then rendered parallel by a concave mirror M_1 placed in the observing station on Mt. Wilson. This parallel beam of light travels a distance of 35 km and falls on another concave mirror M_2 placed at Mt. St Antonio, and it is reflected to a plane mirror *d* placed at the focus of the concave mirror M_2. The ray of light from *d* is rendered parallel after getting reflected by M_2 and travels back to the concave mirror M_1.

After reflections at M_1 and the plane mirrors *e* and *f*, the ray falls on the opposite face a_1 of the octagonal mirror. The final image which is totally reflected by a total reflecting prism P, is viewed through an eye piece E.

When the octagonal mirror is stationary, the image of the slit is seen through the eye piece. When it is rotated the image disappears. The speed of rotation of R is suitably adjusted so that the image is seen again clearly as when R is stationary. The speed of revolution is measured by stroboscope.

Let D be the distance travelled by light from face a to face a_1 and n be the number of rotations made by R per second.

The time taken by R to rotate through 45^{o} or $\frac{1}{8}$ of a rotation = $\frac{1}{8n}$

During this time interval, the distance travelled by the light = D

$$\therefore \text{ The velocity of light } c = \frac{\text{Distance travelled}}{\text{Time taken}} = \frac{D}{\frac{1}{8n}} = 8nD.$$

In general, if the number of faces in the rotating mirror is N, the velocity of light = NnD.

The velocity of light determined by him is 2.99797 × 10^{8} m s^{-1}.

Importance of velocity of light

The value of velocity of light in vacuum is of great importance in science. The following are some of the important fields where the value of velocity of light is used.

Frequency - wavelength relation **:** From the relation $c = \nu\lambda$, the frequency of electromagnetic radiations can be calculated if the wavelength is known and vice versa.

Relativistic mass variation with velocity : Theory of relativity has shown that the mass m of a moving particle varies with its velocity v according to the relation $m = \frac{m_o}{\sqrt{1 - \frac{v^2}{c^2}}}$

$\sqrt{\quad c^2}$

Here m_o is the rest mass of the particle.

Mass - Energy relation **:** E = mc^2 represents conversion of mass into energy and energy into mass. The energy released in nuclear fission and fusion is calculated using this relation.

Measurement of large distance in Astronomy : Light year is a unit of distance used in astronomy. A light year is the distance travelled by light in one year. It is equal to 9.46×10^{15} metre.

Refractive index : The refractive index **μ** of a medium is

given by

$$= \frac{\textit{velocity of light in vacuum}}{\textit{velocity of light in medium}} = \frac{c}{v}$$

9.6 Refraction of light

When a ray of light travels from one transparent medium into another medium, it bends while crossing the interface, separating the two media. This phenomenon is called refraction.

Image formation by spherical lenses is due to the phenomenon of refraction. The laws of refraction at a plane surface are equally true for refraction at curved surfaces also. While deriving the expressions for refraction at spherical surfaces, we make the following assumptions.

The incident light is assumed to be monochromatic and

the incident pencil of light rays is very narrow and close to the principal axis.

9.6.1 Cartesian sign convention

The sign convention followed in the spherical mirror is also applicable to refraction at spherical surface. In addition to this two more sign conventions to be introduced which are:

The power of a converging lens is positive and that of a diverging lens is negative.

The refractive index of a medium is always said to be positive. If two refractions are involved, the difference in their refractive index is also taken as positive.

9.6.2 Refraction at a spherical surface

Let us consider a portion of a spherical surface AB separating two media having refracting indices μ_1 and μ_2 (Fig. 9.17). This is symmetrical about an axis passing through the centre *C* and cuts the surface at *P*. The point *P* is called the pole of the surface. Let *R* be the radius of curvature of the surface.

Consider a point object *O* on the axis in the first medium. Consider two rays *OP* and *OD* originating from *O*. The ray *OP* falls

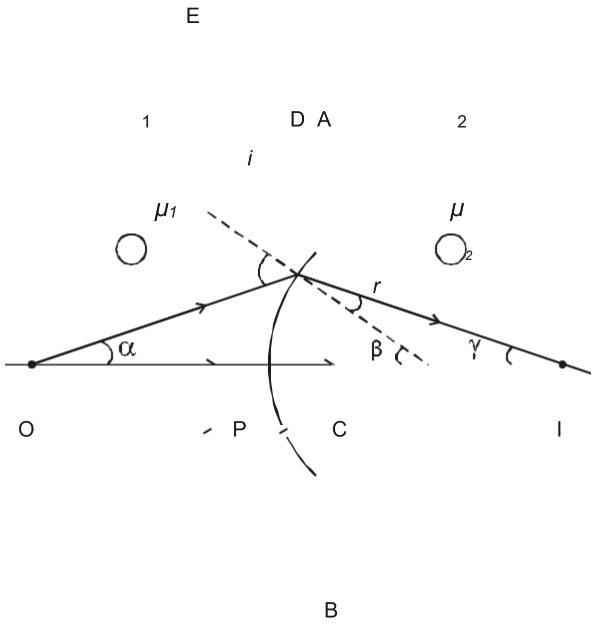


Fig. 9.17 Refraction at a spherical surface

normally on *AB* and goes into the second medium, undeviated. The ray *OD* falls at *D* very close to *P*. After refraction, it meets at the point *I* on the axis, where the image is formed. *CE* is the normal drawn to the point *D*. Let *i* and *r* be the angle of incidence and refraction respectively.

Let $\angle DOP = \alpha$, $\angle DCP = \beta$, $\angle DIC = \gamma$

Since *D* is close to *P*, the angles α, β and γ are all small. From the Fig. 9.17.

$$\tan\alpha = \frac{DP}{PO}, \quad \tan\beta = \frac{DP}{PC} \text{ and } \tan\gamma = \frac{DP}{PI}$$

$$\therefore \alpha = \frac{DP}{PO}, \quad \beta = \frac{DP}{PC} \text{ and } \gamma = \frac{DP}{PI}$$

From the ΔODC, $i = \alpha + \beta$ ...(1)

From the ΔDCI, $\beta = r + \gamma$ or $r = \beta - \gamma$ ...(2)

From Snell's Law, $\frac{\mu_2}{\mu_1} = \frac{\sin i}{\sin r}$ and for small angles of *i* and *r*, we can write, $\mu_1 i = \mu_2 r$ ...(3)

From equations (1), (2) and (3)

we get $\mu_1(\alpha + \beta) = \mu_2(\beta - \gamma)$ or $\mu_1\alpha + \mu_2\gamma = (\mu_2 - \mu_1)\beta$ ... (4)

Substituting the values of α, β and γ in equation (4)

$$\mu_1 \frac{DP}{PO} + \mu_2 \frac{DP}{PI} = (\mu_2 - \mu_1)\frac{DP}{PC}$$

$$\frac{\mu_1}{PO} + \frac{\mu_2}{PI} = \frac{\mu_2 - \mu_1}{PC} \qquad ...(5)$$

As the incident ray comes from left to right, we choose this direction as the positive direction of the axis. Therefore u is negative, whereas v and R are positive substitute $PO = -u$ $PI = +v$ and $PC = +R$ in equation (5),

$$\frac{\mu_1}{-u} + \frac{\mu_2}{v} = \frac{\mu_2 - \mu_1}{R}$$

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{R} \qquad ... (6)$$

Equation (6) represents the general equation for refraction at a spherical surface.

If the first medium is air and the second medium is of refractive index μ, then

$$\frac{\mu}{v} - \frac{1}{u} = \frac{\mu - 1}{R} \qquad ...(7)$$

9.6.3 Refraction through thin lenses

A lens is one of the most familiar optical devices. A lens is made of a transparent material bounded by two spherical surfaces. If the distance between the surfaces of a lens is very small, then it is a thin lens.

As there are two spherical surfaces, there are two centres of curvature C_1 and C_2 and correspondingly two radii of curvature R_1 and R_2. The line joining C_1 and C_2 is called the *principal axis* of the lens. The centre P of the thin lens which lies on the principal aixs is called the optic centre.

9.6.4 Lens maker's formula and lens formula

Let us consider a thin lens made up of a medium of refractive index μ_2 placed in a medium of refractive index μ_1. Let R_1 and R_2 be the radii of curvature of two spherical surfaces *ACB* and *ADB* respectively and *P* be the optic centre.

Consider a point object O on the principal axis. The ray OP falls normally on the spherical surface and goes through the lens undeviated. The ray OA falls at A very close to P. After refraction at the surface ACB the image is formed at I′. Before it does so, it is again refracted by the surface ADB. Therefore the final image is formed at I as shown in Fig. 9.18.

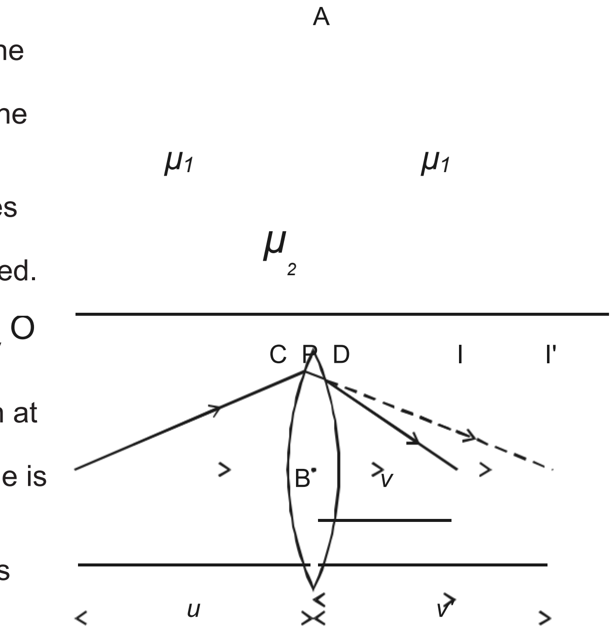


Fig. 9.18 Refraction through a lens

The general equation for the refraction at a spherical surface is given by

$$\frac{\mu_2}{v} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{R} \qquad \text{... (1)}$$

For the refracting surface ACB, from equation (1) we write

$$\frac{\mu_2}{v'} - \frac{\mu_1}{u} = \frac{\mu_2 - \mu_1}{R_1} \qquad \text{... (2)}$$

The image I′ acts as a virtual object for the surface ADB and the final image is formed at I. The second refraction takes place when light travels from the medium of refractive index μ_2 to μ_1.

For the refracting surface ADB, from equation (1) and applying sign conventions, we have

$$\frac{\mu_1}{v} - \frac{\mu_2}{v'} = \frac{\mu_2 - \mu_1}{-R_2} \qquad \text{... (3)}$$

Adding equations (2) and (3) $\frac{\mu_1}{v} - \frac{\mu_1}{u} = (\mu_2 - \mu_1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$

Dividing the above equation by μ_1

$$\frac{1}{v} - \frac{1}{u} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad ...(4)$$

If the object is at infinity, the image is formed at the focus of the lens.

Thus, for $u = \infty$, $v = f$. Then the equation (4) becomes.

$$\frac{1}{f} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad ...(5)$$

If the refractive index of the lens is μ and it is placed in air, $\mu_2 = \mu$ and $\mu_1 = 1$. So the equation (5) becomes

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \qquad ...(6)$$

This is called the *lens maker's formula,* because it tells what curvature will be needed to make a lens of desired focal length. This formula is true for concave lens also.

Comparing equation (4) and (5)

$$\text{we get } \frac{1}{v} - \frac{1}{u} = \frac{1}{f} \qquad \text{... (7)}$$

which is known as the *lens formula*.

9.6.5 Magnification

Let us consider an object OO' placed on the principal axis with its height perpendicular to the principal axis as shown in Fig. 9.19. The ray OP passing through the optic centre will go

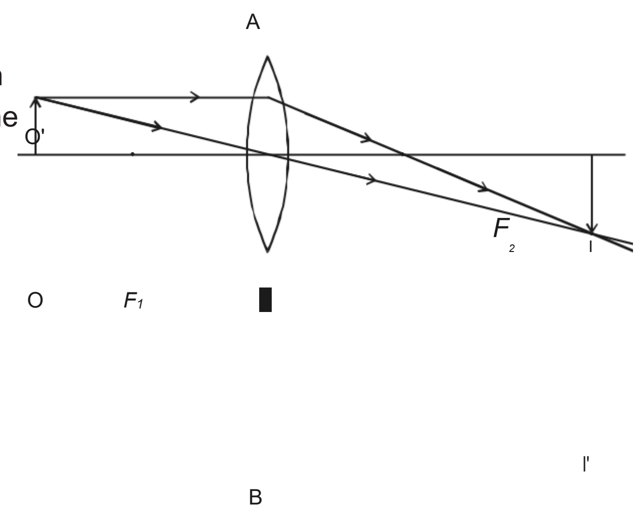


Fig. 9.19 Magnification

undeviated. The ray $O'A$ parallel to the principal axis must pass through the focus F_2. The image is formed where $O'PI'$ and AF_2I' intersect. Draw a perpendicular from I' to the principal axis. This perpendicular II' is the image of OO'.

The linear or transverse magnification is defined as the ratio of the size of the image to that of the object.

$$\textit{Size of the image} \quad II' \quad h_2$$

$\therefore$ Magnification $m = \frac{}{\text{Size of the object}} = \frac{'}{OO} = \frac{}{h_1}$

where h_1 is the height of the object and h_2 is the height of the image.

From the similar right angled triangles $OO'P$ and $II'P$, we have

$$\frac{II'}{OO'} = \frac{PI}{PO}$$

Applying sign convention,

$II' = -h_2$; $OO' = +h_1$;

$PI = +v$; $PO = -u$;

Substituting this in the above equation, we get magnification

$$= \frac{-h_2}{+h_1} = \frac{+v}{-u}$$

$$m = +\frac{v}{u}$$

The magnification is negative for real image and positive for virtual image. In the case of a concave lens, it is always positive.

Using lens formula the equation for magnification can also be obtained as $m = \frac{h_2}{h_1} = \frac{v}{u} = \frac{f - v}{f} = \frac{f}{f + u}$

This equation is valid for both convex and concave lenses and for real and virtual images.

9.6.6 Power of a lens

Power of a lens is a measure of the degree of convergence or divergence of light falling on it. *The power of a lens (P) is defined as the reciprocal of its focal length.*

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

The unit of power is dioptre (*D*) : *1 D* = 1 m^{-1}. *The power of the lens is said to be 1 dioptre if the focal length of the lens is 1 metre.* P is positive for converging lens and negative for diverging lens. Thus, when an optician prescribes a corrective lens of power + 0.5 D, the required lens is a convex lens of focal length + 2 m. A power of -2.0 D means a concave lens of focal length -0.5 m.

9.6.7 Combination of thin lenses in contact

Let us consider two lenses A and B of focal length f_1 and f_2 placed in contact with each other. beyond the focus of the first lens A on the common principal axis. The lens A produces an image at I_1. This image I_1 acts as the object for the second lens B. The final image is produced at I as shown in Fig. 9.20. Since the lenses are thin, a common optical centre P is chosen.

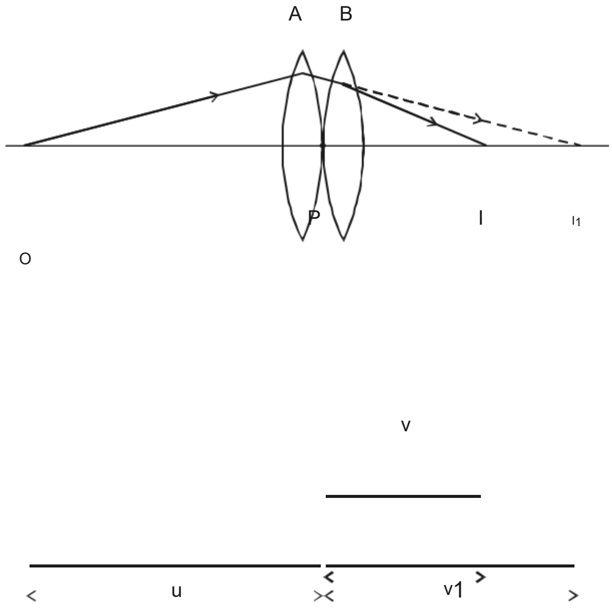


Fig. 9.20 Image formation by two thin lenses

Let $PO = u$, object distance for the first lens (A), $PI = v$, final image distance and $PI_1 = v_1$, image distance for the first lens (A) and also object distance for second lens (B).

For the image I_1 produced by the first lens A,

$$\frac{1}{v_1} - \frac{1}{u} = \frac{1}{f_1} \qquad ...(1)$$

For the final image I, produced by the second lens B,

$$\frac{1}{v} - \frac{1}{v_1} = \frac{1}{f_2} \qquad ...(2)$$

Adding equations (1) and (2),

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{f_1} + \frac{1}{f_2} \qquad ...(3)$$

If the combination is replaced by a single lens of focal length F such that it forms the image of O at the same position I, then

$$\frac{1}{v} - \frac{1}{u} = \frac{1}{F} \qquad ...(4)$$

From equations (3) and (4)

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} \qquad ...(5)$$

This F is the focal length of the equivalent lens for the combination.

The derivation can be extended for several thin lenses of focal lengths f_1, f_2, f_3 ... in contact. The effective focal length of the combination is given by

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \ldots \qquad \ldots(6)$$

In terms of power, equation (6) can be written as

$$P = P_1 + P_2 + P_3 + \ldots. \qquad \ldots(7)$$

Equation (7) may be stated as follows :

The power of a combination of lenses in contact is the algebraic sum of the powers of individual lenses.

The combination of lenses is generally used in the design of objectives of microscopes, cameras, telescopes and other optical instruments.

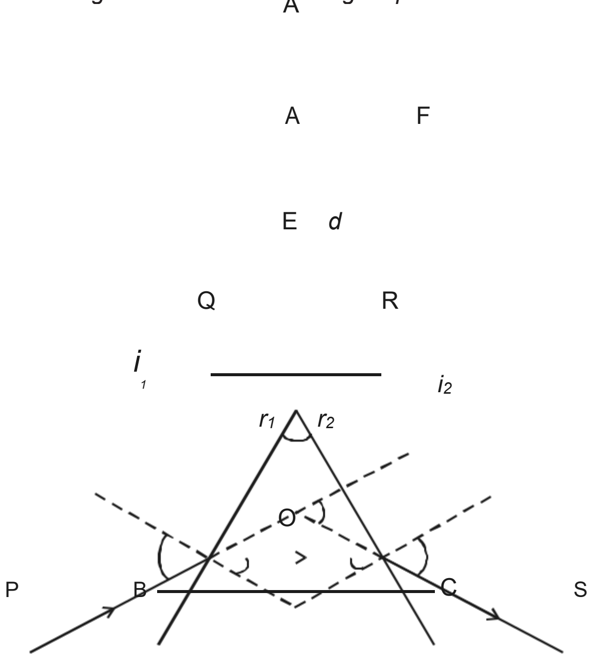
9.7 Prism

A prism is a transparent medium bounded by the three plane faces. Out of the three faces, one is grounded and the other two are polished. The polished faces are called refracting faces. The angle between the refracting faces is called angle of prism, or the refracting angle. The third face is called base of the prism.

Refraction of light through a prism

Fig. 9.21 shows the cross section of a triangular prism *ABC*, placed in air. Let *'A'* be the refracting angle of the prism. *A* ray of light *PQ* incident on the refracting face *AB*, gets refracted along *QR* and emerges along *RS*. The angle of incidence and refraction at the two faces are i_1, r_1, r_2 and i_2 respectively. The angle between the incident ray *PQ* and the emergent ray *RS* is called angle of deviation, *d*.

Fig. 9.21 Refraction through a prism



In the ΔQER, the exterior angle $\angle FER = \angle EQR + \angle ERQ$

$$d = (i_1 - r_1) + (i_2 - r_2)$$

$$\therefore \quad d = (i_1 + i_2) - (r_1 + r_2) \qquad ...(1)$$

In the quadrilateral *AQOR*, the angles at *Q* and *R* are right angles

$$\angle Q + \angle R = 180^o$$

$$\therefore \quad A + \angle QOR = 180^o \qquad ...(2)$$

Also, from the ΔQOR

$$r_1 + r_2 + \angle QOR = 180^o \qquad ...(3)$$

From equation (2) and (3)

$$r_1 + r_2 = A \qquad ...(4)$$

Substituting in (1),

$$d = i_1 + i_2 - A$$

or $$A + d = i_1 + i_2 \qquad ...(5)$$

For a given prism and for a light of given wavelength, the angle of deviation depends upon the angle of incidence.

As the angle of incidence *i* gradually increases, the angle of deviation *d* decreases, reaches a minimum value *D* and then increases. *D* is called the angle of minimum deviation. It will be seen from the graph (Fig. 9.22) that there is only one angle of incidence for which the deviation is a minimum.

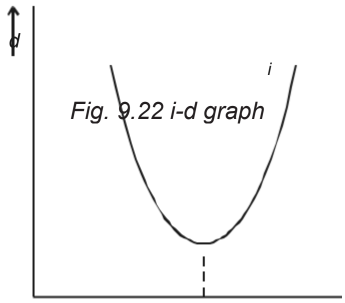


Fig. 9.22 i-d graph

At minimum deviation position the incident ray and emergent ray are symmetric with respect to the base of the prism. (i.e)

the refracted ray *QR* is parallel to the base of the prism.

At the minimum deviation $i_1 = i_2 = i$ and $r_1 = r_2 = r$

from equation (4) $2r = A$ or $r = \frac{A}{2}$

and from equation (5) $2i = A + D$ or $i = \frac{A+D}{2}$

The refractive index is $\mu = \frac{\sin i}{\quad}$

$$\sin r$$

$$\therefore \quad \mu = \frac{\sin\frac{A + D}{2}}{\sin\frac{A}{2}}$$

9.8 Dispersion of light

Dispersion is the splitting of white light into its constituent colours.

This band of colours of light is called its spectrum.

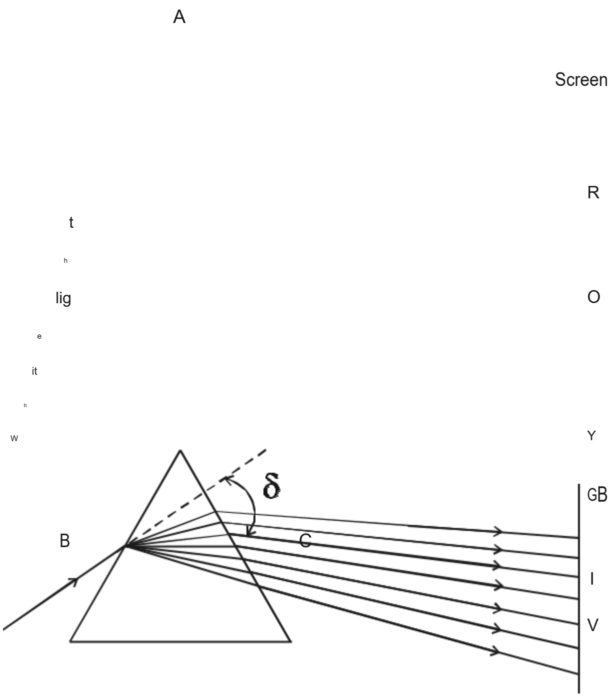


Fig. 9.23 Dispersion of light

In the visible region of spectrum, the spectral lines are seen in the order from violet to red. The colours are given by the word VIBGYOR (Violet, Indigo, Blue, Green, Yellow, Orange and Red) (Fig. 9.23)

The origin of colour after passing through a prism was a matter of much debate in physics. Does the prism itself create colour in some way or does it only separate the colours already present in white light?

Sir Isaac Newton gave an explanation for this. He placed another similar prism in an inverted position. The emergent beam from the first prism was made to fall on the second prism (Fig. 9.24). The resulting emergent beam was found to be white light. The first prism separated the white light into its constituent colours, which were then recombined by the inverted prism to give white light. Thus it can be concluded that the prism does not create any colour but it only separates the white light into its constituent colours.

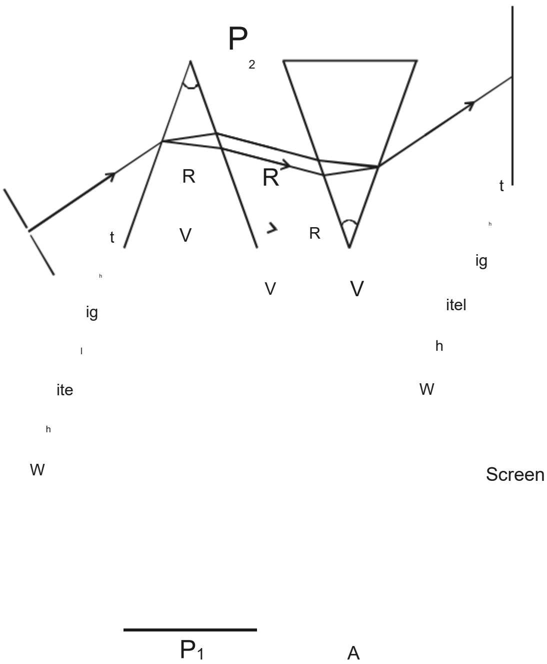


Fig. 9.24 Newton's experiment on dispersion

Dispersion takes place because the refractive index of the material of the prism is different for different colours (wavelengths). The deviation and hence the refractive index is more for violet rays of light than the corresponding values for red rays of light. Therefore the violet ray travels with a smaller velocity in glass prism than red ray. The deviation and the refractive index of the yellow ray are taken as the mean values. Table 9.2 gives the refractive indices for different wavelength for crown glass and flint glass.

Table 9.2 Refractive indices for different wavelengths

(NOT FOR EXAMINATION)

Colour	Wave length (nm)	Crown glass	Flint glass
Violet	396.9	1.533	1.663
Blue	486.1	1.523	1.639
Yellow	589.3	1.517	1.627
Red	656.3	1.515	1.622

The speed of light is independent of wavelength in vacuum. Therefore vacuum is a non-dispersive medium in which all colours travel with the same speed.

9.8.1 Dispersive power

The refractive index of the material of a prism is given by the relation $\mu = \dfrac{\sin\frac{A+D}{2}}{\sin\frac{A}{2}}$

Here A is the angle of the prism and D is the angle of minimum deviation.

If the angle of prism is small of the order of 10^{o}, the prism is said to be small angled prism. When rays of light pass through such prisms the angle of deviation also becomes small.

If *A* be the refracting angle of a small angled prism and δ the angle of deviation, then the prism formula becomes $\mu = \dfrac{\sin\frac{A+\delta}{2}}{\sin\frac{A}{2}}$

For small angles *A* and δ , $\sin\frac{A+\delta}{2} = \frac{A+\delta}{2}$ and $\sin\frac{A}{2} = \frac{A}{2}$

$$\therefore \quad \mu = \frac{\frac{A+\delta}{2}}{\frac{A}{2}}$$

$\overline{2}$

$$\mu A = A + \delta$$

$$\delta = (\mu - 1)A \qquad \text{... (1)}$$

If δ_v and δ_r are the deviations produced for the violet and red rays and μ_v and μ_r are the corresponding refractive indices of the material of the small angled prism then,

for violet light,

$$\delta_v = (\mu_v - 1) A \qquad \text{...(2)}$$

for red light, $\delta_r = (\mu_r - 1)A \qquad \text{...(3)}$

From equations (2) and (3)

$$\delta_v - \delta_r = (\mu_v - \mu_r) A \qquad \text{...(4)}$$

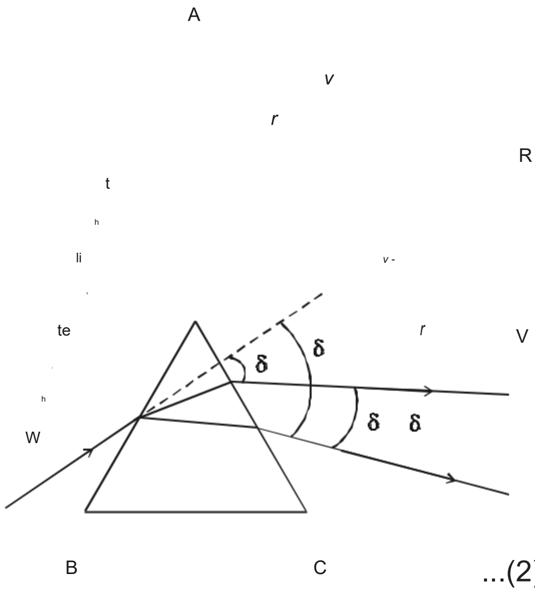


Fig. 9.25 Dispersive power

$\delta_V - \delta_r$ is called the angular dispersion which is the difference in deviation between the extreme colours (Fig. 9.25).

If δ_y and μ_y are the deviation and refractive index respectively for yellow ray (mean wavelength) then,

for yellow light, $\delta_y = (\mu_y - 1) A$ ... (5)

Dividing equation (4) by (5) we get $\dfrac{\delta_V - \delta_r}{\delta_y} = \dfrac{(\mu_V - \mu_r)A}{(\mu_y - 1)A}$

$$\frac{\delta_V - \delta_r}{\delta_y} = \frac{\mu_V - \mu_r}{\mu_y - 1}$$

The expression $\dfrac{\delta_V - \delta_r}{\delta_y}$ is known as the dispersive power of the material of the prism and is denoted by ω.

$$\therefore \omega = \frac{\mu_V - \mu_r}{\mu_y - 1}$$

The dispersive power of the material of a prism is defined as the ratio of angular dispersion for any two wavelengths (colours) to the deviation of mean wavelength.

9.9 Spectrometer

The spectrometer is an optical instrument used to study the spectra of different sources of light and to measure the refractive indices of materials (Fig. 9.26). It consists of basically three parts. They are collimator, prism table and Telescope.

Fig. 9.26 Spectrometer (NEED NOT DRAW IN THE EXAMINATION)

Collimator

The collimator is an arrangement to produce a parallel beam of light. It consists of a long cylindrical tube with a convex lens at the inner end and a vertical slit at the outer end of the tube. The distance between the slit and the lens can be adjusted such that the slit is at the focus of the lens. The slit is kept facing the source of light. The width of the slit can be adjusted. The collimator is rigidly fixed to the base of the instrument.

Prism table

The prism table is used for mounting the prism, grating etc. It consists of two circular metal discs provided with three levelling screws. It can be rotated about a vertical axis passing through its centre and its position can be read with verniers V_1 and V_2. The prism table can be raised or lowered and can be fixed at any desired height.

Telescope

The telescope is an astronomical type. It consists of an eyepiece provided with cross wires at one end of the tube and an objective lens at its other end co-axially. The distance between the objective lens and the eyepiece can be adjusted so that the telescope forms a clear image at the cross wires, when a parallel beam from the collimator is incident on it.

The telescope is attached to an arm which is capable of rotation about the same vertical axis as the prism table. A circular scale graduated in half degree is attached to it.

Both the telescope and prism table are provided with radial screws for fixing them in a desired position and tangential screws for fine adjustments.

9.9.1 Adjustments of the spectrometer

The following adjustments must be made before doing the experiment with spectrometer.

(i) Adjustment of the eyepiece

The telescope is turned towards an illuminated surface and the eyepiece is moved to and fro until the cross wires are clearly seen.

(ii) Adjustment of the telescope

The telescope is adjusted to receive parallel rays by turning it towards a distant object and adjusting the distance between the objective lens and the eyepiece to get a clear image on the cross wire.

(iii) Adjustment of the collimator

The telescope is brought along the axial line with the collimator. The slit of the collimator is illuminated by a source of light. The distance between the slit and the lens of the collimator is adjusted until a clear image of the slit is seen at the cross wires of the telescope. Since the telescope is already adjusted for parallel rays, a well defined image of the slit can be formed, only when the light rays emerging from the collimator are parallel.

