Lecture Based
Modules for Bridge Course in
Physics

All India Council for Technical Education
Nelson Mandela Marg, Vasant Kunj, New Delhi 110 070
www.aicte-india.org

PREFACE

Globalization of the world economy and higher education are driving profound changes in engineering education system. Worldwide adaptation of Outcome Based Education framework and enhanced focus on higher order learning and professional skills necessitates paradigm shift in traditional practices of curriculum design, education delivery and assessment. AICTE has also taken various quality initiatives for strengthening the technical education system in India. These initiatives are essential for promoting quality education in our institutions in the country so that our students passing out from these institutions may match the pace with global standards.

A quality initiative by AICTE is 'Revision of Curriculum'. Recently, AICTE has released an outcome based Model Curriculum for various Undergraduate degree courses in Engineering & Technology which are available on AICTE website. A three-week mandatory induction program is developed as a part of the model curriculum for the first year UG Engineering students which helps students joining the first year of the college from diverse backgrounds to get adjusted in the new environment of the institution.

Education is primarily conceived by students as one simple remembering facts by rote. However, Science education also requires clear understanding of science concepts and a proper logical thinking or a constructive thinking by students. We all know that the students seeking admission in an undergraduate degree engineering program have passed their 10+2 in science but it was felt that a student joining an engineering program after 10+2 require reinforcement of fundamental science concepts i.e. basic science courses in Physics, Chemistry and Mathematics. To support the students, gain better understanding, AICTE decided to initiate the task of development of bridge courses in Physics, Chemistry and Mathematics and it was entrusted to IIT-BHU. These bridge courses aim to accelerate the students' knowledge in these subjects acquired at 10+2 level; and also bridge the gap between the school science syllabus and the level needed to understand their applications to engineering concepts. Therefore, it was decided that after completion of the 3-week mandatory induction programme introduced for the first year UG engineering students, bridge course in basic Physics, Chemistry and Mathematics may be taken up by universities/institutions for the students for the remaining part of the semester. The concerned University/institution has a flexibility to adopt these modules on bridge courses by adjusting teaching hours accordingly.

The lecture based modules in Physics, Chemistry and Mathematics have been developed by a team of respective Course Coordinators from Indian Institute of Technology, Banaras Hindu University. AICTE approved institutions may utilize these modules *'Lecture Based Modules for Bridge Courses - Physics, Chemistry and Mathematics'* for teaching students to help bridge the gap of their studies of 10+2 and UG level.

(Prof. Anil D. Sahasrabudhe)
Chairman, AICTE

ACKNOWLEDGEMENT

Curriculum plays a crucial role in enabling quality learning for our young learners in our society i.e. students. An effective curriculum not only enables a student's learning process & knowledge acquired but also supports students to overcome their inhibitions and aids in their holistic development. AICTE in 2018 released a Model Curriculum for various Undergraduate degree courses in Engineering & Technology. This curriculum is equipped with making students industry ready, allow internships for hands on experience, learn about Constitution of India, Environment science etc. Induction program has been included as a mandatory program for the first year engineering students to get acquainted and get accustomed to this new environment in the college. a curriculum needs to be consistent and sustainable and it has been noticed that students joining an engineering program required to strengthen their concepts in science subjects i..e Physics, Chemistry and Mathematics building a better foundation during the first semester itself. AICTE therefore decided to develop lecture based bridge courses in basic science subjects i..e Physics, Chemistry and Mathematics for students,. The lecture based modules in Physics, Chemistry and Mathematics have been developed by IIT-BHU. This task has been accomplished by a team of respective Course Coordinators under Prof Indrajit Sinha, Department of Chemistry, IIT BHU as Overall Coordinator.

AICTE places on record its acknowledgement and appreciation to Dr. Indrajit Sinha, Department of Chemistry, IIT-BHU as overall coordinator; and respective course coordinators and their team of faculty members at IIT-BHU for developing these lecture based modules for bridge courses:

The faculty team from IIT-BHU:

Department of Physics, IIT, BHU

Prof. B. N. Dwivedi, Prof. O.N. Singh, Prof. D. Giri, Prof. P. Singh, Prof. S. Chatterjee, Prof. R. Prasad, Dr. (Mrs) A. Mohan, Dr. P.C. Pandey, Dr. (Mrs) S. Upadhyay, Dr. A.K. Srivastava, Dr. S.K. Mishra, Dr. A.S. Parmar, Dr. S. Tripathi, Dr. S. Patil, Dr. (Mrs) S. Mishra, Dr P. Dutta, Dr. S.K. Singh, Dr (Mrs) N. Agnihotri

Department of Chemistry, IIT, BHU

Prof. R. B. Rastogi, Prof. A K Mukherjee, Prof. M A Quraishi, Prof. V. Srivastava, Prof. Y.C. Sharma, Prof. D. Tiwari, Prof. K.D. Mandal, Dr. I. Sinha, Dr. S.Singh, Dr. M. Malviya

Department of Mathematical Sciences, IIT, BHU

Prof. L. P. Singh, Prof. Rekha Srivastava, Prof. T Som, Prof. S.K. Pandey Prof. S.K. Upadhyay, Prof. S. Das, Prof. S. Mukhopadhyay, Prof. S. Ram, Prof. K.N. Rai, Dr. A.J. Gupta, Dr. Rajeev, Dr. R.K. Pandey, Dr. V.K. Singh, Dr. Sunil Kumar, Dr. Lavanya Shivkumar, Dr. A Benerjee, Dr. D. Ghosh, Dr. V.S. Pandey

Institutions may adjust teaching hours to utilize these modules *'Lecture based Modules for Bridge Courses - Physics, Chemistry and Mathematics'* to bridge the gap of 10+2 and UG level.

(Prof. Rajive Kumar)
Adviser-I(P&AP), AICTE

PHYSICS MODULES

(For AICTE Approved Colleges)

Prepared by

Department of Physics
Indian Institute of Technology
(Banaras Hindu University)
Varanasi - 221005

Preface

The genesis of this module lies in the Induction Program first conceived and started by IIT(BHU) on 2016 on mass scale for about 1000 students. The fact is that the students are over-burdened and stressed out due to a hectic high school life. To refresh their creative mind, they were exposed to month long diverse credit courses like Physical Education, Human Values and Creative Practices, as well as several non-credit informal activities. In a welcome step the AICTE has proposed to extend this program to the Engineering Colleges affiliated to them.

Infact, purpose of this module is to bridge the gap between what the students need to know before they can start taking the advanced courses in the college level and what they are actually aware of from the intermediate level. Consequently, after the completion of the 3-weeks induction program, it is proposed that (besides other subjects) bridge courses in basic Physics, Chemistry and Mathematics should be taught to these students for the rest of the semester. The bridge courses will cover typical weaknesses of students in science at the 10+2 level.

The modules in Physics are prepared keeping in mind that an hour of discussion will bring all the students in the same stage such that they can cope up with the courses in their college level, that requires the concepts of different topics in Physics. The modules are made as interactive sessions between the students and the instructors. Furthermore, we have discussed those topics which harder to understand. At the end of the discussion teacher may also take a small test to understand how much the students followed the class.

We are very much grateful to all the faculty members (Prof. B. N. Dwivedi, Prof. O.N. Singh, Prof. D. Giri, Prof. P. Singh, Prof. S. Chatterjee, Prof. R. Prasad, Dr. (Mrs.) A. Mohan, Dr. P. C. Pandey, Dr. (Mrs.) S. Upadhyay, Dr. A. K. Srivastava, Dr. S. K. Mishra, Dr. A. S. Parmar, Dr. S. Tripathi, Dr. S. Patil, Dr. (Mrs.) S. Mishra, Dr. P. Dutta, Dr. S. K. Singh and Dr. (Mrs.) N. Agnihotri) in the Department of Physics who devoted their valuable time to prepare the module

This is to mention that that modules are prepared for the students with an objective to create interest among them in the subject. Many materials from the Internet have been adopted to make this lecture more illustrative and elaborative. The materials from the Internet have been utilized solely for educational purpose.

Department of Physics

IIT(BHU), Varanasi

Content

Syllabus

1. **Classical Mechanics**: Centre of Mass, Motion of Centre of mass, Pure Translational and Rotational motion, Torque and angular momentum, Principle of moments (Moment of Inertia), Radius of Gyration, Generalized Motion, Kinematics of rotational motion about a fixed axis.

2. **Mechanical Properties of Solids and Fluids**: Elastic behaviors of solids, Hooke's Law, Young's Modulus, Shear Modulus, Bulk Modulus, Applications of Elastic behaviors of materials, Compressibility, Viscosity, Relative density, Pascal's Law, Streamline Flow, Bernoulli's Principle, Surface Tension, Drops and Bubbles

3. **Waves and Oscillations**: Rectilinear motion, Oscillations or Vibrations, Simple Harmonic Motion, Damped Harmonic motion: Real oscillatory system, Forced or Driven oscillation, TYPES OF WAVES, Superposition of Waves, Reflection and Refraction, Standing Waves and Normal Modes, Beats, Resonance, Doppler's Effect

4. **Electricity and Magnetism:** Physical concepts of gradient, divergence, and curl; Laplacian operator, Concept of electricity and magnetism, Coulomb's law, Electrostatics, Magnetostatics, The Lorentz force, Maxwell's equations

5. **Electromagnetic Signal**: Introduction to Maxwell's equations, The dynamical magnetic field, The dynamical electric field, Electromagnetic Waves

6. **Wave Optics:** Interference of light, Photons, Young's Double Slit Experiment, Huygens's Principle, Diffraction, Diffraction Grating, Polarization

7. **Semiconductor Electronics**: Classification of metals, conductors and semiconductors, Fermi Level, Intrinsic Semiconductor, Extrinsic Semiconductor, *p-n* junction, Semiconductor Diode, Half-Wave Rectifier, Full-Wave Rectifier, Zener diode, Photodiode, Light emitting diode, Junction Transistor

8. **Modern Physics**: Wave nature of light, Particle nature of light: the photon, De Broglie Hypothesis, Experimental confirmation of de Broglie hypothesis (Davisson and Germer's Experiment)

9. **Atomic and Nuclear Physics**: Matters, Atoms, Atomic Theory: Atomic Theory by John Dolton, Atomic Theory by J. J Thompson, Atomic Theory by Ernest Rutherford, Atomic Theory by James Chadwick, Discovery of the Neutron, Bohr's Postulates, Proton, Neutron, Electron, Limitations of Bohr's Theory

Module – I
Classical Mechanics
(Rotational Motion)

Lectures: 02

Objective: In these lectures we will discuss the rotational motion of extended bodies (bodies of finite size made of many particles). We will also discuss the concept of motion of centre of mass for extended bodies.

Prerequisites: To understand the concept described in these lectures, students should know about the motion of single particles. Conservation laws of motion of single particle and concept of vector and scalar product. The topic can be started by asking questions like

(a) What are the Newton's laws of motion?
(b) What is the difference between rectilinear and rotational motion?
(c) Is the velocity of a particle under uniform rotational motion constant?
(d) What happens when we slide a block down an inclined plane.
(e) What is the difference when we slide a roll along the same inclined plane.
(f) The above two motions are same or different.
(g) How to describe the rotation of ceiling fan.

We see mainly three types of motion: Pure translational, Pure rotational and combination of translational and rotational. To start the discussion, we first need to know about the centre of mass.

Centre of Mass:

Motion of an extended body is described by the motion of its centre of mass. Although, no material body is perfectly rigid, many extended bodies can be considered to be a rigid body to solve most of the problems related to motion of extended bodies. A rigid body is one whose geometric shape and size remains unchanged under the action of any external force. Centre of mass of non-homogeneous bodies can be calculated by

$$R = \frac{\sum m_i r_i}{M}$$

where R =Xi+Yj+Zk is the position vector of centre of mass and M is the total mass of the body and summation is over all the particles in the body for example if body consist of five particles than i varies from 1 to 5. For homogeneous bodies (bodies having a continuous distribution of mass), the summation in the formula of centre of mass can be replaced by integration. Hence, for non-homogeneous body's centre of mass is not necessary to coincide with the geometrical centre of the body. But for homogeneous body's centre of mass is the same as the geometrical centre of the body.

Motion of Centre of mass:

The motion of a rigid body, is the motion of the centre of mass. Centre of mass of a system of particles move as if all the mass is concentrated to centre of mass and all the forces are applied to that. This is a consequence of **Newton's third** law of motion. Hence the total linear momentum of the system of particles is the product of the total mass of the system and the velocity of the centre of mass. And Newton's second law is applied in the same way as for a single particle to the centre of mass.

Pure Translational and Rotational motion:

We characterize the motion of a body as pure translational if all the parts of the body have the same translational velocity. Similarly, we will call motion is pure rotational if all the parts will have a same angular velocity.

Question 1: *What is the angular velocity. Is it a scalar or vector (like translational velocity). If it is a vector, what is its direction.*

Question 2: *What are suitable coordinate systems to describe transitional and angular velocities.*

The angular velocity is always defined with respect to a fixed axis of rotation. The relation between translational and angular velocity is defined by a vector product

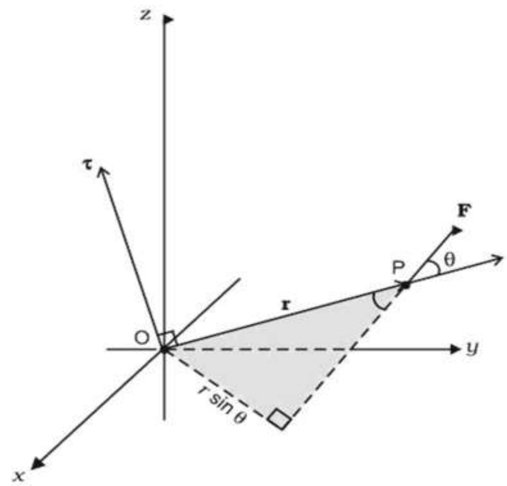
$$v = \omega \times r$$

where, v is translational velocity, ω is the angular velocity and r is the position vector. Hence, direction of v is perpendicular to angular velocity and position vector. Going back to Newton's first law, which says that in the absence of any external force particle is either at rest or moves with constant velocity. A similar kind of statement can be made for rotational motion also. Since angular velocity is analogous to translational velocity. Therefore, the rate of change of angular velocity is similar to the rate of change of translational velocity. Rate of change of angular velocity is called as **angular acceleration.**

Torque and angular momentum:

We know that a force is needed to have a translational motion of a body. We also need to apply a force to have rotational motion. Than question arise that what kind of force or moment of force is required to have rotational motion? Can we apply force to any point and about any axis?

Since the concept of rotational motion is defined with respect to a fixed axis.



Torque acting on a body is perpendicular to the plane containing position **r** and force **F** and its direction is given by the right handed screw rule. (Source NCERT)

The above moment of force or also called as Torque (τ) is a vector product of force and position vector

$$\tau = r \times F$$

Hence the torque will be zero, either if force is zero, is passing through the origin or if direction of position of force applied and direction of force applied is same.

An example can be given by rotating a door attached from the wall from a fixed axis.

Just like torque in rotational motion is analogue of the force of translational, angular momentum (l) has analogue of linear momentum (p). Therefore, for a particle of mass m and linear momentum **p** at a position vector **r** relative to the origin O, angular momentum **l** of the particle with respect to the origin is defined as

$$l = r \times p$$

Angular momentum **l** is again a vector product of two vector **r** and **p.** It will be zero, if either **r** or **p** is zero or they are in the same direction. Similarly, equation $\frac{dp}{dt} = F$ of translational motion is analogous to equation $\frac{dl}{dt} = \tau$ of rotational motion.

Above equations and concepts can be extended to a system of many particles:

Total angular momentum (L) of a rigid body made up of **n** number of particles is the vector sum of the angular momentum of all the particles in the body

$$L = l_1 + l_2 + \cdots \quad + l_n = \sum_{i=1}^{i=n} l_i$$

Where **l**$_i$ is the angular momentum of the ith particle and defined as

$$l_i = r_i \times p_i$$

Hence the rate of change of total momentum L

$$\frac{dl}{dt} = \tau_{ext}$$

where $$\tau_{ext} = \sum_{i=1}^{i=n} r_i \times F_i^{ext}$$

internal torque, which is a vector product of position and internal force is zero, because of Newton's third law of action and reaction and force between two particles lie along the line joining them. Hence similar to translational equilibrium condition in which total force on the body is zero;

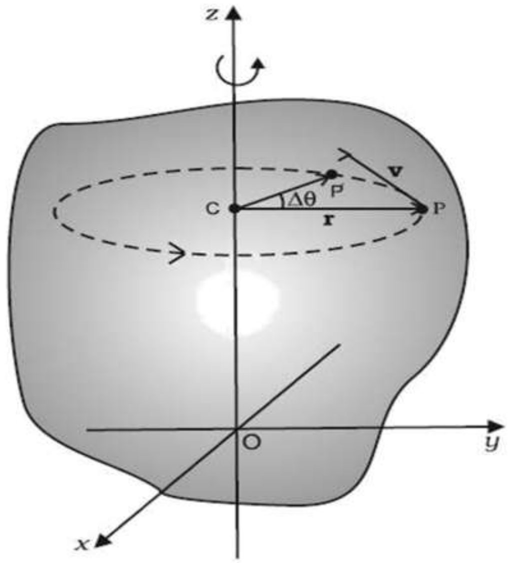
$$F_1 + F_2 + \cdots \ldots \ldots + F_n = \sum_{i=1}^{i=n} F_i = 0$$

We can define rotational equilibrium as total torque acting on a body is zero;

$$\tau_1 + \tau_2 + \cdots \ldots \ldots + \tau_n = \sum_{i=1}^{i=n} \tau_i = 0$$

Principle of moments (Moment of Inertia):

Angular momentum and torque are also called as moment of linear momentum and force. Hence, they are defined with respect to an axis. Imagine rotating a cylinder about two different axes, one passing through the long axis of the cylinder and the other perpendicular to that. We need to apply different force for the two types of rotations. Hence, the force applied for rotational motion of rigid bodies very much depends on the axis of rotation. So far we have learnt about analogous of different quantities in rotational motion with translational motion. Then a question arises what is the analogue of mass of translational motion in rotational motion?