S

(iv) Levelling the prism table

The prism table is adjusted or levelled to be in*A* horizontal position by means of levelling screws and a spirit level.

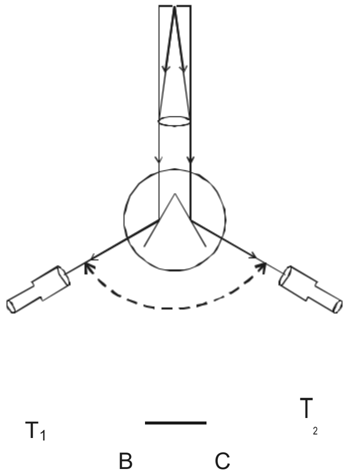
2A

Fig. 9.27 Angle of the prism

9.9.2 Determination of the refractive index of the material of the prism

The preliminary adjustments of the telescope, collimator and the prism table of the spectrometer are made. The refractive index of the prism can be determined by knowing the angle of the prism and the angle of minimum deviation.

(i) Angle of the prism (A)



The prism is placed on the prism table with its refracting edge facing the collimator as shown in Fig 9.27. The slit is illuminated by a sodium vapour lamp.

The parallel rays coming from the collimator fall on the two faces AB and AC.

The telescope is rotated to the position T_1 until the image of the slit, formed by the reflection at the face AB is made to coincide with the vertical cross wire of the telescope. The readings of the verniers are noted. The telescope is then rotated to the position T_2 where the image of the slit formed by the reflection at the face AC coincides with the vertical cross wire. The readings are again noted.

The difference between these two readings gives the angle rotated by the telescope. This angle is equal to twice the angle of the prism. Half of this value gives the angle of the prism A.

(ii) Angle of minimum deviation (D)

The prism is placed on the prism table so that the light from the collimator falls on a refracting face, and the refracted image is observed through the telescope (Fig. 9.28). The prism table is now rotated so that the angle of deviation decreases. A stage comes when the image stops for a moment and if we rotate the prism table further in the same direction, the image is seen to recede and the angle of deviation increases. The vertical cross wire of the telescope is made to coincide with the image of the slit where it turns back. This gives the minimum deviation position. The readings of the verniers are noted. Now the prism is removed and the telescope is turned to receive the direct ray and the vertical cross wire is made to coincide with the image. The readings of the verniers are noted. The difference between the two readings gives the angle of minimum deviation D.

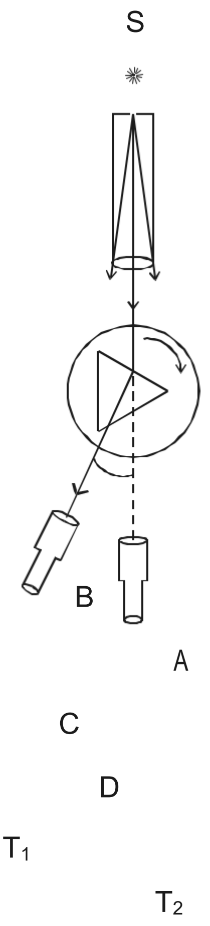


Fig. 9.28 Angle of minimum deviation

The refractive index of the material of the prism μ is calculated using the formula $\mu = \dfrac{\sin\left(\frac{A+D}{2}\right)}{\sin\left(\frac{A}{2}\right)}$.

The refractive index of a liquid may be determined in the same way using a hollow glass prism filled with the given liquid.

9.10 Rainbow

One of the spectacular atmospheric phenomena is the formation of rainbow during rainy days. The rainbow is also an example of dispersion of sunlight by the water drops in the atmosphere.

When sunlight falls on small water drops suspended in air during or after a rain, it suffers refraction, internal reflection and dispersion.

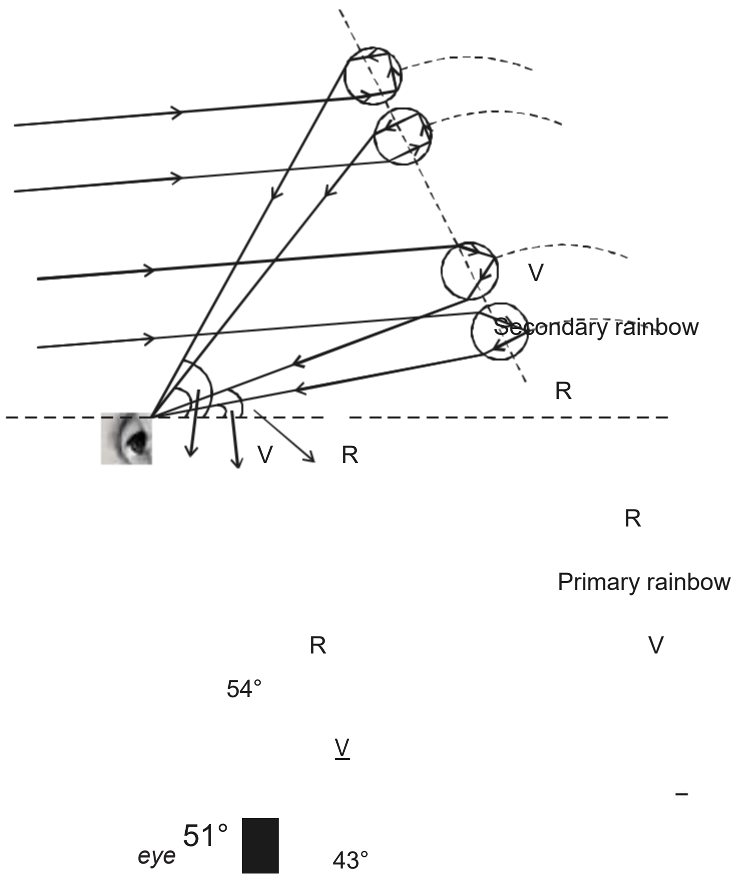


Fig. 9.29 Formation of rainbows

If the Sun is behind an observer and the water drops infront, the observer may observe two rainbows, one inside the other. The inner one is called primary rainbow having red on the outer side and violet on the inner side and the outer rainbow is called secondary rainbow, for which violet on the outer side and red on the inner side.

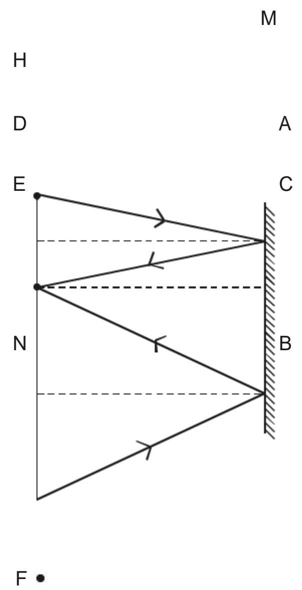
Fig. 9.29 shows the formation of primary rainbow. It is formed by the light from the Sun undergoing one internal reflection and two refractions and emerging at minimum deviation. It is however, found that the intensity of the red light is maximum at an angle of 43^{o} and that of the violet rays at 41^{o}. The other coloured arcs occur in between violet and red (due to other rain drops).

The formation of secondary rainbow is also shown in Fig. 9.31. It is formed by the light from the Sun undergoing two internal reflections and two refractions and also emerging at minimum deviation. In this case the inner red edge subtends an angle of 51^{o} and the outer violet edge subtends an angle of 54^{o}. This rainbow is less brighter and narrower than the primary rainbow. Both primary and secondary rainbows exhibit all the colours of the solar spectrum.

From the ground level an arc of the rainbow is usually visible. A complete circular rainbow may be seen from an elevated position such as from an aeroplane.

Solved Problems

9.1 A man 2 m tall standing in front of a plane mirror whose eye is 1.90 m above the ground. What is the minimum size of the mirror required to see complete image?



Solution :

M	–	*Mirror*
FH	–	*Man*
H	–	*Head*
E	–	*Eye*
F	–	*Feet*

A ray HA from the head, falls at A on the mirror and reflected to E along AE. AD is the perpendicular bisector of HE.

$$\therefore AC = \frac{1}{2} HE = \frac{1}{2} \times 0.10 = 0.05 \text{ m.}$$

A ray FB from the feet, falls at B and reflected to E along BE. BN is the perpendicular bisector of EF.

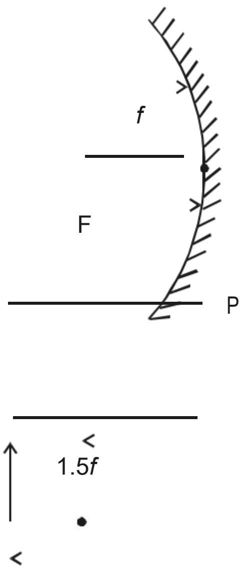
$$\therefore CB = \frac{1}{2} EF = \frac{1}{2} \times 1.90 = 0.95 \text{ m.}$$

The size of the mirror= AC + CB

0.05 m + 0.95 m

Size of the mirror $= 1\ m$

9.2 An object of length 2.5 cm is placed at a distance of 1.5 times the focal length *(f)* from a concave mirror. Find the length of the image. Is the image is erect or inverted?



Data : $f = -f$; $u = -1.5\ f$; $h_1 = 2.5\ cm$; $h_2 = ?$

Solution :

We know,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$$

$$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{-f} - \frac{1}{-1.5\ f}$$

$$\frac{1}{v} = \frac{1}{1.5\,f} - \frac{1}{f}$$

$$v = -3f$$

$$\textit{magnification, } m = -\frac{v}{u} = (-)\frac{-3f}{-1.5\,f} = -2$$

$$\text{But } \frac{h_2}{h_1} = m = -2$$

$$\therefore \quad h_2 = -5\text{ cm}$$

The length of the image is 5.0 cm. The −ve sign indicates that the image is inverted.

9.3 In Michelson's method to determine the velocity of light in air, the distance travelled by light between reflections from the opposite faces of the octagonal mirror is 150 km. The image appears stationary when the minimum speed of rotation of the octagonal mirror is 250 rotations per second. Calculate the velocity of light.

Data :

D = 150 km = 150×10^3 m; n = 250 rps; N = 8; C = ?

Solution :

In Michelson's method, the velocity of light is

$C = NnD$

$C = 8 \times 250 \times 150 \times 10^3$

$C = 3 \times 10^8\ ms^{-1}$

9.4 The radii of curvature of two surfaces of a double convex lens are 10 cm each. Calculate its focal length and power of the lens in air and liquid. Refractive indices of glass and liquid are 1.5 and 1.8 respectively.

Data : *$R_1 = 10$ cm, $R_2 = -10$ cm ; $\mu_g = 1.5$ and $\mu_l = 1.8$*

Solution : In air

$$\frac{1}{f_a} = ({}_a\mu_g - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = (1.5 - 1)\left(\frac{1}{10} + \frac{1}{10}\right)$$

$$f_a = 10\ cm$$

$$P_a = \frac{1}{f_a} = \frac{1}{10 \times 10^{-2}}$$

$$P_a = 10\ dioptres$$

In liquid

$$\frac{1}{f_l} = ({}_l\mu_g - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

$$= \left(\frac{\mu_g}{\mu_l} - 1\right)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \left(\frac{1.5}{1.8} - 1\right)\left(\frac{1}{10} + \frac{1}{10}\right) = -\frac{1}{6} \times \frac{2}{10}$$

$$f_l = -30\ cm$$

$$P_l = \frac{1}{f_l} = -\frac{1}{30 \times 10^{-2}}$$

$$P_l = -3.33\ dioptres$$

9.5 A needle of size 5 cm is placed 45 cm from a lens produced an image on a screen placed 90 cm away from the lens. Identify the type of the lens and calculate its focal length and size of the image.

Data : *h_1 = 5 cm, u = − 45 cm, v = 90 cm, f = ? h_2 = ?*

Solution : *We know that* $\frac{1}{f} = \frac{1}{v} - \frac{1}{u} = \frac{1}{90} - \frac{1}{-45}$

∴ f = 30 cm

Since f is positive, the lens is converging

Since $\frac{h_2}{h_1} = \frac{v}{u}$ $\quad \frac{h_2}{5} = \frac{90}{-45} = -2$

∴ h_2 = −10 cm *(The –ve sign indicates that the image is real and inverted)*

Self evaluation

(The questions and problems given in this self evaluation are only samples. In the same way any question and problem could be framed from the text matter. Students must be prepared to answer any question and problem from the text matter, not only from the self evaluation.)

9.1 The number of images of an object held between two parallel plane mirrors.

(a) infinity *(b) 1*

(c) 3 *(d) 0*

9.2 Radius of curvature of concave mirror is 40 cm and the size of image is twice as that of object, then the object distance is

(a) 20 cm *(b) 10 cm*

(c) 30 cm *(d) 60 cm*

9.3 A ray of light passes from a denser medium strikes a rarer medium at an angle of incidence i. The reflected and refracted rays are perpendicular to each other. The angle of reflection and refraction are r and r'. The critical angle is

(a) $\tan^{-1}(\sin i)$ *(b)* $\sin^{-1}(\tan i)$

(c) $\tan^{-1}(\sin r)$ *(d)* $\sin^{-1}(\tan r')$

9.4 *Light passes through a closed tube which contains a gas. If the gas inside the tube is gradually pumped out, the speed of light inside the tube*

(a) increases *(b) decreases*

(c) remains constant *(d) first increases and then decreases*

9.5 *In Michelson's experiment, when the number of faces of rotating mirror increases, the velocity of light*

(a) decreases *(b) increases*

(c) does not change *(d) varies according to the rotation*

9.6 *If the velocity of light in a medium is (2/3) times of the velocity of light in vacuum, then the refractive index of that medium is.*

(a) 3/2c *(b) 2c/3*

(c) 2/3 *(d) 1.5*

9.7 *Two lenses of power +12 and −2 dioptre are placed in contact. The focal length of the combination is given by*

(a) 8.33 cm	*(b) 12.5 cm*
(c) 16.6 cm	*(d) 10 cm*

9.8 *A converging lens is used to form an image on a screen. When the lower half of the lens is covered by an opaque screen then,*

half of the image will disappear

complete image will be formed

no image is formed

intensity of the image is high

9.9 *Two small angled prism of refractive indices 1.6 and 1.8 produced same deviation, for an incident ray of light, the ratio of angle of prism*

(a) 0.88	*(b) 1.33*
(c) 0.56	*(d) 1.12*

9.10 *Rainbow is formed due to the phenomenon of*

refraction and absorption

dispersion and focussing

refraction and scattering

dispersion and total internal reflection

9.11 *State the laws of reflection.*

9.12 *Show that the reflected ray turns by 2θ when mirror turns by θ.*

9.13 *Explain the image formation in plane mirrors.*

9.14 *Draw graphically the image formation in spherical mirrors with different positions of the object and state the nature of the image.*

9.15 *What is the difference between the virtual images produced by*

(i) plane mirror (ii) concave mirror (iii) convex mirror

9.16 *The surfaces of the sun glasses are curved, yet their power may be zero. Why?*

169

9.17 *Prove the mirror formula for reflection of light from a concave mirror producing (i) real image (ii) virtual image.*

9.18 *With the help of ray diagram explain the phenomenon of total internal reflection. Give the relation between critical angle and refractive index.*

9.19 *Write a note on optical fibre.*

9.20 *Explain Michelson's method of determining velocity of light.*

9.21 *Give the importance of velocity of light.*

9.22 *Derive lens maker's formula for a thin biconvex lens.*

9.23 *Define power of a lens. What is one dioptre?*

9.24 *Establish the relation* $\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$ *of thin lenses in contact.*

9.25 *Derive the relation* $\mu = \dfrac{\sin \frac{A + D}{2}}{\sin \frac{A}{2}}$.

9.26 Does a beam of white light disperse through a hollow prism?

9.27 Derive an equation for dispersive power of a prism.

9.28 Describe a spectrometer.

9.29 Explain how will you determine the angle of the minimum deviation of a prism using spectrometer.

9.30 Write a note on formation of rainbows.

Problems

9.31 Light of wavelength 5000 Å falls on a plane reflecting surface. Calculate the wavelength and frequency of reflected light. For what angle of incidence, the reflected ray is normal to the incident ray?

9.32 *At what distance from a convex mirror of focal length 2.5 m should a boy stand, so that his image has a height equal to half the original height?*

9.33 *In a Michelson's experiment the distance travelled by the light between two reflections from the octagon rotating mirror is 4.8 km. Calculate the minimum speed of the mirror so that the image is formed at the non−rotating position.*

9.34 *If the refractive index of diamond be 2.5 and glass 1.5, then how faster does light travel in glass than in diamond?*

9.35 *An object of size 3 cm is kept at a distance of 14 cm from a concave lens of focal length 21 cm. Find the position of the image produced by the lens?*

9.36 *What is the focal length of a thin lens if the lens is in contact with 2.0 dioptre lens to form a combination lens which has a focal length of −80 cm?*

9.37 *A ray passes through an equilateral prism such that the angle of incidence is equal to the angle of emergence and the later is equal to 3/4 of the angle of prism. Find the angle of deviation.*

9.38 *The refractive indices of flint glass of equilateral prism for 400 nm and 700 nm are 1.66 and 1.61 respectively. Calculate the difference in angle of minimum deviation.*

9.39 *White light is incident on a small angled prism of angle 5^O. Calculate the angular dispersion if the refractive indices of red and violet rays are 1.642 and 1.656 respectively.*

9.40 *A thin prism of refractive index 1.5 deviates a ray by a minimum angle of 5^O. When it is kept immersed in oil of refractive index 1.25, what is the angle of minimum deviation?*

Answers

9.1	(a)	**9.2**	(b)	**9.3**	(b)
9.4	(a)	**9.5**	(c)	**9.6**	(d)
9.7	(d)	**9.8**	(b)	**9.9**	(b)
9.10	(d)				

9.31 *5000 Å ; 6×10^{14} Hz ; 45^{o}*

9.32 *2.5 m*

9.33 *7.8×10^{3} rps*

9.34 *1.66 times*

9.35 *− 8.4 cm*

9.36 *–30.8 cm*

9.37 *30^{o}*

9.38 *4^{o}*

9.39 *0.07^{o}*

9.40 *2^{o}*

UNIT 6

WAVE OPTICS

1. Introduction

- Light is a form of energy .This fact was predicted by Maxwell on theoretical grounds and was verified by Lebedew experimentally in 1901
- Since light is a form of energy, its transmission from one place to another can be understood in terms of transmission of energy
- There are only two modes of propagation of energy through any material medium
    1. energy is carried by stream of material particles travelling with finite velocity
    2. Transfer of energy by wave motion without actual travelling with matter
- First mode of energy transfer leads to Newton's corpuscular theory of light in which he tried to understand travel of light in the straight line assuming that luminous body emits very minute and weightless particles called corpuscles travelling through empty space in straight lines in all directions with the speed of light and carry KE with them
- This corpuscular theory of light can fairly explain the phenomenon of reflection .refraction and rectilinear propagation of light but failed to explain the phenomenon of interference, diffraction and polarization etc
- Second mode of energy transfer leads the wave theory of light which was put forward by Dutch physicist Christian Huygens in 1678
- Huygens suggested that light may be a wave phenomenon produced by mechanical vibrations of an all pervading hypothetical homogenous medium called eather just like those in solids and liguid .This medium was supposed to be mass less with extremely high elasticity and very low density
- At first wave theory of light was not accepted primarily because of Newton's authority and also light could travel through vacuum and waves require a medium to propagate from one point to other
- Wave theory of light first begin to gain acceptance when double slit experiment of Thomas Young in 1801 established that light is indeed a wave phenomenon
- After Young's double slit interference experiment ,many experiments were carried out by scientists involving interference and diffraction of light waves which could only be satisfactory explained by assuming wave model of light
- Later on in nineteenth century Maxwell put forwards his electromagnetic theory and predicted the existence of electromagnetic waves and calculated the speed of EM waves in free space and found that this value was very close to the measured value of speed of light
- He then suggested light must be an EM wave associated with changing electric and magnetic field which result in the propagation of light or EM waves in vacuum .So no material medium (like ether suggested by Huygens) is required for the propagation of light wave from one place to another .This argument established that light is a wave phenomenon
- In this chapter we will study the various phenomenon related to wave nature of light

2) Wave fronts and rays

- Consider the figure given below in which a point source of light S starts a distance or wave in air

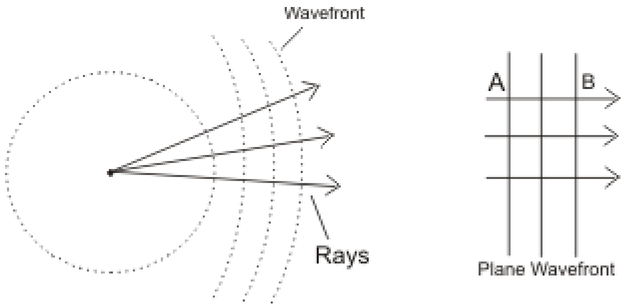


Figure 1.Wavefronts and rays

- These waves will travel in all directions with the same velocity c, which is the velocity of light
- After time t, distance travelled by the wave would be equal to ct and light energy thus reaches the surface of the sphere of radius ct with S as its center as shown in figure 1
- The surface of such a sphere is known as wave front of light at this instant and all the particles forming wave front are in the same phase of vibration
- With the passage of time wave travels farther and new wave fronts are obtained .These are all the surfaces of spheres of center S
- Thus at any instant of time wave front may be defined as the locus of all the particles in the medium which are being distributed at the same instant of time and are in the same phase of vibration
- At points very far away from source S such as A or B the wave fronts are parts of the sphere of very large radius so at any such large distances from source wave fronts are substantially plane
- Rays are defined as normal's to the wave fronts and in case of plane wave fronts ,rays are all parallel to one another as shown in figure 1.

3) Huygens's principle

- Huygens principle of wave propagation is a geometrical description used to determine the new position of a wave front at later time from its given position at any given instant of time. It is based on two principles
    1. Each point on a given or primary wave front acts like a new source sending out disturbance in all directions and are known as secondary wavelets
    2. The envelope or the tangential plane to these secondary wavelets consitutes the new wave front
- To understand the propagation of wave on the basis of these postulates consider the figure given below

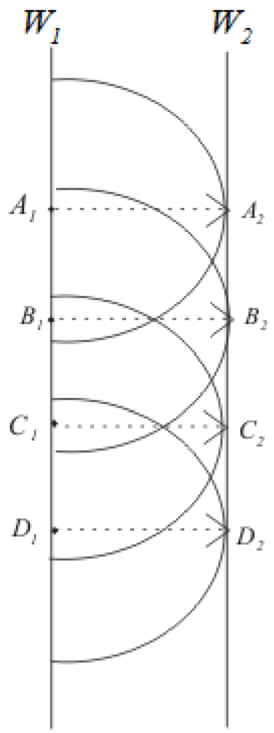


Figure 2. Huygen's geometrical construction for wave propagation

- For simplicity we are considering the simple case of a plane wave
  a) Let at time t=0,W_1 be the wave front which spates those part of the medium which are undisturbed from those where wave has already reached
  b) Each point on W_1 acts as a source of secondary waves by sending out spherical wave of radius vt where v is the velocity of the wave
  c) After some time t, the disturbance in the medium reaches all points within the region covered by all t these secondary waves. The boundary of this region is new wave front W_2 and W_2 is the surface tangent to all spheres and is known as forward envelope of these secondary wave fronts
  d) The secondary wave fro pint A_1 on W_1 touches W_2 at A_2.The line connecting point A_1 and A_2 on wavelength W_1 and W_2 respectively is a ray of length vt.This is the reason why rays are perpendicular to the wave fronts
  e) We can repeat this construction starting with W_2 to get the next wave fronts W_3 at some later time t_2 and so on

- We have explained Huygens construction using a place wave fronts but the construction is more general than our simple example .The wave fronts can have any shape and sped of waves can be different at different places and in different direction's

4) Reflection of and Refraction of plane waves using Huygens's principle

i) Reflection of plane wave at plane surface:-

- Consider the figure given below which shows incident and reflected wave fronts when a plane wave fronts travels towards a plane reflecting surface

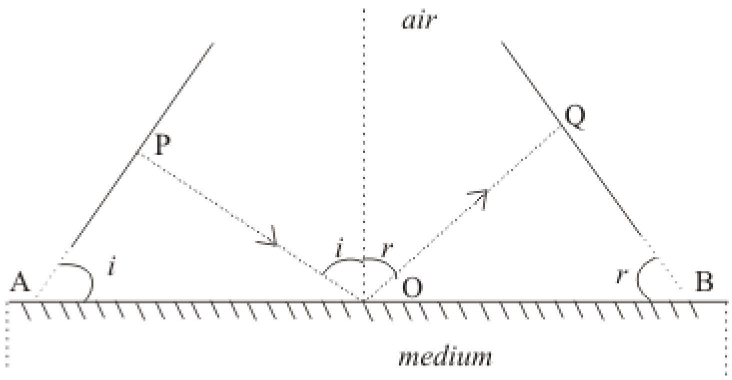


Figure 3. Wavefronts and corresponding waves for reflection at plane reflecting surface

- POQ is the ray normal to both incident and reflected wave fronts
- The angle of incidence i and angle of reflection r are the angles made by incidence and reflected rat respectively with the normal and these are also the angles between the wave fronts and the surface as shown in the figure 3
- The time taken by the ray POQ to travel from incident wave front to then reflected one is Total time from P to Q= $t=PO/v_1 + OQ/v_1$
  where v_1 is the velocity of the wave. From figure (3)

$$t = \frac{AO\sin i}{v_1} + \frac{OB\sin r}{v_1}$$

$$= \frac{OA\sin i + (AB - OA)\sin r}{v_1}$$

$$= \frac{AB\sin r + OA(\sin i - \sin r)}{v_1} \qquad (1)$$

- There can be different rays normal to incident wave front and they can strike plane reflecting surface at different point O and hence they have different values of OA
- Since tome travel by each ray from incident wave front to reflected wave front must be same so, right side of equation (1) must be independent of OA.This conditions happens only if (sini-sin r)=0

or i=r
Thus law of reflection states that angle of incidence i and angle of reflection are always equal

ii) Refraction of plane waves at plane surfaces:-

- Consider the figure given below which shows a plane surface AB separating medium 1 from medium 2
- v_1 be the speed of light in medium 1 and v_2 the speed of light in medium 2

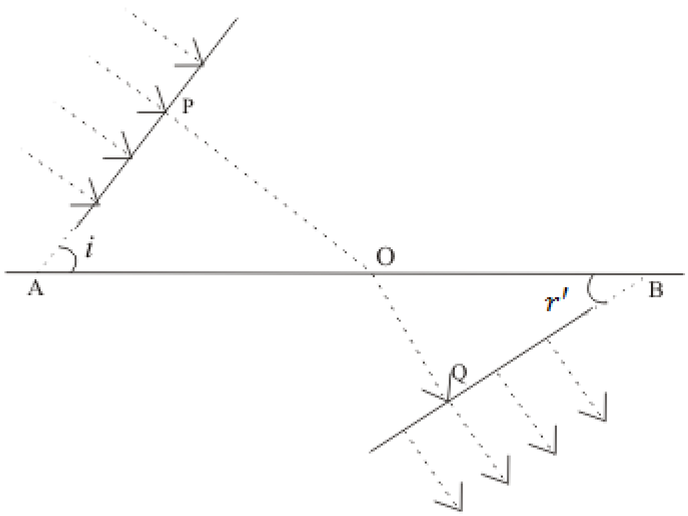


Figure 4. Refraction at plane wavefront and corresponding rays by a plane surface

- Incident and refracted wave front makes angles i and r' with surface AB where r' is called angle of refraction
- Time taken by ray POQ to travel between incident and refracted wave fronts would be

$$t = \frac{PO}{v_1} + \frac{OQ}{v_2}$$
$$= \frac{OA \sin i}{v_1} + \frac{(AB - OA)\sin r'}{v_2}$$
$$= \frac{AC}{v_2}\sin r' + OA\left(\frac{\sin i}{v_1} - \frac{\sin r'}{v_2}\right) \qquad (2)$$

- Now distance OA would be different for different rays .So time t should be independent of any ray we might consider
- This can be achieved only if coefficient of OA in equation (2) becomes equal to zero or

$$\frac{sini}{sinr'} = \frac{v_1}{v_2} = n(constant) \qquad (3)$$

- Equation (3) is nothing but ***Snell's law of refraction*** where n is called the reflective index of second medium with respect to the first medium.

iii) Refractive index

- The ratio of phase velocity of light c in vacuum to its value v_1 in a medium is called the
  r $\mu_1 = n_1 = {}^{c}/_{v_1}$ (4)

- When light travels from medium 1 to medium 2,what we measure is the refractive index of medium 2 relative to medium 1 denoted by n_{12} (or μ_{12}).Thus

  $$n_{12} = \frac{sini}{sinr'} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \qquad (5)$$

  where n_1 is refractive index of medium 1 with respect to vacuum and n_2 is refractive index of medium 2 w.r.t. vacuum
- When light travels from one medium to another the frequency v=1/T remains same i.e. $v_1 = v_2$
- Since the velocities of light v_1 and v_2 are different is different medium ,the wavelength λ_1 and λ_2 are also different i.e.,

  $$v = \nu\lambda$$
  $$\Rightarrow \frac{v_1}{v_2} = \frac{\nu_1\lambda_1}{\nu_2\lambda_2} \qquad (6)$$

  the wavelength of light in the medium is directly proportional to phase velocity and hence inversely proportional to the refractive index

5) Principle of Superposition of waves

- When two or more sets of waves travel through a medium and cross one another the effects produced by one are totally independent of the
- At any instant the resultant displacement of a particle in the medium depends on the phase difference between the waves and is the algebraic sum of the displacement it would have at the same instant due to each separate set. This is known as the principle of superposition of waves and forms the basis of whole theory of interference of waves discovered by Young in 1801

- If at any instant $\mathbf{y_1}$,$\mathbf{y_2}$,$\mathbf{y_3}$,--- are the displacements due to different waves present in the medium then according to superposition principle resultant displacement **y** at any instant would be equal to the vector sum of the displacements ($\mathbf{y_1}$,$\mathbf{y_2}$,$\mathbf{y_3}$) due to the individual waves i.e.,

  $$\boldsymbol{y} = \boldsymbol{y_1} + \boldsymbol{y_2} + \boldsymbol{y_3} + \cdots$$

- The resultant displacement of the particles of the medium depends on the amplitude ,phase difference and frequency of the superposing waves
- Consider two waves of same frequency f and wavelength λ travelling through a medium in the same direction and superpose at any instant of time say t
- Equation of these waves at time t is

  $$\mathbf{y}_1 = \mathbf{a}_1 \sin\left[\frac{2\pi}{\lambda}(ft - x)\right]$$

  $$\mathbf{y}_2 = \mathbf{a}_2 \sin\left[\frac{2\pi}{\lambda}(ft - x) + \phi\right] \qquad \text{---(7)}$$

  where φ is the phase difference between the waves

- According to principle of superposition of waves ,resultant displacement of particles equals

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$$

$$\mathbf{y} = \mathbf{a}_1 \sin\left[\frac{2\pi}{\lambda}(ft - x)\right] + \mathbf{a}_2 \sin\left[\frac{2\pi}{\lambda}(ft - x) + \phi\right] \qquad \text{-----(8)}$$

Now from trigonometry identity

$$\sin(A + B) = sinAcosB + cosAsinB$$

So,

$$\sin\left[\frac{2\pi}{\lambda}(ft - x) + \phi\right] = \sin\frac{2\pi}{\lambda}(ft - x)\cos\phi + \cos\frac{2\pi}{\lambda}(ft - x)\sin\phi$$

Putting it in equation (8) we find

$$y = a_1 \sin\left[\frac{2\pi}{\lambda}(ft - x)\right] + a_2 \sin\left[\frac{2\pi}{\lambda}(ft - x)\right]\cos\varphi + a_2 \cos\frac{2\pi}{\lambda}(ft - x)\sin\varphi \qquad \text{---(}$$

Let us suppose

$$Acos\theta = a_1 + a_2cos\phi \ and \ Asin\theta = a_2sin\phi$$

Putting them in equation (9) we have

$$y = A\cos\theta\sin\frac{2\pi}{\lambda}(ft - x) + A\sin\theta\cos\frac{2\pi}{\lambda}(ft - x) = A\sin\left[\frac{2\pi}{\lambda}(ft - x) + \theta\right]$$

where

$$A = \sqrt{a_1^2 + a_2^2 + 2a_1a_2\cos\phi} \qquad \text{----- (11)}$$

$$\theta = \tan^{-1}\left[\frac{a_2\sin\phi}{a_1 + a_2\cos\phi}\right] \qquad \text{---- (12)}$$

- We know that intensity of waves is proportional to its amplitude i.e.

$$I \propto A^2 = a_1^2 + a_2^2 + 2a_1a_2cos\phi$$

For maximum intensity

$$\cos\phi = 1$$

$$\phi = 2n\pi$$

$$n = 0,1,2.....$$

Therefore,

$$I_{max} = (a_1 + a_2)^2 \qquad \text{--- (13)}$$

For $a=a_1=a_2$

$$I_{max}=4a^2$$

For minimum intensity

$$\cos\varphi = -1$$

$$\phi = (2n+1)\pi$$

$$n = 0,1,2....$$

$$I_{min} = (a_1 - a_2)^2 \qquad \text{--- (14)}$$

For $a_1 = a_2 = a$

$$I_{min}=0$$

6) Interference of light waves

- Interference of light wave is the modification in distribution of light energy obtained by superposition of two or more waves
- At some points where crest of the one wave falls on the crest of another ,resultant amplitude is maximum
- At some points where crest of one wave falls on trough of another, the resultant amplitude become minimum and hence intensity of the light is minimum
- At points, where the resultant intensity of light is maximum ,the interference is said to be constructive
- At points where resultant intensity of light is minimum ,interference is said to be destructive

7) Coherent Sources

- Coherent sources are those sources of light which emit continuous light waves of same wavelength ,same frequently and are in same phase or have a constant phase difference
- For observing interference phenomenon ,coherence of waves is a must
- For light waves emitted by two sources of light to remain coherent ,the initial phase difference between waves should remain constant in time. If the phase difference changes continuously or randomly with time then the sources are incoherent
- Two independent sources of light are not coherent and hence cannot produce interference because light beam is emitted by millions of atoms radiating independently so that phase difference between waves from such fluctuates randomly many times per second
- Two coherent sources can be obtained either by the source and obtaining its virtual image or by obtaining two virtual images of the same source. This is because any change in phase in real source will cause a simultaneous and equal change it its image
- Generally coherence in interference is obtained by two methods
  i) Division of wave front where wave front is divided into two parts by reflection ,refraction or diffraction and those two parts reunite at a small angle to produce interference such as in

case of Young Double slit experiment ,Fresnel bi-prism .
ii) Division of amplitude whose amplitude of a section of wave front is divided into two parts and reunited later to produce interference such as in case of interference due to thin films

- Laser light is almost monochromatic with light spreading and two independent laser sources can produce observable interference pattern

8) Conditions for sustained interference of light waves

- Two sources should continuously emit waves of same wavelength or frequency
- The amplitudes of the two interfering waves should be equal or approximately equal in order to reduce general illumination
- The sources of light must be coherent sources
- Two sources should be very narrow as a broad source is equivalent to large number of narrow sources lying side by side which causes loss of interference pattern resulting general illumination
- Two sources emitting set of interfering beams must be placed very close to each other so that wavelength interact at very small angles

9)Young Double slit experiment

- Young in 1801 demonstrated interference phenomenon through double slit experiment
- In his experiment ,he divided a single wave front into two and these two slit wave fronts acts as if they emerged from two sources having fixed relationship
- when these two waves were allowed to interfere ,they produce a sustained interference pattern
- In his original experiment he illuminates a pin hole S using a light source and light diverging from pinhole which contains two sets of pinholes S_1 and S_2 equidistant from S and very close to one another as shown below in the figure

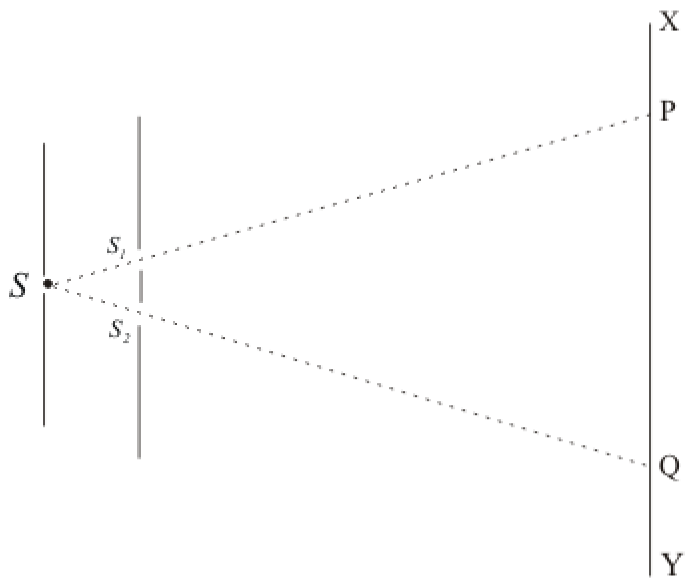


Figure 5. Young's experiment on interference

- Two two sets of spherical waves coming out of the pin holes S_1 and S_2 were coherent and interfered with each other to form a symmetrical pattern of varying intensity on screen XY
- This interference pattern disappear when any one of the pinholes S_1 or S_2 is closed

- Young used the superposition principle to explain the interference pattern and by measuring the distance between the fringes he managed to calculate the wavelength of light.

10) Theory of interference fringes

- In young's double slit experiment ,light wave produce interference pattern of alternate bright and dark fringes or interference band
- To find the position of fringes, their spacing and intensity at any point P on screen XY .Consider the figure given below

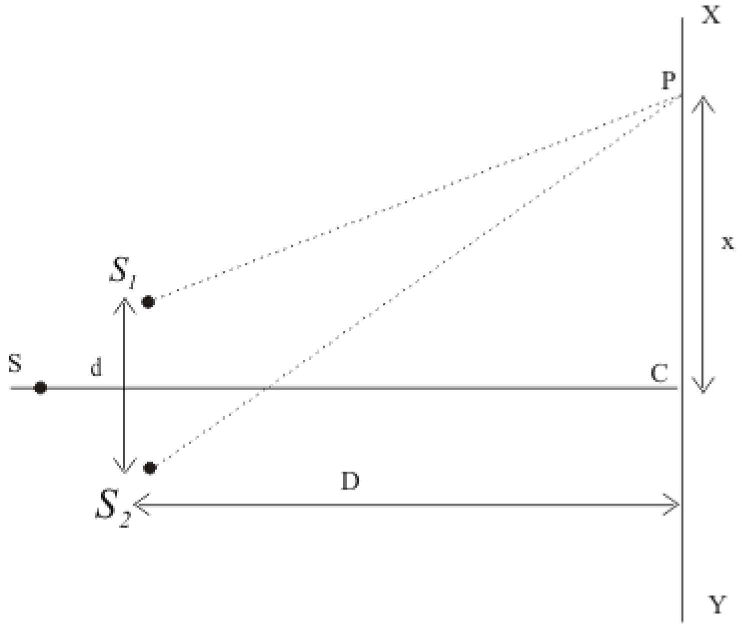


Figure 6. Interference in Young's double slit experiment

- Here S_1 or S_2 two pin holes of YDS interference experiment and position of <u>maxima</u> and minima can be determined on line XOY parallel to Y-axis and lying on the plane parallel to S,S_1 or S_2
- Consider a point P on XY plane such that CP = x.The nature of interference between two waves reaching point P depends on the path difference S_2P-S_1P

- from figure (6)

$$S_1P^2 = D^2 + (x-\frac{d}{2})^2 \quad = D^2\left[1+\frac{(x-d/2)^2}{D^2}\right] \qquad \text{---- (15)}$$

$$S_2P^2 = D^2\left[1+\frac{(x+d/2)^2}{D^2}\right] \qquad \text{--- (16)}$$

$$S_2P^2 - S_1P^2 = (x+\frac{d}{2})^2 - (x-\frac{d}{2})^2 \quad = 2xd \qquad \text{---- (17)}$$

$$(S_2P - S_1P)(S_2P + S_1P) = 2xd$$

$$S_2P - S_1P = \frac{2xd}{(S_2P + S_1P)} \qquad \text{---- (18)}$$

for x, d<<< D , S_1P+S_2P =2D
with negligible error included , path difference would be

$$S_2P - S_1P = \frac{2xd}{2D} = \frac{xd}{D} \qquad (19)$$

And corresponding phase difference between wave is

$$\phi = path\ difference \times \frac{2\pi}{\lambda}$$

$$phase\ difference = \frac{2\pi}{\lambda} \times \frac{xd}{D} \qquad (20)$$

i) Condition of bright fringes(constructive interference)

- If the path difference (S_2P-S_1P) is even multiple of λ/2,the point P is bright

$$\therefore \frac{xd}{D} = \frac{2n\lambda}{2}$$

$$or, x = \frac{n\lambda D}{d} \qquad (21)$$

- Equation (21) gives the condition for bright fringes or constructive interference

ii) Condition for dark fringes (destructive interference)

- I ark. So,

$$\therefore \frac{xd}{D} = \frac{(2n-1)\lambda}{2}$$

$$or, x = \frac{(2n-1)\lambda D}{d} \qquad (22)$$

- Equation (22) gives the condition for dark fringes or destructive interference
- From equations (21) and (22) ,we can get position of alternate bright and dark fringes respectively
- Distance between two consecutive bright fringes is given by

$$x_{n-1} - x_n = \frac{D}{2d}(n-1)\lambda - \frac{D}{2d}n\lambda = \frac{D}{d}\lambda$$

And for dark fringes

$$x_{n-1} - x_n = \frac{D}{2d}(2n+1)\lambda - \frac{D}{2d}(2n-1)\lambda = \frac{D}{d}\lambda$$

- Thus the distance between two successive dark and bright fringes is same. This distance is known as fringe width and is denoted by β. Thus

$$\beta = \frac{D}{d}\lambda \qquad (23)$$

11) Displacement of fringes

- when a film of thickness t and refractive index μ is introduced in the path of one of the source of light ,then fringe shift occur as optical path difference changes

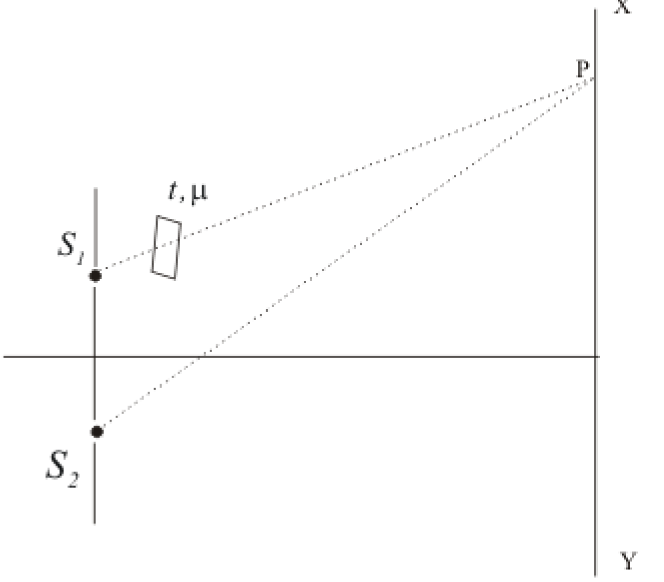


Figure 7. Thin film is introduced in path of one of source of light

- Time required by light to reach from S_1 to point P

  $$= \frac{S_1P - t}{c} + \frac{t}{v}$$

  where v=c/μ

  $$T = \frac{S_1P + t(\mu - 1)}{c}$$

- Hence equivalent path that is covered by light in air is $S_1P+t(\mu-1)$
- Optical path difference at P

  $$= S_2P - [S_1P + \mu t - t]$$
  $$= S_2P - S_1P - [\mu -]t$$
  $$= \frac{xd}{D} - [\mu -]t$$

  Therefore n^{th} fringe shift is given by

  $$\Delta x = \frac{D(\mu - 1)t}{d}$$

  as,

  $$\beta = \frac{D}{d}\lambda$$
  $$\Delta x = \frac{\beta(\mu - 1)t}{\lambda}$$

  where λ is the wavelength of the wave

UNIT15

1. Introduction

- It is a common observation with the waves of all kind that they bend round the edge of an obstacle
- Light like other waves also bends round corners but in comparison to sound waves small bending of light is due to very short wavelength of light which is of the order of 10^{-5}
- This effect of bending of beams round the corner was first discovered by grimed (Italy 1618-1663)
- We now define diffraction of light as the phenomenon of bending of light waves around the corners and their spreading into the geometrical shadows
- Fresnel then explained that the diffraction phenomenon was the result of mutual interference between the secondary wavelets from the same dif wave front

- Thus we can explain diffraction phenomenon using Huygens's principle
- The diffraction phenomenon are usually divided into two classes

  i) Fresnel class of diffraction phenomenon where the source of light and screen are in general at a finite distance from the diffracting aperture

  ii) Fruanhofer class of diffraction phenomenon where the source and the screen are at infinite distance from the aperture, this is easily achieved by placing the source on the focal plane of a convex lens and placing screen on focal plane of another convex lens. This class of diffraction is simple to treat and easy to observe in practice

- Here in this chapter we will only be considering fraunhofer class diffraction by a single slit

2. Fraunhofer Diffraction by single slit

- Let us first consider a parallel beam of light incident normally on a slit AB of width 'a' which is of order of the wavelength of light as shown below in the figure

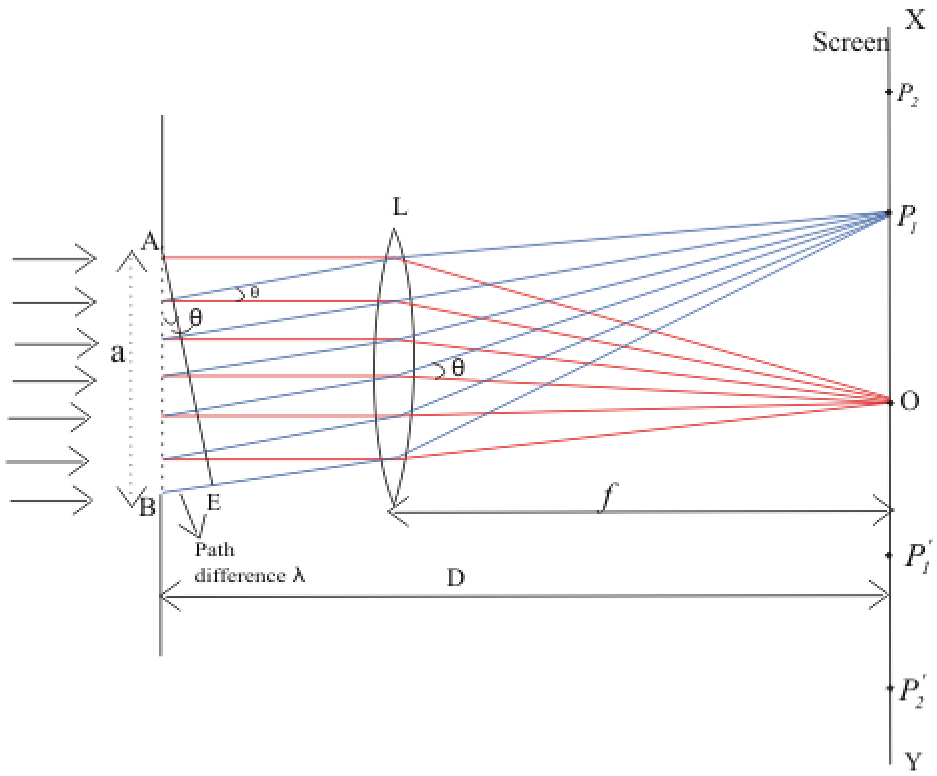


Figure 1. Fraunhofer diffraction of a plane wave at single slit

- A real image of diffraction pattern is formed on the screen with the help of converging lens placed in the path of the diffracted beam
- All the rays that starts from slit AB in the same phase reinforce each other and produce brightness at point O on the axis of slit as they arrive there in the same phase

- The intensity of diffracted beam will be different in different directions and there are some directories where there is no light

- Thus diffraction pattern on screen consists of a central bright band and alternate dark and bright bands of decreasing intensity on both sides

- Now consider a plane wave front PQ incident on the narrow slit AB. According to Huygens principle each point t on unblocked portion of wavefront PQ sends out secondary wavelets in all directions

- Their combined effect at any distant point can be found y summing the numerous waves arriving there from the principle of superposition

- Let C be the center of the slit AB.The secondary waves, from points equidistant from center C of the clit lying on portion CA and CB of wave front travel the same distance in reaching O and hence the path difference between them is zero

- These waves reinforce each other and give rise to the central maximum at point O

i) Condition for minima

- We now consider the intensity at point P_1 above O on the screen where another set of rays diffracted at a angle θ have been bought to focus by the lens and contributions from different elements of the slits do not arise in phase at P_1
- If we drop a perpendicular from point A to the diffracted ray from B,then AE as shown in figure constitutes the diffracted wavefront and BE is the path difference between the rays from the two edges A and B of the slit.

- Let us imagine this path difference to be equal to one wavelength.

- The wavelets from different parts of the slit do not reach point P_1 in the phase because they cover unequal distance in reaching P_1.Thus they would interfere and cancel out each other effect. For this to occur

  $BE=\lambda$
  Since $BE=AB\sin\theta$
  $a\sin\theta=\lambda$
  or $\sin\theta=\lambda/a$
  or $\theta=\lambda/a$ ---(1)
  As angle of diffraction is usually very small so that
  $\sin\theta=\theta$

- Such a point on screen as given by the equation (1) would be point of secondary minimum
- It is because we have assume the slit to be divided into two parts, then wavelets from the corresponding points of the two halves of the slit will have path difference of #955;/2 and wavelets from two halves will reach point P_1 on the screen in a opposite phase to produce minima

- Again consider the point P_2 in the figure 1 and if for this point path difference $BE=2\lambda$,then we can imagine slit to be divided into four equal parts

- The wavelets from the corresponding points of the two adjacent parts of the slit will have a path difference of $\lambda/2$ and will mutually interfere to cancel out each other

- Thus a second minimum occurs at P_2 in direction of θ given by
  θ=2θ/a
- Similarly n_{th} minimum at point P_n occurs in direction of θ given by
  θ_n=nθ/a ---(2)

ii) Positions of maxima

- If there is any point on the screen for which path difference
  BN=asinθ=3θ/2
  Then point will be position of first secondary maxima

- Here we imagine unblocked wavefront to be divided into three equal parts where the wavelets from the first two parts reach point P in opposite phase thereby cancelling the e effects of each other

- The secondary waves from third part remain uncancelled and produce first maximum at the given point

- we will get second secondary maximum for BN=5θ/2 and n_{th} secondary maxima for
  BN=(2n+1)θ/2 =asinθ_n ---(3)

  where n=1,2,3,4..

- Intensity of these secondary maxima is much less then central maxima and falls off rapidly as move outwards

- Figure below shows the variation of the intensity distribution with their distance from the center of the central maxima

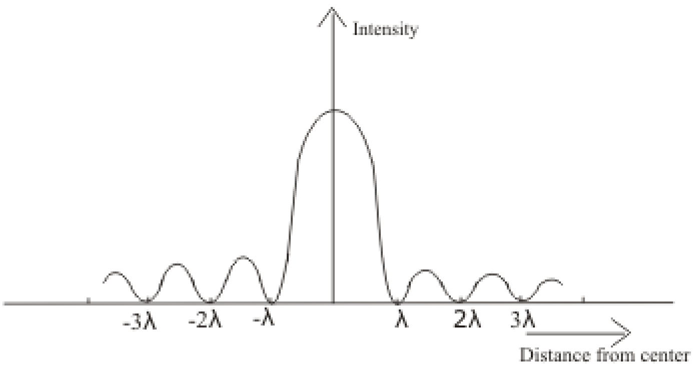


Figure 2. Intensity distribution in the diffraction due to single slit

3) Resolving power

- When two objects are very close to each other they may appear as one and we might not see them as separate objects just by magnifying them

- To separate two objects which are very close together, optical instrument such as telescope ,microscope,prism,grating etc are employed

- The separation of such close object is termed as resolution and the ability of an optical instrument to produce distinctly separate images of two close objects is called its resolving power

- Every optical instrument has a limit up to which it can produce distinctly separate images to two objects placed very close to each other

- After detailed study of the intensity of diffraction pattern of two very close point objects ,lord Rayleigh suggested that the two objects will be just resolved when central maximum of diffraction pattern of first object lies on first secondary minimum of diffraction pattern of second object

- The minimum distance between two point object which can just appear to be as separate by optical instrument is called the limit of resolution of the instrument

- we will discuss about the resolving power of optical instrument when we study exclusively about optical instruments

4) Polarization of light

- Waves are generally of two types

  i) Longitudinal waves: In case of longitudinal waves ,particles of the medium oscillates along the direction of the propagation of the waves
  ii) Transverse waves: In this case direction of oscillation of particles is perpendicular to the direction of propagation of waves

- we already know that light is an EM wave in which electric and magnetic field vector vary sinosoidally,perpendicular to each other as well as perpendicular to the direction of propagation of light wave

- This shows that light waves consists of transverse waves
- The fact that light consists of transverse waves can be confirmed in the experiments in which beams of lights were allowed to pass through Polaroid which are artificial crystalline materials that allow lights vibrations to pass through only in a particular plane

- We would now observe the light passing through two Polaroid A and B placed one behind another in from of source of light as shown below in the figure

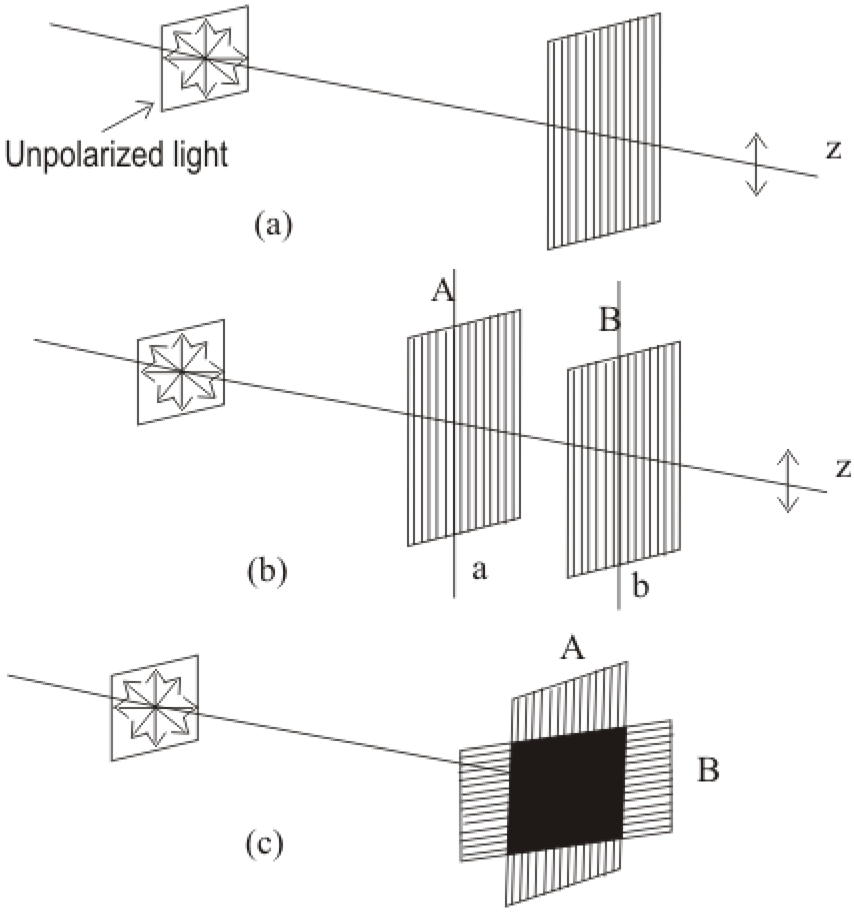


Figure 3. Plane polarization of light

- when axes a and b of Polaroid A and B respectively are parallel to each other then the light through Polaroid B appears slightly darker intensity of light is reduced after being transmitted from Polaroid A

- Since axis of both the Polaroid are parallel to each other so light transmitted from Polaroid A is transmitted as it is by Polaroid B
- Now if start rotating Polaroid B about the z-axis ,one will observe the variation of intensity i.e. light passing through crystal B becomes darker and darker and disappears at one stage

- This happens when axis a of Polaroid A is perpendicular to axis b of the Polaroid B as shown in fig 3(b)

- Again on rotating Polaroid B in same direction light reappears and becomes brightest when the axis a and b are again parallel

- This simple experiment proves that light consists of transverse waves

- Here in this experiment Polaroid A acts as polarizer and the beam transmitted through polarizer is linearly polarized and second Polaroid acts as analyzer

5) Vibrations in unpolarized and polarized light

- A light wave is a transverse wave having vibration at right angles to the direction of propagation
- An ordinary beam of light consists of million of lights waves each with its own plane of vibration so it have wave vibrating in all possible plane with equal probability .Hence an ordinary beam of light in unpolarized
- If we consider the light beam being propagated in a direction perpendicular to the plane of paper while its vibrations are in the plane of the paper then figure 4 given below shows that vibrations in ordinary lights occurs in every plane perpendicular to the direction of propagation of light and are in the plane of the paper

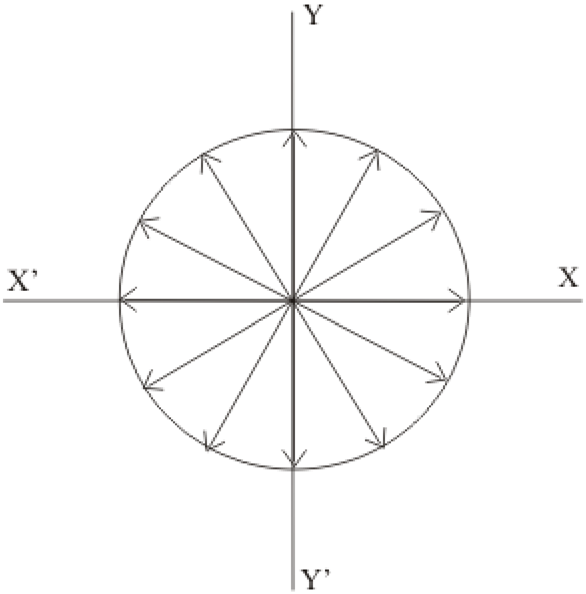


Figure 4. Vibrations of the beam of unpolarised light

- As it can be seen from the figure that amplitude of vibrations are all equal
- When such a beam of unpolaroized light is incident on a Polaroid the emergent light is linearly polarized with vibrations in a particular directions
- The direction of vibrations of beam transmitted by the Polaroid depends on the orientation of the Polaroid
- Consider the vibrations in ordinary light when it is incident on the polaroid as shown in figure 3(a). Each vibrations can be resolved into two components, one in a direction parallel to a which is the direction of transmission of light through polaroid and the other direction m perpandicular to a as shown in the figure given below.