Let us calculate the kinetic energy of a rotating body made of many particles. Kinetic energy of its i[th] particle

$$k_i = \frac{1}{2} m_i v_i^2 = \frac{1}{2} m_i r_i^2 \omega_i^2$$

Hence total kinetic energy, K

$$K = \sum_{i=1}^{i=n} k_i = \frac{1}{2} \sum_{i=1}^{i=n} m_i\, r_i^2 \omega_i^2$$

$$K = \frac{1}{2} I \omega^2$$

where

$$I = \sum_{i=1}^{i=n} m_i\, r_i^2$$

Where I- is called the moment of inertia and it is a characteristic of rigid body and axis through which it is rotated. Hence, the moment of inertia of a rigid body depends on the axis of rotation.

Question: *About which axis would a uniform cube have its minimum moment of inertia?*

Two theorems: Theorem of perpendicular and parallel axes.

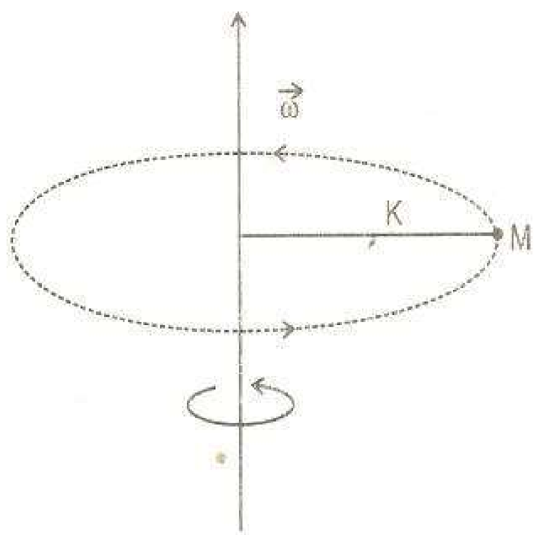
The two theorems and details can be obtained from standard text books. They are useful to calculate the moment of inertia of different bodies.

Radius of Gyration

As discussed above moment of inertia of a rigid body of mass M rotating about an axis passing through the body and perpendicular to its horizontal plane is

$$I = \sum_{i=1}^{i=n} m_i r_i^2$$

If whole mass 'M' of the body is at a radial distance 'K' from the axis of rotation and its moment of inertia remains the same about that axis of rotation then 'K' is called radius of body as shown in below figure



The moment of inertia of the body is then given by

$$I = MK^2$$

$$K = \sqrt{I/M}$$

Table: Analogy between translational motion and rotational motion

Linear motion	Rotational motion
Mass (M)	Moment of Inertia (I)
Translational displacement (s)	Angular displacement (ө)
Translational velocity (v)	Angular velocity (ω)
Translational acceleration (a)	Angular acceleration (α)
Force (F) =Ma	Torque (τ) = I α
Linear momentum (p) =Mv	Angular momentum (L)=I ω
Translational kinetic energy (E)= ½ Mv^2	Angular kinetic energy (E_R)= ½ $I\omega^2$

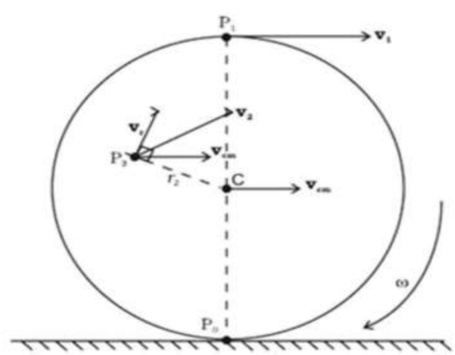
Generalised Motion:

With pure translational motion of rigid body, each part of the body moves with same linear velocity, motion of the centre of mass is enough to explain. But in more general situations, complete motion of the system of particles is described: (i) motion of the centre of mass and (ii) motion of the body about the centre of mass. Example is the motion of a roll sliding along an inclined plane. Although the centre of mass is having translational motion, but each part of the body is having rotation about the fixed axis passing through the centre of mass.

Kinematics of rotational motion about a fixed axis:

Kinematics of rotational motion is similar to translational motion. Analogue to translational motion, rotational motion about a fixed axis, conserve the angular momentum in the absence of external torque. An example can be taken from the rotatory motion on a merry go round.

Rolling motion: Rolling motion is a combination of both translational and rotational motion. Let us consider the rolling motion (without slipping) of a disc on a level surface as shown in below figure. At any instant the point of contact P_0 of the disc with the surface is at rest. The centre of mass of the disc moves with velocity, v_{cm}. The disc rotates with angular velocity, ϖ about its axis which passes through C ; $v_{cm} = R\,\varpi$ where R is the radius of the disc.



The total Kinetic energy (K) of rolling motion is the sum of kinetic energy of translational motion of centre of mass and kinetic energy of rotational motion:

$$K = \frac{1}{2} I\omega^2 + \frac{1}{2} m v_{cm}^2$$

where ϖ is the angular velocity of rotation and I is the moment of inertia about the axis of rotation. For example, if we consider rolling motion of a flat disc. Then I is the moment of inertia about an axis passing through the centre and perpendicular to the plane of the disc. Hence, for disc of radius R and mass M, moment of inertia I:

$$I = MR^2/2$$

Conclusions: The discussion should end with giving some homework exercises;

(a) Three different kinds of motion.
(b) Calculating centre of mass of systems of few particles.
(c) Direction of the torque acting with a given force and axis of force applied.
(d) Some examples of the application of conservation of linear and angular momentum.
(e) Moment of inertia of different bodies about different axes.

Few Questions:

1. A rigid body consisting of N point masses m_i is rotating about an axis passing through the origin with an angular velocity **ω.** Obtain relations connecting the angular momentum of the rigid body with angular velocity.
2. If a rigid body, with on point fixed, rotates with an angular velocity **ω** has an angular momentum L. Show that the kinetic energy is ½ L.**ω**.
3. A rigid body s rotating about an axis through the origin. Obtain relations between the components of total angular momentum and angular velocity.
4. Discuss three different kinds of motion.
5. Calculating centre of mass of few systems of particles.
6. Direction of torque acting with a given force and axis of force applied.
7. Some examples of application of conservation of linear and angular momentum.

References

1. Classical Mechanics: H. Goldstein
2. Introduction Classical Mechanics: David Morin

Module – II
Mechanical Properties of Solids and Fluids

Lectures: 03

Objective:

In this module we will discuss the mechanical properties of solids and fluids. The students will learn different mechanical properties and their uses.

Pretest Questions:

1. In how many states matters exist?
2. What are the differences between solids, liquids and gases?
3. What do you mean by mechanical properties?
4. What are the utility to measure the mechanical properties?
5. Which is more elastic- steel or diamond?
6. Explain why the temperature of a wire under tension will change if it snaps suddenly?
7. Springs are usually made of steel but not of copper. Why?
8. On the basis of moduli of elasticity, distinguish between solid, liquid and gaseous substances.
9. In the case of an elastic body which one is more fundamental –stress or strain?
10. ”Within elastic limit the poison’s ratio depends only on the nature of the material but not on the stress applied”-explain.
11. A steel wire with a greater diameter can withstand a greater load. Why?
12. Can a steel wire be elongated to twice its initial length by hanging a load from its end?

1. Mechanical Properties of Solids

❖ In the general sense, a rigid body is defined as a hard solid object with a definite shape and size. But, from the practical point of view, it is found that rigid bodies can be stretched, compressed and bent when a sufficiently large external force is applied on it. For example, a rigid steel bar can be deformed on the application of suitable force. This means that solid bodies are not perfectly rigid.

➢ **Elasticity**
The property of a body, by which the body opposes any change in its shape and size when the deforming forces are applied and recovers its original state as soon as deforming force is removed, is known as **elasticity** and the deformation caused is known as **elastic deformation**.

➢ **Plasticity**
The property by the virtue of which bodies do not recover its original state as soon as deforming force is removed, is known as **plasticity** and the deformation caused is known as **plastic** deformation.

ELASTIC BEHAVIOUR OF SOLIDS

- Each atom or molecule in a solid is surrounded by atoms or molecules and are bonded together by interatomic or intermolecular forces due to which they stay in a stable equilibrium position.
- Atoms or molecules are displaced from their equilibrium positions on the application of deforming/applied forces.
- On the removal of deforming/applied force interatomic forces drive the atoms or molecules back to their original positions and hence it recovers its original state (i.e. Shape and size). This mechanism is very well visualized by a model of spring-ball system shown in the Fig. 1.1 (Source NCERT book).

If we displace the black ball from its equilibrium position, the restoring force in the spring ball system tries to bring the ball back to its original position. Thus elastic behavior of solid can be visualized via from the microscopic level.

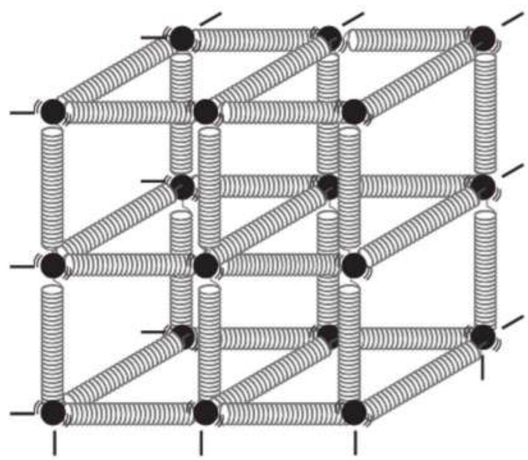


Fig. 1.1 Spring-ball model (Source: NCERT book).

STRESS AND STRAIN

- When we apply a deforming force on the body, a restoring force is developed in the body which is equal in magnitude but opposite in direction to the applied deforming force. This restoring force per unit area is known as stress which is given by

$$\text{Magnitude of the stress} = F/A \tag{1.1}$$

- The SI unit of stress is Nm^{-2} or Pascal (Pa).

There are three ways {longitudinal (a), shear (b & c) and hydraulic stress (d)} in which a solid may change its dimensions when an external force acts on it. These are shown in Fig. 1.2(a, b, c & d) (Source NCERT book).

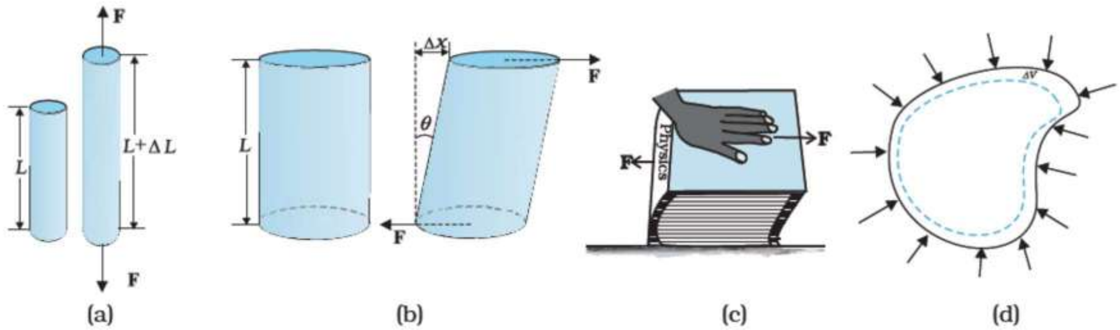


Fig. 1.2 (a) A cylindrical body under tensile stress elongates by ΔL (b) Shearing stress on a cylinder deforming it by an angle θ (c) A body subjected to shearing stress (d) A solid body under a stress normal to the surface at every point (hydraulic stress). The volumetric strain is ΔV/V, but there is no change in shape (Source: NCERT book).

In Fig.1.2 (a), a cylinder is stretched by two equal forces applied normal to its cross-sectional area. The restoring force per unit area in this case is called **tensile stress**. If the cylinder is compressed under the action of applied forces, the restoring force per unit area is known as **compressive stress**. Tensile or compressive stress is also known as longitudinal or normal stress.

The normal tensile or compressive stress is expressed as

$$\sigma_{tensile} = \frac{F_{tensile}}{Cross\ \text{sectional area (A)}}$$
$$\sigma_{compressive} = \frac{F_{compressive}}{Cross\ \text{sectional area (A)}} \tag{1.2}$$

In both the cases, there is a change in the length of the cylinder. The change in the length ΔL to the original length L of the body (cylinder in this case) is known as **longitudinal strain or normal strain ε**.

$$\text{Longitudinal strain, } \boldsymbol{\varepsilon} = \Delta L / L \tag{1.3}$$

- On the application of applied tangential force, there is a relative displacement Δx between opposite faces of the cylinder as shown in the Fig. 1.2(b). The shear stress (τ) acting on the body is defined as

$$\tau = \frac{F_{\text{tangential}}}{A} \tag{1.4}$$

- Shearing strain $\Delta x / L = \tan\theta$ (1.5)

- where θ is the angular displacement of the cylinder from the vertical (original position of the cylinder). Usually θ is very small i.e. $\theta \le 10°$, tan θ is nearly equal to angle θ (See Fig. 2(b & c)).

$$\text{Thus, shearing strain} = \tan\theta \approx \theta. \tag{1.6}$$

- A solid body when placed in the fluid under high pressure is compressed uniformly on all sides as the force applied by the fluid acts in perpendicular direction on all the points on the surface which decreases its volume without any change of its geometrical shape. This develops internal restoring forces that are equal and opposite to the forces applied by the fluid and is known as hydraulic stress.

The **hydraulic/volume strain** and is defined as the ratio of change in volume (ΔV) of the body to the original volume (V).

$$\text{Volume strain} = \Delta V/V \tag{1.7}$$

HOOKE'S LAW

- Hooke, in 1679, showed experimentally that for small deformations stress and strain are proportional to each other.

$$\text{Stress} \propto \text{Strain}$$
$$\text{Stress} = \text{E x strain} \tag{1.8}$$

where E is the proportionality constant and is known as 'modulus of elasticity'.

STRESS-STRAIN CURVE

Typical stress strain curve for a metal is shown in Fig. 1.3. This graph is plotted between the stress (which is equal in magnitude to the applied force per unit area) and the strain produced.

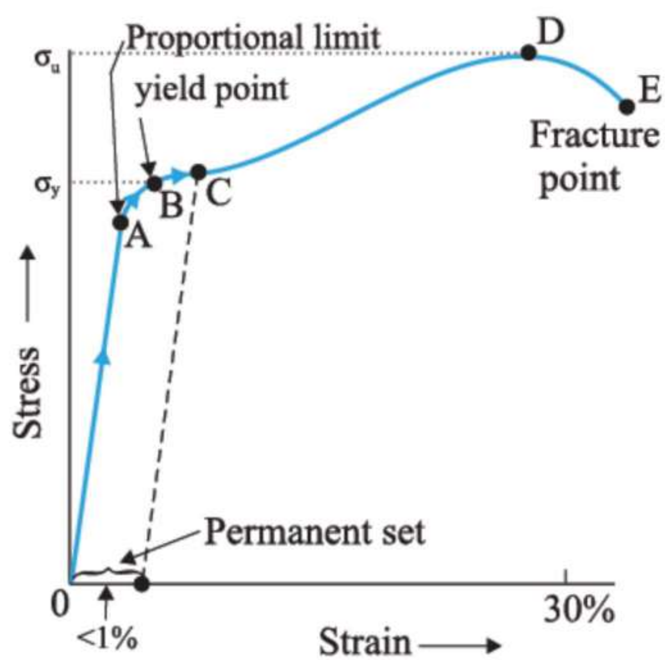


Fig. 1.3 A typical stress-strain curve for a metal (Source: NCERT book).

- The curve is linear in the region between O to A (Hooke's law is obeyed in this region and body behaves as elastic). The relationship between stress and strain in this initial region is not only *linear* but also *proportional.* Beyond point *A*, the proportionality between stress and strain no longer exists; hence the stress at *A* is called the **proportional limit**.

- With an increase in stress beyond the proportional limit, the strain begins to increase more rapidly for each increment in stress. Consequently, the stress-strain curve has a smaller and smaller slope, until, at point *B*, the curve becomes horizontal.

- Point B is the **yield point** (also known as **elastic limit**) and the corresponding stress is **yield strength** (σ_y) of the material.

- Above point B the strain increases rapidly even for a small change in the stress (The portion between B and D). In this region the body does not regain its original dimension and when the stress is made zero, the strain is not zero. Thus the material is said to have a **permanent set** and the deformation is said to be **plastic deformation** (See Fig. 1.3).

- After the point D, additional strain is produced even by a small applied force and fracture occurs at point E (See Fig. 1.3).

- The ratio of stress and strain, in the proportional region within the elastic limit of the stress-strain curve (region OA in Fig. 1.3) is called **modulus of elasticity** and is characteristic of the material.

- It is of great importance to know the elastic limit for applications so that we can avoid the region of plastic deformation which may create problems in designing devices.

Young's Modulus

- The ratio of tensile (or compressive) stress (σ) to the longitudinal strain (ε) is defined as **Young's modulus** and is denoted by the symbol *Y*.

$$Y = \sigma/\varepsilon \tag{1.9}$$

$$Y = (F/A)/(\Delta L/L) = (\mathrm{FL})/(A\Delta L) \tag{1.10}$$

- Dimension of Young's modulus is same as that of stress *i.e.*, Nm^{-2} or Pascal (Pa) as strain is dimensionless quantity.

Table 1.1: Young's moduli, elastic limit and tensile strengths of some materials (Source: NCERT book).

Substance	Young's modulus 10^9 N/m^2 σ_y	Elastic limit 10^7 N/m^2 %	Tensile strength 10^7 N/m^2 σ_u
Aluminium	70	18	20
Copper	120	20	40
Iron (wrought)	190	17	33
Steel	200	30	50
Bone			
(Tensile)	16		12
(Compressive)	9		12

- It is evident from the data of materials given in the Table 1.1 that for metals Young's moduli are large, therefore, they require a large force to produce small change in length.

Question 1: Why steel is more elastic than copper, brass and aluminium?

Solution

Elasticity of a material is also viewed as the resistance offered by the material against the deformation. Higher the resistance offered by the materials against the deformation for a given applied load, higher will be the elasticity of a material. Experimentally, it has been found that to increase the length of a thin steel wire of 0.1 cm^2 cross-sectional area by 0.1%, a force of 2000 N is required. The force required producing the same strain in aluminum, brass and copper wires with same cross-sectional area are 690 N, 900 N and 1100 N respectively. It indicates that steel offers higher resistance as compared to aluminum, brass and copper for the same amount of strain, hence steel is more elastic than copper, brass and aluminum. Due to the same reason steel is preferred in structural designs.

1.7 SHEAR MODULUS

- The ratio of shear stress to shear strain is called the *shear modulus* of the material (G). It is also called the *modulus of rigidity*.

$$G = \text{shearing stress } (\tau)/\text{shearing strain}$$

$$G = (F/A)/(\Delta x/L)$$

$$= (FL)/(A\Delta x) \qquad (1.12)$$

Using Eq. (1.5)

$$G = (F/A)/\theta$$

$$= F/(A\theta) \qquad (1.13)$$

The shearing stress τ can also be expressed as

$$\tau = G\,\theta \qquad (1.14)$$

SI unit of shear modulus is Nm^{-2} or Pa. The shear moduli of a few common materials are given in Table 1.2.

Table 1.2: Shear moduli *(G)* of some common materials (Source: NCERT book).

Material	G (10^9 Nm^{-2} or GPa)
Aluminium	25
Brass	36
Copper	42
Glass	23
Iron	70
Lead	5.6
Nickel	77
Steel	84
Tungsten	150
Wood	10

It is evident from the Table 1.1 and 1.2 that shear modulus (or modulus of rigidity) is in general less than Young's modulus (from Table 1.1 and 1.2).

Bulk Modulus

- The ratio of hydraulic stress to the corresponding hydraulic strain is called *bulk modulus(B)*.

$$B = -p/(\Delta V/V) \quad (1.15)$$

- Negative sign indicates that with an increase in pressure, a decrease in volume occurs.

- SI unit of bulk modulus is the same as that of pressure *i.e.*, Nm^{-2} or Pa.

- The reciprocal of the bulk modulus is called *compressibility* (*k*). It is defined as the fractional change in volume per unit increase in pressure.

$$k = (1/B) = -(1/\Delta p).(\Delta V/V) \quad (1.16)$$

Table 1.3: The bulk moduli of a few common materials (Source: NCERT book)

Material **Solids**	B (10^9 N m^{-2} or GPa)
Aluminium	72
Brass	61
Copper	140
Glass	37
Iron	100
Nickel	260
Steel	160
Liquids	
Water	2.2
Ethanol	0.9
Carbon disulphide	1.56
Glycerine	4.76
Mercury	25
Gases	
Air (at STP)	1.0×10^{-4}

Question 2: It can be seen from the data given in Table 1.3 that the bulk moduli for solids are much larger than for liquids, which are again much larger than the bulk modulus for gases (air)?

Solution

The bulk modulus of a material is inversely proportional to the compressibility. Since the interatomic bonds between the neighboring atoms in solid are strongest as compared to the liquid and gases, the solids are least compressible. The interatomic bond in the gases are weakest, therefore they are the most compressible substance. Gases are about a million times more compressible than solids. Hence, the bulk modulus of solids is much larger than liquids and gases.

Table 1.4: Summary of stress, strain and various elastic moduli (Source: NCERT book)

Type of stress	Stress	Strain	Change in shape	Change in volume	Elastic modulus	Name of modulus	State of Mater
Tensile or compressive	Two equal and opposite forces perpendicular to opposite faces ($\sigma = F/A$)	Elongation or compression parallel to force direction ($\Delta L/L$) (longitudinal strain)	Yes	No	$Y = (FL)/(A\,\Delta L)$	Young's modulus	Solid
Shearing	Two equal and opposite forces parallel to oppoiste surfaces [forces in each case such that total force and total torque on the body vanishes ($\sigma_s = F/A$)	Pure shear, θ	Yes	No	$G = (F\theta)/A$	Shear modulus	Solid
Hydraulic	Forces perpendicular everywhere to the surface, force per unit area (pressure) same everywhere.	Volume change (compression or elongation ($\Delta V/V$)	No	Yes	$B = -p/(\Delta V/V)$	Bulk modulus	Solid, liquid and gas

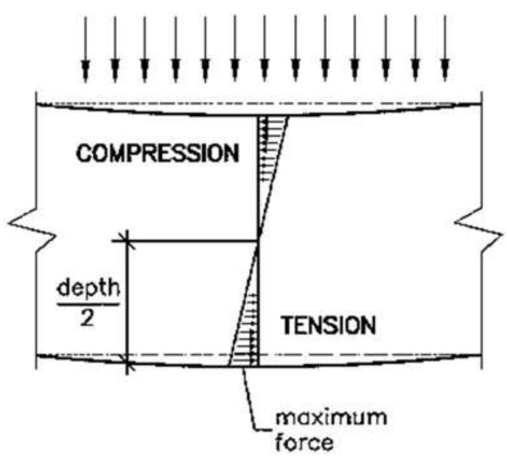
1.9 APPLICATIONS OF ELASTIC BEHAVIOUR OF MATERIALS

The elastic behavior of materials plays an important role in designing a building, the structural design of the columns, beams and supports. One needs to have the knowledge of the elastic behaviour of materials, so that we do not exceed the limit which can bring trouble.

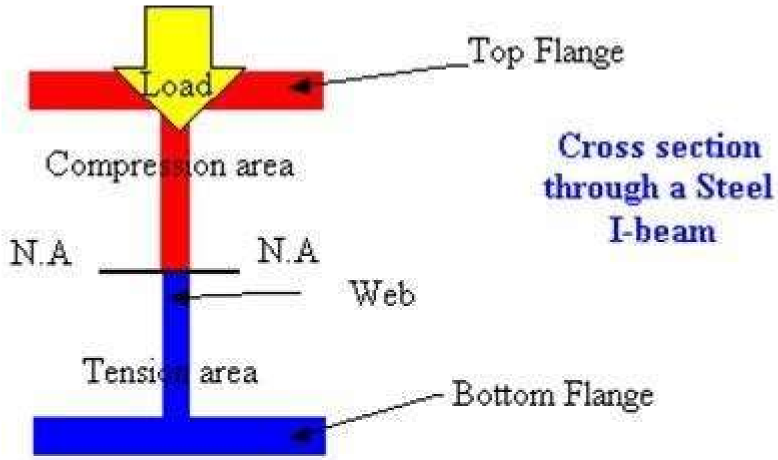
Question 3: Have you ever thought why the beams used in construction of bridges, as supports etc. have a cross-section of the type I?

Solutions

When a beam bends the top of the beam is in compression and the bottom is in tension.



These forces are greatest at the very top and very bottom. So to make the stiffest beam with the least amount of material you would want the material to be only at the top and bottom sides. However you still need to connect them together or they would just be two separate plates and would not be stiff at all. So you put a web in the middle to connect them and make them work together. The resulting shape is the traditional "I-beam" or wide flange beam.



N.A is a neutral surface area, where the stresses are zero.

Some solved examples:

Question 1: Cranes used for lifting and moving heavy loads from one place to another have a thick metal rope to which the load is attached. The rope is pulled up using pulleys and motors. Suppose we want to make a crane, which has a lifting capacity of 40 tonnes or metric tons (1 metric ton = 1000 kg). How thick should the steel rope be?

Solutions

We obviously want that the load does not deform the rope permanently. Therefore, the extension should not exceed the elastic limit. From Table 1.1, we find that mild steel has yield strength (S_y) of about 300×10^6 N m^{-2}. Thus, the area of cross-section (A) of the rope should at least be

$$A \geq W/Sy = Mg/Sy \tag{1.17}$$
$$= (4 \times 10^4 \text{kg} \times 10 \text{ms}^{-2})/(300 \times 10^6 \text{N m}^{-2})$$
$$= 13.3 \times 10^{-4} \text{ m}^2$$

corresponding to a radius of about 2.06 cm for a rope of circular cross-section. A single wire of this radius would practically be a rigid rod. So the ropes are always made of a number of thin wires braided together, like in pigtails, for ease in manufacture, flexibility and strength.

Question 2: Steel rod of length 2.0 m and radius 15 mm has been stretched along its length by a force of 200 kN. Calculate (a) stress, (b) elongation and (c) strain on the rod. The Young's modulus of steel is 2.0×10^{11} Nm^{-2}.

Solution

(a) Stress = F/A = 200×10^3 N/$3.14 \times (15 \times 10^{-3}\text{m})^2 = 2.83 \times 10^8$ N m^{-2}.

(b) Y = stress/strain= (F/A)/Δl/L

Now Δl (elongation) = (F/A) (L/Y) = [(200×10^3N)x2m] / [$3.14 \times (15 \times 10^{-3})^2 \times 2 \times 10^{11}$]

2.83mm.

(c) Strain = Δl/L = $2.83/2 \times 10^3$ = 1.415.

Question 3: A steel wire of length 6 m and cross-section 5.0×10^{-5} m^2 stretches by same amount as a copper wire of length 4 m and cross-section 6.0×10^{-5} m^2 under a given load. What is the ratio of the Young's modulus modulus of steel to that of the copper?

Solution

Y = stress/strain= $(Mg/A)/\Delta l/L = Mg\ L/A\Delta l$
Now let Ys and Yc be the Young's moduli, Ls and Lc the original length's and As and Ac the cross-sectional area of the steel and copper wires respectively. Since, the load Mg and the stretching Δl are same for the two wires respectively. Since, the load Mg and the stretching Δl are same for two wires, we have from the above equation

Ys/Yc=Ls/As x Ac/Lc = 6 m x ($6.0 \times 10^{-5} m^2$) / (5.0×10^{-5} m^2) x 4 m = 1.8.

Question 4: Given data:
Initial volume = 200.5 litre, Increase in pressure *p* = 200.0 atm (1 atm = 1.013×10^5 Pa), final volume = 200.0 litre. Compare the bulk modulus of water with that of air (at constant temperature). Explain in simple terms why the ratio is so large.

Solution
B = volume stress/volume strain = - change in pressure/volume strain = $-p/\Delta v/V$
Here, for water:
p = 200.0 atm = 200.0 x (1.013×10^5 Pa) = 2.026×10^7 Pa, V = 200.5 litre
and ΔV = 200.0 litre – 200.5 litre = -0.5 litre.
$B_{water} = -2.026 \times 10^7$ Pa/-0.5 litre / 200.5 litre = 8.12×10^9 Pa.
The bulk modulus of air at STP is 1.0×10^{-4} Pa.
$B_{water}/B_{air} = 8.12 \times 10^9/1.0 \times 10^{-4} = 8.12 \times 10^{13}$.

Question 5: Calculate the pressure required to stop the increase in volume of a copper block when it is heated from 60^0C to 90^0C. Coefficient of linear expansion of copper is $\alpha = 8.0 \times 10^{-6}$ per^{0}C and bulk modulus of elasticity is 1.3×10^{11} Nm^{-2}.

Solution: Here, we require coefficient of volume expansion of copper (γ) = 3 α (Coefficient of linear expansion of copper.
Now, increase in volume is related to starting volume with coefficient of linear expansion by the relation by relation $\Delta V = V \times \gamma \times (\Delta t = t_2 - t_1)$
Volume strain is given by $\Delta V/V = \gamma \times (t_2 - t_1)$
Bulk modulus is
B = - change in pressure(p)/volume strain = $p/\gamma((t_2 - t_1)$
Which gives $p = -B\ \gamma\ ((t_2 - t_1)$
On substituting the given values, we have
$P = -(1.3 \times 10^{11})\ x3x\ 8.0 \times 10^{-6}\ (90-60)$
$p = 9.36 \times 10^7$ N m^{-2}.

Question 6: A piece of copper having rectangular cross-section of 13 mm x 18 mm is pulled in tension with 675000 N force, producing only elastic deformation. Calculate the resulting strain. Modulus of rigidity of copper = 4.20×10^{10}Pa.

Solution: The modulus of rigidity of the material of the body is given by
η= shearing stress/ shearing strain = $(F/A)/\theta$, where F is the tangential force applied and A the area of cross section of body. Thus shearing strain $\theta = F/A\eta$
On substituting the values we get the shearing strain
$\theta = 675000\ N/(2.34 \times 10^{-4} m^2)(4.20 \times 10^{10} N/m^2)$
$\theta = 0.06868$ radians
$\theta = 3.93^0$

2. MECHANICAL PROPERTIES OF FLUIDS

What is Fluid?
Answers
Specifically, a fluid is defined as a substance that deforms continuously when acted on by a shearing stress of any magnitude. A shearing stress (force per unit area) is created whenever a tangential force acts on a surface as shown in Fig.1.

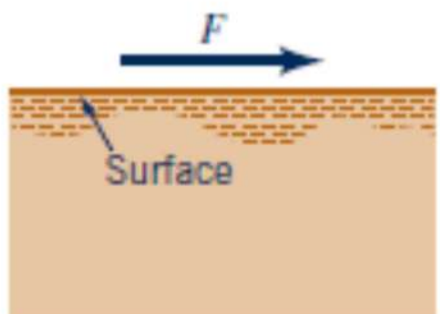


Figure1: Action of shearing force on the surface of liquid
Reference: Munson B.R "Fundamental of Fluid Mechanics" 7th edition, John Wiley & Sons, 2013.

Any shear stress applied to a fluid, no matter how small, will result in motion of that fluid. The fluid moves and deforms continuously as long as the shear stress is applied.
Given the definition of a fluid above, there are two classes of fluids, *liquids* and *gases.*

How are fluids different from solids?
We have a general, vague idea of the difference. A solid is "hard" and not easily deformed, whereas a fluid is "soft" and is easily deformed. Although quite descriptive, these casual observations of the differences between solids and fluids are not very satisfactory from a scientific or engineering point of view. A closer look at the molecular structure of materials reveals that matter that we commonly think of as a solid (steel, concrete, etc.) has densely spaced molecules with large intermolecular cohesive forces that allow the solid to maintain its shape, and to not be easily deformed. However, for matter that we normally think of as a liquid (water, oil, etc.), the molecules are spaced farther apart, the intermolecular forces are smaller than for solids, and the molecules have more freedom of movement. Thus, liquids can be easily deformed (but not easily compressed) and can be poured into containers or forced through a tube. Gases (air, oxygen, etc.) have even greater molecular spacing and freedom of motion with negligible cohesive intermolecular forces, and as a

consequence are easily deformed (and compressed) and will completely fill the volume of any container in which they are placed. Both liquids and gases are fluids. .

Based on the above discussions, a *solid can resist a shear stress by a static deformation; a fluid cannot*. Any shear stress applied to a fluid, no matter how small, will result in motion of that fluid. The fluid moves and deforms continuously as long as the shear stress is applied.

Zero compressibility: An ideal liquid is incompressible, that is on pressing the liquid there is no change in its volume (or density). Most of the liquids may be considered approximately incompressible because on pressing them the change in their volume is negligible.

Fluids offer very little resistance to shear stress; hence there are enormous changes in shape of fluids on application of very small shear stress. A small shearing stress can cause change in shape of fluid many times more than that of a solid.

Zero viscosity: An ideal liquid is non-viscous, that is when there is relative motion between different layers of the liquid then there is no tangential frictional force in between the layers. In practice there is some viscosity in all liquids (and gases). It is less in gases and larger in liquids.
Fluids in common life

PRESSURE

Question: Why a sharp needle when pressed against our skin pierces it and our skin, remains intact when a blunt object (say the back of a spoon) is pressed against it with the same force?

Solution

Actually the area on which the force acts on skin is the major reason for piercing. Smaller the area on which the force acts, greater is the impact. This concept is known as pressure.

- If F is the magnitude the normal force on the area A then the pressure P is defined as the normal force acting per unit area.

$$\mathbf{P} = \mathbf{F}/A \qquad (1.1)$$

- Pressure is the force acting normal to the area under consideration and is a scalar quantity.

The SI unit of pressure is Nm^{-2} or pascal (Pa). Another common unit of pressure in which it is usually expressed is atmosphere (atm), i.e. the pressure exerted by the atmosphere at sea level (1 atm = 1.013×10^5 Pa).

From equation (1.1), we can see that pressure is inversely proportional to the area. Smaller the area, higher will be the pressure acting on the surface for the same applied force. A sharp needle possesses smaller contact area as compared to the any blunt object; therefore, it pierces more in our skin, when applied with the same force.

RELATIVE DENSITY

- The density of water at 4°C (277 K) is 1.0×10^3 kg m^{-3}. The relative density of a substance is the ratio of its density to the density of water at 4°C.

$$\text{Relative density} = \frac{\text{Density of a substance}}{\text{Density of water (at } 4^0C)} \qquad (1.2)$$

- Relative density is a dimensionless positive scalar quantity. For example, the relative density of mercury is 13.6. Its density is 13.6×10^3 kg m^{-3}.

The densities of some common fluids are displayed in Table 1.1.

Table 1. Densities of some common fluids at STP (STP means standard temperature (0^0C) and 1 atm pressure (Source: NCERT Book).