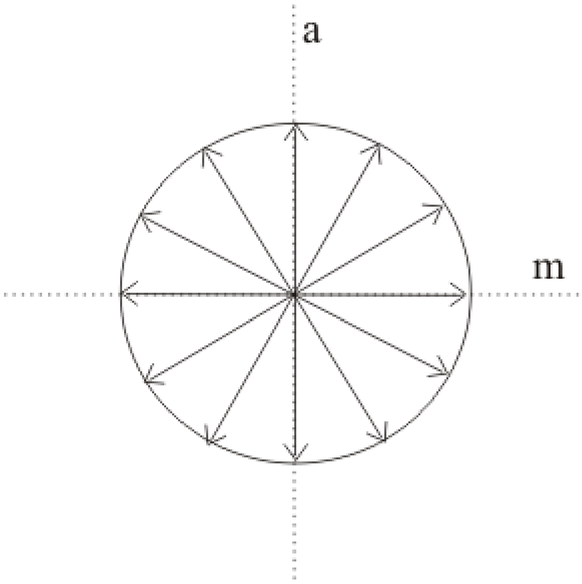


Figure 5. Plane polarised waves by selective absorption

- Polaroids absorbs the light due to vibrations parallel to m , known as ordinary rays but allow light due to other vibration , known as extra ordinary rays, to pass through.
- So the plane polarised light due to vibrations in one plane is produced as shown in figure 3(a).
- The Polaroid absorbs the light vibration along a particular direction and the component at right angles to it is allowed to pass through the Polaroid.
- This selective absorption of light vibrations along a particular direction is also shown by certain natural crystals for example tourmaline crystal

6) Polarization of reflection

- This simple method of obtaining plane polarized light by reflection was discovered by malus in 1808
- We found that when a beam of light is reflected from the surface of a transparent medium like glass or water, the reflected light is partially polarized and degree of the polarization varies with angle of incidence
- The percentage of polarized light is greatest in reflected beam when light beam is incident on the transparent medium with an incident angle equal to the angle of polarization
- For ordinary glass with refractive index =1.52 ,angle of polarization is 57.5^0
- Figure below shows the polarization of light by reflection

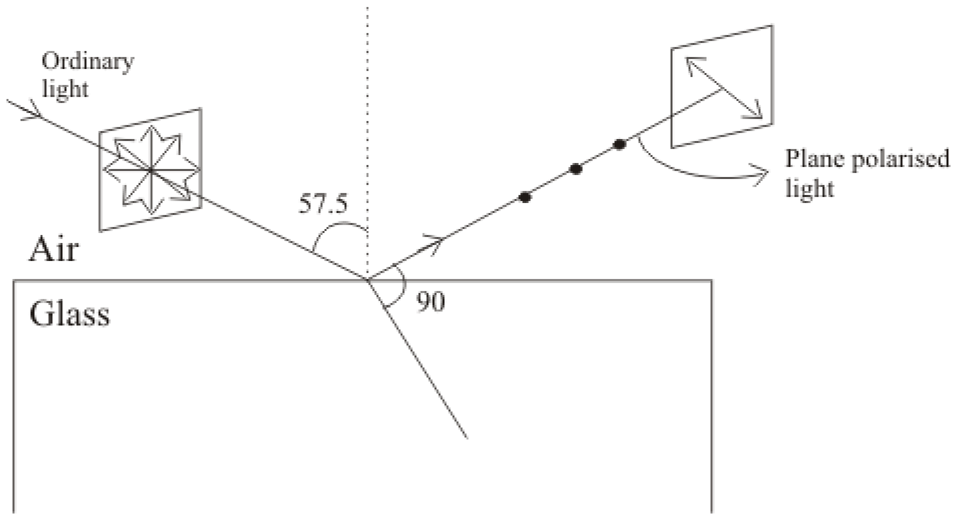


Figure 6. Plane polarised light by reflection

- we can use a Polaroid as an analyzer to show that reflected light is plane polarized . we rather say that reflected light is partially plane polarized
- the examination of transmitted light shows the variation in intensity indicating that the light is partially polarized
- The vibrations of this plane polarized reflected light are found to be perpendicular to the plane of incidence and therefore ,the reflected light is said to be plane polarized in the plane of incidence

7) polarization by scattering

- when an unpolarized beam of light is allowed to pass through a medium containing gas or molecules it gets scattered and the beam scattered at 90^0 to the incident beam is plane polarized (having vibrations in one plane). This phenomenon is called polarization by scattering
- The blue light we receive from sky is partially polarized ,although our eye can not distinguish it from an umpolaroized light but if we view it through a Polaroid which can be rotated we can clearly see it to be as partially polarized
- this is nothing but the sunlight that has been scattered when it encountered the molecules of the earth surface

8) Polarization and electric vector

- we know that a transverse EM wave contains both electric and magnetic field and electric field in light wave is perpendicular to the direction of propagation of light wave
- We generally define direction of light vibration to be that of the electric vector **E**
- For an ordinary unpolarised beam electric vector keeps changing its direction in random manner

- when light is allowed to pass through a Polaroid the emergent light is plane polarized with its electric vector vibrating in a particular direction

- The direction of vector of emergent beam depends on the orientation of the Polaroid and the plane of polarization is designed as the plane containing E-vector and light ray

- Figure below shows the plane polarized light due to i) a vertical E-vector ii) a horizontal E-vector

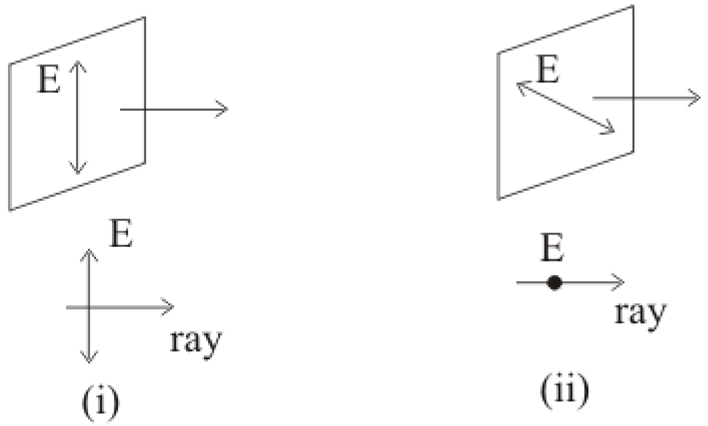


Figure 7. Electric vectors in plane polarised light

9)Law of Malus

- consider the figure given below in which an unpolarized light passes through a Polaroid P_1 and then through Polaroid P_2 making an angle θ with y ax-s

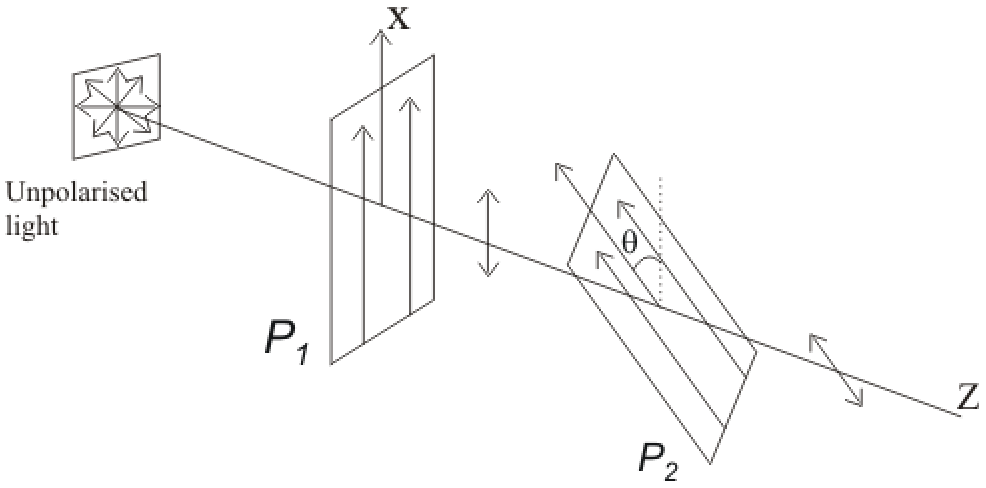


Figure 8 . Illustrating Malus Law

- After light propagating along x-direction passes through the Polaroid P_1 ,the electric vector associated with the polarized light will vibrate only along y-axis
- Now if we allow this polarized beam to again pass through a polarized P_2 making an angle θ with y-axis then if E_0 is the amplitude of incident electric field on P_2 then amplitude of wave emerging from P_2 would be $E_0\cos\theta$ and hence intensity of emerging beam would be

  $I=I_0\cos^2\theta$ ---(4)

  where I_0 is the intensity of beam emerging from P_2 when $\theta=0$

- This equation (4) represents the law of malus
- Thus when a plane polarized light is incident on an analyzer ,the emerging light varies in accordance with the equation (4) where θ is the angle between the planes of transmission of the analyzer and the polarizer

10) Brewster's law

- Brewster law is a simple relationship between angle of maximum polarization and the refractive index of the medium
- Brewster's law is given by relationship
  $\mu=\tan i$ --(5)

  where i is the Polarizing angle
  μis the index of refraction
  It is clear from equation (5) that when light is incident at polarizing angle then reflected ray is at right angles to the reflected ray

UNIT 7

1. Discovery of nucleus

- In 1897 J.J. Thomson discovered the electron in the rays emitted from the cathode of discharge tube filled with gas at low temperatures.
- Again 1910 Thomson suggested a model for describing atom , known as 'Thomson's atomic model' which suggests that atom consists of positively charged sphere of radius 10^{-8}cm in which electrons were supposed to be embedded.
- Thomson atomic model failed as it could not give convincing explanation for several phenomenon such as, spectrum
  of atoms, alpha particle scattering and many more.
- In 1909 Gieger and Marsden employed α-particles (Helium ion) as projectile to bombard thin metallic foil.
- According to Thomson atomic model since all positive charge of atom was neutralized by the negatively charged electrons, there would be rare event for an α-particle to suffer a very large deflection , as expected force of repulsion would not be very strong.
- Surprisingly experiments of Gieger and Marsden showed large deflections of alpha particles that were many orders of magnitude and more common then expected.
- This result of Gieger and Marsden α-particle scattering experiment was explained by Sir Rutherford in 1911.
- Rutherford proposed a new atomic model in which electrons were located at much greater distance from the positive charge.
- Rutherford proposed that all the positive charge , and nearly all the mass of the atom, was concentrated in an extremely small nucleus.
- The electrons were supposed to be distributed around the nucleus in a sphere of atomic radius nearly equal to 10^{-8}cm.
- In explaining this experiment Rutherford made simple assumptions that both the nucleus and α-particles (Helium ion) were point electrical charges and the repulsive force between them is given by Coulomb's inverse square law at all distances of separation.
- These assumptions made by Rutherford were not valid if α-particle approaches the nucleus to a distance comparable with the diameter of the nucleus.
- From this experiment there emerged a picture of internal structure of atoms and it also confirmed the existence of the atomic nucleus.
- Approximate values for size and electrical charge of nucleus were calculated using data of various scattering experiments.

2. Nuclear Composition

- Atomic nuclei are build up of protons and neutrons.
- Nucleus of hydrogen atom contains only single proton.
- Charge on a proton is $+1.6\times10^{-19}$ C and its mass is 1836 times greater then that of electron.
- Neutrons are uncharged particles and mass of a neutron is slightly greater then that of a proton.

- Neutrons and protons are jointly called nucleons.
- Number of protons in nuclei of an element is equal to the number of electrons in neutral atom of that element.
- All nuclei of a given element does not have equal number of neutrons for example99.9 percent of hydrogen nuclei contains only one proton , some contain one proton and one neutron and a very little fraction contains one proton and two neutrons.
- Elements that have same number of protons but differ in number of neutrons in their nucleus are called ISOTOPES.
- Hydrogen isotope deuterium is stable but tritium is radioactive and it decays to changes into an isotope of helium.

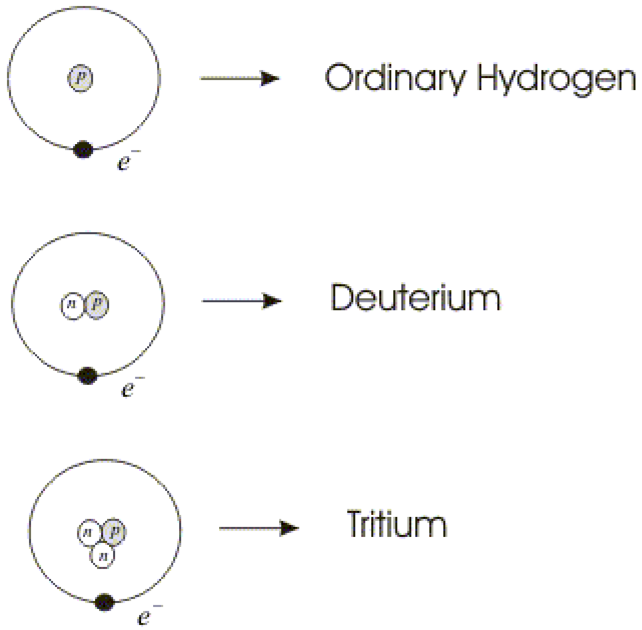


Figure :- Isotopes of Hydrogen

- In heavy water instead of ordinary hydrogen deuterium combines with oxygen.
- Symbol for nuclear species follows the pattern $^{A}X_{Z}$ where
  X= Chemical symbol of element
  Z= Atomic number of element or number of protons in the nucleus of that element.
  A= Mass number of nuclide or number of nucleons in the nucleus. A=Z+N where N is the number of neutrons in the nucleus.
- In symbolic form
  (1) hydrogen = $^{1}H_{1}$ and Deuterium = $^{2}H_{1}$
  (2) Chlorine isotopes are $^{35}Cl_{17}$ and $^{37}Cl_{17}$

3. Atomic mass

- Atomic masses refer to the masses of neutral atoms , not of bare nuclei i.e., an atomic mass always includes the masses of all its electrons.
- Atomic masses are expressed in mass units (u).
- One atomic mass unit is defined as one twelfth part of the mass of $^{12}C_{6}$ atom.
- So the mass of $^{12}C_{6}$, the most abundant isotope of carbon is 12u.

- Value of a mass unit is
  $1u=1.66054 \times 10^{-27} Kg$
- We now calculate the energy equivalent of mass unit. We know that Einstein's Mass-Energy relation is
  $\Delta E=\Delta mc^2$
  here,
  $\Delta m = 1.60 \times 10^{-27}$ Kg and
  $c = 3 \times 10^8$ m/s
  therefore
  $\Delta E = (1.60 \times 10^{-27}) \times (3 \times 10^8)^2$
  $=1.49 \times 10^{-10}$ J
  but $1eV = 1.6 \times 10^{-19}$ J
  therefore,

$$\Delta E = \frac{1.49 \times 10^{-10}}{1.60 \times 10^{-19}}$$

- or,
  $\Delta E = .931 \times 10^9$ eV
  $\Delta E = 931$ MeV
  Thus 1 amu = 931 MeV
- Mass of proton is 1.00727663 u which is equal to 1.6725×10^{-27}kg or 938.26 MeV.
- Mass of neutron is 1.0086654 u which is equal to 1.6748×10^{-27}kg or 939.55 MeV.

4. Isobars and Isotones

- Nuclei with same A but different Z are known as Isobars for example $^{40}K_{19}$ and $^{40}Ca_{20}$ share same mass number 40 but differs in one unit of Z.
- Although isobaric atoms share same mass number but they differ slightly in their masses.
- This very slight difference in masses of isobaric atoms is related to difference between energies of two atoms since small mass difference corresponds to considerable amount of difference in energies.
- Nuclei with same number of neutrons but different number of protons are called Isotones for example $^{198}Hg_{80}$ and $^{198}Au_{79}$

5. Size of nucleus

- First estimate of size of nucleus was provided by Rutherford scattering experiment.
- In Rutherford's scattering experiment incident alpha particles gets deflected by the target nucleus as long as the distance approached by the alpha particles does not exceeds 10^{-14}m and Coulomb's law remains consistent.
- Apart from Rutherford's scattering experiment various other experiments like fast electrons and neutron scattering experiments were performed to determine the nuclear dimensions.
- Since electrons interact with nucleus only through electric forces so electron scattering experiments gives information on distribution of charge in the nucleus.
- A neutron interacts with nucleus through nuclear forces so neutron scattering provides information on distribution of nuclear matter.
- It was found that the volume of a nucleus is directly proportional to the number of nucleons it contains which is its mass number A.

- If R is the nuclear radius then relationship between R and A is given as
  $R=R_0A^{1/3}$
  Where value of $R_0 \cong 1.2 \times 10^{-15} \cong 1.2$ fm and is known as nuclear radius parameter.
- Since R^3 is proportional to A this implies that density of nucleas ($\rho = m/V$) is a constant independent of A for all nuclei.
- The density of nuclear matter is approximately of the order of $!^{17}$ Kg/m^3 and is very large compared to the density of ordinary matter.

1) Introduction

- Phenomenon of radioactivity was first discovered by A.H.Bacquerel in 1896 while studying fluorescence and phosphorence of compounds irradiated by visible light
- these phosphorescent materials glow in dark after being exposed to visible light
- while conducting experiment on uranium salts, he found that uranium salts has a capability to blacken the photographic plate kept in a dark place wrapped through a paper
- Subsequent experiments showed that radioactivity is a nuclear phenomenon in which an unstable nucleus under goes a decay process referred as radioactive decay
- There are three types of radioactivity decays that occur in nature .These are α decay ,β decay and γ decay.
- We now define radioactive decay as the process by which unstable atomic nucleus looses energy by emitting ionizing particles or radiations (α,β and γ rays)
- Radioactive decay of an atomic nucleus is a spontaneous process and can occur without any interaction of other particles outside the atom
- This process of radioactive decay is random and we can not predict whether a given radioactive atom will emit radiations at a particular instant of time or not
- Phenomenon of radioactivity is observed in heavy elements like uranium and unstable isotopes like carbon 14

2) Properties of radioactive decay

- Radioactive rays ionize the surrounding air and affect photographic plate
- Radioactive rays acts differently on different biological cells and tissues
- A beam of radioactive rays from a radium sample into three components in presence of strong magnetic or electric fields

I.The alpha rays(particles)

- The alpha particles are nuclei of helium atoms
- Alpha particles was first identified by Rutherford and Royds in 1909 by spectroscopic method where they found traces of helium in an originally pure sample of Radon gas which is an α emitter.
- Examples of α decay are

(A) $^{222}Rn_{86} \rightarrow {}^{4}He_{2} + {}^{218}Po_{84}$

(B) $^{238}U_{92} \rightarrow {}^{4}He_{2} + {}^{234}Th_{90}$

- α rays can be stopped by thin sheet of paper.
- α rays can cause intense ionization in air.
- Any group of α particles emitted from same type of nuclei always have definite energy and definite velocity.
- Most α particles are emitted with velocities between ˜ $1.5x10^7$ and ˜ $2.2x10^7$.
- The α particles cover a definite distance in a material without any loss of intensity and suddenly in a small distance they are absorbed completely.
- The distance α rays travel within a given material is called their range in that material.

II.The beta rays(particles)

- β particles are identical with electrons.
- They have mass 1⁄1836 of mass of proton.
- examples of β decay are

  (A) $^{234}Th_{90} \rightarrow {}^{234}Pa_{91} + e^-$

  (B) $^{210}Bi_{83} \rightarrow {}^{210}Po_{84} + e^-$

  (C) $^{14}C_6 \rightarrow {}^{14}N_7 + e^-$

- Mass number and charge are conserved and the daughter product moves one place up in the periodic table, as loss of negative charge by nucleus implies gain of positive charge.
- β rays cause much less ionization in air , but are ˜ 100 times more penetrating then α rays.
- β rays can penetrate a aluminum sheet of few mm thickness.
- A particular β active element emits β particles with energies varying between zero and a certain maximum.
- This maximum energy is called end point energy.

III. The Gamma rays:-

- They are part of EM spectrum $\lambda_\gamma < \lambda_{X\text{-rays}}$
- γ ray photons are more energetic and more penetrating then X-rays photons
- λ_γ rages between $1.7X10^{-8}$ cm and $4.0X10^{-6}$ cm
- Ionization due to Gamma rays is a photoelectric effect
- Owing to their large energies ,the Gamma rays photons can dislodge electrons not only from outer orbits(valence orbits on conduction band) of atoms but also from the inner orbits
- Besides photoelectric effect ,gamma rays loose energy by

  i) Compton effect ,in which the gamma photon collides with an electron and gets scattered with a shift in wavelength

  $$\Delta\lambda = \frac{h}{m_0 c}(1 - \cos\alpha)$$

  ii) Pair production ,in which a gamma photon is converted into a pair consisting of an electron and a positron(particle having mass and charge equal to electron but carrying positive change)

3) Law of radioactive Decay

- Radioactivity is a nuclear phenomenon

- When a nucleus disintegrates by emitting a particle (α and β) or by capturing an electron from the atomic shell(K-shell) ,the process is called radioactive decay. This decay process is spontaneous.
- Let us take a radioactive sample containing N_0 at time t=0 i.e, at the beginning. We wish to calculate the number N of these nuclei left after time t.
- The number of nuclei of a given radioactive sample disintegrating per sec is called the activity of that sample is
  dN/dt=rate of decrease of nuclei with time=Activity of sample at time t --(1)
- Experimentally it is found that the activity at any instant of time t is directly proportional to the number N of parent type nuclei present at that time

$$-\frac{dN}{dt}\alpha\, N$$

or

$$-\frac{dN}{dt}=\lambda N \qquad \text{--(2)}$$

Where λ > 0 is proportionality constant and negative sign indicates that N decreases as t increases

- From equation (2) we get

$$\lambda=\left(\frac{-dN}{N}\right) \qquad \text{----(3)}$$

i.e. ,λ is fractional change in N per sec
=> λ is not merely a proportionality constant ,but it gives us the probability of decay per unit interval of time

- Hence λ is called the probability constant or decay constant or disintegration constant
- dN is the no of parent nuclei that decay between t and t+dt and we have taken N as continuous variable
- From (2)

$$\int_{N_0}^{N}\frac{dN}{N}=-\int_{0}^{t}\lambda dt$$

$$N=N_0e^{-\lambda t} \qquad \text{----(4)}$$

N_0=No of radioactive nuclei at t=0

- From (4) we see that law of radioactive decay is exponential in character

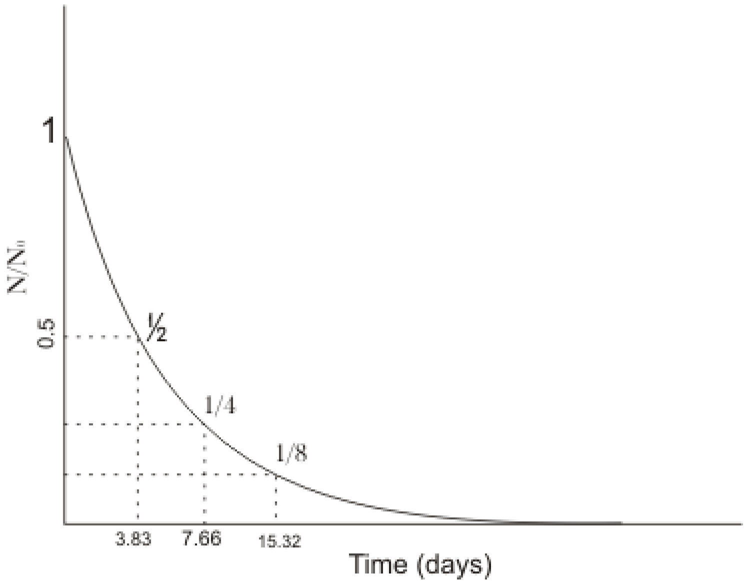


Figure 1. Decay curve for radon

- From figure it can be noted that only half the amount of radon present initially after 3.83 days and 1/4 after 7.66 days and so on
- Plot shows that in a fixed time interval a fixed fraction of the amount of radioactive substance at the beginning of interval decays
- This faction is independent of the amount of radioactive substance and depends only on the interval of the time
- The decay constant λ is a characteristics of radioactive substance and it depends in no way on the amount of the substance present

a) Half Life

- Time interval during which half of a given sample of radioactive substance decays is called its half life. It is denoted by T

$$\frac{N}{N_0}\frac{1}{2} = e^{-\lambda T}$$

$$e^{\lambda T} = 2$$

$$\lambda T = \log_e 2 = .693$$

$$T = \frac{.693}{\lambda} \qquad \text{----(5)}$$

b) Mean Life τ

- Individual radio atomic atoms may have life spans between zero and infinity
- Average or mean life τ is defined as
  τ=Total life time of all nuclei in a given sample/Total no of nuclei in that sample --(6)

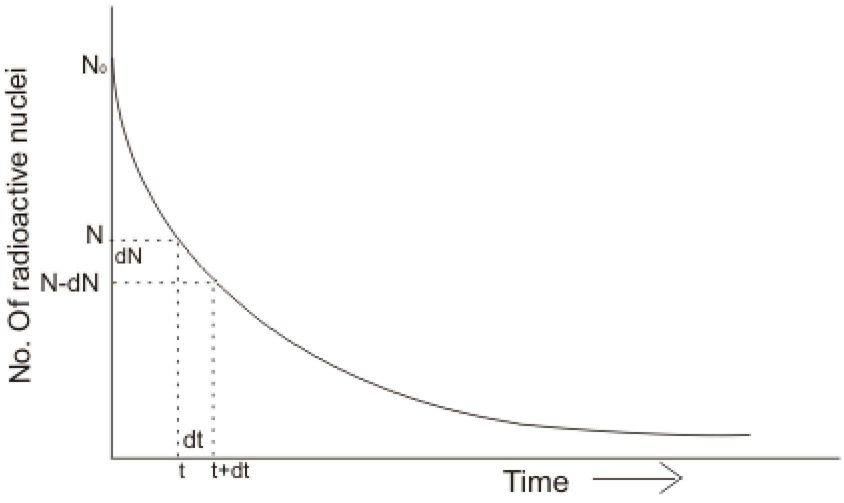


Figure 2. This curve shows how number of nuclei dN decays in time dt

- From curve one can see that each of dN number of radioactive nuclei has lived a life of t sec i.e. the total life span of a dN nuclei is (dN.t) sec . Therefore equation (6) can be written as

$$\tau = \frac{\int_{N_0}^{0} t dN}{\int_{N_0}^{0} dN}$$

$$= \frac{-N_0 \lambda \int_0^\infty t e^{-\lambda t} dt}{-N_0}$$

$$= \lambda \int_0^\infty t e^{-\lambda t} dt$$

which in integration by parts becomes

$$\tau = \frac{1}{\lambda} \qquad \text{----(7)}$$

4) Unit of activity

- The most commonly used unit is the curie
- Curie was originally based on the rate of decay of a gram of radium
- There are 3.7×10^{10} disintegrations per sec per gram of radium .This no is taken as a standard
  => One curie=3.7×10^{10} disintegrations per sec
- One curie of activity is very strong source of radiation
  => 1 milli curie=1mCi=10^{-3} Ci
  1 microcurie=1μCi=10^{-6} Ci
- Another unit of activity is Rutherford
  1rd=10^{6} dis/sec

- Activity $|dN/dt|=\lambda N=.693N/T$
  => A very short lived substance gives rise to large activity ,even it is present in minute quantities
- The SI unit of radioactivity recently proposed is Becquerel (Bq) which is defined as activity done to one disintegration per sec hence
  $1ci=3.7X10^{10}$ bq
  =37G bq

5) Alpha decay:

- Nucleus before the decay is called parent nucleus and after the decay is called daughter nucleus
- In Alpha decay, the parent nucleus $^{A}X_{Z}$ emits an α particle ($=^{4}He_{2}$) leaving behind a daughter nucleus of four mass unit less and two charge units less i.e. $^{A-4}X_{Z-2}$
  $$^{A}X_{Z} -> ^{A-4}Y_{Z-2} + ^{4}He_{2}$$
- α decay shift the element two places to the left in the periodic tables of elements ex
  $$^{226}Ra_{88} -> ^{222}Rn_{86} + \alpha \quad (1600\ Y)$$
  $$^{239}Pu_{94} -> ^{235}U_{92} + \alpha \quad (24000Y)$$
- All nuclides of A >= 210 and Z > 83 tends to decay by α emission
- ^{209}Bi is the heaviest stable nuclide in nature
- α decay in heavy nucleus occur because a too heavy nucleus becomes unstable due to coulomb repulsion and by emitting an α particle the nucleus decrease its A and Z to moves towards stability
- Now the rest mass energy of parent nucleus $^{A}X_{Z}$ is greater then the sum of rest mass energies of $^{A-4}X_{Z-2}$ and $^{4}He_{2}$
- The difference between the rest mass energies of initial constituents and final products is called Q-value of the process
- For α decay process ,Q value is
  $Q=[m_p -(m_d+m_\alpha)]c^2$
  where m_p -> Mass of parent nucleus $_{Z}{}^{A}X$
  m_d -> Mass of parent nucleus $_{Z-2}{}^{A-4}X$
  m_α -> Mass of parent nucleus $_{2}{}^{4}He$

6) β Decay

- There are two types of β decay , β^- and β^+
- In β decay an nucleus decay spontaneously emitting an electron or positron
- Under β^- decay one of the neutrons in the parent nucleus gets transformed into a proton and in the process an electron and an antineutrino are emitted
  $n-> p+e^-+v^-$
- The daughter nucleus thus formed in β^- decay would be an element one place to the right of the parent in the periodic table of elements
- Examples of β^- decay
  $$^{24}Na_{11} \rightarrow ^{24}Mg_{12}(stable) + e^- + \bar{\nu} \qquad (T_{1/2} = 15.03h)$$
  $$^{32}P_{15} -> ^{32}S_{16}(stable) + e^- + \bar{\nu} \qquad (T_{1/2} = 14.28days)$$
- β^- is common over entire range of nuclides and amongst the naturally occurring heavy radioactive nuclides and in fission products

- In β^+ decay one the protons of the parent nucleus gets transformed into a neutron emitting a positron and neutrino
  p->n+e^++v
- In β^+ decay the daughter nucleus would be one place to the left of parent nuclei in the periodic table
- Examples of β^+ decay

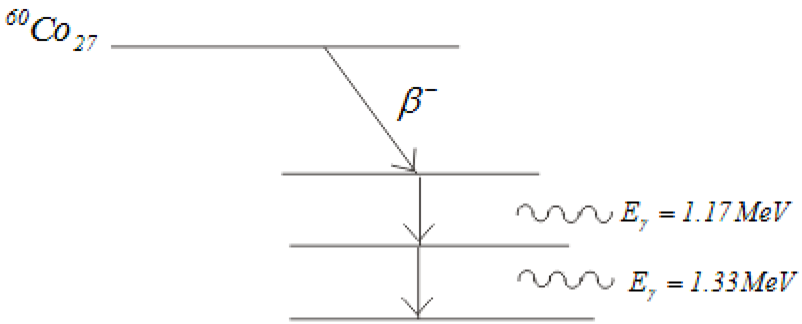
$$^{22}Na_{11} -> ^{22}Ne_{10}(stable) + e^+ + \nu \qquad (T_{1/2} = 2.62 years)$$

$$^{45}Ti_{22} -> ^{45}Sc_{21}(stable) + e^+ + \nu \qquad (T_{1/2} = 3.08 h)$$

- In both β^+ and β^- symbol v^- and v represents antineutrino and neutrino
- Both antineutrino (v^-) and neutrino(v) are charge less and nearly less particles and interact very weakly with matter which make their detection very difficult
- In these β decay(β^+ and β^-) mass number A of nucleus remain same after the decay

7) γ Decay

- After alpha or beta decay processes it is common to find the daughter nucleus to be in an excited state
- Just like atoms ,nucleus also have energy levels
- So an nucleus in excited state can make transitions from higher energy levels to lower one by the emission of electro magnetic radiation
- The energy difference in allowed energy levels of a nucleus are of the order of Mev and the photons emitted by nuclei have energies of the order of Mev and are called γ rays
- As an example,β decay of $^{60}Co_{27}$ nucleus gets transformed into $^{60}Ni_{28}$ nucleus in excited state which then de -excites to its ground state by successive emission of 1.17 Mev and 1.33 Mev gamma rays as shown in energy level diagram,



UNIT8

9. Electronic Devices

An electronic device controls the flow of electron and is main building block of electronic circuits. In years before second world war, vacuum tubes were used in the process of generation, amplification and transmission of electrical signal. In 1947, Bell laboratories developed the first transistor based on research of Shockley, Bardeen and Brattain. However, transistor radios are not developed until late 1950's due to existing huge stock of vacuum tubes and 1950 onwards transistor as an electronic device replaced vacuum tubes in many fields.

A solid state electronic device essentially consists of a semiconducting material. In this chapter we will discuss semiconductors and some semiconductor devices (electronic devices).

9.1 Semiconductors

It has been observed that certain materials like germanium, silicon etc. have resistivity between good conductors like copper and insulators like glass. These materials are known as semiconductors. A material which has resistivity between conductors and insulators is known as semiconductor. The resistivity of a semiconductor lie approximately between 10^{-2} and $10^{4}\,\Omega\,m$ at room temperature. The resistance of a semiconductor decreases with increase in temperature over a particular temperature range. This behaviour is contrary to that of a metallic conductor for which the resistance increases with increase in temperature.

The elements that are classified as semiconductors are Si, Ge, In, etc. Germanium and silicon are most widely used as semiconductors.

9.1.1 Energy band in solids

In the case of a single isolated atom, there are various discrete energy levels. In solids, the atoms are arranged in a systematic space lattice and each atom is influenced by neighbouring atoms. The closeness of atoms results in the intermixing of electrons of neighbouring atoms. Due to this, number of permissible energy levels increases. Hence in the case of a solid, instead of a single energy level associated with single atom, there will be bands of energy levels. A set of such closely packed energy levels is called an energy band. The bands of energy levels are referred to the entire solid as a whole and not to the single atom.

The concept of energy bands can be understood from Fig 9.1a and Fig 9.1b. The energy levels of a single isolated atom of silicon are shown in Fig 9.1a. Each silicon atom has 14 electrons, two of which occupy K shell, 8 occupy the L shell and 4 occupy the M shell. The electrons in the M shell are distributed as 2 electrons in the subshell 3s and 2 electrons in the subshell 3p. This subshell 3p is partially filled because it can accommodate a total of 6 electrons. The completely filled levels are known as core levels and the electrons filling these levels are called core electrons. The electrons in the outermost level are called valence electrons. The partially filled outermost level is valence level and the permitted levels which are vacant are known as conduction levels.

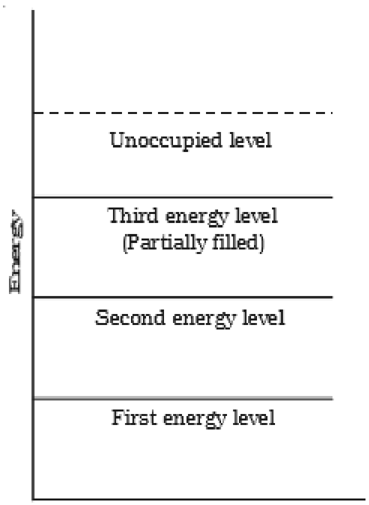


Fig 9.1a Energy levels of a single isolated atom

In a solid, there are large number of atoms, which are very close to each other. The energy of s or p level is of the order of eV, therefore the levels are very closely spaced. The first orbit electrons form a band called first energy band. Similarly second orbit electrons form second energy band and so on as shown in Fig 9.1b.