Fluid	ρ (kg m^{-3})
Water	1.00×10^3
Sea water	1.03×10^3
Mercury	13.6×10^3
Ethyl alcohol	0.806×10^3
Whole blood	1.06×10^3
Air	1.29
Oxygen	1.43
Hydrogen	9.0×10^{-2}
Interstellar space	$\approx 10^{-20}$

Pascal's Law

Pascal's law states that the pressure in a fluid at rest is the same at all points if they are at the same height.

Variation of Pressure with Depth

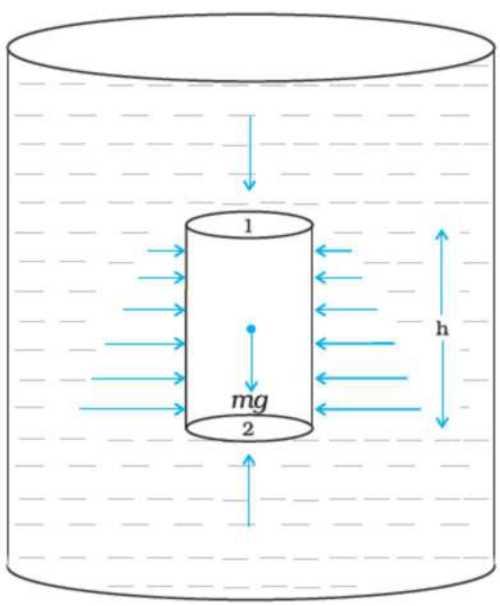


Figure 2: Fluid under gravity. The effect of gravity is illustrated through pressure on a vertical cylindrical column (Source: NCERT Book).

$$P_2 - P_1 = \rho g h \qquad (1.3)$$

Pressure difference depends between the two points in a fluid depends on the following factors:

1. Vertical distance h between the points (1 and 2).
2. Density of the fluid ρ.
3. Acceleration due to gravity g.

If we want to find the pressure at a depth (h) from the surface of the water

$$P = P_a(\text{atmospheric pressure}) + \rho g h \qquad (1.4)$$

The pressure P, at any depth h vertically below the surface of a liquid at atmospheric pressure is greater than atmospheric pressure by an amount ρgh. The excess of pressure, P-Pa, at depth h is is known as gauge pressure at that point.

Q4) Consider three vessels A, B and C as shown in Fig.3 of different shapes. What is the pressure at the bottom of A, B and C?

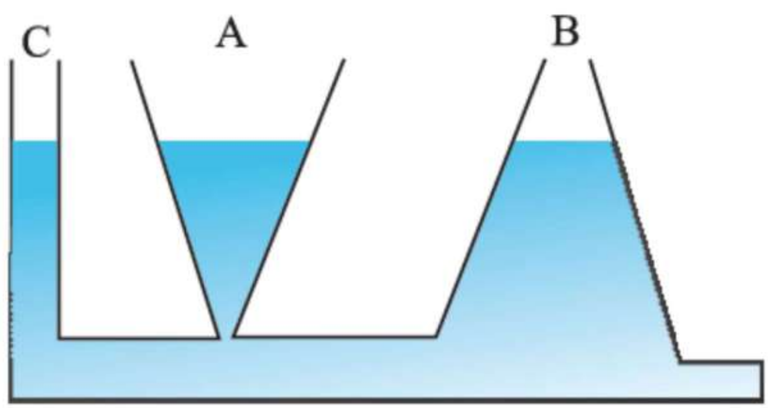


Figure 3: Illustration of hydrostatic paradox. The three vessels A, B and C contain different amounts of liquids, all up to the same height will have same pressure at the base (Source: NCERT Book).

Solution:

We can see from the expression $P = P_a$ (atmospheric pressure) + ρgh, that the area is not coming into consideration which implies that the cross sectional or base area is not important factor in tuning the pressure in this case. Secondly, height of the liquid column is deciding factor at the bottom where the other things are same. Thus the pressure at the base is same in all the three vessels A, B and C.

Hydraulic Machines

Pascal's law for transmission of fluid states that whenever we apply a pressure on any part of a fluid contained in a vessel, it is transmitted undiminished and equally in all directions.

- e.g. Hydraulic lift and hydraulic brakes.

By changing the force at A_1 and hence the pressure $P = F_1/A_1$, the platform of the other cylinder with area A_2 can be moved up or down with a larger force $F_2 = PA_2 = F_1A_2/A_1$. Thus, we find that the applied force has been increased by a factor of A_2/A_1 which is called the mechanical advantage of the device (See Fig. 4)

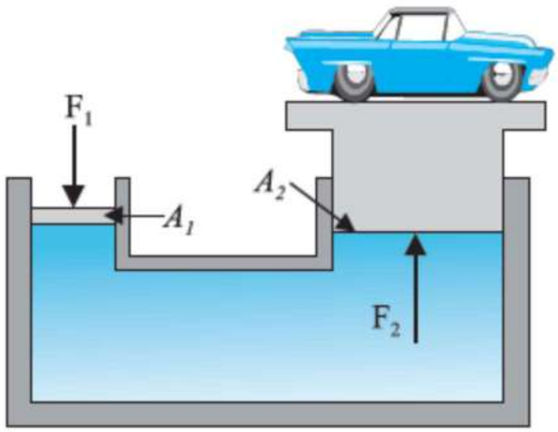


Figure 4: Schematic diagram illustrating the principle behind the hydraulic lift, a device used to lift heavy loads (Source: NCERT Book).

STREAMLINE FLOW

- If the velocity of each particle passing through a point remains constant with time, the flow of the fluid is said to be steady at that point.
- When the particular particle moves from one point to another it can change as its velocity, but every other particle which passes the second point behaves exactly in the same way as the previous particle which has just passed that same point.
- The paths of the particles do not cross each other.
- The path taken by a fluid particle under a steady flow is a streamline.
  Property of streamline flow
- Two streamlines do not cross each other. If the map of flow is stationary in time, it is known to be steady.

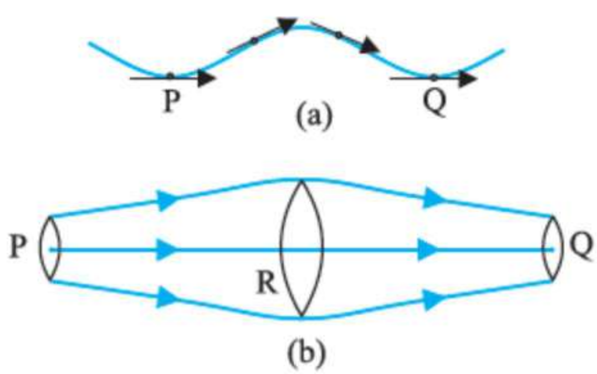


Figure 5: The meaning of streamlines. (a) A typical trajectory of a fluid particle. (b) A region of streamline flow (Source: NCERT Book).

- Steady flow is achieved at low flow speeds.
- Beyond a limiting value, called critical speed, this flow loses steadiness and becomes turbulent.

BERNOULLI'S PRINCIPLE

It states that when an incompressible and non-viscous liquid flows in a streamlined motion from one point to another, then at every point of its path total energy per unit volume (kinetic energy + gravitational potential energy) summed up to pressure remains a constant.

$$P + 1/2\rho\, v^2 + \rho g h = \text{constant} \tag{1.5}$$

APPLICATIONS

When the air flows with high speed through the nozzle of the carburetor of an automobile, the pressure is lowered at the narrow neck and the petrol (gasoline) is sucked up in the chamber as per Bernoulli's theorem to provide the correct mixture of air to fuel necessary for combustion (See Fig. 6).

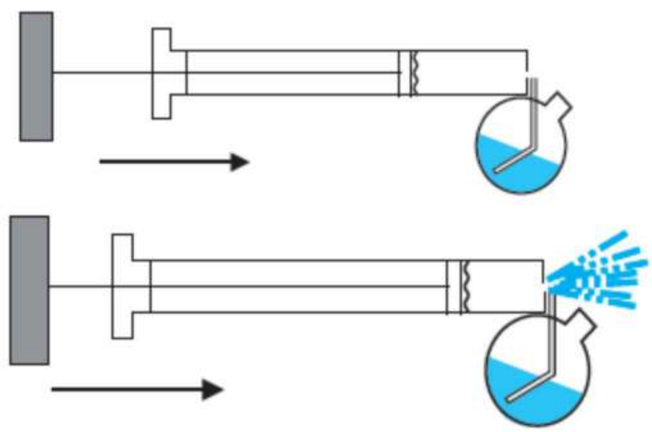


Figure 6: The spray gun. Piston forces air at high speeds causing a lowering of pressure at the neck of the container (Source: NCERT Book).

Bunsen burner, atomisers and sprayers used for perfumes or to spray insecticides are few examples which work on the same principle

VISCOSITY

- Practically most of the fluids in general are not ideal and offer some resistance to motion between the layers of liquid. This resistance to the motion of fluid between the layers is like an internal friction analogous to friction when a solid move on a surface. This phenomenon is called viscosity.

Question: Why when a fluid flows in a pipe or a tube, then velocity of the liquid layer along the axis of the tube is maximum and decreases gradually as we move towards the walls where it becomes zero.

Solution:

This is due to the internal forces offered by layers of the fluid and the fluid and solid at the end of the pipe. Figure 7 shows a velocity profile of a fluid flow in a pipe. It clearly indicates that flow velocity is maximum at the centerline and minimum at the pipe walls.

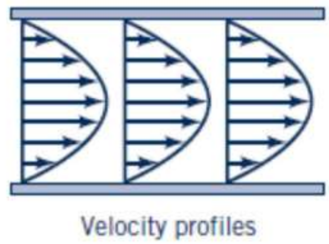


Figure 7: Velocity profile of a fluid through a pipe.
Reference: Munson B.R "Fundamental of Fluid Mechanics" 7th edition, John Wiley & Sons, 2013.

- In general thin liquids like water, alcohol etc. are less viscous as they offer less resistance between the layers than the thick liquids like coal tar, blood, glycerin etc.
- The viscosity of liquids decreases with temperature while it increases in the case of gases.

SURFACE TENSION

- Surface tension is defined as force per unit length or surface energy per unit area acting on the interface between the liquid and the bounding surface.
- Surface tension arises due to excess potential energy of the molecules on the surface (i.e. interface of two substances) as they are not surrounded on all sides by molecules in comparison to the molecules in the interior where the intermolecular distances are such that it is attracted to all the surrounding molecules resulting in a negative potential energy.
- Surface energy is present at the interface separating two substances of which at least one of the substance is a fluid.

REYNOLDS NUMBER

- When the fluid flows at larger rate, the flow becomes turbulent (velocity of the fluids at any point in space varies rapidly and randomly with time) and no longer remains laminar (the velocities at different points in the fluid may have different magnitudes but their directions are parallel).
- Osborne Reynolds defined a dimensionless number, whose value gives us an idea whether the flow would be turbulent or stream line. This number is called the Reynolds number expressed as

$$Re = \rho vd/\eta \tag{1.6}$$

Here ρ is the density of the fluid flowing with a speed v, d stands for the dimension of the pipe, and h is the viscosity of the fluid.

- Re is a dimensionless number.
- $Re < 1000$ fluid flow is streamline or laminar.

- $Re > 2000$ fluid flow is turbulent.
- $1000 < Re < 2000$ fluid flow becomes unsteady.

Re can also be expressed as $Re = \rho v^2/\eta v/d = \rho A v^2 /\eta A v/d$ (1.7)

$= \text{inertial force/force of viscosity.}$

- Thus Re represents the ratio of inertial force (force due to inertia) to viscous force.

Angle of Contact

Contact angle θ appears when a liquid interface intersects with a solid surface, as in Figure 8. The surface of liquid in contact with another medium is in general curved at the plane of contact.

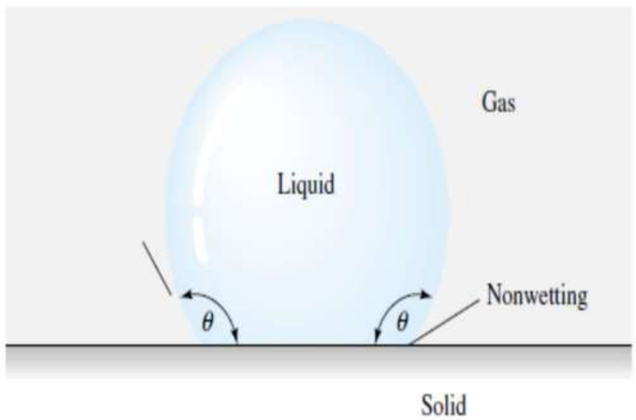


Figure 8: Contact-angle effects at liquid-gas-solid interface
Reference: White F.M. "Fluid Mechanics", 4th edition, Mc Graw Hill, 2004.

If the contact angle is less than 90°, the liquid is said to *wet* the solid; if $\theta > 90°$, the liquid is termed *nonwetting*. For example, water wets soap but does not wet wax. Water is extremely wetting to a clean glass surface, with $\theta = 0°$.

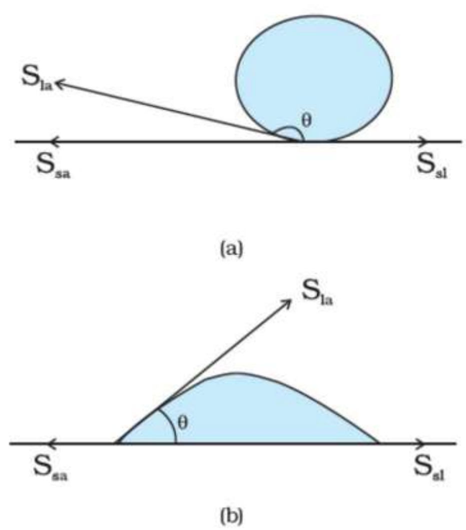


Figure 9: Different shapes of water drops with interfacial tensions (a) on a lotus leaf (b) on a clean plastic plate (Source: NCERT Book).

- You can get the idea whether a liquid will spread on the surface of a solid or it will form droplets on it by looking θ. For example, by observing θ you can tell that the water forms droplets on leaf of lotus as shown in Fig. 9 (a) while spreads over a clean plastic plate as shown in Fig. 9(b).

Let us consider three interfacial tensions denoted by S_{la}, S_{sa} & S_{sl} respectively at the three interfaces, liquid-air, solid-air and solid-liquid shown in Fig. 9 (a) and (b). Surface forces between the three interfaces must be in equilibrium and so from the Fig. 9(b) we have

$$S_{la} \cos\theta + S_{sl} = S_{sa} \tag{1.8}$$

- The angle of contact will be obtuse if $S_{sl} > S_{la}$ which is the case for water-leaf interface while it is acute if $S_{sl} < S_{la}$ and is the case for water-plastic interface.
- When θ an obtuse angle then molecules of the liquid is are attracted strongly to each other and weakly to those of solid. It costs a lot of energy to create a liquid-solid surface, and so the liquid does not wet the solid e.g. water on a waxy or oily surface.
- When θ is an acute angle then molecules of the liquid are strongly attracted to those of the solid e.g. water spreads on glass or on plastic.

APPLICATIONS

- Soaps, detergents and dying substances are wetting agents as they make the angle of contact become small, so that when they are added they may penetrate well and work effectively.
- Water proofing agents on the other hand are added in the materials so that a large angle of contact forms between the water and fibres.

Drops and Bubbles

- Free liquid drops and bubbles are spherical is a result of surface tension if the effects of gravity can be neglected.
  E.g. high-speed spray or jet gives drops, soap bubbles blown by children's.
  Why are drops and bubbles spherical? What keeps soap bubbles stable?
- Liquid air interface has energy, and we know that for a given volume minimum energy is the one with the least surface area. The sphere has this property.
  Check which one has least surface area with same volume? Sphere or a Cube?
- If gravity and other forces (e.g. air resistance) can be neglected, liquid drops would be spherical.
- Due to surface tension pressure inside (P_i) a spherical drop Fig. 10(a) is more than the pressure outside (P_o).
- Consider a spherical drop of radius r is in equilibrium and its radius is increased by Δr. Then the extra surface energy is

$$[4\pi(r + \Delta r)^2 - 4\pi r^2]S_{la} = 8\pi r \Delta r S_{la} \tag{1.9}$$

The energy cost of increment in surface area is balanced by the gain in energy via expansion due to the pressure difference (Pi – Po) between the two sides of the bubble i.e. inside and outside. The work done is given by

$$W = (P_i - P_o)\, 4\pi r^2 \Delta r \tag{1.10}$$

So from equation 1.9 and 1.10 we have,

$$(P\text{i} - P\text{o}) = (2\ S_{\text{la}}/r) \quad (1.11)$$

- In general, for a liquid-gas interface, the convex side (Inside the drop, a cavity in a fluid or a bubble) has a higher pressure than the concave side. For example, an air bubble in a liquid would have higher pressure inside it than outside. See Fig 10 (b).

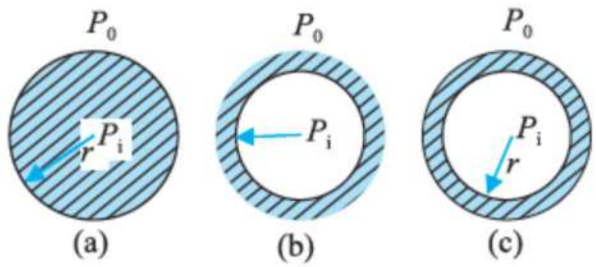


Figure 10: Drop, cavity and bubble of radius r (Source: NCERT Book).

A bubble Fig 10 (c) differs from a drop and a cavity; as in this case it has two interfaces and so the overall surface area multiplies by 2.

So for a bubble $(P\text{i} - P\text{o}) = (4\ S_{\text{la}}/\ r)$ (1.12)

This is the reason we have to blow hard to form a bubble to create little extra air pressure inside.

Capillary Rise

In spite of gravity the water level modifies rises (or falls) up in a narrow tube due to the pressure difference across a curved liquid-air interface.