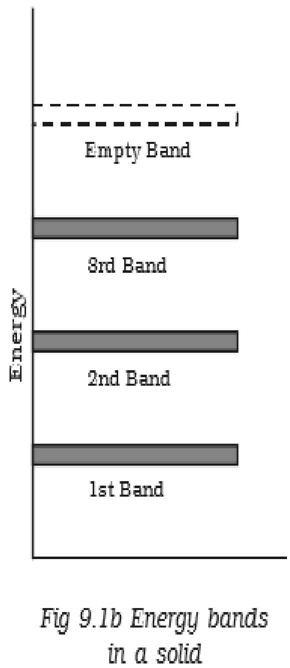


Fig 9.1b Energy bands in a solid

9.1.2 Valence band, conduction band and forbidden energy gap

The atoms of a solid are arranged in a regular repeated geometric pattern and the electrons of the atom revolve around the nucleus in certain permitted energy levels. The electrons in the inner shells are strongly bound to the nucleus. A band which is occupied by the valence electrons or a band having highest energy is defined as valence band (Fig 9.2). The valence band may be partially or completely filled. This band can never be empty.

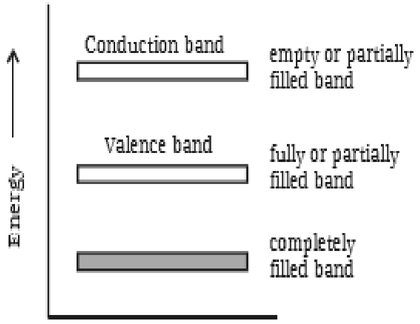


Fig 9.2 Valence band and conduction band

In some materials, the valence electrons are loosely attached to the nucleus. Even at room temperature, some of the valence electrons can leave the valence band. These are called as free electrons. They are responsible for conduction of current in a conductor and are henceforth called as conduction electrons. The band occupied by these electrons is called conduction band. This band may be an empty band or partially filled band.

The separation between valence band and conduction band is known as forbidden energy gap. If an electron is to be transfered from valence band to conduction band, external energy is required, which is equal to the forbidden energy gap.

9.1.3 Insulators, semiconductors and conductors

Insulators

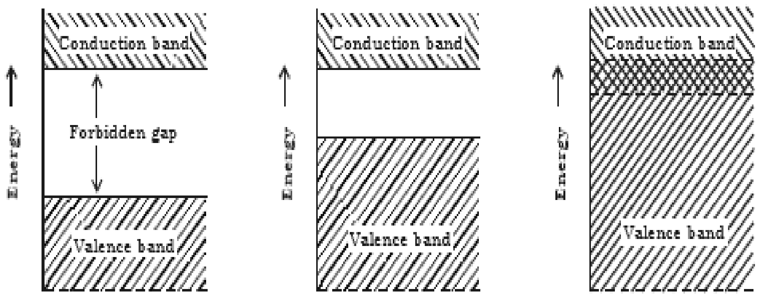
In an insulator, the forbidden energy gap is very large (Fig 9.3a). In general, the forbidden energy gap is more than 3eV and almost no electrons are available for conduction. Therefore, a very large amount of energy must be supplied to a valence electron to enable it to move to the conduction band. In the case of materials like glass, the valence band is completely filled at 0 K. The energy gap between valence band and conduction band is of the order of 10 eV. Even in the presence of high electric field, the electrons cannot move from valence band to conduction band. If the electron is supplied with high energy, it can jump across the forbidden gap. When the temperature is increased, some electrons will move to the conduction band. This is the reason, why certain materials, which are insulators at room temperature become conductors at high temperature. The resistivity of insulator approximately lies between 10^{11} and $10^{16}\,\Omega$ m.

Semiconductors

In semiconductors (Fig 9.3b), the forbidden gap is very small. Germanium and silicon are the best examples of semiconductors. The forbidden gap energy is of the order of 0.7eV for Ge and 1.1eV for Si. There are no electrons in the conduction band. The valence band is completely filled at 0 K. With a small amount of energy that is supplied, the electrons can easily jump from the valence band to the conduction band. For example, if the temperature is raised, the forbidden gap is decreased and some electrons are liberated into the conduction band. The conductivity of a semiconductor is of the order of 10^2 mho m^{-1}.

Conductors

In conductors, there is no forbidden gap available, the valence and conduction band overlap each other (Fig 9.3c). The electrons from valence band freely enter into the conduction band. Due to the overlapping of the valence and conduction bands, a very low potential difference can cause the continuous flow of current.



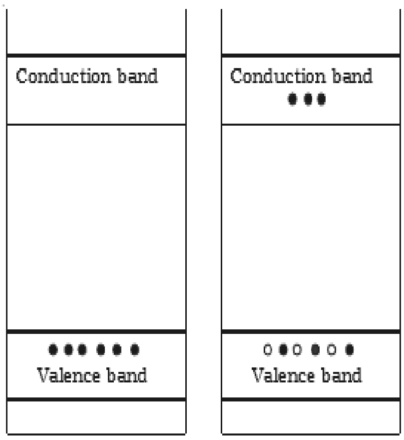
(a) Insulator *(b) Semiconductor* *(c) Conductor*

Fig 9.3 Energy band of solids

9.1.4 Electrons and holes in semiconductors

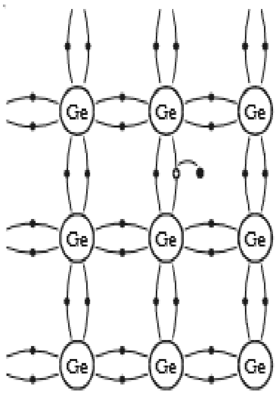
Fig 9.3b shows the energy band diagram of an intrinsic semiconductor (pure semiconductor). Fig 9.4a and Fig 9.4b represent charge carriers at absolute zero temperature and at room temperature respectively.

The electrons in an intrinsic semiconductor, which move in to the conduction band at high temperatures are called as intrinsic carriers. In the valence band, a vacancy is created at the place where the electron was present, before it had moved in to the conduction band. This vacancy is called hole. Fig 9.4c helps in understanding the creation of a hole. Consider the case of pure germanium crystal. It has four electrons in its outer or valence orbit. These electrons are known as valence electrons. When two atoms of germanium are brought close to each other, a covalent bond is formed between the atoms. If some additional energy is received, one of the electrons contributing to a covalent bond breaks and it is free to move in the crystal lattice.



(a) (b)

Fig 9.4a&b Electrons and holes in semiconductors



(c)

Fig 9.4c Formation of a hole

While coming out of the bond, a hole is said to be created at its place, which is usually represented by a open circle. An electron from the neighbouring atom can break the covalent bond and can occupy this hole, creating a hole at another place. Since an electron has a unit negative charge, the hole is associated with a unit positive charge. The importance of hole is that, it may serve as a carrier of electricity in the same manner as the free electron, but in the opposite direction.

9.1.5 Intrinsic semiconductor

A semiconductor which is pure and contains no impurity is known as an intrinsic semiconductor. In an intrinsic semiconductor, the number of free electrons and holes are equal. Common examples of intrinsic semiconductors are pure germanium and silicon.

The forbidden energy gap is so small that even at ordinary room temperature, there are many electrons which possess sufficient energy to cross the forbidden energy gap and enter into the conduction band. Schematic band diagram of an intrinsic semiconductor at room temperature is represented in Fig 9.5.

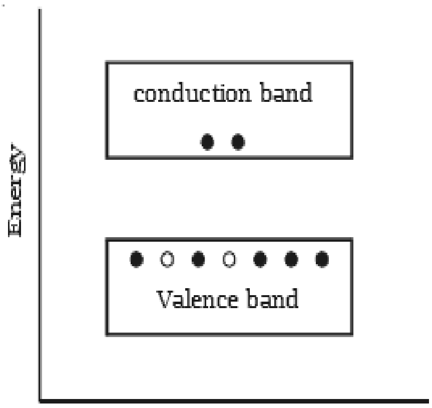


Fig 9.5 Energy band diagram of an intrinsic semiconductor

9.1.6 Doping a semiconductor

Electrons and holes can be generated in a semiconductor crystal with heat energy or light energy. But in these cases, the conductivity remains very low. The efficient and convenient method of generating free electrons and holes is to add very small amount of selected impurity inside the crystal. The impurity to be added is of the order of 100 ppm (parts per million). The process of addition of a very small amount of impurity into an intrinsic semiconductor is called doping. The impurity atoms are called dopants. The semiconductor containing impurity atoms is known as impure or doped or extrinsic semiconductor.

There are three different methods of doping a semiconductor.

(i) The impurity atoms are added to the semiconductor in its molten state.

(ii) The pure semiconductor is bombarded by ions of impurity atoms.

(iii) When the semiconductor crystal containing the impurity atoms is heated, the impurity atoms diffuse into the hot crystal.

Usually, the doping material is either pentavalent atoms (bismuth, antimony, phosphorous, arsenic which have five valence electrons) or trivalent atoms (aluminium, gallium, indium, boron which have three valence electrons). The pentavalent doping atom is known as donor atom, since it donates one electron to the conduction band of pure semiconductor. The trivalent atom is called an acceptor atom, because it accepts one electron from the pure semiconductor atom.

9.1.7 Extrinsic semiconductor

An extrinsic semiconductor is one in which an impurity with a valency higher or lower than the valency of the pure semiconductor is added, so as to increase the electrical conductivity of the semiconductor.

Depending upon the type of impurity atoms added, an extrinsic semiconductor can be classified as N-type or P-type.

(a) N-type semiconductor

When a small amount of pentavalent impurity such as arsenic is added to a pure germanium semiconductor crystal, the resulting crystal is called N-type semiconductor.

Fig 9.6a shows the crystal structure obtained when pentavalent arsenic impurity is added with pure germanium crystal. The four valence electrons of arsenic atom form covalent bonds with electrons of neighbouring four germanium atoms. The fifth electron of arsenic atom is loosely bound. This electron can move about almost as freely as an electron in a conductor and hence it will be the carrier of current. In the energy band picture, the energy state corresponding to the fifth valence electron is in the forbidden gap and lies slightly below the conduction band (Fig 9.6b). This level is known as the donor level.

When the fifth valence electron is transferred to the conduction band, the arsenic atom becomes positively charged immobile ion. Each impurity atom donates one free electron to the semiconductor. These impurity atoms are called donors.

In N-type semiconductor material, the number of electrons increases, compared to the available number of charge carriers in the intrinsic semiconductor. This is because, the available larger number of

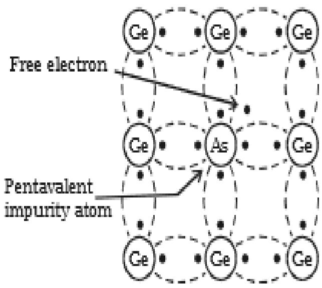


Fig 9.6a N-type semiconductor

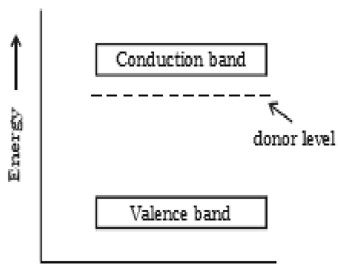


Fig 9.6b Energy band diagram of N-type semiconductor

electrons increases the rate of recombination of electrons with holes. Hence, in N-type semiconductor, free electrons are the majority charge carriers and holes are the minority charge carriers.

(b) P-type semiconductor

When a small amount of trivalent impurity (such as indium, boron or gallium) is added to a pure semiconductor crystal, the resulting semiconductor crystal is called P-type semiconductor.

Fig 9.7a shows the crystal structure obtained, when trivalent boron impurity is added with pure germanium crystal. The three valence electrons of the boron atom form covalent bonds with valence electrons of three neighbourhood germanium atoms. In the fourth covalent bond, only one valence electron is available from germanium atom and there is deficiency of one electron which is called as a hole. Hence for each boron atom added, one hole is created. Since the holes can accept electrons from neighbourhood, the impurity is called acceptor. The hole, may be filled by the electron from a neighbouring atom, creating a hole in that position from where the electron moves. This process continues and the hole moves about in a random manner due to thermal effects. Since the hole is associated with a positive charge moving from one position to another, this is called as P-type semiconductor. In the P-type semiconductor, the acceptor impurity produces an energy level just above the valence band. (Fig 9.7b). Since, the energy difference between acceptor energy level and the valence band is much smaller, electrons from the valence band can easily jump into the acceptor level by thermal agitation.

In P-type semiconductors, holes are the majority charge carriers and free electrons are the minority charge carriers.

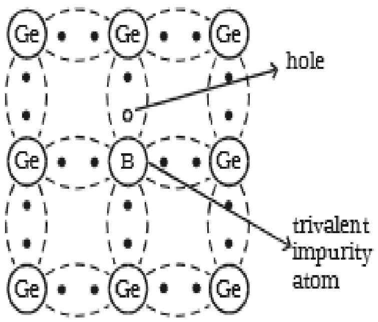


Fig 9.7a P-type semiconductor

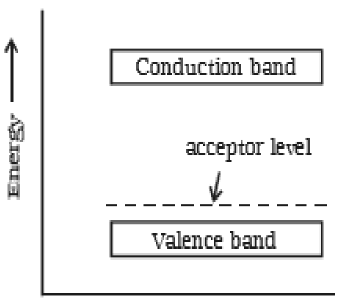


Fig 9.7b Energy band diagram of a P-type semiconductor

9.2 PN Junction diode

If one side of a single crystal of pure semiconductor (Ge or Si) is doped with acceptor impurity atoms and the other side is doped with donor impurity atoms, a PN junction is formed as shown in Fig 9.8. P region has a high concentration of holes and N region contains a large number of electrons.

As soon as the junction is formed, free electrons and holes cross through the junction by the process of diffusion. During this process, the electrons crossing the junction from N-region into the P region, recombine with holes in the P-region very close to the junction. Similarly holes crossing the junction from the P-region into the N-region, recombine with electrons in the N-region very close to the junction. Thus a region is formed, which does not have any mobile charges very close to the junction. This region is called depletion region. In this region, on the left side of the junction, the acceptor atoms become negative ions and on the right side of the junction, the donor atoms become positive ions (Fig 9.8).

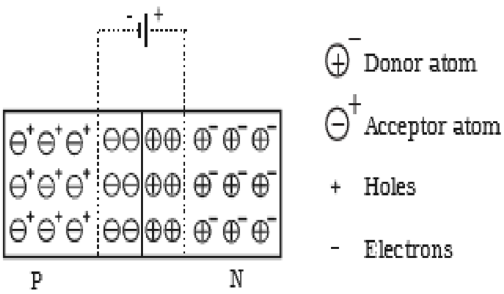


Fig 9.8 P N Junction diode

An electric field is set up, between the donor and acceptor ions in the depletion region. The potential at the N-side is higher than the potential at P-side. Therefore electrons in the N-side are prevented to go to the lower potential of P-side. Similarly, holes in the P-side find themselves at a lower potential and are prevented to cross to the N-side. Thus, there is a barrier at the junction which opposes the movement of the majority charge carriers. The difference of potential from one side of the barrier to the other side is called potential barrier. The potential barrier is approximately 0.7V for a silicon PN junction and 0.3V for a germanium PN junction. The distance from one side of the barrier to the other side is called the width of the barrier, which depends upon the nature of the material.

9.2.1 Forward biased PN junction diode

When the positive terminal of the battery is connected to P-side and negative terminal to the N-side, so that the potential difference acts in opposite direction to the barrier potential, then the PN junction diode is said to be forward biased.

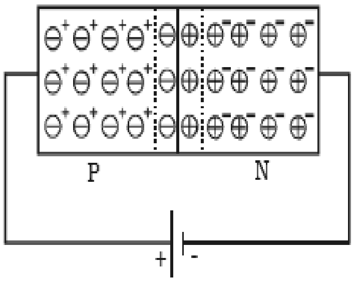


Fig 9.9 Forward biased PN junction diode

When the PN junction is forward biased (Fig 9.9), the applied positive potential repels the holes in the P-region, and the applied negative potential repels the electrons in the N-region, so the charges move towards the junction. If the applied potential difference is more than the potential barrier, some holes and free electrons enter the depletion region.

Hence, the potential barrier as well as the width of the depletion region are reduced. The positive donor ions and negative acceptor ions within the depletion region regain electrons and holes respectively. As a result of this, the depletion region disappears and the potential barrier also disappears. Hence, under the action of the forward potential difference, the majority charge carriers flow across the junction in opposite direction and constitute current flow in the forward direction.

9.2.2 Reverse biased PN junction diode

When the positive terminal of the battery is connected to the N-side and negative terminal to the P-side, so that the applied potential difference is in the same direction as that of barrier potential, the junction is said to be reverse biased.

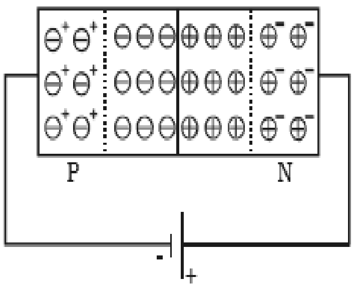


Fig 9.10 Reverse biased PN junction diode.

When the PN junction is reverse biased (Fig 9.10), electrons in the N region and holes in the P-region are attracted away from the junction.

Because of this, the number of negative ions in the P-region and positive ions in the N-region increases. Hence the depletion region becomes wider and the potential barrier is increased.

Since the depletion region does not contain majority charge carriers, it acts like an insulator. Therefore, no current should flow in the external circuit. But, in practice, a very small current of the order of few microamperes flows in the reverse direction. This is due to the minority carriers flowing in the opposite direction. This reverse current is small, because the number of minority carriers in both regions is very small. Since the major source of minority carriers is, thermally broken covalent bonds, the reverse current mainly depends on the junction temperature.

9.2.3 Symbol for a semiconductor diode

The diode symbol is shown in Fig 9.11. The P-type and N-type regions are referred to as P–end and N–end respectively. The arrow on the diode points the direction of conventional current.

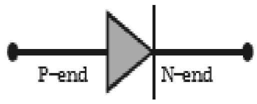
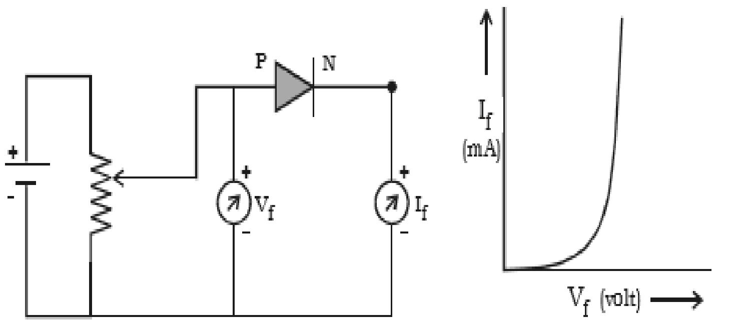


Fig 9.11 Circuit symbol for a semiconductor diode

9.2.4 Forward bias characteristics



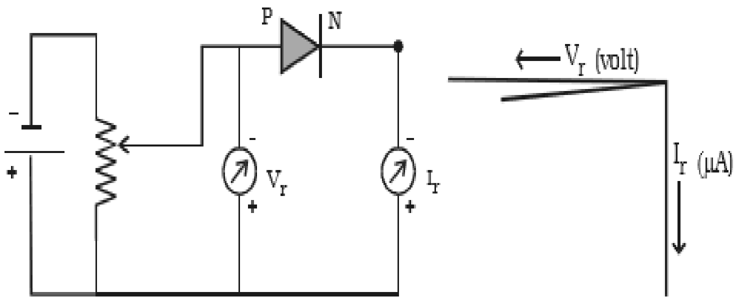
(a) Diode circuit–Forward bias *(b) Forward characteristics*

Fig 9.12 Forward bias characteristics of a diode

The circuit for the study of forward bias characteristics of PN junction diode is shown in Fig 9.12a. The voltage between P–end and N–end is increased from zero in suitable equal steps and the corresponding currents are noted down. Fig 9.12b shows the forward bias characteristic curve of the diode. Voltage is the independent variable. Therefore, it is plotted along X–axis. Since, current is the dependent variable, it is plotted against Y–axis. From the characteristic curve, the following conclusions can be made. (i) The forward characteristic is not a straight line. Hence the ratio V/I is not a constant (i.e) the diode does not obey Ohm's law. This implies that the semiconductor diode is a non-linear conductor of electricity. (ii) It can be seen from the characteristic curve that initially, the current is very small. This is because, the diode will start conducting, only when the external voltage overcomes the barrier potential (0.7V for silicon diode). As the voltage is increased to 0.7 V, large number of free electrons and holes start crossing the junction. Above 0.7V, the current increases rapidly. The voltage at which the current starts to increase rapidly is known as cut-in voltage or knee voltage of the diode.

9.2.5 Reverse bias characteristics

The circuit for the study of reverse bias characteristics of PN junction diode is shown in Fig 9.13a. The voltage is increased from zero in suitable steps. For each voltage, the corresponding current readings are noted down. Fig 9.13b shows the reverse bias characteristic curve of the diode. From the characteristic curve, it can be concluded that, as voltage is increased from zero, reverse current (in the order of microamperes) increases and reaches the maximum value at a small value of the reverse voltage. When the voltage is further increased, the current is almost independent of the reverse voltage upto a certain critical value. This reverse current is known as the reverse saturation current or leakage current. This current is due to the minority charge carriers, which depends on junction temperature.



(a) Diode circuit–Reverse bias *(b) Reverse characteristics*

Fig 9.13 Reverse bias characteristics of a diode

9.3 PN junction diode as rectifier

The process in which alternating voltage or alternating current is converted into direct voltage or direct current is known as rectification. The device used for this process is called as rectifier. The junction diode has the property of offering low resistance and allowing current to flow through it, in the forward biased condition. This property is used in the process of rectification.

9.3.1 Half wave rectifier

A circuit which rectifies half of the a.c wave is called half wave rectifier.

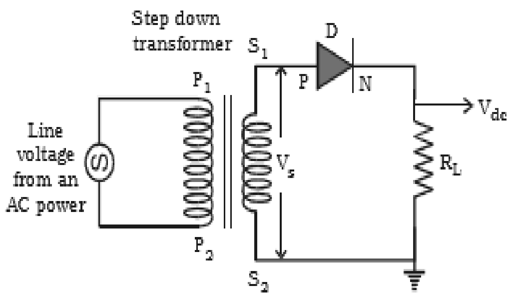


Fig 9.14 Half wave rectifier

Fig 9.14 shows the circuit for half wave rectification. The a.c. voltage (V_s) to be rectified is obtained across the secondary ends S_1 S_2 of the transformer. The P-end of the diode D is connected to S_1 of the secondary coil of the transformer. The N-end of the diode is connected to the other end S_2 of the secondary coil of the transformer, through a load resistance R_L. The rectified output voltage V_{dc} appears across the load resistance R_L.

During the positive half cycle of the input a.c. voltage V_s, S_1 will be positive and the diode is forward biased and hence it conducts.

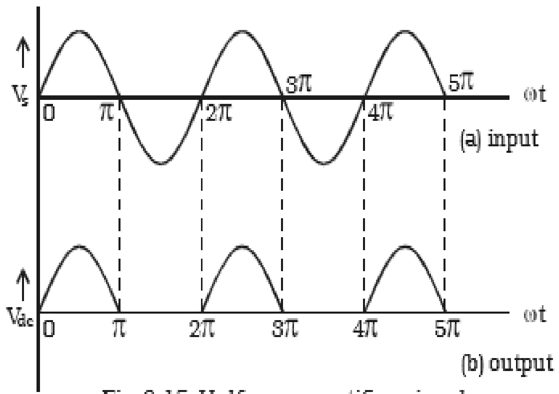


Fig 9.15 Half wave rectifier signals

Therefore, current flows through the circuit and there is a voltage drop across R_L. This gives the output voltage as shown in Fig 9.15.

During the negative half cycle of the input a.c. voltage (V_s), S_1 will be negative and the diode D is reverse biased. Hence the diode does not conduct. No current flows through the circuit and the voltage drop across R_L will be zero. Hence no output voltage is obtained. Thus corresponding to an alternating input signal, unidirectional pulsating output is obtained.

The ratio of d.c. power output to the a.c. power input is known as rectifier efficiency. The efficiency of half wave rectifier is approximately 40.6%

9.3.2 Bridge rectifier (Not for examination)

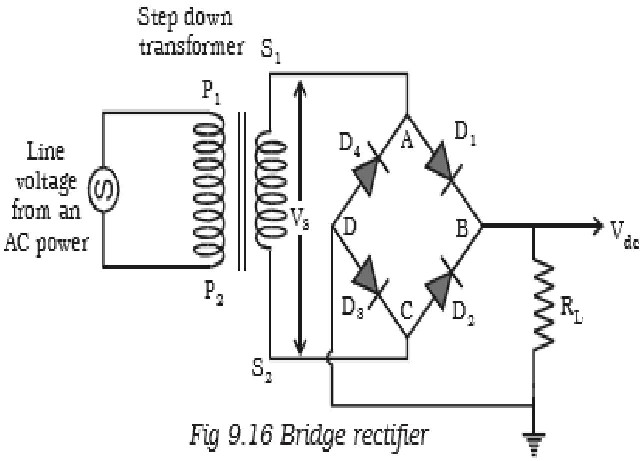


Fig 9.16 Bridge rectifier

A bridge rectifier is shown in Fig 9.16. There are four diodes D_1, D_2, D_3 and D_4 used in the circuit, which are connected to form a network. The input ends A and C of the network are connected to the secondary ends S_1 and S_2 of the transformer. The output ends B and D are connected to the load resistance R_L.

During positive input half cycle of the a.c. voltage, the point A is positive with respect to C. The diodes D_1 and D_3 are forward biased and conduct, whereas the diodes D_2 and D_4 are reverse biased and do not conduct. Hence current flows along S_1ABDCS_2 through R_L. During negative half cycle, the point C is positive with respect to A. The diodes D_2 and D_4 are forward biased and conduct, whereas the diodes D_1 and D_3 are reverse biased and they do not conduct. Hence current flows along S_2CBDAS_1 through R_L. The same process is repeated for subsequent half cycles. It can be seen that, current flows through R_L

in the same direction, during both half cycles of the input a.c. signals. The output signal corresponding to the input signal is shown in Fig 9.17. The efficiency of the bridge rectifier is approximately 81.2%.

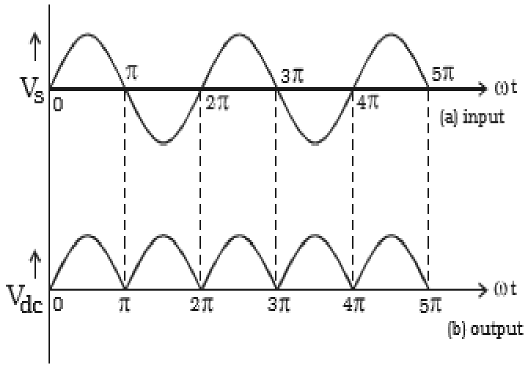


Fig 9.17 Full wave bridge rectifier signals

9.3.3 Filter circuits and regulation property of the power supply (Not for examination)

Both in half wave and full wave rectifiers, it is observed that the output voltage across R_L varies from zero to a maximum value. Eventhough, unidirectional current through R_L is obtained, the output voltage fluctuates. This fluctuation in output voltage is not desirable, when pure d.c. voltage is required. Hence they must be removed or smoothened. This can be achieved with the help of suitable networks called filters such as capacitor filter, inductor filter etc., and we can get almost a steady d.c. voltage. But this steady d.c. output voltage from a rectifier is not constant due to the following reasons.

(i) As the load varies, the d.c. output voltage is not constant. That is, as the current drawn from the rectifier increases, the output voltage decreases and vice versa. The variation of d.c. output voltage as a function of d.c. load current is called regulation.

The percentage of regulation = $\dfrac{V_{no\ load} - V_{load}}{V_{load}} \times 100$

(ii) The d.c. output voltage varies directly as the a.c. input voltage to the rectifier. The line voltage from a.c. power (220 V) may not be a constant and may vary from 200 V to 240 V. Hence the d.c. output voltage will also vary. To overcome these difficulties, Zener diodes are used as regulators and are used along with rectifier and filter circuits. They are called 'regulated power supplies'.

9.4 Breakdown mechanisms

There are two mechanisms which give rise to the breakdown of a PN junction under reverse bias condition. They are (i) avalanche breakdown and (ii) zener breakdown.

(i) Avalanche breakdown : When both sides of the PN junction are lightly doped and the depletion layer becomes large, avalanche breakdown takes place. In this case, the electric field across the depletion layer is not so strong. The minority carriers accelerated by the field, collide with the semiconductor atoms in the crystal. Because of this collision with valence electrons, covalent bonds are broken and electron hole pairs are generated. These charge carriers, so produced acquire energy from the applied potential and in turn produce more and more carriers. This cumulative process is called avalanche multiplication and the breakdown is called avalanche breakdown.

(ii) Zener breakdown : When both sides of the PN junction are heavily doped, consequently the depletion layer is narrow. Zener breakdown takes place in such a thin narrow junction. When a small reverse bias is applied, a very strong electric field is produced across the thin depletion layer. This field breaks the covalent bonds, extremely large number of electrons and holes are produced, which give rise to the reverse saturation current (Zener current). Zener current is independent of applied voltage.

9.5 Zener diode

Zener diode is a reverse biased heavily doped semiconductor (silicon or germanium) PN junction diode, which is operated exclusively in the breakdown region.

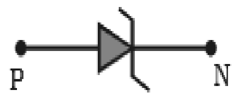


Fig 9.18 Symbol for Zener diode

The symbol of a Zener diode is shown in Fig 9.18. For normal operation of a Zener diode, in breakdown region, the current through the diode should be limited by an external circuit. Hence the power dissipated across the junction is within its power-handling capacity. Unless this precaution is observed, a large current will destroy the diode.

The V-I characteristic curve for the Zener diode is shown in Fig 9.19. It can be seen from the figure, that, as the reverse voltage applied to the PN junction is increased, at a particular voltage, the current increases enormously from its normal cut off value. This voltage is called zener voltage or breakdown voltage (V_z).

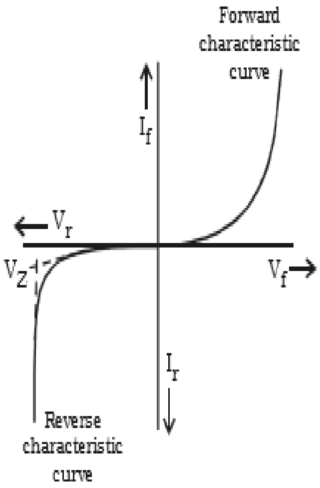


Fig 9.19 V - I characteristics of a Zener diode.

9.6 Zener diode as voltage regulator

To maintain a constant voltage across the load, even if the input voltage or load current varies, voltage regulation is to be made. A Zener diode working in the breakdown region can act as voltage regulator.

The circuit in which a Zener diode is used for maintaining a constant voltage across the load R_L is shown in Fig 9.20. The Zener diode in reverse biased condition is connected in parallel with the load R_L. Let V_{dc} be the unregulated dc voltage and V_z be Zener voltage (regulated output voltage). R_s is the current limiting resistor. It is chosen in such a way that the diode operates in the breakdown region.

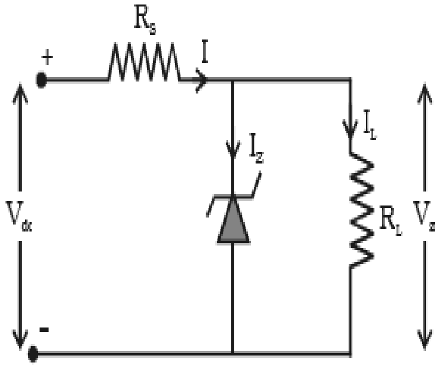


Fig 9.20 Zener diode as a voltage regulator

Inspite of changes in the load current or in the input voltage, the Zener diode maintains a constant voltage across the load. The action of the circuit can be explained as given below.