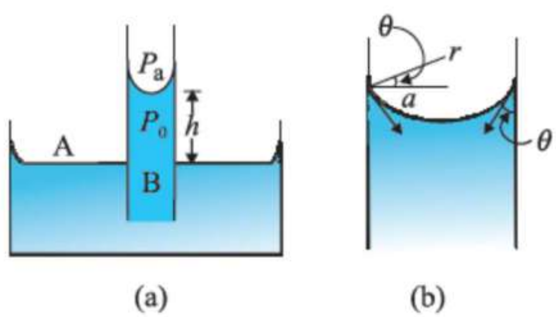


Figure 11: Capillary rise, (a) Schematic picture of a narrow tube immersed water. (b) Enlarged picture near interface (Source: NCERT Book).

- The contact angle between water and glass is acute which gives a concave surface of water in the capillary.
- There is a pressure difference between the two sides of the top concave surface of water at the meniscus (air-water interface).

This is given by $2S/r = 2S/(a\sec\theta)$ (here r and a are radius of meniscus and capillary respectively

$= (2S/a)\cos\theta)$ (1.13)

From the figure 11 it is clear that the pressure just below the plane surface of water outside the tube is also P (at point A) but below the meniscus inside the tube (at point B) is *P-2S/r*. Since we know that the pressure at all points in the same level of water must be same so the points A and B just below the surface of water should be at same pressure. In order to make up the deficiency of pressure, *2S/r* below the meniscus, water from outside capillary begins to flow into the capillary and stops at certain height h when it becomes equal to 2S/*r*. This gives

$$h \rho g = (2S \cos\theta)/a \tag{1.14}$$
$$h = (2S \cos\theta)/a \rho g$$

Thus we see that the capillary rise is due to surface tension and is larger, for a smaller radius of capillary (a).

Question 1: A hydraulic automobile lift is designed to life cars with a maximum mass of 6,000 kg. The area of cross-section of the piston carrying the load is 500 cm^2. What maximum pressure would the smaller piston have to bear? Take g = 9.8 m/s^2**.**

Solutions:
Maximum mass applied on the bigger piston having larger area (i.e. 500 cm^2) is 6,000 kg. Then the pressure on this piston is

$$P = F/A = (6{,}000 \times 9.8)\ N / 500 \times 10^{-4} m^2 = 11.76 \times 10^5\ N\ m^{-2}.$$

Now from Pascal's law pressure is transmitted unchanged through the liquid of the hydraulic to the piston of smaller cross section, hence the smaller piston would bear a pressure of 11.76×10^5 N m^{-2}.

Question 2: Find at a depth of 2,000 m in an ocean (a) absolute pressure, (b) gauge pressure, (c) force acting on on a window of area 600 cm^2 of a submarine whose interior is maintained at sea-level atmospheric pressure. Given: atmospheric pressure = 1.01×10^5 Pa, density of sea water = 1.03×10^3 kg m^{-3} and g=10m/s^2.

Solutions:
(a) Absolute pressure at a depth h is

$$\begin{aligned} P_h &= P + h \rho g \\ &= (1.01 \times 10^5\ N\ m^{-2}) + \{2000 \times (1.03 \times 10^3\ kg\ m^{-3}) \times 10\ N\ kg^{-1}\} \\ &= \{(1.01 \times 10^5) + (2.06 \times 10^7)\}\ N\ m^{-2} \\ &= \{(1.01 \times 10^5) + (206 \times 10^5)\}\ N\ m^{-2} \\ &= 207.01 \times 10^5\ Pa \approx 207.01\ atm. \end{aligned}$$

(b) The gauge pressure is

$$P_h - P = h \rho g = 206 \times 10^5\ Pa = 206\ atm.$$

(c) The pressure outside the submarine is P + h ρ g and that inside is P 9 given). Hence net pressure acting on the window is gauge pressure i.e. h ρ g (206×10^5 N m^{-2}). The area of the window is 600cm^2 = 0.06 m^2. Hence the force acting on the window is

$$\begin{aligned} F &= \text{pressure x area} = (206 \times 10^5\ N\ m^{-2}) \times 0.06\ m^2 \\ &= 12.36 \times 10^5\ N. \end{aligned}$$

Question 3: A soap film is on a rectangular wire ring of size 6 cm x 6 cm. If the size of the film is changed to 6 cm x 8 cm, then calculate the work done in this process. The surface tension of soap solution is 3.0×10^{-2} N m^{-1}.

Solutions:

Initial surface-area of the of the film = 36 cm^2
Final surface-area of the film = 48 cm^2
Increase in surface-area = $(48 \times 10^{-4}) - (36 \times 10^{-4}) = 12 \times 10^{-4}\ m^2$

The film has two surfaces. Hence net increase in surface-area of the film is

$\Delta A = 2 \times 12 \times 10^{-4}\ m^2$
Now, work done = surface tension x increase in area
$= 3.0 \times 10^{-2}\ N\ m^{-1} \times 24 \times 10^{-4}\ m^2$
$= 7.2 \times 10^{-5}$ J

Question 4: The surface tension of a soap solution is 0.030 N m^{-1}. How much work is required to force a bubble of 2.0 cm radius from this solution?

Solutions

Work done in blowing a bubble is stored in the form of energy in the surface of the bubble (inside and outside). The area of the 2 surfaces is

$A = 2 \times 4\pi r^2 = 8\pi r^2$, here r is radius of the bubble.
$r = 2.0$ cm $= 2.0 \times 10^{-2}$ m
$A = 8 \times 3.14 \times (2.0 \times 10^{-2})^2 = 10.048 \times 10^{-3}\ m^2$

Now, work done = energy of the extended area = surface tension x area
$= 0.030\ N\ m^{-1} \times 10.048 \times 10^{-3}\ m^2$
$= 30.144 \times 10^{-5}$ J.

Question 5: A drop of mercury has a radius of 4.00 mm at room temperature. The surface tension of mercury at that temperature is $4.65 \times 10^{-1}\ N\ m^{-1}$. Find excess pressure inside the drop and the total pressure inside the drop. The atmospheric pressure is $1.01 \times 10^5\ Nm^{-2}$ (or Pa).

Solutions.

Excess pressure inside the drop is given by

$p = 2T/R = 2(4.65 \times 10^{-1})/(4.00 \times 10^{-3}) = 232.5$ Pa
Atmospheric pressure $P = 1.01 \times 10^5$ Pa.
Total pressure inside the drop = $P + p = (1.01 \times 10^5)$ Pa + (0.002325×10^5) Pa
$= 1.012325 \times 10^5$ Pa.

Question 6: (i)What is the excess pressure inside a soap bubble of radius 5.00 mm at room temperature (20^0C)? The surface tension of soap solution at 20^0C is 2.50 x 10-2 Nm^{-1}, (ii) If an air bubble of the same radius were formed at a depth of 40.0 cm inside a vessel containing soap solution of relative density 1.20, then what would be the pressure inside the air bubble?

Solutions

(i) Excess pressure inside a soap bubble

$P = 4T/R = 4(2.50 \times 10^{-2}\ N\ m^{-1})/(5.00 \times 10^{-3} m) = 20.0\ Nm^{-2} = 20.0$ Pa.

(ii) For an air bubble of same radius inside the soap solution, the excess pressure inside the bubble is

$P = 2T/R = 10.0$ Pa.

If P be the atmospheric pressure, then the pressure outside the air bubble at a depth h in a soap solution of density ($\rho = 1.20 \times 10^3$) kg m^{-3} is

$$P' = P + h \rho g$$
$$= (1.01 \times 10^5) \text{ Pa} + (40.0 \times 10^{-2} \text{m})(1.20 \times 10^3 \text{ kgm}^{-3})(9.8 \text{ N kg}^{-1})$$
$$= (1.01 \times 10^5) \text{ Pa} + (0.047 \times 10^5) \text{ Pa}$$
$$= 1.057 \times 10^5 \text{ Pa}$$

Total pressure inside the soap bubble is

$$= P' + p$$
$$= (1.057 \times 10^5) \text{ Pa} + 10.0 \text{ Pa.}$$
$$= 1.057 \times 10^5 \text{ Pa}$$
$$\approx 1.06 \times 10^5 \text{ Pa}$$

Question 7: Mercury has an angle of contact equal to 140^0 with glass. A narrow tube of glass of radius 2.0 mm is dipped in a trough of mercury. By what amount does the mercury dip down in the tube relative to the liquid surface outside? Surface tension of mercury at the temperature of the experiment is 0.465 N m^{-1}. Density of mercury = 13.6×10^3 kgm^{-3}.

Solutions

Height of liquid column of surface tension T and density ρ in a glass capillary tube of radius r is given by

$$h = 2T \cos\theta / r \rho g \text{ , where } \theta \text{ is the angle of contact of liquid-glass.}$$

Substituting the values we get

$$h = \{2 \times 0.465 \times (- 0.766)\}/\{(2.0 \times 10^{-3}) \times (13.6 \times 10^3) \times 9.8\}$$
$$= -2.67 \times 10^{-3} \text{ m} = -2.67 \text{ mm}$$

The negative sign indicates that the mercury is depressed in the glass capillary tube.

Question 8: Water flows steadily through a horizontal pipe of varying diameter. The pressure of water is 2.0 cm of mercury column at a point where velocity of flow is 0.50 m s^{-1}. Find the pressure at another point where the velocity of flow is 0.75 m s^{-1}. The densities of mercury and water are 13.6×10^3 and 10^3 kgm^{-3} respectively. Take g = 9.8 N kg^{-1}.

Solutions

Using Bernoulli's theorem: $P_1 + 1/2\rho v_1^2 + \rho g h_1 = P_2 + 1/2\rho v_2^2 + \rho g h_2$ for a horizontal flow of water, we have

$$P_1 + 1/2\rho v_1^2 = P_2 + 1/2\rho v_2^2.$$

Here P_1 = 2.0 cm of Hg = 0.02 m of Hg = $0.02 \times (13.6 \times 10^3) \times 9.8$ N m^{-2}, $v_1 = 0.50$ m s^{-1}, $v_2 = 0.75$ m s^{-1} and $\rho = 10^3$ kg m^{-3}.

$$\text{Now, } P_2 = P_1 - ½ \rho (v_2^2 - v_1^2)$$
$$= 0.02 \times (13.6 \times 10^3) \times 9.8 - ½ \times 10^3 \times [(0.75)^2 - (0.50)^2]$$
$$= 2.509 \times 10^3 \text{ N m}^{-2}$$

Question 9: The piston and nozzle of syringe have diameters 10mm and 4mm respectively. The syringe filled with water is held horizontally at a height of 2 m above the ground. The piston is pushed with a constant speed of 0.25 m s^{-1}. Find the horizontal range of water jet (coming out of the nozzle) on the ground. Take g = 10 m s^{-2}.

Solutions

Let v be the horizontal with which water comes out of the nozzle. From equation of continuity (Av = constant), we have

$$\pi\ (5 \times 10^{-3}\ m)^2 \times 0.25\ m\ s^{-1} = \pi\ (2 \times 10^{-3}\ m)^2 \times v$$

Solving, we get $v = 1.56\ m\ s^{-1}$

Water coming out of nozzle will follow a projectile path. If t be the time taken coming out of the nozzle to hit the ground, we have

$$x = vt \qquad \text{(i)}$$

$$\text{and } y = \tfrac{1}{2} g t^2 \qquad \text{(ii)}$$

where x is the horizontal range covered and y is the vertical height of the nozzle above the ground. From eq. (ii) we have

$$t = (2\ y / g)^{1/2}$$

$$t = 0.63\ s$$

From eq. (i), the horizontal range is

$$x = 1.56\ m\ s^{-1} \times 0.63\ s = 0.99\ m$$

Question 10: An oil drop falls through air with a terminal velocity of $5 \times 10^{-4}\ m\ s^{-1}$. (a) Calculate the radius of drop. (b) What will be the terminal velocity of a drop of half of this radius? Given viscosity of air = $1.8 \times 10^{-5}\ N\ s\ m^{-2}$, density of oil = $900\ kg\ m^{-3}$. Neglect density of air compared to that of oil. Take $g = 9.8\ N\ kg^{-1}$.

Solutions

(a) When an oil drop (radius r, density ρ) falls through air (density σ, viscosity η) the terminal velocity attained by the drop is

$$v = 2/9\ r^2 (\rho - \sigma)\ g/\ \eta \qquad \text{(i)}$$

Ignoring σ compared to ρ and soling for r, we have

$$r = (9\eta v/2\rho g)^{1/2}$$

On substituting values we have

$$r = 2.14 \times 10^{-6}\ m.$$

(b) As is clear from equation. 9(i), $v \propto r^2$, therefore the terminal velocity of drop of half the above radius will be one-fourth the above velocity, that is,

$$(5 \times 10^{-4}\ m\ s^{-1})/4 = 1.25 \times 10^{-4}\ m\ s^{-1}.$$

Post Module questions:

1. On application of a pressure of $21 kg/cm^2$, the volume of 1 lit. of an oil decreases by 840mm3.Calculate the bulk modulus and compressibility of the oil.

2. If the tension in a wire increases gradually to 6kg, the elongation of the wire becomes 1.13 mm. Calculate the work done.

3. A 3kg mass is hanging from one end of a vertical wire of length of 2 m and of diameter 0.5 mm. Due to this mass, elongation produced in the wire is 2.38mm Find the Young's modulus of copper.

4. Six external forces, each of magnitude F, are applied on all the faces of a unit cube. Considering its elastic modulus, calculate the longitudinal strain and the volume strain on the unit cube.

5. A steel bar of breadth and depth both 1 cm, is supported on two knife edges 1 m apart. A load of 1.5 kg at the center of the bar depresses that point by 2.51 mm. What will be the Young's modulus of steel?

6. Two bodies M and N of equal mass are hung separately from two light weight springs. Force constants of the springs are k_1 and k_2.The bodies are set to vibrate so that their maximum velocities are equal. Find the ratio of the amplitudes of vibration of the two bodies.

7. The poisson's ratio of a wire is σ. Show that if e is the longitudinal strain due to an applied force, the volume strain will be e.(1-2σ).

8. Young's moduli of two rods of equal length and equal cross section are y_1 and y_2.These rods are joined end to end forming a composite rod system. Prove that equivalent Young's modulus of the composite system of rods $=2(y_1.y_2)/(y_1+y_2)$.

References

1. General Properties of Matter: C. J. Smith
2. General Properties: D. S. Mathur

....

Part of the chapter has been taken from the NCERT book

Module - III
Oscillations and Waves

Lectures: 03

Pre Test Questions

1. (a) What is a periodic Phenomenon? Give few examples.

   (b) Define Oscillation. What is the difference between mechanical and non-mechanical Oscillation?

   (c) What is Phase? Define phase constant.

2. Define angular frequency and how it is related with time period. Give the relation. Give a physical relation showing the relation between them.

3. What is Simple Harmonic Motion? What is a simple pendulum? How can you relate the motion of a simple pendulum to Simple Harmonic Motion? Write the differential equation linking acceleration to displacement.

4. What are Waves? What is a pulse? What are the different types on waves?

5. What are Standing Waves? How do you define harmonics of standing waves?

6. What is Doppler's Effect?

7. Define Resonance. What do mean by Beats?

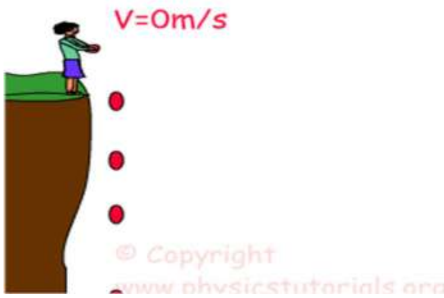
8. What is meant by Damping? Distinguish between under damping and over damping.

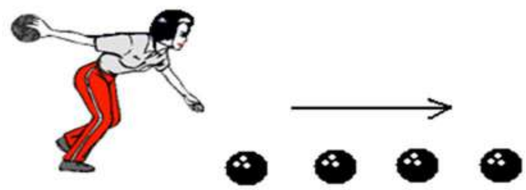
9. Distinguish between rarefaction and compression, and crest and trough.

10. How does a wave pulse interact for the following cases:
    i) When it encounters a fixed boundary?
    ii) When it encounters a free boundary?

Types of Motion: We very often come across the following types of motions

1. Rectilinear motion: Motion of a particle in a straight line





Examples: Motion of ball in straight line, a body that falls freely in vertical direction under the influence of gravity etc.

Periodic Motion

A motion that repeats itself at regular intervals of time is called **periodic motion**.

Swinging Pendulum and clock

In both the cases, the motions repeat after certain interval of time. Such a motion that repeats after certain interval of time is known as periodic motion. The body is displaced from a fixed point and it is given a small displacement, a force comes into action that tries to bring it to its equilibrium point, giving rise to oscillation or vibration.

Every oscillatory motion is a periodic, but every periodic motion need not be oscillatory.

Oscillations

- Oscillatory motion is a to and fro motion about a mean position and periodic motion repeats at regular intervals of time.
- All oscillatory motions are periodic but all periodic motions are not oscillatory.

Oscillations or **Vibrations**

There is no significant difference between oscillations and vibrations.

- Low frequency periodic motions are called as oscillation (like the oscillation of a branch of a tree),
- High frequency Periodic motions are called as vibration (like the vibration of a string of a musical instrument).

Periodic Motions which can be represented by Sinusoidal waveform (Sine or Cosine wave) are known as harmonic motion

Simple Harmonic Motion

- ***Simple Harmonic Motion (SHM)*** is a specialized form of periodic motion
- Periodic vibration around an equilibrium position
- Restoring force must be
    - proportional to displacement from equilibrium
    - in the direction of equilibrium

There are two types of SHM that will be discussed.