(i) load current varies, input voltage is constant : Let us consider that the load current increases. Zener current hence decreases, and the current through the resistance R_s is a constant. The output voltage is $V_z = V_{dc} - IR_s$, since the total current I remains constant, output voltage remains constant.

(ii) input voltage varies : Let us consider that the input voltage V_{dc} increases. Now the current through Zener increases and voltage drop across R_s increases in such a way that the load voltage remains the same. Thus the Zener diode acts as a voltage regulator.

9.7 Light Emitting Diode (LED)

A light emitting diode (LED) is a forward biased PN junction diode, which emits visible light when energized.

When a junction diode is forward biased, electrons from N-side and holes from P-side move towards the depletion region and recombination takes place. When an electron in the conduction band recombines with a hole in the valence band, energy is released. In the case of semiconducting materials like gallium arsenide (GaAs), gallium phosphide (GaP) and gallium – arsenide phosphide (GaAsP), a greater percentage of energy is given out in the form of light. If the semiconductor material is transluscent, light is emitted and the junction becomes a light source (turned ON). The LED is turned ON, when it is forward biased and it is turned OFF, when it is reverse biased. The colour of the emitted light will depend upon the type of the material used. By using gallium arsenide phosphide and gallium phosphide, a manufacturer can produce LEDs that radiate red, green, yellow and orange. Fig 9.21 shows the symbol of LED. LEDs are used for instrument displays, calculators and digital watches.

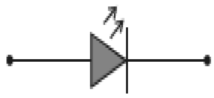


Fig 9.21 Symbol of LED

9.8 Junction transistor

A junction transistor is a solid state device. It consists of silicon or germanium crystal containing two PN junctions. The two PN junctions are formed between the three layers. These are called base, emitter and collector.

(i) Base (B) layer : It is a very thin layer, the thickness is about 25 microns. It is the central region of the transistor.

(ii) Emitter (E) and Collector (C) layers : The two layers on the opposite sides of B layer are emitter and collector layers. They are of the same type of the semiconductor.

An ohmic contact is made to each of these layers. The junction between emitter and base is called emitter junction. The junction between collector and base is called collector junction.

In a transistor, the emitter region is heavily doped, since emitter has to supply majority carriers. The base is lightly doped. The collector region is lightly doped. Since it has to accept majority charge carriers, it is physically larger in size. Hence, emitter and collector cannot be interchanged.

The construction of PNP and NPN transistors are shown in Fig 9.22a and Fig 9.22b respectively.

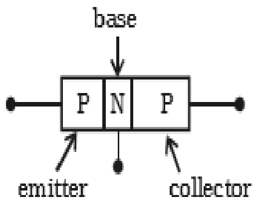


Fig 9.22a Construction of PNP transistor

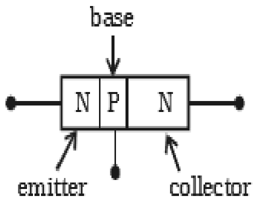


Fig 9.22b Construction of NPN transistor

For a transistor to work, the biasing to be given are as follows :

(i) The emitter-base junction is forward biased, so that majority charge carriers are repelled from the emitter and the junction offers very low resistance to the current.

(ii) The collector-base junction is reverse biased, so that it attracts majority charge carriers and this junction offers a high resistance to the current.

9.9 Transistor circuit symbols

The circuit symbols for a PNP and NPN transistors are shown in Fig 9.23. The arrow on the emitter lead pointing towards the base represents a PNP transistor. When the emitter-base junction of a PNP transistor is forward biased, the direction of the conventional current flow is from emitter to base. NPN transistor is represented by arrow on the emitter lead pointing away from the base. When the emitter base junction of a NPN transistor is forward biased, the direction of the conventional current is from base to emitter.

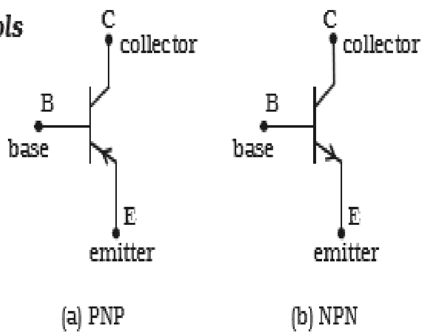


Fig 9.23 Symbol for transistors.

9.9.1 Working of a PNP transistor

A PNP transistor is like two PN junction diodes, which are placed back-to-back. At each junction, there is a depletion region which gives rise to a potential barrier. The external biasing of the junction is provided by the batteries V_{EE} and V_{CC} as shown in Fig. 9.24. The emitter base junction is forward biased and the collector base junction is reverse biased.

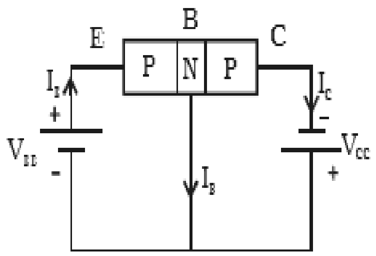


Fig 9.24 PNP Transistor Action

Since the emitter-base junction is forward biased, a large number of holes cross the junction and enters the base. At the same time, very few electrons flow from the base to the emitter. These electrons, when they reach emitter, recombine with an equal number of holes in the emitter. The loss of total number of holes in the emitter is made by flow of an equal number of electrons from the emitter to the positive terminal of the battery. The flow of holes from the emitter to base gives rise to emitter current I_E. In the emitter, I_E is due to the flow of holes. But in the external circuit the current is due to the flow of electrons from the emitter to the positive terminal of the battery V_{EE}. The holes diffuse through the base. These holes take a very small time to flow through this region before they reach the depletion region. During this time, a very small number of holes recombine with an equal number of electrons in the base. Because the base is lightly doped and very thin, this number is very small. The loss of total number of electrons per second is made up by the flow of an equal number of electrons from the negative terminal of V_{EE} into the base. The flow of these electrons contribute the base current I_B.

The remaining numbers of holes, which do not undergo recombination process in the base, reach the collector. These are neutralised by an equal number of electrons flowing from the negative terminal of the battery V_{CC} into the collector. At the same time, an equal number of electrons flows from the negative terminal of V_{EE} and reach the positive terminal of V_{CC}. The flow of holes per second from the base to the collector gives rise to the collector current I_c from the base to the collector. In the external circuit, it is due to the flow of electrons from the negative terminal of the battery V_{CC} into the collector.

Applying Kirchoff's current law to the circuit, the emitter current is the sum of collector current and base current.

i.e $\quad I_E = I_B + I_C$

This equation is the fundamental relation between the currents in a transistor circuit.

This equation is true regardless of transistor type or transistor configuration.

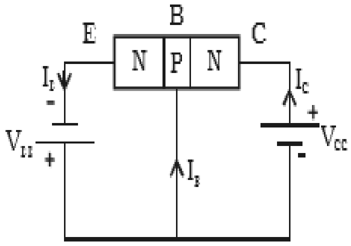


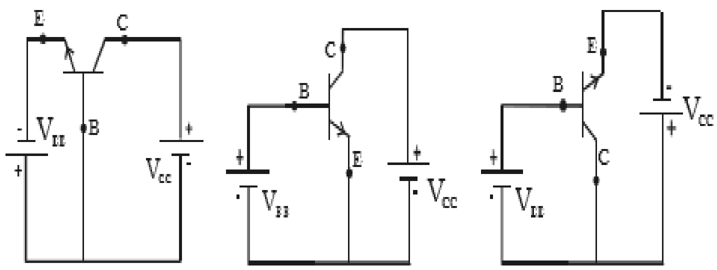
Fig 9.25 NPN transistor action

The action of NPN transistor (Fig 9.25) is similar to that of PNP transistor.

9.9.2 Transistor circuit configurations

There are three types of circuit connections (called configurations or modes) for operating a transistor. They are (i) common base (CB) mode (ii) common emitter (CE) mode and (iii) common collector (CC) mode.

The term common is used to denote the lead that is common to the input and output circuits. The different modes are shown in Fig 9.26 for NPN transistor.



(a) CB mode *(b) CE mode* *(c) CC mode*

Fig 9.26 Three modes of transistor circuit

In a similar way, three configurations can be drawn for PNP transistor.

9.9.3 Current amplification factors α and β and the relation between them

The current amplification factor or current gain of a transistor is the ratio of output current to the input current. If the transistor is connected in common base mode, the current gain $\alpha = \frac{I_C}{I_E}$ and if the transistor is connected in common emitter mode, the current gain $\beta = \frac{I_C}{I_B}$. Fig 9.27 shows a NPN transistor connected in the common base and common emitter configurations. Since, 95% of the

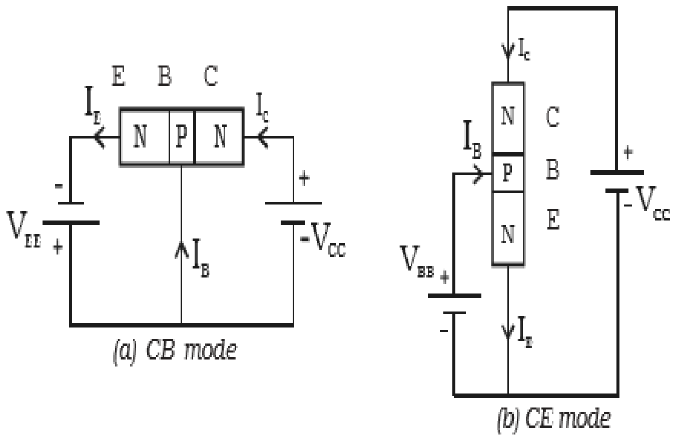


Fig 9.27 CB and CE modes of an NPN transistor

injected electrons reach the collector, the collector current is almost equal to the emitter current. Almost all transistors have α, in the range 0.95 to 0.99.

We know that

$$\alpha = \frac{I_C}{I_E} = \frac{I_c}{I_B + I_C} \qquad (\because\ I_E = I_B + I_C)$$

$$\frac{1}{\alpha} = \frac{I_B + I_C}{I_C} = \frac{I_B}{I_C} + 1$$

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

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

Usually β lies between 50 and 300. Some transistors have β as high as 1000.

9.9.4 Characteristics of an NPN transistor in common emitter configuration

The three important characteristics of a transistor in any mode are (i) input characteristics (ii) output characteristics and (iii) transfer characterstics.

The circuit to study the characteristic curves of NPN transistor in common emitter mode is as shown in Fig 9.28.

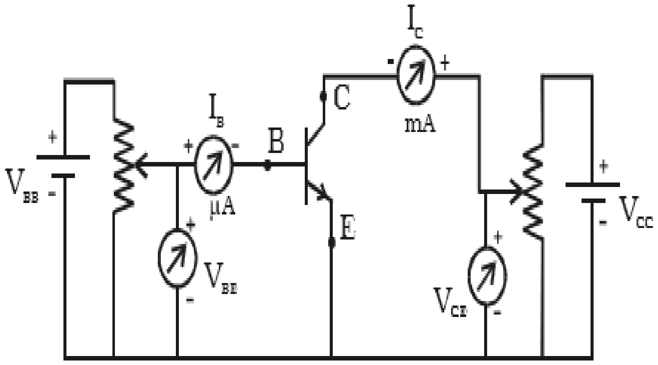


Fig 9.28 Transistor circuit in CE mode.

(i) Input characteristics

Input characteristic curve is drawn between the base current (I_B) and voltage between base and emitter (V_{BE}), when the voltage between collector and emitter (V_{CE}) is kept constant at a particular value. V_{BE} is increased in suitable equal steps and corresponding base current is noted. The procedure is repeated for different values of V_{CE}. I_B values are plotted against V_{BE} for constant V_{CE}. The input characteristic thus obtained is shown in Fig 9.29.

The input impedance of the transistor is defined as the ratio of small change in base – emitter voltage to the corresponding change in base current at a given V_{CE}.

$$\therefore \text{ Input impedance, } r_i = \left(\frac{\Delta V_{BE}}{\Delta I_B}\right)_{V_{CE}}$$

The input impedance of the transistor in CE mode is very high.

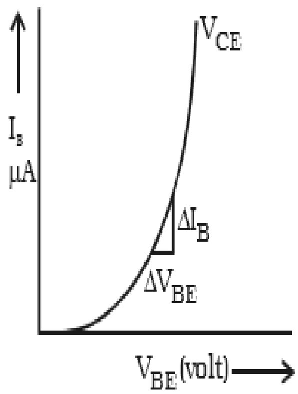


Fig 9.29 Input characteristics

(ii) Output characteristics

Output characteristic curves are drawn between I_C and V_{CE}, when I_B is kept constant at a particular value.

The base current I_B is kept at a constant value, by adjusting the base emitter voltage V_{BE}. V_{CE} is increased in suitable equal steps and the corresponding collector current is noted. The procedure is repeated for different values of I_B. Now, I_c versus V_{CE} curves are drawn for different values of I_B. The output characteristics thus obtained are

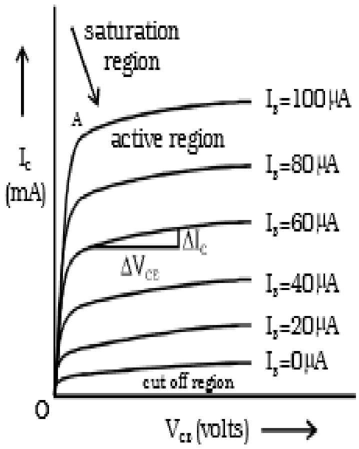


Fig 9.30 Output characteristics

represented in Fig 9.30. The three regions of the characteristics can be discussed as follows :

Saturation region : The initial part of the curve (ohmic region, OA) is called saturation region. (i.e) The region in between the origin and knee point. (Knee point is the point, where I_c is about to become a constant).

Cut off region : There is very small collector current in the transistor, even when the base current is zero (I_B = 0). In the output characteristics, the region below the curve for I_B = 0 is called cut off region. Below the cut off region, the transistor does not function.

Active region : The central region of the curves is called active region. In the active region, the curves are uniform. In this region, E-B junction is forward biased and C-B junction is reverse biased.

The output impedance r_o is defined as the ratio of variation in the collector emitter voltage to the corresponding variation in the collector current at a constant base current in the active region of the transistor characteristic curves.

$$\therefore \text{ output impedence, } r_o = \left(\frac{\Delta V_{CE}}{\Delta I_C}\right)_{I_B}$$

The output impedance of a transistor in CE mode is low.

(iii) Transfer characteristics

The transfer characteric curve is drawn between I_C and I_B, when V_{CE} is kept constant at a particular value. The base current I_B is increased in suitable steps and the collector current I_C is noted down for each value of I_B. The transfer characteristic curve is shown in Fig 9.31.

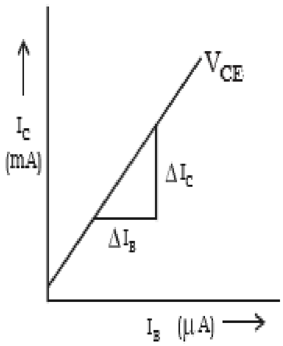


Fig 9.31 Transfer characteristic curve

The current gain is defined as the ratio of a small change in the collector current to the corresponding change in the base current at a constant V_{CE}.

$\therefore$ current gain, $\beta = \left(\frac{\Delta I_C}{\Delta I_B}\right)_{V_{CE}}$

The common emitter configuration has high input impedance, low output impedance and higher current gain when compared with common base configuration.

9.10 Transistor as a switch

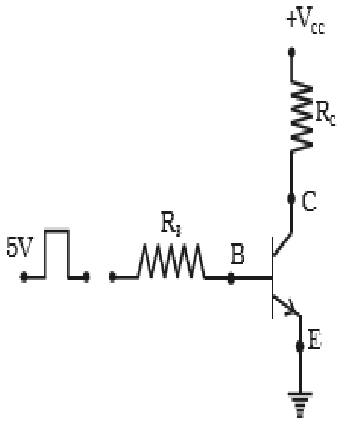


Fig 9.32 NPN Transistor as a switch.

Transistors are widely used in switching operations. In the Fig 9.32, NPN transistor is connected in common emitter configuration and a resistor R_B is connected in series with the base. The load resistance R_c is connected in series with the collector. A pulse type waveform is applied as the input to the transistor through R_B. When the input is high, base emitter junction is forward biased and current flows through R_B into the base. The values of R_B and R_C are chosen in such a manner that the base current flowing, is enough to saturate the transistor. When the transistor is saturated, it is said to be ON (maximum current). When the input is low (i.e.) at 0 V, the base emitter junction is not forward biased. So, no base current flows. Hence the transistor is said to be OFF.

9.11 Transistor amplifier

The important function of a transistor is the amplification. An amplifier is a circuit capable of magnifying the amplitude of weak signals. The important parameters of an amplifier are input impedance, output impedance, current gain and voltage gain. A good design of an amplifier circuit must possess high input impedance, low output impedance and high current gain.

9.11.1 Operating point

For the given values of the load resistance R_c and supply voltage V_{cc}, two points A (V_{CC}, 0) and B $\left(0, \frac{V_{cc}}{R_C}\right)$ are located on the axes of V_{CE}

and I_C respectively, of the output characteristics of the transistor (Fig 9.33). Joining A and B, load line AB is obtained. The point of intersection Q of this line in the active region of the output characteristics with a suitable value of the base current I_B, such that the output voltage is symmetrical is called operating point or quiescent point for the amplifier. $I_{B(Q)}$ is the input base current at the operating point. $V_{CE(Q)}$ and $I_{C(Q)}$ are the collector to emitter voltage and the collector current respectively at the operating point.

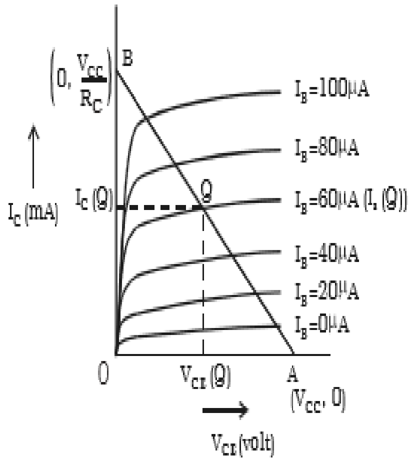


Fig 9.33 load line and operating point

9.11.2 Working

A basic circuit of an amplifier in common emitter mode with NPN transistor is shown in Fig 9.34. The emitter-base junction is forward biased by a supply voltage V_{BB}. The input ac signal to be amplified is applied between base and emitter of the transistor. R_C is the load resistance.

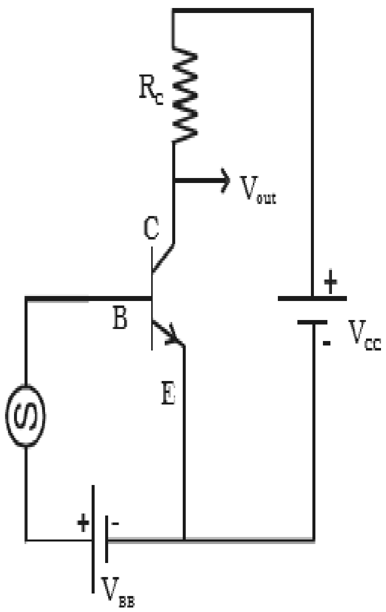


Fig 9.34 Transistor amplifier

The amplifying action of a transistor can be explained as follows. When the a.c. signal is not applied, the base current is available in small quantity in microamperes, which is represented by OP and the corresponding collector current in milliamperes is represented by PQ (Q is the operating point). When the ac signal voltage is applied, the potential difference between the base and emitter changes continuously. This results in increase of base current (I_B) from OP to OA, then decrease of base current from OA to OP (during positive half cycles of the input a.c. voltage) and then to PB and once again increase from OB to OP (during negative half cycle of the input a.c. voltage) for each cycle of the input signal voltage. This variation in base current is reflected in the collector current as shown in Fig 9.35. The collector current (I_C) increases from PQ to AA_1, falls from AA_1 to BB_1 and again increases from BB_1 to PQ. Thus a variation in the base current in micro-amperes produces a corresponding variation in the collector current in milliamperes. This produces a corresponding potential difference across R_C. The increase of potential difference across R_c makes a decrease in the output voltage.

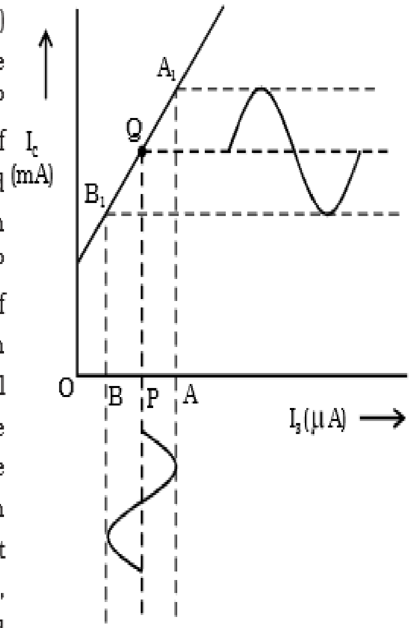


Fig 9.35 Variation in output current (I_C) with variation in input current (I_B)

Therefore, there is always a phase reversal of 180º between the input and output voltages in CE amplifier.

9.12 Transistor biasing (Not for examination)

In order to amplify the input signal using a transistor, the signal is to be applied at an operating point called Q point in the active region. Once the operating point is established, its position should not change. If the Q point shifts near the saturation line or near cut off region of the output characteristics, the signal will be distorted after amplification.

The proper selection of operating point of a transistor and maintenance of proper emitter voltage during the passage of the signal is known as transistor biasing.

The most commonly used methods of obtaining transistor biasing are (i) base bias, (ii) base bias with emitter feedback, (iii) base bias with collector feedback and (iv) voltage divider bias.

The principle involved in all these types is to obtain the required base current corresponding to the operating point under zero signal conditions.

In all the bias circuits except voltage divider bias, the collector current depends on the current gain (β) of the transistor. But β of a transistor is very sensitive to temperature changes. For this reason, it is desirable to have a bias circuit whose action is independent of β. The requirement is met by the voltage divider bias circuit.

9.12.1 Voltage divider bias

This is the most widely used method of providing bias and stabilization to a transisitor. In this method, two resistances R_1 and R_2 are connected across the supply voltage V_{CC} (Fig 9.36) and provide biasing. The emitter resistance R_E provides stabilization. The voltage drop across R_2 forward biases the base emitter junction. This causes the base current and hence collector current to flow in zero signal conditions.

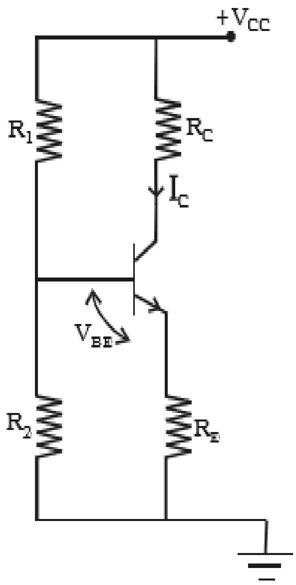


Fig 9.36 Voltage divider bias

The stabilization provided by R_E can be explained as follows. Since β is very sensitive to temperature changes, the collector current I_C increases with rise in temperature. Consequently, it can be seen that I_E increases. This will cause the voltage drop across emitter resistance R_E to increase. The voltage drop across $R_2 = V_{BE} + V_{R_E}$. As voltage drop across R_2 is independent of I_C, V_{BE} decreases. This decreases I_B and the reduced value of I_B tends to bring back I_C to the original value. Hence any variation of β will have no effect on the operating point.

9.13 Single stage CE amplifier

Fig 9.37 shows a single stage CE amplifier. The different circuit elements and their functions are described as follows.

(i) Biasing circuit : The resistances R_1, R_2 and R_E form the biasing

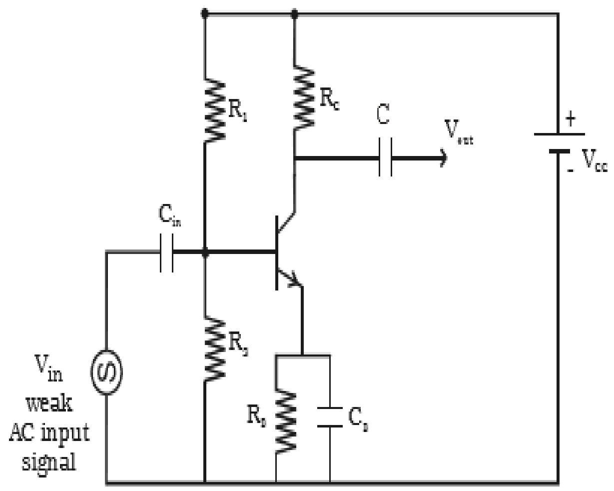


Fig 9.37 Single stage CE amplifier

and stabilization circuit.

(ii) Input capacitance C_{in} : This is used to couple the signal to the base of the transistor. If this is not used, the signal source resistance will come across R_2 and thus change the bias. The capacitor C_{in} allows only a.c. signal to flow.

(iii) Emitter bypass capacitor C_E : This is connected in parallel with R_E to provide a low reactance path to the amplified a.c. signal. If it is not used, then amplified a.c. signal flowing through R_E will cause a voltage drop across it, thereby shifting the output voltage.

(iv) Coupling capacitor C : This is used to couple the amplified signal to the output device. This capacitor C allows only a.c. signal to flow.

Working

When a weak input a.c. signal is applied to the base of the transistor, a small base current flows. Due to transistor action, a much larger a.c. current flows through collector load R_C, a large voltage appears across R_C and hence at the output. Therefore, a weak signal applied to the base appears in amplified form in the collector circuit. Voltage gain (A_v) of the amplifier is the ratio of the amplified output voltage to the input voltage.

Frequency response and bandwidth

The voltage gain (A_v) of the amplifier for different input frequencies can be determined. A graph can be drawn by taking frequency (f) along X–axis and voltage gain (A_v) along Y–axis. The frequency response curve obtained will be of the form as shown in Fig 9.38. It can be seen that the gain decreases at very low and very high frequencies, but it remains constant over a wide range of mid-frequency region.

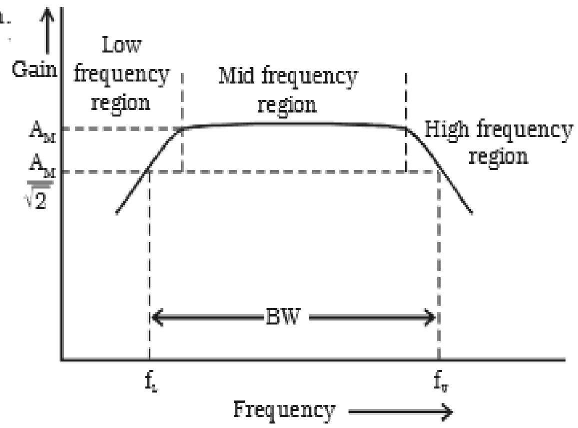


Fig 9.38 Frequency response curve

Lower cut off frequency (f_L) is defined as the frequency in the low frequency range at which the gain of the amplifier is $\frac{1}{\sqrt{2}}$ times the mid frequency gain (A_M). Upper cut off frequency (f_U) is defined as the frequency in the high frequency range at which the gain of the amplifier is $\frac{1}{\sqrt{2}}$ times the mid frequency gain (A_M).

Bandwidth is defined as the frequency interval between lower cut off and upper cut off frequencies. $\therefore$ $BW = f_U - f_L$

9.14 Transistor oscillators

An oscillator may be defined as an electronic circuit which converts energy from a d.c. source into a periodically varying output. Oscillators are classified according to the output voltage, into two types viz. sinusoidal and non-sinusoidal oscillators. If the output voltage is a sine wave function of time, the oscillator is said to be sinusoidal oscillator. If the oscillator generates non-sinusoidal waveform, such as square, rectangular waves, then it is called as non-sinusoidal oscillator (multivibrator). The oscillators can be classified according to the range of frequency as audio-frequency (AF) and radio-frequency (RF) oscillators.

Sinusoidal oscillators may be any one of the following three types:

(i) LC oscillato (ii) RC oscillat (iii) Crystal oscillators

9.14.1 Generation of sinusoidal waves by a tuned LC circuit

Sinusoidal oscillators consist of two main sections : a frequency determining device and maintaining device. A resonant LC network can be used as frequency determining device. The frequency maintaining device is a transistor amplifier with its power supply. The amplifier must have sufficient gain to compensate for the attenuation of the frequency determining section and must introduce required phase shift for positive feedback.

If a capacitor of capacitance C and an inductor of inductance L are connected in parallel, then such a circuit represents an oscillatory circuit.

Let us consider a fully charged capacitor C connected with an inductance L as shown in Fig 9.39a. When the charged capacitor is connected to inductance L, the capacitor will discharge, sending current through L and induce magnetic field as shown in Fig 9.39b. Thus the electrostatic energy stored in the capacitor has been converted into electromagnetic energy associated with inductance L.

When the capacitor is completely discharged, the induced magnetic field begins to collapse, sending current in the same direction. The capacitor C is now charged with opposite polarity (Fig 9.39c). In this case, energy associated with magnetic field is converted

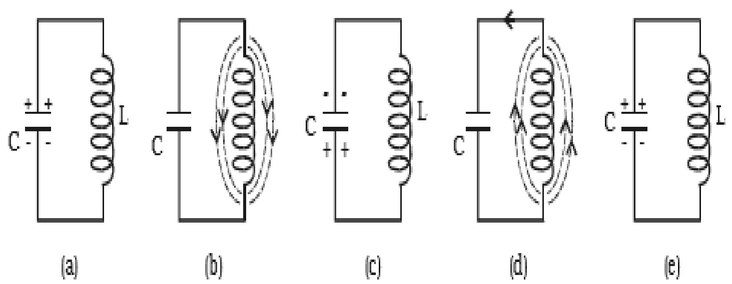


Fig 9.39 Tuned LC circuit

into electrostatic energy. This energy is stored in the capacitor. Once the capacitor is completely charged, it begins to discharge in the reverse direction producing again a magnetic field across L in the opposite direction (Fig 9.39d). Again the magnetic field will collapse and will charge the capacitor. The circuit returns to the original state. (Fig 9.39e). This charging and discharging process results in oscillating current and hence electrical oscillations are set up in the LC circuit. When a LC circuit is used to store energy, it is called tank circuit. The frequency of oscillations is given by,

$$f = \frac{1}{2\pi\sqrt{LC}}$$

If there are no power losses in the LC circuit, then the electrical oscillations will continue for indefinite time. But, in practice, there is some power loss during each cycle of oscillation, as some resistance is always associated with a given LC circuit. Hence the amplitude of oscillations decreases gradually and becomes zero, when all the energy is consumed as losses. Therefore, damped oscillations are produced in the circuit (Fig 9.40). I_{max} represents the maximum current flowing through the circuit.

In order to make the oscillations undamped, energy must be supplied to the circuit at the same rate, at which it is dissipated. The energy supplied should be in phase with oscillations set up in LC circuit. The applied energy should have the same frequency as that of

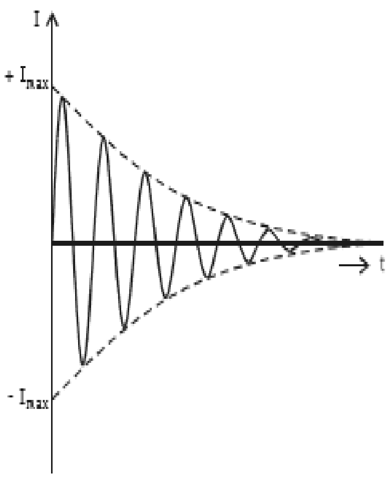


Fig 9.40 damped oscillations

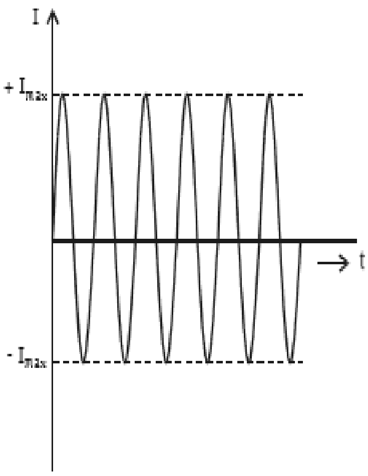


Fig 9.41 undamped oscillations

oscillations in the tank circuit. If these conditions are fulfilled, the circuit will produce continuous undamped oscillations (Fig 9.41).

9.14.2 Essentials of LC oscillator

Fig 9.42 shows the block diagram of an oscillator. Its essential components are (i) tank circuit, (ii) amplifier and (iii) feedback circuit.

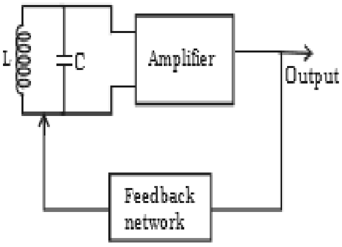


Fig 9.42 Oscillator block diagram

(i) Tank circuit : It consists of inductance coil (L) connected in parallel with capacitor (C). The frequency of oscillations in the circuit depends upon the values of inductance coil and capacitance of the capacitor.

(ii) Amplifier : The transistor amplifier receives d.c. power from the battery and changes it into a.c. power for supplying to the tank circuit.

(iii) Feedback circuit : It provides positive feedback (i.e.) this circuit transfers a part of output energy to LC circuit in proper phase, to maintain the oscillations.

9.14.3 LC oscillators

A transistor can work as an LC oscillator to produce undamped oscillations of any desired frequency, if tank and feedback circuits are properly connected to it. There are different LC oscillators used in electronic circuits, of which, the working principle of Colpitt's oscillator is discussed here.

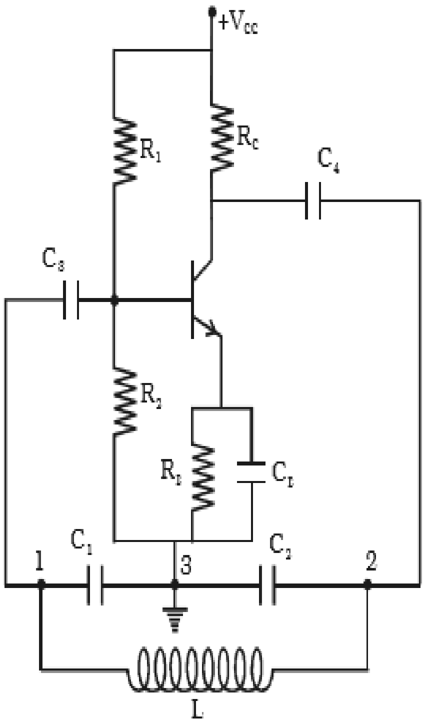


Fig 9.43 Colpitt's oscillator

Colpitt's oscillator

The circuit diagram of Colpitt's oscillator is shown in Fig 9.43. The resistance R_1, R_2 and R_E provide the sufficient bias for the circuit.

The frequency determining network is the parallel resonant circuits consisting of capacitors C_1, C_2 and the inductor L. The junction of C_1 and C_2 is earthed. The function of the capacitor C_4 is to block d.c and provide an a.c. path from the collector to the tank circuit. The voltage developed across C_1 provides the positive feedback for sustained oscillations.

Working

When the collector supply voltage is switched on, a transient current is produced in the tank circuit and damped harmonic oscillations are produced. The oscillations across C_1 are applied to the base emitter junction and appear in the amplified form in the collector circuit. If terminal 1 is at positive potential with respect to terminal 3 at any instant, then terminal 2 will be at negative potential with respect to 3, since 3 is grounded. Hence points 1 and 2 are 180° out of phase. The amplifier produces further phase shift of 180°. Thus the total phase shift is 360°. In other words, energy supplied to the tank circuit is in phase with the oscillations and if $A\beta = 1$, oscillations are sustained in the circuit.

The frequency of oscillations is given by $f = \frac{1}{2\pi\sqrt{LC}}$

where $C = \frac{C_1C_2}{C_1 + C_2}$

$$\therefore \quad f = \frac{1}{2\pi}\sqrt{\frac{(C_1 + C_2)}{LC_1C_2}}$$

9.15 Integrated circuit (IC)

An integrated circuit (IC) consists of a single – crystal chip of silicon, containing both active (diodes and transistors) and passive (resistors, capacitors) elements and their interconnections. ICs have the following advantages over the discrete components:

(i) Extremely small in size

(ii) Low power consumption

(iii) Reliability

(iv) Reduced cost

(v) Very small weight

(vi) Easy replacement

ICs offer a wide range of applications and they are broadly classified as digital ICs and linear ICs*. Two distinctly different IC technologies have been employed which are monolithic and hybrid technology.

In monolithic integrated circuits, all circuit components both active and passive elements and their inter connections are made on the top of a single silicon chip. The monolithic circuit is ideal for applications in the situations, where identical currents are received in large quantities. Hence it provides lowest cost per unit and highest order of reliability. In hybrid circuits, separate component parts are attached to a ceramic substrate and the components are interconnected by means of either metallization pattern or wire bonds.

Typical chip sizes range from about 40 × 40 mils (a mil is 0.001 inch) to about 300 × 300 mils depending on the complexity of the circuit. Any number of components from very few in number to thousands can be fabricated on a single chip. The integrated circuits are available in Dual-in-line package (DIP).

Note: ***Digital ICs :*** *The integrated circuits which process the digital signals are called digital ICs.*

Linear ICs : *The integrated circuits which process the analog signals are called linear ICs.*

9.16 Digital electronics

The term digital is derived from the way in which computers perform operations using digits. Initially, applications of digital electronics were confined to computer systems. Nowadays, digital techniques are applied in many areas, such as telephony, radar, medical instruments, navigation and military systems etc. Digital Electronics involves circuits and systems in which there are only two possible states which are represented by voltage levels. Other circuit conditions such as current levels, open or closed switches can also represent the two states.

Analog signal

The signal current or voltage is in the form of continuous, time varying voltage or current (sinusoidal). Such signals are called continuous or analog signals. A typical analog signal is shown in Fig 9.44.

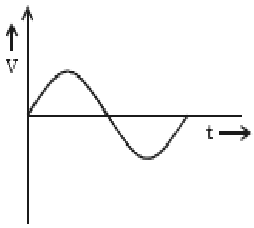


Fig 9.44 Analog signal

Digital signal and logic levels

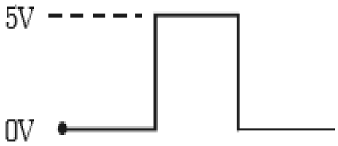


Fig 9.45 Digital Signal

A digital signal (pulse) is shown in Fig 9.45. It has two discrete levels, 'High' and 'Low'. In most cases, the more positive of the two levels is called HIGH and is also referred to as logic 1. The other level becomes low and also called logic 0. This method of using more positive voltage level as logic 1 is called a positive logic system. A voltage 5V refers to logic 1 and 0 V refers to logic 0. On the other hand, in a negative logic system, the more negative of the two discrete levels is taken as logic 1 and the other level as logic 0. Both positive and negative logic are used in digital systems. But, positive logic is more common of logic gates. Hence we consider only positive logic for studying the operation of logic gates.

9.16.1 Logic gates

Circuits which are used to process digital signals are called logic gates. They are binary in nature. Gate is a digital circuit with one or more inputs but with only one output. The output appears only for certain combination of input logic levels. Logic gates are the basic building blocks from which most of the digital systems are built up. The numbers 0 and 1 represent the two possible states of a logic circuit. The two states can also be referred to as 'ON and OFF' or 'HIGH and LOW' or 'TRUE and FALSE'.

9.16.2 Basic logic gates using discrete components

The basic elements that make up a digital system are 'OR', 'AND' and 'NOT' gates. These three gates are called basic logic gates. All the possible inputs and outputs of a logic circuit are represented in a table called TRUTH TABLE. The function of the basic gates are explained below with circuits and truth tables.

(i) OR gate

An OR gate has two or more inputs but only one output. It is known as OR gate, because the output is high if any one or all of the inputs are high. The logic symbol of a two input OR gate is shown in Fig 9.46a.

The Boolean expression to represent OR gate is given by Y= A+B (+ symbol should be read as OR)

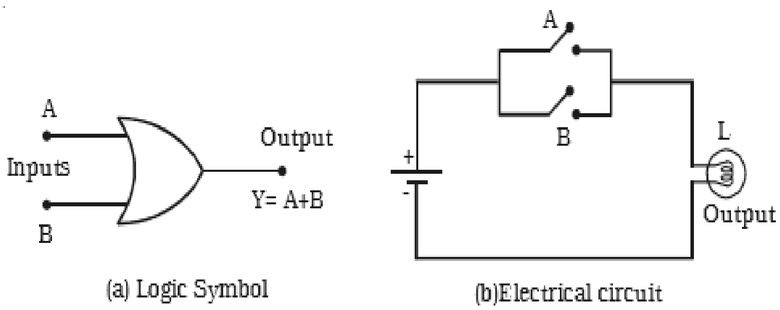


Fig 9.46 OR gate

The OR gate can be thought of like an electrical circuit shown in Fig 9.46b, in which switches are connected in parallel with each other. The lamp will glow if both the inputs are closed or any one of them is closed.

Diode OR gate

Fig 9.47 shows a simple circuit using diodes to build a two input OR gate. The working of this circuit can be explained as follows.

Case (i) A = 0 and B = 0

When both A and B are at zero level, (i.e.) low, the output voltage will be low, because the diodes are non-conducting.

Case (ii) A = 0 and B = 1

When A is low and B is high, diode D_2 is forward biased so that current flows through R_L and output is high.

Case (iii) A = 1 and B = 0

When A is high and B is low, diode D_1 conducts and the output is high.

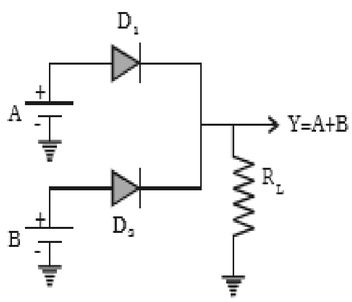


Fig 9.47 OR gate using diodes

Table 9.1 Truth table of OR gate

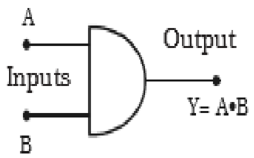
Inputs		Output
A	B	Y = A + B
0	0	0
0	1	1
1	0	1
1	1	1

Case (iv) A = 1 and B = 1

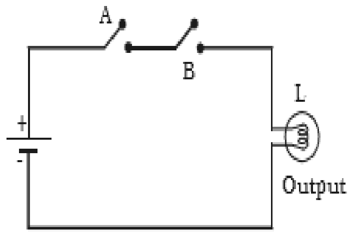
When A and B both are high, both diodes D_1 and D_2 are conducting and the output is high. Therefore Y is high. The OR gate operations are shown in Table 9.1.

(ii) AND gate

An AND gate has two or more inputs but only one output. It is known as AND gate because the output is high only when all the inputs are high. The logic symbol of a two input AND gate is shown in Fig 9.48a.



(a) Logic symbol



(b) Electrical Circuit

Fig 9.48 AND gate

The Boolean expression to represent AND gate is given by Y = A·B (· should be read as AND)

AND gate may be thought of an electrical circuit as shown in Fig 9.48b, in which the switches are connected in series. Only if A and B are closed, the lamp will glow, and the output is high.

Diode AND gate

Fig 9.49 shows a simple circuit using diodes to build a two-input AND gate. The working of the circuit can be explained as follows :

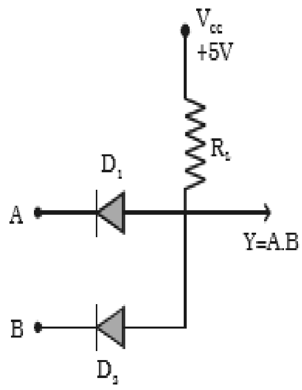


Fig 9.49 AND gate using diodes

Table 9.2 Truth table of AND gate

Inputs		Output
A	B	Y = A · B
0	0	0
0	1	0
1	0	0
1	1	1

Case (i) A = 0 and B = 0

When A and B are zero, both diodes are in forward bias condition and they conduct and hence the output will be zero, because the supply voltage V_{CC} will be dropped across R_L only. Therefore Y = 0.

Case (ii) A = 0 and B = 1

When A = 0 and B is high, diode D_1 is forward biased and diode D_2 is reverse biased. The diode D_1 will now conduct due to forward biasing. Therefore, output Y = 0.

Case (iii) A = 1 and B = 0

In this case, diode D_2 will be conducting and hence the output Y = 0.

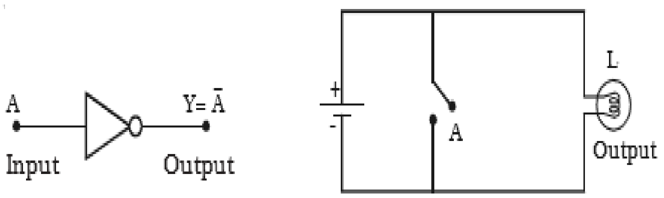
Case (iv) A = 1 and B = 1

In this case, both the diodes are not conducting. Since D_1 and D_2 are in OFF condition, no current flows through R_L. The output is equal to the supply voltage. Therefore Y = 1.

Thus the output will be high only when the inputs A and B are high. The Table 9.2 summarises the function of an AND gate.

(iii) NOT gate (Inverter)

The NOT gate is a gate with only one input and one output. It is so called, because its output is complement to the input. It is also known as inverter. Fig 9.50a shows the logic symbol for NOT gate.



(a) logic symbol *(b) Electrical circuit*

Fig 9.50 NOT gate

The Boolean expression to represent NOT operation is Y = $\bar{A}$.

The NOT gate can be thought of like an electrical circuit as shown in Fig 9.50b. When switch A is closed, input is high and the bulb will not glow (i.e) the output is low and vice versa.

Fig 9.51 is a transistor in CE mode, which is used as NOT gate. When the input A is high, the transistor is driven into saturation and hence the output Y is low. If A is low, the transistor is in cutoff and hence the output Y is high. Hence, it is seen that whenever input is high, the output is low and vice versa. The operation of NOT gate is shown in Table 9.3.

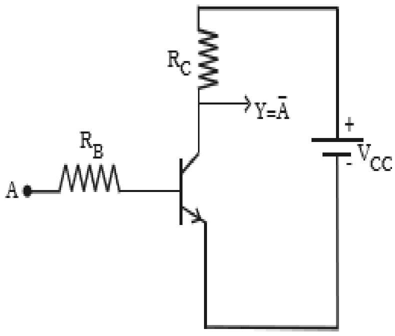


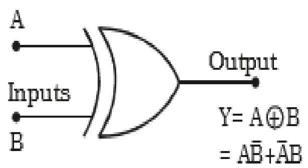
Fig 9.51 NOT gate using transistor

Table 9.3 Truth Table of NOT gate

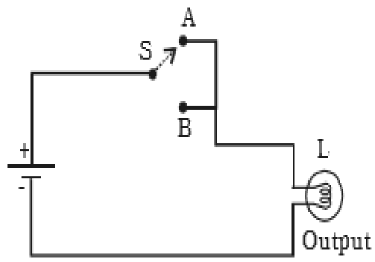
Input	Output
A	$Y = \overline{A}$
0	1
1	0

9.16.3 Exclusive OR gate (EXOR gate) [Not for examination]

The logic symbol for exclusive OR (EXOR) gate is shown in Fig 9.52a.



(a) Logic symbol



(b) Electrical circuit

Fig 9.52 Exclusive OR gate

The Boolean expression to represent EXOR operation is

$$Y = A \oplus B = A\overline{B} + \overline{A}B$$

EXOR gate has an output 1, only when the inputs are complement to each other. The equivalent switching circuit is shown in Fig 9.52b.

Switch positions A and B will individually make the lamp to be ON. But the combination of A and B is not possible.

The EXOR operation is represented in Table 9.4.

Table 9.4 Truth table of EXOR gate

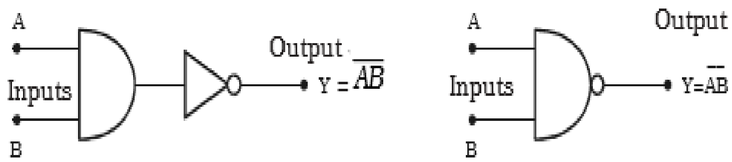
Inputs		Output
A	B	$Y = A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

9.16.4 NAND gate

This is a NOT–AND gate. It can be obtained by connecting a NOT gate at the output of an AND gate (Fig 9.53a).

The logic symbol for NAND gate is shown in Fig 9.53b.

The Boolean expression to represent NAND Operation is Y = $\overline{AB}$



(a) AND gate followed by NOT *(b) Logic Symbol*

Fig 9.53 NAND gate

NAND gate function is reverse of AND gate function. A NAND gate will have an output, only if both inputs are not 1. In other words, it gives an output 1, if either A or B or both are 0. The operation of a NAND gate is represented in Table 9.5.

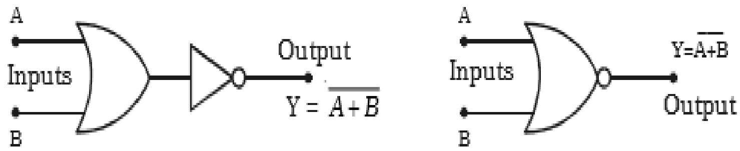
Table 9.5 Truth table of NAND gate

Inputs		Output Y = $\overline{AB}$
A	B	
0	0	1
0	1	1
1	0	1
1	1	0

9.16.5 NOR gate

This is a NOT–OR gate. It can be made out of an OR gate by connecting an inverter at its output (Fig 9.54a).

The logic symbol for NOR gate is given in Fig 9.54b.



(a) OR gate followed by NOT *(b) Logic Symbol*

Fig 9.54 NOR gate

The Boolean expression to represent NOR gate is Y = $\overline{A+B}$

The NOR gate function is the reverse of OR gate function. A NOR gate will have an output, only when all inputs are 0. In a NOR gate, output is high, only when all inputs are low. The NOR operation is represented in Table 9.6.

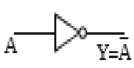
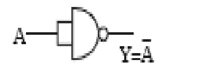
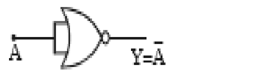
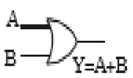
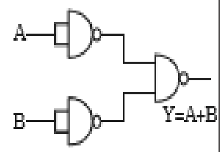
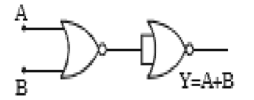
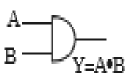
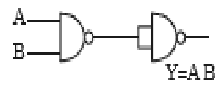
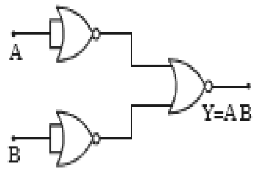
Table 9.6 Truth table of NOR gate

Inputs		Output
A	B	$Y = \overline{A+B}$
0	0	1
0	1	0
1	0	0
1	1	0

Note: ***NAND and NOR as Universal gates***

NAND and NOR gates are called Universal gates because they can perform all the three basic logic functions. Table 9.8 gives the construction of basic logic gates NOT, OR and AND using NAND and NOR gates.

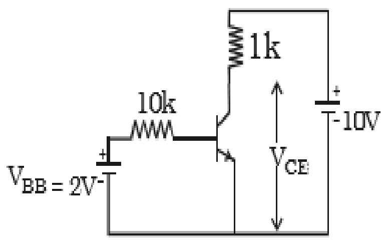
Table 9.8 Substituting NAND / NOR gates

Logic function	Symbol	Circuits using NAND gates only	Circuits using NOR gates only
NOT			
OR			
AND			

Solved Problems

9.1 The current gain β of the silicon transistor used in the circuit as shown in figure is 50. (Barrier potential for silicon is 0.69 V)

Find (i) I_B (ii) I_E (iii) I_C and (iv) V_{CE}



Data V_{BB} = 2 V, V_{CC} = 10 V; β = 50 R_B = 10 kΩ ; R_C = 1 kΩ

The barrier potential for silicon transistor V_{BE} = 0.69 V

Solution : $V_{BB} = I_B R_B + V_{BE}$

$$I_B = \frac{V_{BB} - V_{BE}}{R_B} = \frac{2 - 0.69}{10 \times 10^3} = 131\ \mu A$$

Current gain $\beta = \frac{I_C}{I_B}$

$I_C = I_B\ \beta = 131 \times 10^{-6} \times 50 = 6.5$ mA

Emitter current $I_E = I_C + I_B = 6.5$ mA + 131 μA

= 6.5 mA + 0.131 mA = 6.631 mA

$V_{CC} = V_{CE} + I_C R_C$

$V_{CE} = V_{CC} - I_C R_C$

$= 10 - (6.5 \times 10^{-3} \times 1 \times 10^3)$

= 3.5 V

9.2 A transistor is connected in CE configuration. The voltage drop across the load resistance (R_C) 3 kΩ is 6 V. Find the base current. The current gain α of the transistor is 0.97

Data : Voltage across the collector load resistance (R_C) = 6 V

α = 0.97; R_C = 3 kΩ

Solution : The voltage across the collector resistance is, $R_C = I_C R_C = 6V$

Hence, $I_C = \frac{6}{R_C} = \frac{6}{3 \times 10^3} = 2mA$

Current gain $\beta = \frac{\alpha}{1-\alpha} = \frac{0.97}{1-0.97} = 32.33$ ∴ $I_B = \frac{I_C}{\beta} = \frac{2 \times 10^{-3}}{32.33} = 61.86\ \mu A$

Self evaluation

(The questions and problems given in this self evaluation are only samples. In the same way any question and problem could be framed from the text matter. Students must be prepared to answer any question and problem from the text matter, not only from the self evaluation.)

9.1 *The electrons in the atom of an element which determine its chemical and electrical properties are called*

(a) valence electrons (b) revolving electrons

(c) excess electrons (d) active electrons

9.2 *In an N–type semiconductor, there are*

(a) immobile negative ions (b) no minority carriers

(c) immobile positive ions (d) holes as majority carriers

9.3 *The reverse saturation current in a PN junction diode is only due to*

(a) majority carriers (b) minority carriers

(c) acceptor ions (d) donor ions

9.4 *In the forward bias characteristic curve, a diode appears as*

(a) a high resistance (b) a capacitor

(c) an OFF switch (d) an ON switch

9.5 *Avalanche breakdown is primarily dependent on the phenomenon of*

(a) collision (b) ionisation

(c) doping (d) recombination

9.6 *The colour of light emitted by a LED depends on*

(a) its reverse bias (b) the amount of forward current

(c) its forward bias (d) type of semiconductor material

9.7 *The emitter base junction of a given transistor is forward biased and its collector–base junction is reverse biased. If the base current is increased, then its*

(a) V_{CE} will increase (b) I_C will decrease

(c) I_C will increase (d) V_{CC} will increase.

9.8 *Improper biasing of a transistor circuit produces*

(a) heavy loading of emitter current

(b) distortion in the output signal

(c) excessive heat at collector terminal

(d) faulty location of load line

9.9 An oscillator is

(a) an amplifier with feedback

(b) a convertor of ac to dc energy

(c) nothing but an amplifier

(d) an amplifier without feedback

9.10 In a Colpitt's oscillator circuit

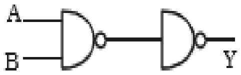
(a) capacitive feedback is used

(b) tapped coil is used

(c) no tuned LC circuit is used

(d) no capacitor is used

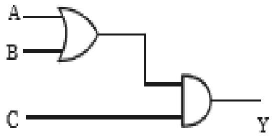
9.11 The following arrangement performs the logic function of _____ gate



(a) AND (b) OR

(c) NAND (d) EXOR

9.12 If the output (Y) of the following circuit is 1, the inputs A B C must be



(a) 0 1 0 (b) 1 0 0

(c) 1 0 1 (d) 1 1 0

9.13 Describe the valence band, conduction band and forbidden energy gap with the help of energy level diagram.

9.14 Describe the energy band structure of insulator, semiconductor and conductor.

9.15 What do you understand by intrinsic and extrinsic semiconductor?

9.16 What is rectification?

9.17 Explain the working of a half wave diode rectifier.

9.18 What is zener breakdown?

9.19 Describe the construction of Zener diode.

9.20 Explain with necessary circuit how zener diode can be used as a voltage regulator.

9.21 Describe the working of PNP and NPN transistors.

9.22 Deduce the relation between α and β of a transistor.

9.23 *Explain an experiment to determine the characteristics of a transistor in CE configuration. Explain how the transistor parameters can be evaluated.*

9.24 *Why is a transistor called as current amplification device?*

9.25 *Why CE configuration is preferred over CB configuration for operating transistor as an amplifier?*

9.26 *Describe the working of a transistor amplifier.*

9.27 *Define bandwidth of an amplifier.*

9.28 *Sketch the circuit of Colpitt's oscillator. Explain its working.*

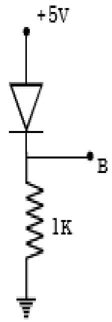
9.29 *Give the function of 'OR' and 'NAND' gates.*

9.30 *What are universal gates? Why are they called so?*

Problems

9.31 *The base current of the transistor is 50 μA and collector current is 25 mA. Determine the values of β and α.*

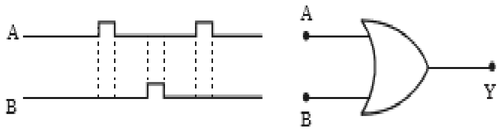
9.32 *Find the voltage at the point B in the figure (Silicon diode is used).*



9.33 *The gain of the amplifier is 100. If 5% of the output voltage is fed back into the input through a negative feed back network. Find out the voltage gain after feed back.*

9.34 *Determine the frequency of oscillations in a Colpitt's oscillator if C_1 = 0.01 μF, C_2 = 0.03 μF and L = 100 mH.*

9.35 *If the two waveforms shown in figure are applied to the OR gate. What is the resulting output wave form?*



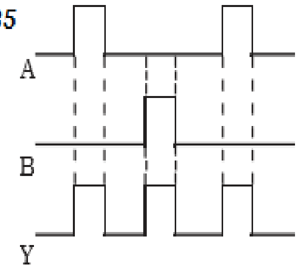
Answers

9.1 *(a)*	**9.2** *(c)*	**9.3** *(b)*	**9.4** *(d)*	**9.5** *(a)*
9.6 *(d)*	**9.7** *(c)*	**9.8** *(b)*	**9.9** *(a)*	**9.10** *(a)*
	9.11 *(a)*	**9.12** *(c)*		

9.31 *500 ; 0.998* **9.32** *+ 4.31 V*

9.33 *16.66* **9.34** *5815 Hz*

9.35



UNIT 9

10. Communication System

A simple greeting "Asalamualiakum" from one person at one location, to be conveyed effectively and, clearly without noise to another person at another location is called communication. Communication needs some elements (devices) which are used to send, process and receive a message (information). A collection of elements which works together to establish a communication between the sender and receiver is called a communication system. The main examples of the communication system include Radio broadcasting, telephone, telegraph, mobile, Edison telegraph, computer and TV cable etc. The sources of an input or message signal include audio files (mp3, mp4, and GIFs) , human voice, e-mail messages, TV picture, and electromagnetic radiation. So one can say two or more people communicating with each other by using sound signals (human voice) is also known as the communication system.

10.1 Basic Elements of Communication System

(Block Diagram of Communication system)

The basic components of a communication system are information source, input transducer, transmitter, communication channel, receiver, output transducer, and destination.

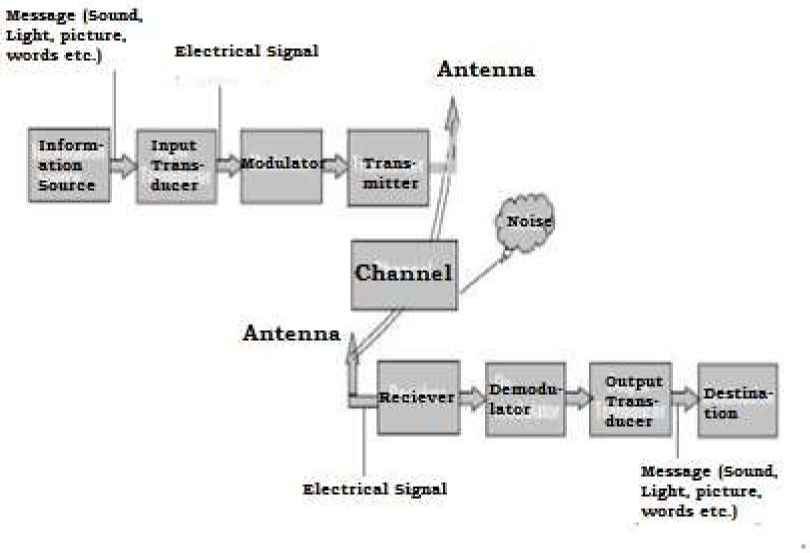


Fig 10 Block Diagram of Communication system

Information Source or message

As we know that the communication system establishes the communication bridge between the sender (transmitter) and receiver. To establish this communication bridge between the sender and receiver, first, we need an information to send. This information originates in the information source.

The information generated by the source may be in the form of sound (human speech), picture (image source), words (plain text in some particular language)

For example, if you are talking with your friend on a phone, you are considered as the information source who generates information in the form of sound.

For beginners to analog communication, it's important to understand the difference between message and information. The message is the part of a communication which involves sending information from source to destination. Information is a meaningful data that the receiver consumes.

Input Transducer

If you want to talk (communicate) with your friend who is sitting beside you, then you can directly talk with him by using voice signals (sound signals). But if the same friend is farther away from you, then you can't directly communicate with him by using voice signals (sound signals) because sound signals cannot travel larger distances. So in order to overcome this problem and transmit information to larger distances, first we need to convert this sound signal into another form of signal (electrical signal or light signal) which travel larger distances. The device which is used to convert this sound signal into another form of signal is called transducer.

A transducer is a device which converts one form of energy or signal into another form of energy or signal. The transducer is present at the input side and output side of the communication system. The transducer that is present at the input side of the communication system is called input transducer. Generally, the input transducer converts the non-electrical signal (sound signal or light signal) into an electrical signal. The best example of an input transducer is the microphone which is placed between the information source and the transmitter section. A microphone is a device which converts your voice signals (sound signals) into electrical signals.

Modulator

Translates the input signal to a higher frequency spectrum and also modulates (camouflages) the signal to combat noise (Amplitude Modulation, Freq Modulation, Phase Modulation, PCM, Delta Modulation, ASK, FSK, PSK, QPSK, QAM, GMSK, etc). The output can be analog or digital (thro A/D converters).

Transmitter

The transmitter is a device which converts the signal produced by the source into a form that is suitable for transmission through antena over a given channel or medium. Transmitters use a technique called modulation to convert the electrical signal into a form that is suitable for transmission over a given channel or medium. Modulation is the main function of a transmitter.

When we send the signal to larger distances, it undergoes various circumstances which makes the signal weak. In order to send the signals to larger distances, without the effect of any external interferences or noise addition and without getting faded away, it has to undergo a process called modulation. Modulation increases the strength of a signal without changing the parameters of the original signal. Thus the resulted signal overcomes the various effects which make it to become weak.