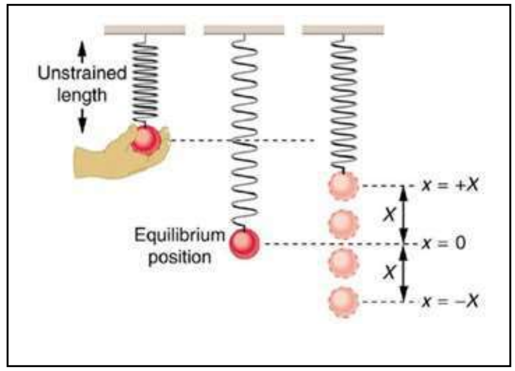
- Mass-Spring System
- Pendulum

- Simple harmonic motion (SHM) is a special kind of periodic motion occurs in mechanical system where net force acting on an object is proportional to the displacement of the object from its equilibrium position and the force is always directed towards the equilibrium position.
- In order for mechanical oscillation to occur, a system must posses two quantities: ***elasticity*** and ***inertia***.

- When the system is displaced from its equilibrium position, the *elasticity* provides a *restoring force* such that the system tries to return to equilibrium.
- The *inertia* property causes the system to *overshoot* equilibrium. This constant play between the elastic and inertia properties is what allows oscillatory motion to occur.

The natural frequency of the oscillation is related to the elastic and inertia properties

An oscillating system is a mass connected to a rigid foundation with a spring.



Example: Spring mass system

- An oscillating system is a mass connected to a rigid foundation with a spring.
- The spring constant k provides the elastic restoring force, and the inertia of the mass m provides the overshoot.

By applying Newton's second law *F=ma* to the mass, one can obtain the equation of motion for the system:

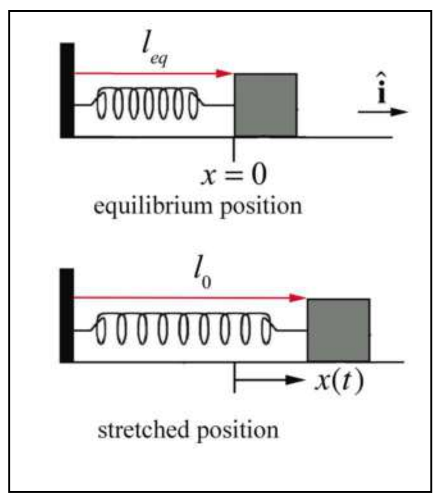
$$F = -kx \Rightarrow m\frac{d^2x}{dt^2} = -kx \Rightarrow \frac{d^2x}{dt^2} + \frac{k}{m}x = 0 \Rightarrow \frac{d^2x}{dt^2} + \omega_o^2 x = 0$$

$\omega_o = \sqrt{\frac{k}{m}}$ is the natural frequency

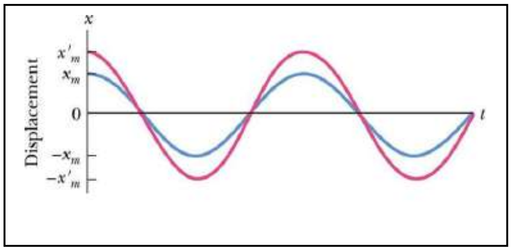
The solution of the wave equation

$$x(t) = x_m \cos(\omega_o t + \phi)$$

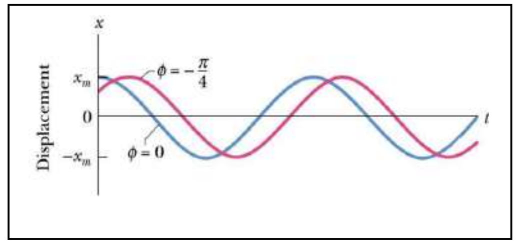
where x_m is the amplitude of the oscillation, and φ is the *phase constant* of the oscillation.



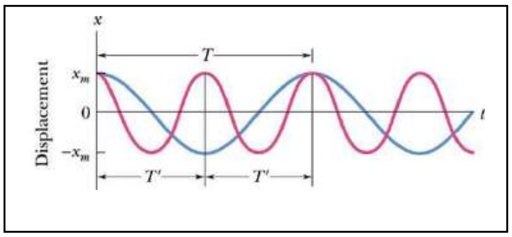
Waves with different amplitudes



Waves with different phase



Waves with different time period

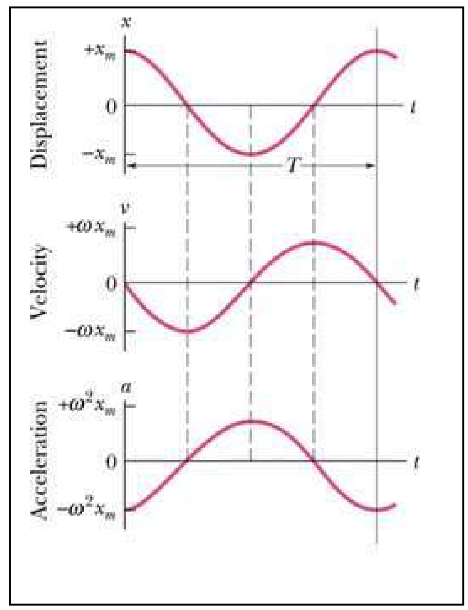


The period of the oscillatory motion is defined as the time required for the system to start one position, complete a cycle of motion and return to the starting position.

$$T = \frac{2\pi}{\omega_o} = 2\pi\sqrt{\frac{m}{k}}$$

$$v(t) = -\omega_o x_m \sin(\omega_o t + \phi)$$

$$a(t) = -\omega_o^2 x_m \cos(\omega_o t + \phi)$$



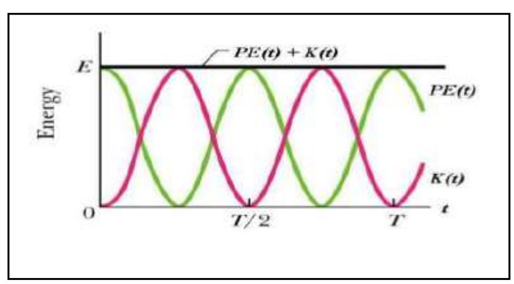
Energy in Simple Harmonic Motion

Kinetic energy (*K*) of the particle executing SHM

$$\mathrm{K} = \frac{1}{2}\mathrm{mv}^2$$

$$\mathrm{K} = \frac{1}{2}\mathrm{m}\omega^2\mathrm{A}^2\sin^2(\omega t + \varphi)$$

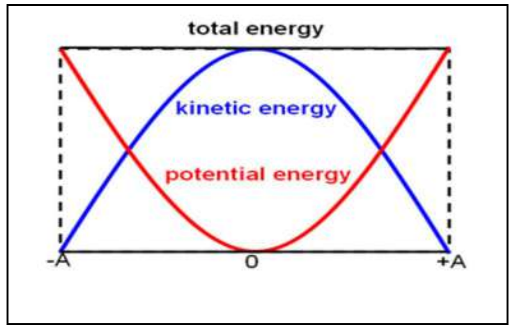
$$\mathrm{K} = \frac{1}{2}\mathrm{kA}^2\sin^2(\omega t + \varphi)$$



associated potential energy

$$U = \frac{1}{2}\ kx^2$$

$$U = \ \frac{1}{2}kA^2 \cos^2(\ \omega t + \varphi)$$



Total energy, E, of the system is,

$$E = U + K = \frac{1}{2}kA^2$$

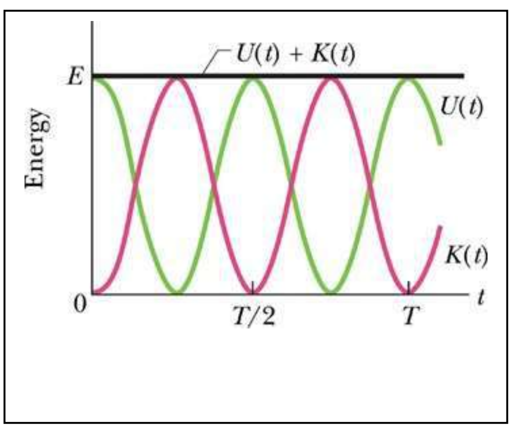
Throughout oscillation, KE continually being transformed into PE and *vice versa, but TOTAL ENERGY remains constant*

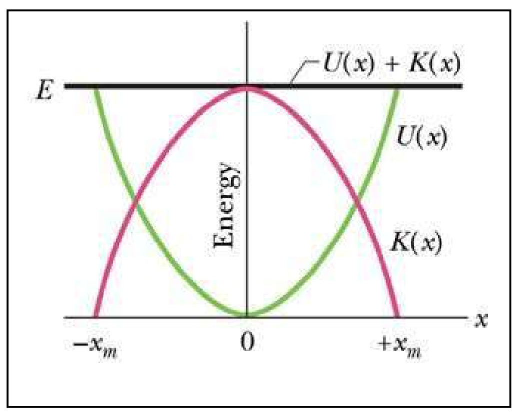
As the system oscillates, the total mechanical energy in the system trades back and forth between potential and kinetic energies.

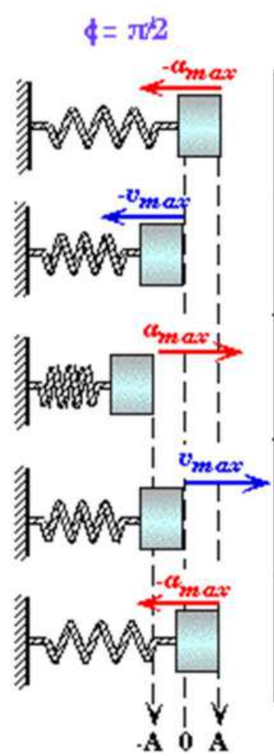
$$PE = \frac{1}{2}kx^2 = \frac{1}{2}kx_m^2 \cos^2(\omega_o t + \phi)$$
$$KE = \frac{1}{2}mv^2 = \frac{1}{2}m\omega_o^2 x_m^2 \cos^2(\omega_o t + \phi)$$

$$PE + KE = \frac{1}{2}kx_m^2$$
$$= \frac{1}{2}m\omega_o^2 x_m^2 = \frac{1}{2}mv_m^2$$

The total energy in the system, however, remains constant, and depends only on the spring constant and the maximum displacement (or mass and maximum velocity $v_m = \omega x_m$)

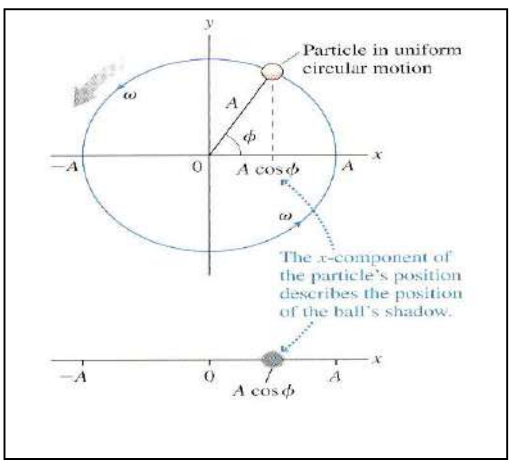






Time t	Position $x(t)$	Velocity $v(t)$	Accel $a(t)$	KE	PE
0	x_{max}	0	$-a_{max}$	0	PE_{max}
$\frac{T}{4}$	0	$-v_{max}$	0	KE_{max}	0
$\frac{T}{2}$	$-x_{max}$	0	a_{max}	0	PE_{max}
$\frac{3T}{4}$	0	v_{max}	0	KE_{max}	0
T	x_{max}	0	$-a_{max}$	0	PE_{max}

Simple Harmonic Motion and Uniform Circular Motion



Displacement of oscillating object = projection on x-axis of object undergoing circular motion

$x(t) = A\cos\theta$

For rotational motion with angular frequency ω, displacement at time t:

$x(t) = A\cos(\omega t + \varphi)$

φ= angular displacement at t=0 (phase constant)

A = amplitude of oscillation (= radius of circle)

Velocity and Acceleration in Simple Harmonic Motion

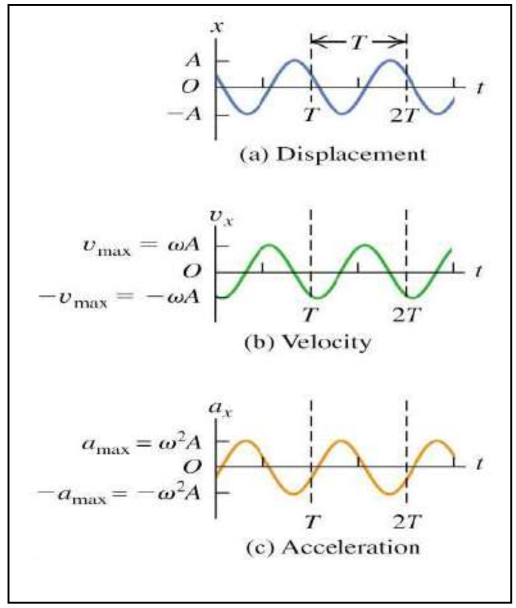
Simple harmonic motion is the projection of uniform circular motion on a diameter of the circle in which the latter motion takes place.

Displacement

$$x(t) = A\cos(\omega t + \varphi)$$

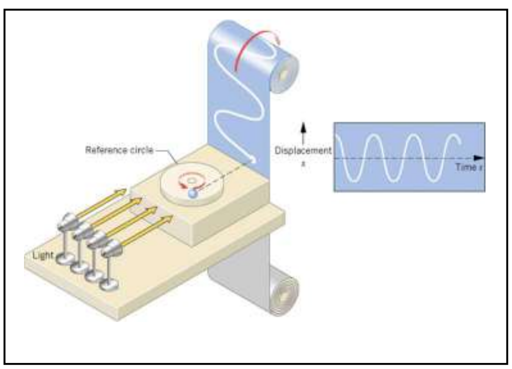
$$\text{Velocity}: v(t) = \frac{dx}{dt} = -\omega A\sin(\omega t + \varphi)$$

$$\text{Acceleration}: a(t) = \frac{d^2x}{dt^2} = -\omega^2 A\cos(\omega t + \varphi)$$



Circular Motion

- Uniform Circular motion projected in one dimension is SHM

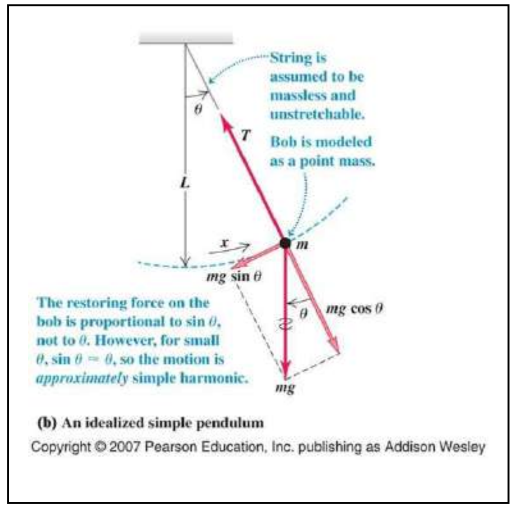


The ball mounted on the turntable moves in uniform circular motion, and its shadow, projected on a moving strip of film, executes simple harmonic motion.

Simple Pendulum

A pendulum consists of an object hanging from the end of a string or rigid rod pivoted about the point. The object is displaced (a small displacement; about 5-10^0) to one side and allowed to oscillate. If the object has negligible size and the string or rod is massless, then the pendulum is called a simple pendulum.

Restoring force



(b) An idealized simple pendulum

Copyright © 2007 Pearson Education, Inc. publishing as Addison Wesley

$$F = -mg\sin\theta$$

$$F \approx -mg\theta = -\mathrm{mg}\frac{x}{L} \quad \text{for small angles}$$

$$\therefore \mathrm{F} \propto x \qquad \text{SHM motion}$$

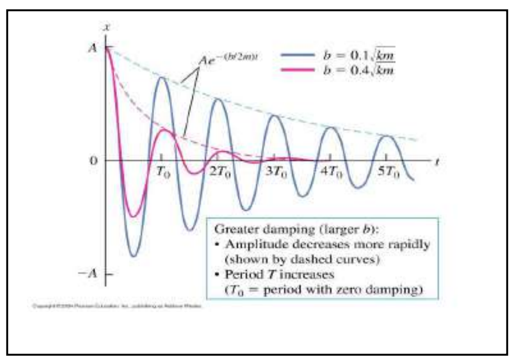
Damped Harmonic motion: Real oscillatory system

Have you ever thought why a simple pendulum or spring mass system comes to rest when they are kept in oscillatory motion. Ideally, the oscillatory motion should continue forever.

It is because of the resistance created by air, i.e. air works as damping medium which always apposes the motion or in other words we can say that the damping force is always in opposite direction to the restoring force.

In general, it is found that the damping force is proportional to the velocity of the oscillatory body.

Damped Oscillations

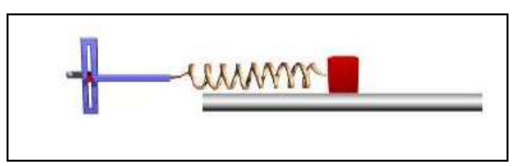


For damped oscillations, simplest case is when the damping force is proportional to the velocity of the oscillating object

Equation of motion:

$$m\frac{d^2x}{dt^2} = -kx - b\frac{dx}{dt}$$

Forced or Driven oscillation



- The natural frequency is the frequency at which it will oscillate if there is no driving and damping forces.

What is a wave?

A disturbance or variation that transfers energy progressively from point to point in a medium and that may take the form of an elastic deformation or of a variation of pressure, electric potential, temperature or more.

The Human Wave

The human wave is the disturbance (people jumping up and sitting back down), and it travels around the stadium. However, none of the individual people in the stadium are carried around with the wave as it travels - they all remain at their seats.