Communication Channel

The communication channel is a medium through which the signal travels.

or

The communication channel is a wired or wireless medium through which the signal (information) travels from source (transmitter) to destination (receiver).

or

The communication channel is a wired or wireless medium that is used to send the signal from the source (transmitter) to the destination (receiver).

or

The communication channel is a wired or wireless medium that connects the transmitter and receiver for sending the signal.

Communication channels are divided into two categories: wired and wireless. Some examples of wired channels include co-axial cables, fiber optic cables, and twisted pair telephone lines. Examples of wireless channels are air, water, and vacuum.

Although channel provides a way for communication, it has one drawback. The communication channel reduces the signal strength (attenuates the signal) that carries the information. This reduction in signal strength is mainly caused by the addition of external noise, physical surroundings, and travel distance. Thus the signal received by the receiver is very weak. To compensate this signal loss, amplifiers (the device that amplifies the signal strength) are used at both the transmitter and the receiver side.

Noise

Noise is an unwanted signal that enters the communication system via the communication channel and interferes with the transmitted signal. The noise signal (unwanted signal) degrades the transmitted signal (signal containing information).

Receiver

The receiver is a device that receives the signal (electrical signal) from the channel through antenna and converts the signal (electrical signal) back to its original form (light and sound) which is understandable by humans at the destination. TV set is a good example of a receiver. TV set receives the signals sent by the TV transmitting stations and converts the signal into a form which is easily understandable by the humans who are watching TV.

Output Transducer

The transducer that is present at the output side of the communication system is called output transducer. Generally, the output transducer converts the electrical signal into a non-electrical signal (sound signal, light signal, or both sound and light signal). The best example of an output transducer is the loudspeaker which is placed between the receiver section and the destination. The loudspeaker converts the electrical signals into sound signals which are easily understandable by the humans at the destination.

Destination

The destination is the final stage in the communication system. Generally, humans at some place are considered as the destination. A destination is a place where humans consume the information. For example, if you are watching TV, you are considered as the destination.

10.2 Signals – Bandwidth:

The bandwidth of a signal is defined as the difference between the upper and lower frequencies of signal (or frequency band). It is also defined as the amount of data transmitted in a fixed amount of data. knowing time and frequency are inversely dependent. Band width is usually expressed in bits per second (bps). Megabits per second (Mbps) is its large unit. In a communication the message signal can be voice, music, picture or computer data. Each of the them has different frequency ranges. For example,

The speech signals frequency range from 300Hz to 3100Hz for telephonic conversation. Therefore, requires *bandwidth* = 3100 -300 = 2800 Hz.

Knowing that the audible range of frequencies extends from 20Hz to 20kHz. Any music requires *bandwidth* of approximately 20kHz because of high frequencies produced by musical instruments.

Video signals for transmission of picture require *bandwidth of* 4.2 MHz

The Television signal which contains both voice and picture is usually allocated a *bandwidth* of 6MHz for transmission

10.2.1 Bandwidth of Transmission Medium

The bandwidth of a signal is defined as the difference between the upper and lower frequencies of signal.

Different types of transmission media offers different bandwidth. For example,

Coaxial cables, widely used wire medium offers *bandwidth* of approximately 750 MHz

Communication through free space using radio waves offers wide range from hundreds of kHz to few GHz (large bandwidth).

Optical fibres are used in the frequency range of 1THz to 1000 THz (huge bandwidth).

As mentioned earlier, to avoid mixing of signals, allotting a band of frequencies to a specific transmitter is in practise

The International Telecommunication Union allots and administers this frequency allocation

Services like FM Broadcast, Television, Cellular Mobile Radio and Satellite communication operate under fixed frequency bands

10.3 Propagation of electromagnetic waves

The propagation of electromagnetic waves depend on the properties of the waves and the environment. Radio waves ordinarily travel in straight lines except where the earth and its atmosphere alter their path. The useful ranges of the electromagnetic spectrum for communication are summarised in Table 10.1.

Radio wave is propagated from the transmitting to the receiving antenna mainly in three different ways depending on the frequency of the wave. They are :

(i) Ground (surface) wave propagation

(ii) Space wave propagation

(iii) Sky wave (or) ionospheric propagation

Table 10.1 Ranges of electromagnetic spectrum used for communication (NOT FOR EXAMINATION)

Name	Frequency	Wavelength
Extremely Low Frequencies (ELF)	30-300 Hz	$10^7 - 10^6$ m
Voice Frequencies (VF)	300-3000 Hz	$10^6 - 10^5$ m
Very Low Frequencies (VLF)	3-30 kHz	$10^5 - 10^4$ m
Low Frequencies (LF)	30-300 kHz	$10^4 - 10^3$ m
Medium Frequencies (MF)	300 kHz – 3 MHz	$10^3 - 10^2$ m
High Frequencies (HF)	3 – 30 MHz	$10^2 - 10$ m
Very High Frequencies (VHF)	30 – 300 MHz	10 – 1 m
Ultra High Frequencies (UHF)	300 MHz – 3 GHz	$1 - 10^{-1}$ m
Super High Frequencies (SHF)	3 – 30 GHz	$10^{-1} - 10^{-2}$ m
Extremely High Frequencies (EHF)	30 – 300 GHz	$10^{-2} - 10^{-3}$ m

10.3.1 *Ground (surface) wave propagation*

Ground or surface waves are the radio waves which travel along the surface of the earth as shown in Fig 10.1. Ground wave propagation takes place when the transmitting and receiving antennas are close to the ground. Ground wave propagation is of prime importance only for medium and long wave signals. All medium wave signals received during the daytime use surface wave propagation.

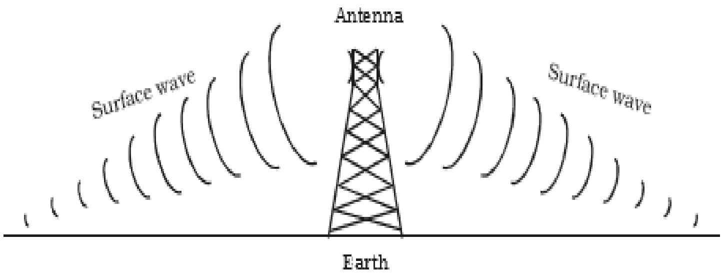


Fig 10.1 Ground or surface wave radiation from an antenna

10.3.2 *Space wave propagation*

Radio waves propagated through the troposphere of the Earth are known as space waves. Troposphere is the portion of the Earth's atmosphere which extends upto 15 km from the surface of the Earth. Space wave usually consists of two components as shown in Fig 10.2.

(i) A component which travels straight from the transmitter to the receiver.

(ii) A component which reaches the receiver after reflection from the surface of the Earth.

Space wave propagation is particularly suitable for the waves having frequency above 30 MHz.

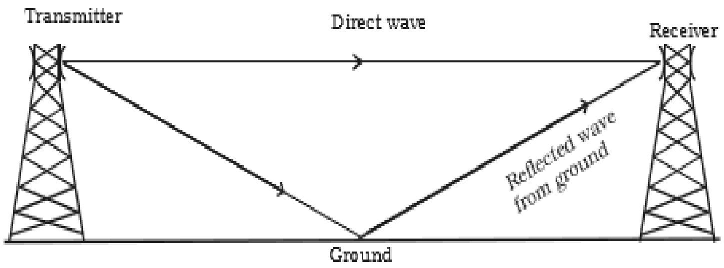


Fig 10.2 Space wave propagation

10.3.3 Sky wave (or) ionospheric propagation

The ionosphere is the upper portion of the atmosphere, which absorbs large quantities of radiant energy like ultra violet rays, cosmic rays etc., from the sun, becoming heated and ionised. This ionised region contains free electrons, positive and negative ions.

Radio waves in the short wave band, radiated from an antenna at large angles with ground, travel through the atmosphere and encounters the ionised region in the upper atmosphere. Under favourable circumstances, the radiowaves get bent downwards due to refraction from the different parts of the ionised region and again reach the earth at a far distant point. Such a radio wave is called the sky wave and such a propagation of radio wave is known as sky wave propagation or ionospheric propagation. Long distance radio communication is thus possible through the sky wave propagation.

Reflection of electromagnetic waves by ionosphere

The electromagnetic waves entering into the ionosphere, are reflected by the ionosphere. In fact, the actual mechanism involved is refraction. The refractive indices of the various layers in the ionosphere do not remain constant and it varies with respect to electron density and the frequency of the incident wave. As the ionisation density increases for a wave approaching the given layer at an angle, the refractive index of the layer is reduced. Hence, the incident wave is gradually bent farther and farther away from the normal as shown in Fig 10.3 until some point. When the electron density is large, the angle of refraction becomes 90° and the wave, then travel towards the Earth.

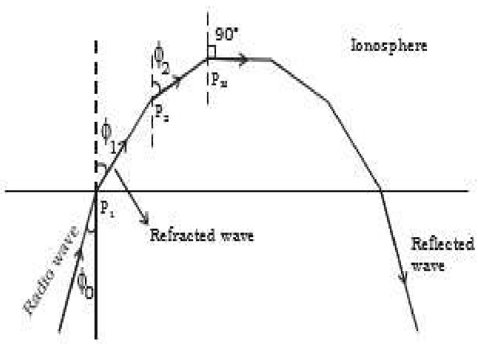


Fig 10.3 Refraction of the radio wave in ionosphere

Skip distance and skip zone

In the skywave propagation, for a fixed frequency, the shortest distance between the point of transmission and the point of reception along the surface is known as the *skip distance.*

When the angle of incidence is large for the ray R_1 as shown in Fig. 10.4, the sky wave returns to the ground at a long distance from

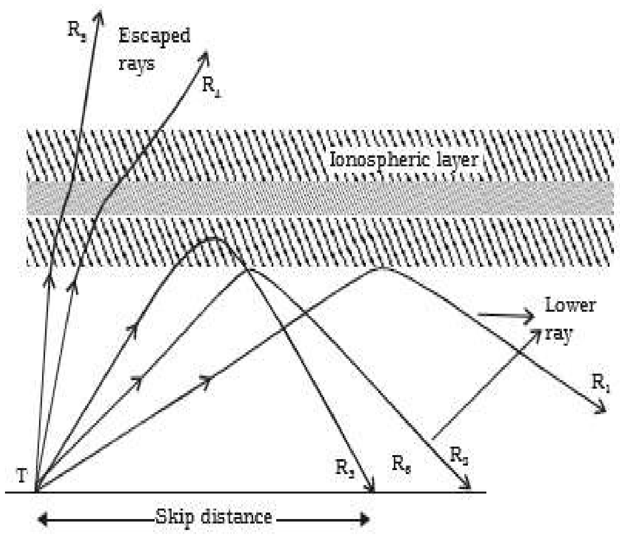


Fig 10.4 Travel of radio waves at different angles of incidence

the transmitter. As this angle is slowly reduced, naturally the wave returns closer and closer to the transmitter as shown by the rays R_2 and R_3. If the angle of incidence is now made significantly less than that of ray R_3, the ray will be very close to the normal to be returned to the Earth. If the angle of incidence is reduced further, the radio waves penetrate through the layer as shown by the rays R_4 and R_5. For a particular angle of incidence, the distance between the point of transmission and the point of reception is minimum. The minimum distance between the transmitter and the ray like R_3 which strikes the Earth is called as the skip distance.

As we move away from the transmitter, the ground wave becomes lesser and lesser significant. A stage comes when there is no reception due to the ground waves. This point lies somewhere in the skip distance. The region between the point where there is no reception of ground waves and the point where the sky wave is received first is known as skip zone. In the *skip zone*, there is no reception at all.

10.4 Modulation

In radio broadcasting, it is necessary to send audio frequency signal (eg. music, speech etc.) from a broadcasting station over great distances to a receiver. The music, speech etc., are converted into audio signals using a microphone. The energy of a wave increases with frequency. So, the audio frequency (20 – 20000 Hz) is not having large amount of energy and cannot be sent over long distances. The radiation of electrical energy is practicable only at high frequencies e.g. above 20 kHz. The high frequency signals can be sent through thousands of kilometres with comparatively small power.

Therefore, if audio signal is to be transmitted properly, the audio signal must be superimposed on high frequency wave called carrier. The resultant waves are known as modulated waves and this process is called as modulation. This high frequency wave (Radio frequency wave) is transmitted in space through antenna. At the receiver end, the audio signal is extracted from the modulated wave by the process called demodulation. The audio signal is then amplified and reproduced into sound by the loud speaker.

A high frequency radio wave is used to carry the audio signal. On adding the audio signal to carrier, any one of the characteristics namely amplitude or frequency or phase of the carrier wave is changed in accordance with the intensity of the audio signal. This process is known as modulation and may be defined as the process of changing amplitude or frequency or phase of the carrier wave in accordance with the intensity of the signal. Some of the modulation process namely, (i) amplitude modulation, (ii) frequency modulation and (iii) phase modulation are discussed.

10.4.1 . Need for Modulation

Need for modulation is because of following reasons:

(i) To separate signal from different transmitters

Audio frequencies are within the range of 20 Hz to 20 kHz. Without modulation all signals at same frequencies from different transmitters would be mixed up. There by giving impossible situation to tune to any one of them. In order to separate the various signals, radio stations must broadcast at different frequencies.

Each radio station must be given its own frequency band. This is achieved by frequency translation as a result of modulation process.

(ii) To reduce Size of the antenna

For efficient transmission the transmitting antennas should have length at least equal to a quarter of the wavelength of the signal to be transmitted. For an electromagnetic wave of frequency 15 kHz, the wavelength λ is 20 km and one-quarter of this will be equal to 5 km. Obviously, a vertical antenna of this size is practically impossible. On the other hand, for a frequency of 1 MHz, this height is reduced to 75 m.

Also, the power radiated by an antenna of length l is proportional to (l/λ). This shows that for the same antenna length, power radiated is large for shorter wavelength. Thus, our signal which is of low frequency must be translated to the high frequency spectrum of the electromagnetic wave. This is achieved by the process of modulation.

(iii) To avoids mixing of signals

If the baseband sound signals are transmitted without using the modulation by more than one transmitter, then all the signals will be in the same frequency range i.e. 0 to 20 kHz . Therefore, all the signals get mixed together and a receiver can not separate them from each other . Hence, if each baseband sound signal is used to modulate a different carrier then they will occupy different slots in the frequency domain (different channels). Thus, modulation

(iv) Increase the Range of Communication

The frequency of baseband signal is low, and the low frequency signals can not travel long distance when they are transmitted. they get heavily attenuated .The attenuation reduces with increase in frequency of the transmitted signal, and they travel longer distance .

The modulation process increases the frequency of the signal to be transmitted . Therefore, it increases the range of communication.

(v) To improves quality of reception

With frequency modulation (FM) and the digital communication techniques such as PCM, the effect of noise is reduced to a great a extent. This improves quality of reception

(vi) To reduce bandwidth.

10.4.2 Amplitude modulation (AM)

When the amplitude of high frequency carrier wave is changed in accordance with the intensity of the signal, the process is called *amplitude modulation.*

In the amplitude modulation, only the amplitude of the carrier wave is changed. The frequency and the phase of the carrier wave remains constant. Fig 10.5 shows the principle of amplitude modulation.

Fig 10.5a shows the audio electrical signal of frequency f_s. Fig 10.5b shows a carrier wave of constant amplitude with frequency f_c. Fig 10.5c is the amplitude modulated wave. It is to be noted that the amplitudes of both positive and negative half cycles of carrier wave are changed in accordance with the signal. Thus the amplitude of the modulated wave possesses the frequency of the audio signal wave.

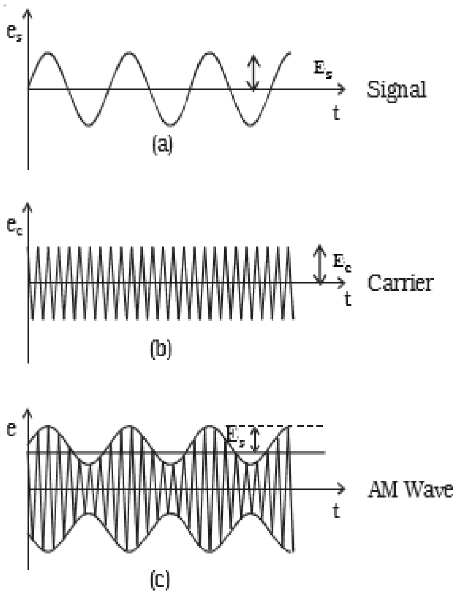


Fig 10.5 Amplitude modulation

Modulation factor

An important term in amplitude modulation is modulation factor which describes the extent to which the amplitude of the carrier wave is changed by the audio signal. It is defined as the ratio of the change of amplitude in carrier wave after modulation to the amplitude of the unmodulated carrier wave.

i.e. modulation factor, $m = \dfrac{\text{Amplitude change of carrier wave after modulation}}{\text{Amplitude of carrier wave before modulation}}$

$$m = \frac{\text{Signal amplitude}}{\text{Carrier amplitude}}$$

Modulation factor determines the strength and quality of the transmitted signal. When the modulation factor $m < 1$, the amount of carrier amplitude variation is small (Fig 10.6a). Consequently, the audio signal being transmitted will not be very strong. When the modulation factor $m > 1$, distortion is produced in the transmitted wave as shown in Fig 10.6 b. Hence, the signal wave is not exactly reproduced. For effective modulation, the degree of modulation should never exceed 100 %.

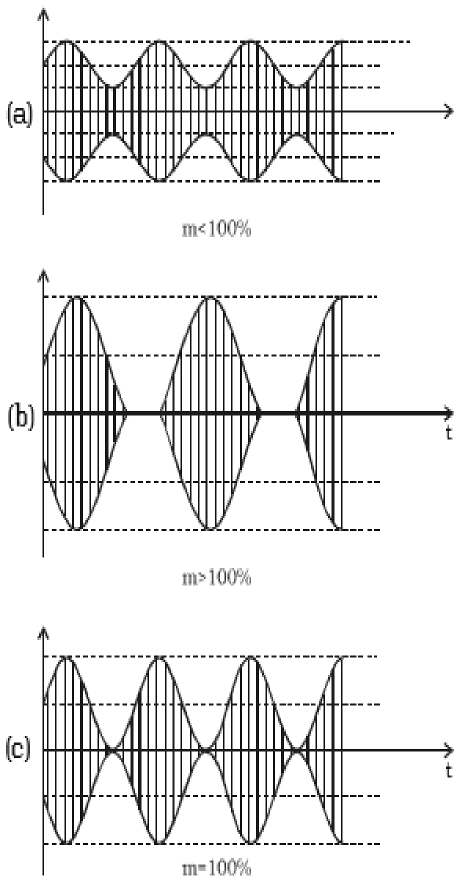


Fig 10.6 Amplitude modulated waves for different modulation factors

Analysis of amplitude modulated wave (Not for examination)

A carrier wave may be represented as,

$$e_c = E_c \cos \omega_c t \qquad \dots (1)$$

where e_c, E_c and ω_c represent the instantaneous voltage, amplitude and angular frequency of the carrier wave respectively.

In amplitude modulation, the amplitude E_c of the carrier wave is varied in accordance with the intensity of the audio signal as shown in Fig 10.5. The modulating signal may be represented as,

$$e_s = E_s \cos\omega_s t \qquad \dots (2)$$

where e_s, E_s and ω_s represent instantaneous voltage, amplitude and angular frequency of the signal respectively.

Amplitude modulated wave is obtained by varying E_c of equation (1) in accordance with E_s. Thus, amplitude modulated wave is,

$$e = (E_c + E_s \cos\omega_s t) \cos\omega_c t$$

$$e = E_c \left[1+\left(\frac{E_s}{E_c}\right)\cos \omega_s t\right] \cos\omega_c t = E_c \left[1+m \cos \omega_s t\right]\cos\omega_c t$$

where m is the modulation factor which is equal to $\frac{E_s}{E_c}$.

$$\therefore\ e = E_c\cos \omega_c t + mE_c\cos \omega_c t \,.\, \cos\omega_s t \qquad \dots(3)$$

$$= E_c \cos \omega_c t + \frac{mE_c}{2}\left[2\cos \omega_c t \cos \omega_s t\right]$$

$$= E_c \cos \omega_c t + \frac{mE_c}{2}\left[\cos (\omega_c + \omega_s) t + \cos (\omega_c - \omega_s) t\right]$$

$$= E_c \cos \omega_c t + \frac{mE_c}{2}\cos (\omega_c + \omega_s) t + \frac{mE_c}{2} \cos (\omega_c - \omega_s) t \quad \dots(4)$$

This expression shows that the modulated wave contains three components:

(i) $E_c \cos \omega_c t$: This component is same as the carrier wave.

(ii) $\frac{mE_c}{2}\cos (\omega_c + \omega_s)t$: This component has a frequency greater than that of the carrier and is called as the Upper Side Band (USB).

(iii) $\frac{mE_c}{2}\cos (\omega_c - \omega_s)t$: This component has a frequency lesser than that of the carrier and is called as the Lower Side Band (LSB).

Frequency spectrum

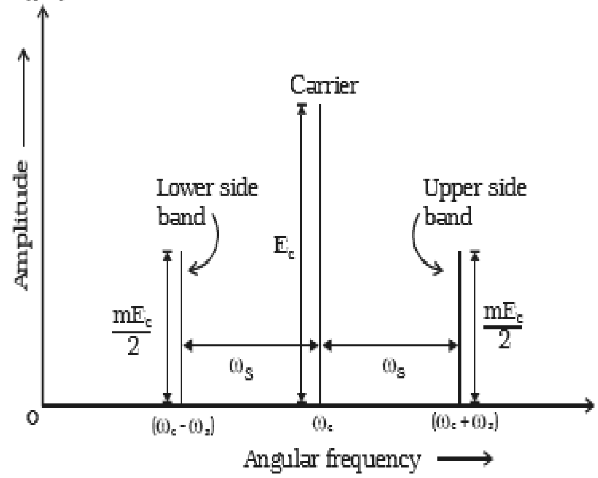


Fig 10.7 Plot of frequency spectrum of amplitude modulated voltage.

The lower side band term and upper side band term are located in the frequency spectrum on either side of the carrier at a frequency interval of ω_s as shown in Fig 10.7. The magnitude of both the upper and lower side bands is $\frac{m}{2}$ times the carrier amplitude E_c. If the modulation factor m is equal to unity, then each side band has amplitude equal to half of the carrier amplitude.

Bandwidth

In an AM wave, the bandwidth is from $(\omega_c - \omega_s)$ to $(\omega_c + \omega_s)$ i.e twice the signal frequency. In the preceding section, it is assumed that the modulating signal is composed of one frequency component only. However, in a broadcasting station, the modulating signal is the human voice or music which contains waves with a frequency range of 300 – 3000 Hz. Each of these waves has its own side bands. The upper side band (USB), in fact, contains all sum components of the signal and carrier frequency whereas lower side band (LSB) contains the difference components, as shown in Fig 10.8.

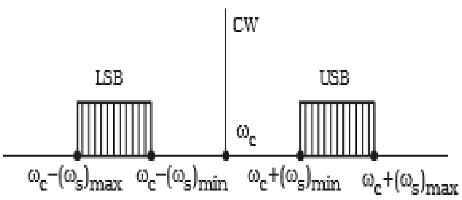


Fig 10.8 Channel width

The channel width is given by the difference between extreme frequencies i.e. between maximum frequency of USB and minimum frequency of LSB.

$\therefore$ Channel width = 2 × maximum frequency of the modulating signal

$$= 2 \times (f_s)_{max}$$

10.4.3 Production of Amplitude Modulated Wave

Block diagram of a simple modulator for obtaining an AM signal is shown in the figure given below

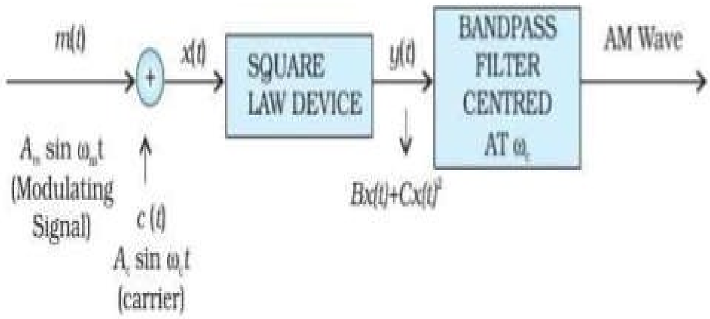


Fig 10.9 Block diagram of simple modulator for attaining AM wave

Here the modulating signal $A_m Sin\omega_m t$ is added to the carrier signal $A_c Sin\omega_c t$ to produce the signal $x(t)$. This signal $x(t) = A_m Sin\omega_m t + A_c Sin\omega_c t$ is passed through a square law device which is a non-linear device which produces an output $y(t) = Bx(t) + Cx(t)^2$, where B and C are constants.

The modulated signal cannot be transmitted as such. The modulated signal is fed to an antenna of appropriate size for radiation as shown in figure 10.9.

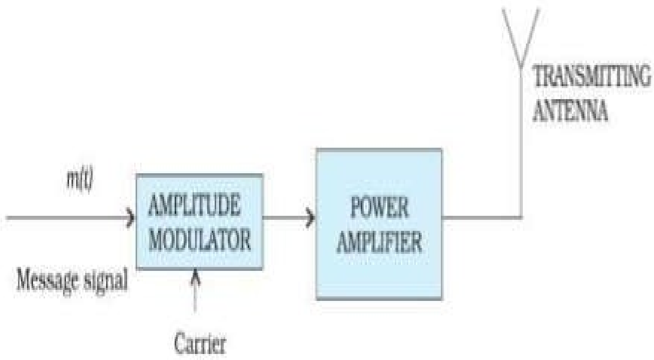


Fig 10.9 AM wave transmission through antenna

10.4.4 Detection of Amplitude Modulated Wave

Detection is the process of recovering the modulating signal from the modulated carrier wave. The block diagram of a typical receiver is shown in the fig 10.10.

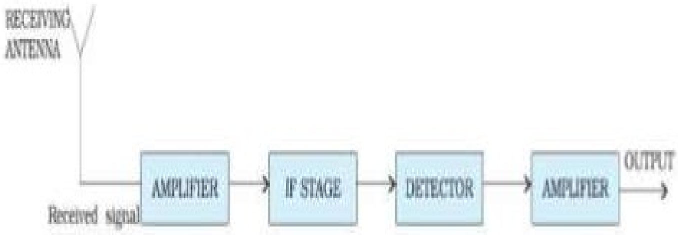


Fig 10.10 Block diagram of receiver to detect AM wave

The transmitted message signal gets attenuated while propagating through the channel. The receiving antenna is, therefore, to be followed by an amplifier & a detector. In addition, to facilitate further processing, the carrier frequency is usually changed to a lower frequency by what is called an intermediate frequency (IF) stage preceding the detection. The detected signal may not be strong enough to be made use of and hence is required to be amplified. The block diagram of a detector for AM signal, the quantity on y-axis can be current or voltage

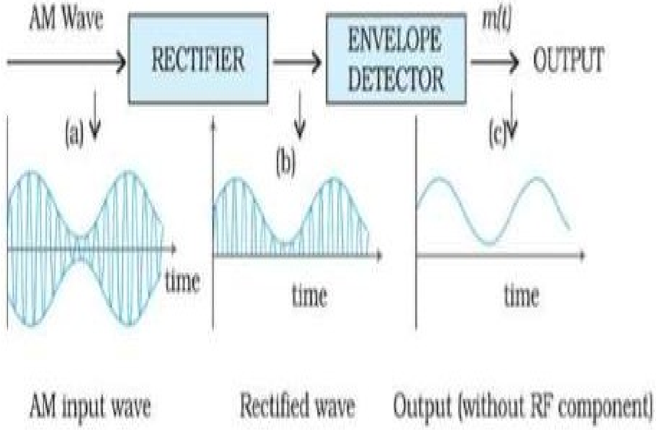


Fig 10.11 Block diagram of AM signal detector

The modulated signal of the form given in Fig 10.11a is passed through a rectifier to produce the output shown in Fig 10.11b. This envelope of the signal (b) is message signal. In order to retrieve back $m(t)$, the signal is passed through an envelope detector (which may consist of a simple RC circuit), see Fig 10.11c.

Solved problems

10.1 A 10 MHz sinusoidal carrier wave of amplitude 10 mV is modulated by a 5 kHz sinusoidal audio signal wave of amplitude 6 mV. Find the frequency components of the resultant modulated wave and their amplitude.

Data: Frequency of the carrier = f_c = 10 MHz

Frequency of the signal = f_s = 5 kHz = 0.005 MHz

Amplitude of the carrier signal = E_c = 10 mV

Amplitude of the audio signal = E_s = 6 mV

Frequency components of modulated wave = ?

Amplitude of the components in the modulated wave = ?

Solution : The modulated carrier wave contains the following frequencies :

(i) Original carrier wave of frequency = f_c = 10 MHz

(ii) Upper side band frequency, $f_c + f_s$ = 10 + 0.005

= 10.005 MHz

(iii) Lower side band frequency $f_c - f_s$ = 10 – 0.005

= 9.995 MHz

The modulation factor is,

$$m = \frac{E_s}{E_c} = \frac{6}{10} = 0.6$$

$\therefore$ Amplitude of USB = Amplitude of LSB = $\frac{mE_c}{2} = \frac{0.6 \times 10}{2}$ = 3mV

10.2 An FM signal has a resting frequency of 105 MHz and highest frequency of 105.03 MHz when modulated by a signal. Determine (i) frequency deviation and (ii) carrier swing.

Data : Resting frequency (f) = 105 MHz

Frequency of the signal (f_s) = 5 kHz

Highest frequency of the modulated wave, (f_m) = 105.03 MHz

Frequency deviation = Δf = ? Carrier swing (CS) = ?

Solution : Frequency deviation $(\Delta f) = f_m - f$

Δf = 105.03 – 105 = 0.03 MHz

Carrier swing = 2 × Δf = 2 × 0.03 = 0.06 MHz = 60 kHz

Self evaluation

(The questions and problems given in this self evaluation are only samples. In the same way any question and problem could be framed from the text matter. Students must be prepared to answer any question and problem from the text matter, not only from the self evaluation.)

10.1 High frequency waves follow

(a) the ground wave propagation

(b) the line of sight direction

(c) ionospheric propagation

(d) the curvature of the earth

10.2 The main purpose of modulation is to

(a) combine two waves of different frequencies

(b) acquire wave shaping of the carrier wave

(c) transmit low frequency information over long distances efficiently

(d) produce side bands

10.3 In amplitude modulation

(a) the amplitude of the carrier wave varies in accordance with the amplitude of the modulating signal.

(b) the amplitude of the carrier wave remains constant

(c) the amplitude of the carrier varies in accordance with the frequency of the modulating signal

(d) modulating frequency lies in the audio range

10.4 In amplitude modulation, the band width is

(a) equal to the signal frequency

(b) twice the signal frequency

(c) thrice the signal frequency

(d) four times the signal frequency

10.5 In phase modulation

(a) only the phase of the carrier wave varies

(b) only the frequency of the carrier wave varies.

(c) both the phase and the frequency of the carrier wave varies.

(d) there is no change in the frequency and phase of the carrier wave

10.6 *What are the different types of radio wave propagation?*

10.7 *Explain the ground wave propagation.*

10.8 *Explain the wave propagation in ionosphere.*

10.9 *What is the necessity of modulation?*

10.10 *Explain amplitude modulation.*

10.11 *Define modulation factor.*

10.12 *Define bandwidth.*

10.13 *What are the limitations of amplitude modulation?*

10.14 *Explain the functional block diagrams of production and detection of AM wave*

Answers

10.1 *(c)* **10.2** *(c)* **10.3** *(a)* **10.4** *(b)*

10.5 *(c)*