Waves in Everyday Life: Examples

- Disturbance produced in pond by throwing a stone creates ripples which move outward.
- Sound: Type of wave that moves through matter and then vibrates our eardrums so we can hear.
- Visible Light: Kind of wave that is made up of photons.
- Radio and TV Signals etc.

TYPES OF WAVES

(a) Mechanical waves
   - Requires medium for propagation
   - Governed by Newton's laws
   - Example: Water waves, sound waves, seismic waves, etc.

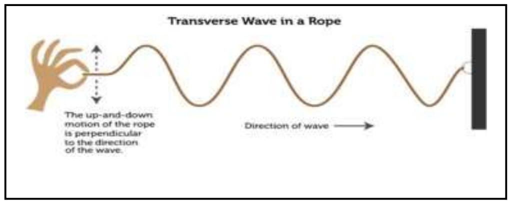
(b) Electromagnetic waves
   - Do not require any medium for their propagation
   - Example: Visible and ultraviolet light, Radio waves, Microwaves, X-rays etc.

(c) Matter waves
   - wave associated with the motion of a particle of atomic or subatomic size (electrons, protons, neutrons, other fundamental particles, and even atoms and molecules)

Waves differ from one another in the manner the particles of medium oscillate (or vibrate) with reference to the direction of propagation.

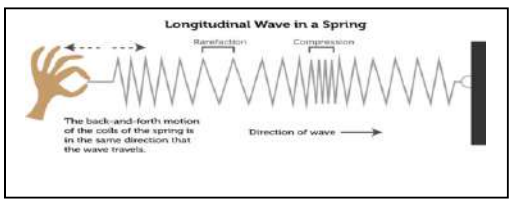
Transverse Wave

A wave in which the particles of the medium vibrate at right angles to the direction of propagation of wave, is called a transverse wave.



Longitudinal waves

A wave in which the particles of the medium vibrate in the same direction in which wave is propagating, is called a longitudinal wave.



Wave Parameters

- The amplitude A, is half the height difference between a peak and a trough.
- The wavelength λ, is the distance between successive peaks (or troughs).
- The period T, is the time between successive peaks (or troughs).
- The wave speed c, is the speed at which peaks (or troughs) move.
- The frequency ν, (Greek letter "nu") measures the number of peaks (or troughs) that pass per second.

A wave is a disturbance that travels from one location to another, and is described by a wave function that is a function of both space and time. If the wave function was sine function then the wave would be expressed by

$$y = A\sin(\omega t \pm kx)$$

The negative sign is used for a wave traveling in the positive *x* direction and the positive sign is used for a wave traveling in the negative *x* direction.

$$k = \frac{2\pi}{\lambda}$$
$$\omega = 2\pi\upsilon$$

Wave Speed

- The speed of a wave depends on the medium through which the wave moves.
- The speed, wavelength, and frequency are related.

$$\mathrm{v} = \frac{\omega}{k} = \frac{\lambda}{T} = \lambda v$$

- Speed of a Transverse Wave on Stretched String

$$\mathrm{v} = \sqrt{\frac{T}{\mu}}$$

Where μ is linear mass density of a string, is the mass *m* of the string divided by its length *l*.

- Speed of a Longitudinal Wave Speed of Sound

$$\mathrm{v} = \sqrt{\frac{B}{\rho}}$$

where B is bulk modulus and ρ is density of the medium.

- Speed of a longitudinal wave in an ideal gas

$$\mathrm{v} = \sqrt{\frac{Y}{\rho}}$$

where *Y* is the Young's modulus of the material of the bar.

- Speed of a longitudinal wave in an ideal gas

$$\mathrm{v} = \sqrt{\frac{P}{\rho}}$$

where *P* is the Pressure of the ideal gas. Above eqn. is known as Newton relation.

Laplace correction

For adiabatic processes the ideal gas satisfies the relation,

PV^{γ} = constant

where γ is the ratio of two specific heats, $\gamma = C_p/C_v$

$$v = \sqrt{\frac{\gamma P}{\rho}}$$

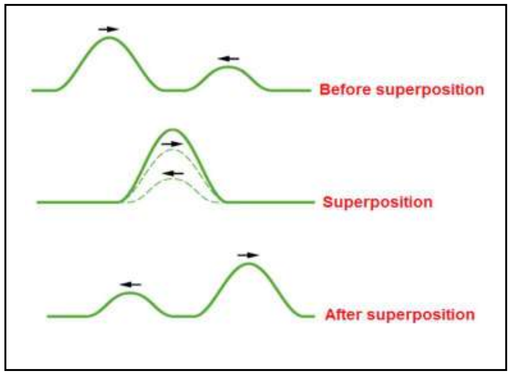
This modification of Newton's formula is referred to as the Laplace correction.

Waves meet matter

- Waves do not travel through the same medium.
- Waves can be reflected, refracted or absorbed.
- Sound wave traveling though a long corridor is reflected back as echo.
- The speed of waves depend on the elastic properties of the medium through which it travels.
- When a wave encounters different medium where the wave speed is different, the wave will change directions.
- Sound wave travelling through different medium undergoes change in speed.
- Refraction enables sound to travel faster along the ground at night than during the say.
- When waves approach a shore with a gradual depth profile, e.g. a beach, most of the wave energy is absorbed by the breaking of waves.
- When waves approach a hard vertical surface, e.g. a concrete wall or a dock, most of the energy is converted into waves moving in the opposite direction, a reflection. Of course the reflected waves are superimposed onto the original waves,

Superposition of Waves

The principle of superposition may be applied to waves whenever two (or more) waves travelling through the same medium at the same time. The waves pass through each other without being disturbed. The net displacement of the medium at any point in space or time, is simply the sum of the individual wave displacements



$y(x, t) = y_1(x, t) + y_2(x, t)$

Reflection of Waves

- When a wave encounters a boundary, it will reflect back.

The way in which it reflects will vary depending on whether it encounters a fixed or free boundary.

Fixed Boundary

A fixed boundary is when a wave encounters a fixed surface. This would occur for a rope attached to a wall.

Free Boundary

A free boundary occurs when, for example, a rope is attached to a post and is free to move up and down at the end.

Reflection and Refraction

- **Reflection**
    - Waves bounce back off of a surface that they encounter.
    - A travelling wave, at a rigid boundary or a closed end, is reflected with a phase reversal but the reflection at an open boundary takes place without any phase change.
- **Refraction**

    Waves bend and pass through a surface that they encounter.

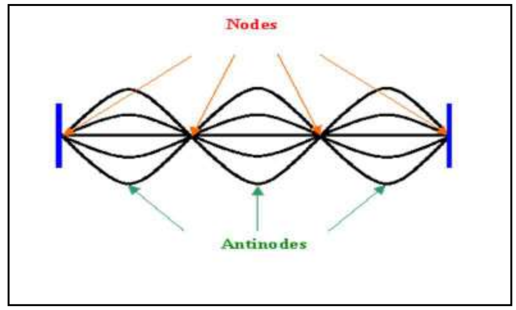
Standing Waves and Normal Modes

- Standing wave oscillates with time but appears to be fixed in its location

wave pattern that results when two waves of the same frequency, wavelength, and amplitude travel in opposite directions and interfere

Node

A point in a standing wave that always undergoes complete destructive interference and therefore is stationary



Antinode

A point in a standing wave, halfway between two nodes, at which the largest amplitude occurs

For Nodes

$kx=n\pi$ where n= 0,1,2,3,…

distance of $\lambda/2$ or half a wavelength separates two consecutive nodes.

For Antinodes

$kx=(n+1/2)\ \pi$ where n= 0,1,2,3,…

distance of $\lambda/2$ or half a wavelength separates two consecutive antinodes.

- For a stretched string of length L, fixed at both ends, the two ends of the string have to be nodes

$\lambda = 2L/n$, for n = 1, 2, 3, … etc

The corresponding frequencies can be represented as

$\nu = n\ v/2L$, for n = 1, 2, 3, … etc.

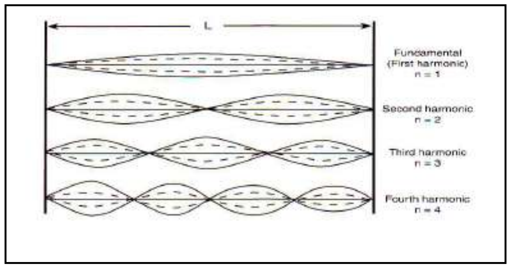
- For a system closed at one end, with the other end being free the closed end will be node while open end will be antinode

$\lambda = 2L/(n+1/2)$, for n = 1, 2, 3, … etc

The corresponding frequencies can be represented as

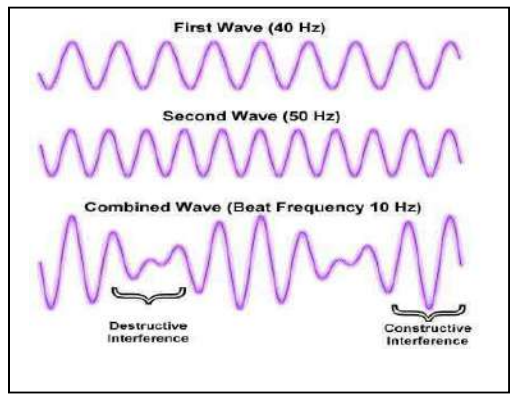
$\nu = (n+1/2)\ v/2L$, for n = 1, 2, 3, … etc.

The oscillation mode with that lowest frequency is called the fundamental mode or the first harmonic The second harmonic is the oscillation mode with n = 2. The third harmonic corresponds to n = 3 and so on. The collection of all possible modes is called the harmonic series and n is called the harmonic number.



Beats

The phenomenon of regular rise and fall in the intensity of sound, when two waves of nearly equal frequencies travelling along the same line and in the same direction superimpose each other is called beats.



One rise and one fall in the intensity of sound constitutes one beat and the number of beats per second is called beat frequency. It is given as:
$\nu_b = (\nu_1 - \nu_2)$
where ν_1 and ν_2 are the frequencies of the two interfering waves; ν_1 being greater than ν_2

Resonance

Resonance occurs in an oscillating system when the driving frequency happens to equal the natural frequency. For this special case the amplitude of the motion becomes a maximum.

An example is trying to push someone on a swing so that the swing gets higher and higher. If the frequency of the push equals the natural frequency of the swing, the motion gets bigger and bigger.

Doppler's Effect

The phenomena of apparent change in frequency of source due to a relative motion between the source and observer is called Doppler's effect.

- When the source and observer are moving toward each other, the frequency heard by the observer is higher than the frequency of the source.
- When the source and observer are moving away from each other, the frequency heard by the observer is lower than the frequency of the source.

Moving Observer and Moving Source

- Moving Source
  → change in λ → Δν
- Moving Observer
  → change in relative velocity → Δν
- Moving Source and Observer
  → change in λ and relative velocity → Δν

Source moving; Observer Stationary

$$v = v_o(1+\frac{v_s}{v})^{-1}$$
$$= v_o(1-\frac{v_s}{v})$$

Observer Moving; Source Stationary

$$v = v_o(1+\frac{v_o}{v})$$

Both Source and Observer moving

$$v = v_o(\frac{v+v_o}{v+v_s})$$

Post Test Questions

1. An object undergoing simple harmonic motion about x = 0 has a displacement of 3 cm and a velocity of 0 at t = 0. The period of oscillation is 2 seconds. Find: (i) the phase constant (ii) the amplitude of the oscillation (iii) the acceleration at t = 1 second.

2. A 150-gram toy is in simple harmonic motion on the end of a horizontal spring with force constant 300 N m −1 . When the toy is 1.2 cm from its equilibrium position, it is observed to have a speed of 30 cm s −1 . Find: (i) the total energy of the toy at any point in its motion (ii) the amplitude of its motion (iii) the maximum speed of the toy (iv) the maximum force on the toy.

3. A string on a guitar has a linear mass density of 3 g m−1 and is 63 cm long. It is tuned to have a fundamental frequency of 196 Hz.
   (a) What is the tension in the tuned string?
   (b) Calculate the wavelengths of the first three harmonics. Sketch the transverse displacement of the string as a function of x for each of these harmonics.

4. A block of mass 200 g is attached to a horizontal spring with k = 0.85 N m −1 . When in motion, the system is damped by a force that is linear in velocity, with γ = 0.2 kg s −1 .
   (a) Write the differential equation of motion for the system.
   (b) Show that the system is underdamped. Calculate the oscillation period and compare it to the natural period.
   (c) How long does it take for the oscillating block to lose 99.9% of its total mechanical energy? How many cycles does this correspond to? By what factor does the amplitude decrease during this time?

5. Two aeroplanes A and B are approaching each other and their velocities are 108 km/h and 144 km/h respectively. The frequency of a note emitted by A as heard by the passengers in B is 1170 Hz. Calculate the frequency of the note heard by the passengers in A. Velocity of sound = 350 m s–1.

6. A car is travelling along a road. A stationary policeman observes that the frequency ratio of the siren of the car is 5/4 as it passes. What is the speed of the car? [Velocity of sound in air = 333 m s–1]

7. A source of sound frequency 256Hz is moving rapidly towards a wall with a velocity of 5m/s. How many beats per second will be heard if sound travels at a speed of 330m/s ?

8. A wire of length 140 cm and mass 0.52×10- kg is stretched by means of a load of 16 kg. Calculate the frequency of the fundamental node

9. For the following decibel levels, determine the corresponding sound intensity levels in W/m^2.
   a. 50 dBel
   b. 90 dBel
   c. 110 dBel

10. A 1.65-meter length string is forced to vibrate in its fifth harmonic. Determine the locations of the nodal positions. Express the locations as a distance measured from one of the ends of the string.

11. A steel piano wire is pulled to a tension of 448 N and has a mass density of 0.00621 kg/m. The string is 61.8 cm long and vibrates at its fundamental frequency.
   a. Determine the speed at which vibrations travel through the wire.
   b. Determine the wavelength of the standing wave pattern for the fundamental frequency.
   c. Determine the frequency of its vibrations.

References

1. Advanced Acoustics: D. P. Raychaudhuri
2. Oscillations and waves: S. Bharadwaj

Module – IV
Electricity, Magnetism, and Maxwell's Equations

Lectures: 03

Prerequisite Questions

To understand the electromagnetics and Maxwell's equations, we need to have the basic knowledge of certain aspects of physics (e.g., electricity, magnetism) and mathematics (e.g., vector calculus). We put forward some prerequisite questions as follows:

(i) What are the concepts of total and partial differentiation? How you will relate them.

(ii) What are the various vector operations? Explain the operation of scalar-vector, and vector-vector multiplication. Illustrate their geometrical representation.

(iii) What will be the result when in the scalar-vector and vector-vector multiplication you will replace one vector by □.

(iv) What is an electrostatic field? Describe it by taking suitable examples. How does it vary with the distance from the source charge? How do you describe its behavior? What do you require to explain its behaviour at $r=0$?

(v) Can you explain the existence of magnetic charges? If not, why? If yes, how?

(vi) How the moving charges or flowing current generate the magnetic field? Does it follow the charge conservation?

(vii) What is the generic difference between induced electric field in Faraday's experiment, and the same in the electrostatic field?

(viii) What is the physics wise concept of the displace current?

Physical concepts of gradient, divergence, and curl

After introducing scalar and vector fields, physical concepts of gradient, divergence and curl are presented in an illustrative manner to enable beginners better understand the physical meaning of Maxwell's equations. Laplacian operator is also briefly discussed.

1.1. Scalar and vector fields

A field is a function that describes a physical quantity at all points in space. A field which is characterized at each point in space by a single number is a **scalar field**, for example, temperature T(x,y,z,t) in Kelvin, density ρ(x,y,z,t) in kg m^{-3}; pressure P(x,y,z,t) in N m^{-2}, etc. If these fields vary with time, they are static scalar fields. But if they do vary with time, they are time-varying scalar fields.

A **vector field** is a vector-valued function that is associated with a vector at each point in its domain. For example, the flow velocity (m s^{-1}) in a fluid can be represented by a field of flow at each and every point in space. A velocity vector **v**(x,y,z,t) describes the velocity of the fluid at a point, i.e., **v**(x,y,z,t). If the flow is steady, then velocity vector **v** does not depend on time but only on space. Another example of a vector field is the gradient, ∇V(x,y,z), of a scalar-valued function V(x,y,z). The electric field **E** (x,y,z,t), measured in N C^{-1} or V m^{-1}, and the magnetic field **B**(x,y,z,t), measured in tesla, are also the vector fields.

1.2. Electric and magnetic fields

Electric field and magnetic field are defined in terms of the forces experienced by an electrical charge. For instance, if we bring a charge q in an electric field **E**, the charge experiences a force q**E**. We also know that a charge moving with velocity **v** in a magnetic field **B** experiences a magnetic force q(**v**×**B**). This means that there is still something there (i.e., a field) even if we remove the charge.

If a charge is located at the point (x, y, z) at the time t, it experiences the force $F = q(E + v \times B)$. We can associate E and B vectors with every point (x,y,z) in space. The electric and magnetic fields are, then, viewed as vector functions in terms of x, y, z and t. Since E or B can be specified at every point in space, we call it 'an electromagnetic field'. Electromagnetic (EM) fields are produced by charges. The relationships between the values of the fields at one point and the values at a nearby point can be fully described in the form of differential equations. Therefore, the knowledge of vector calculus is mandatory on understanding the electromagnetic theory.

1.3. The concept of gradient

While climbing a hill, we experience a direction of steepest ascent at right angles to the contour lines of constant gravitational potential. If we climb at an angle θ to this direction, we climb more slowly by a factor $\cos\theta$. Likewise, we can describe a scalar function of position $V(x,y,z)$ by surfaces of constant V. At a point P, there is a direction in which V increases most rapidly as indicated in Figure 1.1. This direction is perpendicular to the surface of constant V through P. A vector with magnitude equal to the rate of increase of V in this direction is known as the gradient of V (or ∇V). The component of the vector, i.e., ∇V in any direction gives the rate of increase of V in that direction. This is because, at an angle θ to the direction of ∇V, the distance traversed for a given small increment of V is increased in the ratio $1 : \cos\theta$ as shown in Figure 1.1, so that the rate of increase of V in this direction is $\cos\theta$ times less than in the direction of ∇V. Using cartesian coordinates (x,y,z), the rates of increase of V in the directions of coordinate axes are $\partial V/\partial x$, $\partial V/\partial y$, $\partial V/\partial z$. These are, therefore, the cartesian components of the vector, i.e., $\nabla V = \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z}\right)$.

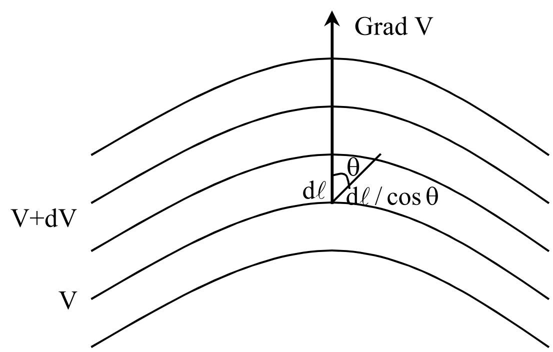


Figure 1.1. Concept of gradient.

Let us illustrate the concept of gradient by considering an example of an electrostatic field. The work done in carrying a unit test charge from one point x to another point x + Δx, is the potential difference between the two points, i.e.,

$$\Delta W = V(x+\Delta x, y, z) - V(x, y, z) = \frac{\partial V}{\partial x}\Delta x .$$

But the work done against the field for the same path is

$$\Delta W = -\oint \mathbf{E}\cdot d\ell = -E_x\,\Delta x .$$

Comparison of these equations shows that

$$E_x = -\frac{\partial V}{\partial x} .$$

Similarly,

$$E_y = -\frac{\partial V}{\partial y} , \quad \text{and} \quad E_z = -\frac{\partial V}{\partial z} .$$

or $\qquad \mathbf{E} = -\nabla V.$

The electric field is given in magnitude and direction by the negative gradient of the electric potential, and the lines of electric force intersect the equipotential surfaces at right angles.

Problem. Show that ∇V is a vector perpendicular to the surface $V(x,y,z)$ = constant.

Let $\mathbf{r} = ix + jy + kz$ be the position vector of any point $P(x,y,z)$ on the surface. Then $d\mathbf{r} = idx + jdy + kdz$ lies in the tangent plane to the surface at P.

But $$dV = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz = 0$$

or $$\left(i\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y} + k\frac{\partial V}{\partial z}\right).\ (idx + jdy + kdz) = 0,$$

i.e., $\nabla V.d\mathbf{r} = 0$, so that ∇V is perpendicular to $d\mathbf{r}$ and therefore to the surface.

Problem. Show that the maximum rate of change of V takes place in the direction of ∇V.

Let $V(x,y,z)$ and $V(x+\Delta x, y+\Delta y, z+\Delta z)$ be the potential at two neighbouring points $P(x,y,z)$ and $Q(x+\Delta x, y+\Delta y, z+\Delta z)$ and $d\ell$ is the distance between these points. Now $\Delta V = \frac{\partial V}{\partial x}\Delta x + \frac{\partial V}{\partial y}\Delta y + \frac{\partial V}{\partial z}\Delta z +$ infinitesimals of order higher terms.

Then $$\lim_{\Delta\ell \to 0}\frac{\Delta V}{\Delta \ell} = \frac{dV}{d\ell} = \frac{\partial V}{\partial x}\frac{dx}{d\ell} + \frac{\partial V}{\partial y}\frac{dy}{d\ell} + \frac{\partial V}{\partial z}\frac{dz}{d\ell}.$$

$$= \left(i\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y} + k\frac{\partial V}{\partial z}\right)\cdot\left(i\frac{dx}{d\ell} + j\frac{dy}{d\ell} + k\frac{dz}{d\ell}\right)$$

$$= \nabla V\cdot\frac{d\mathbf{r}}{d\ell}.$$

Note that $\frac{dV}{d\ell} = \nabla V\cdot\frac{d\mathbf{r}}{d\ell}$ is the projection of ∇V in the direction $\frac{d\mathbf{r}}{d\ell}$. This projection will be a maximum when ∇V and $\frac{d\mathbf{r}}{d\ell}$ have the same direction. Then the maximum value of $\frac{dV}{d\ell}$ takes place in the direction of ∇V and its magnitude is $|\nabla V|$.

1.4. The concept of divergence

The divergence of a vector field **D** at a point P is a measure of how much the field diverges or converges from that point. Figures 1.3 a,b,c show that the divergence of a vector field at a point P has (a) positive divergence because the vector field diverges or spreads out (source), (b) negative divergence because the vector field converges (sink), and (c) zero divergence because the vector field neither diverges nor converges (no source, no sink). Let us now deal with it in some detail.

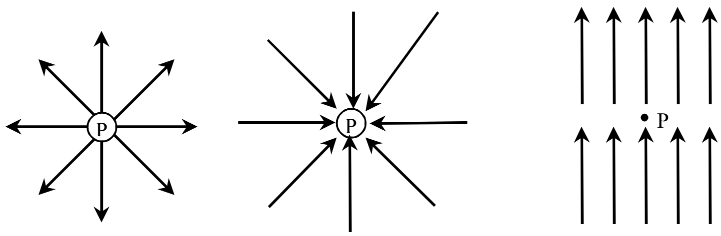


Figure 1.3. Illustration of the divergence of a vector field at P:
(a) positive divergence, (b) negative divergence, and (c) zero divergence.

Let us consider a rectangular element of volume of dimensions ($\Delta x,\Delta y,\Delta z$) with one corner located at the point (x,y,z) as shown in Figure 1.4. Let (D_x,D_y,D_z) be the components of the vector **D**. The x-component of the flux of the vector **D** through a face of area $\Delta y\ \Delta z$ of the volume element is $D_x\ \Delta y\ \Delta z$. The excess of this flux at the x+Δx face over that of the x-face will be

$$\Delta x \frac{\partial}{\partial x}(D_x\ \Delta y\ \Delta z) = \Delta x\ \Delta y\ dz \frac{\partial D_x}{\partial x}.$$

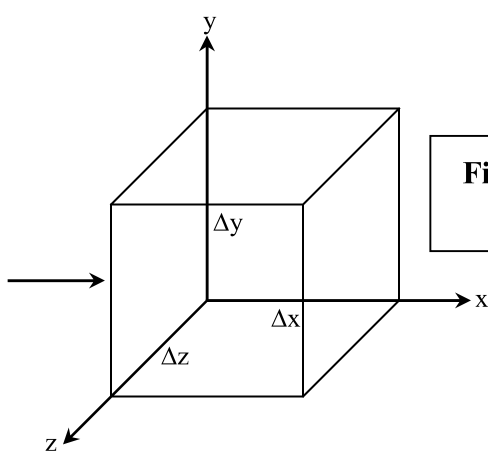


Figure 1.4. Calculation of flux out of a volume element.

The same argument applies to the other two pairs of faces of the volume element. The total flux of **D** out of the volume element is, therefore, given by

$$\int_s \mathbf{D}\cdot d\mathbf{s} = \Delta x\ \Delta y\ \Delta z\left(\frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z}\right)$$

or

$$\frac{\int_s \mathbf{D}\cdot d\mathbf{s}}{\Delta v} = \frac{\partial D_x}{\partial x}+\frac{\partial D_y}{\partial y}+\frac{\partial D_z}{\partial z} = \nabla\cdot\mathbf{D}\ .$$

This means that the outward flux from the surface of a volume element is equal to the divergence of the vector multiplied by the volume of the element. We now see the physical meaning of the divergence of a vector. It is outflow or inflow of the vector per unit volume.

Divergence of **D** is connected to the flux of **D** out of a volume element. For any finite volume, therefore, the total flux from a volume is the sum of the fluxes out of each part. Thus, we can integrate the divergence over the entire volume, i.e.,

$$\int_s \mathbf{D} \cdot d\mathbf{s} = \int_V \nabla \cdot \mathbf{D}\, dV$$

where s is any closed surface and V is the volume inside it. This theorem is named after Gauss, and is known as Gauss' theorem or divergence theorem which connects the surface integral to the volume integral.

Relation between the Partial Differentiation and Total Differentiation: Partial differentiation is used for the multi-variant fields, e.g., f(x,y,z). It includes that how the field varies w.r.t. a particular variable while others are treated as a constant. An example is given as follows:

$$\Delta f(x,y,z) = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \frac{\partial f}{\partial z}\Delta z, \text{ and } \frac{\partial^2 f}{\partial x\, \partial y} = \frac{\partial^2 f}{\partial y\, \partial x}$$

The first equation is true only in the limit that Δx, Δy, and Δz tend to zero.

Problem. Show that (a) $\nabla \cdot \nabla V = \nabla^2 V$, and (b) $\nabla^2(1/r) = 0$

(a) $$\nabla \cdot \nabla V = \left(i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}\right) \cdot \left(i\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y} + k\frac{\partial V}{\partial z}\right)$$

$$= \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)V = \nabla^2 V.$$

(b) $$\nabla^2(1/r) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)(x^2+y^2+z^2)^{-1/2}.$$

Now $$\frac{\partial}{\partial x}(x^2+y^2+z^2)^{-1/2} = -x(x^2+y^2+z^2)^{-3/2},$$

and $$\frac{\partial^2}{\partial x^2}(x^2+y^2+z^2)^{-1/2} = \frac{\partial}{\partial x}(-x(x^2+y^2+z^2)^{-3/2})$$

$$= 3x^2(x^2+y^2+z^2)^{-5/2} - (x^2+y^2+z^2)^{-3/2}$$

$$= \frac{2x^2 - y^2 - z^2}{(x^2+y^2+z^2)^{5/2}}.$$

Similarly, we can evaluate the other terms and sum them up to get

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\left(\frac{1}{\sqrt{x^2+y^2+z^2}}\right) = 0$$

The equation $\nabla^2 V = 0$ is known as Laplace equation and V = 1/r is a solution of this equation.

1.5. The concept of curl

The curl of a vector field **D** at a point P is viewed as a measure of the circulation or how much the field curls around that point. For example, Figure 1.5a shows that the curl of a vector field around P is directed out of the page, while Figure 1.5b shows a vector field with zero curl. We will now deal with it in some detail.

Figure 1.5. (a) curl at P points out of the page; (b) curl at P is zero.

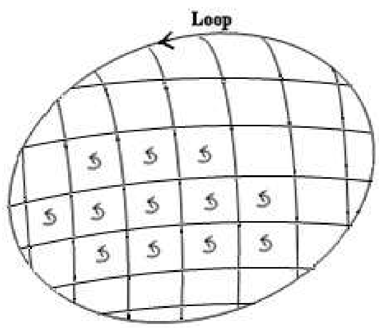


Figure 1. 5 (c). Some surface bounded by the loop is chosen. The surface is divided into a large number of small loops so that each loop becomes approximately a square. The circulation around the loop is the sum of the circulations around the little loops.

Let us find out the circulation of a vector field **D**. We break our loop into a large number of small loops so that each small loop becomes approximately a square. We can find out the circulations around all of the little squares. Then we take their sum to get the circulation around the loop. Let us now consider that the little square lies in the xy-plane as shown in Figure 1.5.

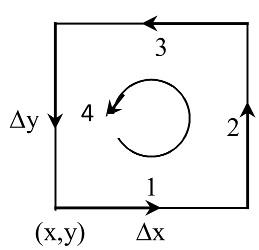


Figure 1.6. Circulation around an element square.

Starting at the point (x, y), we go around in the direction shown by the arrows as shown in Fig. 1.6. The whole line integral will then become

$$\oint \mathbf{D} \cdot d\ell = D_x(1)\,\Delta x + D_y(2)\,\Delta y - D_x(3)\,\Delta x - D_y(4)\,\Delta y.$$

Now $D_x(1)\,\Delta x - D_x(3)\Delta x$ will be zero to the first approximation. To be more accurate, however, we must consider the rate of change of D_x with respect to y, then we have

$$D_x(1)\,\Delta x - \left(D_x(1) + \frac{\partial D_x}{\partial y}\Delta y \right)\Delta x = -\frac{\partial D_x}{\partial y}\Delta x\,\Delta y\ .$$

Similarly, the other two terms in the circulation can obviously be written as

$$D_y(2)\Delta y - D_y(4)\,\Delta y = (D_y + \frac{\partial D_y}{\partial x})\,\Delta y - D_y\Delta y = \frac{\partial D_y}{\partial x}\Delta x\,\Delta y\ .$$

Thus the circulation around the element square becomes

$$\left(\frac{\partial D_y}{\partial x} - \frac{\partial D_x}{\partial y} \right)\Delta x\,\Delta y\ ,$$

which is a z-component of $\nabla \times \mathbf{D}$, i.e., normal to the surface element:

$$\nabla \times \mathbf{D} = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ D_x & D_y & D_z \end{vmatrix}$$

We can, therefore, write the circulation around a differential square as

$$\oint \mathbf{D}\cdot d\ell = (\nabla \times \mathbf{D})_n\ ds = (\nabla \times \mathbf{D})\cdot d\mathbf{s}\ .$$

We can fill in the loop with any convenient surface and add the circulations around a set of infinitesimal squares in this surface, i.e.,

$$\oint \mathbf{D}\cdot d\ell = \int_s \nabla \times \mathbf{D}\cdot d\mathbf{s}\ .$$

This is Stokes' theorem which connects the line integral to the surface integral.

.

Problem. If $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, show that $\boldsymbol{\omega} = \frac{1}{2}$ curl $\mathbf{v}$ where $\boldsymbol{\omega}$ is a constant vector.

$$\nabla \times \mathbf{v} = \nabla \times (\boldsymbol{\omega} \times \mathbf{r}) = \nabla \times \begin{vmatrix} i & j & k \\ \omega_1 & \omega_2 & \omega_3 \\ x & y & z \end{vmatrix}$$

$$= \nabla \times [(\omega_2 z - \omega_3 y)\,i - (\omega_1 z - \omega_3 x)\,j + (\omega_1 y - \omega_2 x)\,k]$$

$$= 2\,(\omega_1 i + \omega_2 j + \omega_3 k) = 2\boldsymbol{\omega}$$

Therefore, $\boldsymbol{\omega} = \frac{1}{2}\,\nabla \times \mathbf{v} = \frac{1}{2}$ curl $\mathbf{v}$

This problem simply shows that the curl of a vector field exhibits rotational properties of the field. If the field is due to a moving fluid, for instance, then a paddle wheel placed at various points in the field would tend to rotate in regions where $\nabla \times \mathbf{v}$ is non-zero. However, if $\nabla \times \mathbf{v} = 0$ in the region, there would be no rotation, and the field is then called irrotational. A field which is not irrotational is also called a vortex field.

1.6. Laplacian operator

Notice (i) the gradient of a scalar field is a vector field, (ii) the divergence of a vector field is a scalar field, and (iii) the curl of a vector field is a vector field. We can construct from the del operator a natural differential operator that creates a scalar field from a scalar field. If we apply the gradient operator to a scalar field to give a vector field, and then apply the divergence operator to this result, we get a scalar field. This is what we call "div grad" of a scalar field, and is given by

$$\nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} .$$

We denote this operator by the symbol ∇^2, and call it as the Laplacian operator after Laplace who studied physical applications of scalar fields (such as the potential of an inverse-square force law) that satisfy the equation $\nabla^2 V = 0$,

i.e., $$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0 .$$

This formula has wide applications in science and technology, e.g., in the context of potential fields (such as the electrostatic potential in an electric field, and the velocity potential in a frictionless ideal fluid), in Poisson's equation $\nabla^2 V = -\rho/\varepsilon_0$.

We have considered above the Laplacian of a scalar. Since the Laplacian operator ∇^2 is a scalar operator, we can define the Laplacian of a vector. In this context, $\nabla^2\mathbf{A}$ should not be viewed as the gradient of **A**, which makes no sense. Rather, $\nabla^2\mathbf{A}$ is defined as the gradient of the divergence of **A** minus the curl of the curl of **A**, i.e.,

$$\nabla^2\mathbf{A} = \nabla(\nabla.\mathbf{A}) - \nabla \times \nabla \times \mathbf{A}.$$

Problem. For a vector field **A**, show that $\nabla.\nabla \times \mathbf{A} = 0$, that is, the divergence of the curl of any vector field vanishes.

$$\nabla \cdot \nabla \times \mathbf{A} = \left(i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z} \right) \cdot \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} = 0$$

Problem. For a scalar field V, show that $\nabla \times \nabla V = 0$, that is, the curl of the gradient of any scalar field vanishes.

$$\nabla \times \nabla V = \left(i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z} \right) \times \left(i\frac{\partial V}{\partial x} + j\frac{\partial V}{\partial y} + k\frac{\partial V}{\partial z} \right) = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial z} \end{vmatrix} = 0 .$$

This chapter begins with the experimental laws of electricity and magnetism. These laws (e.g., Coulomb, Ampère, and Faraday) are described in terms of physical contents and their mathematical representations. Taken together, they suggest no new effects beyond the original experiments they represent. It is only when the displacement current is added that new physics emerges. This physics includes the prediction of the existence of electromagnetic waves which follow from Maxwell's equations and transport energy and momentum through empty space by means of electromagnetic fields.

2.1. Coulomb's law

All matter is made up of fundamental particles of which most important are the neutron (mass = 1.67×10^{-27} kg and charge = zero), the proton (mass = 1.67×10^{-27} kg and charge = 1.60×10^{-19} C), and the electron (mass = 9.1×10^{-31} kg and charge = -1.60×10^{-19} C). The ratio of the charge of a proton to the charge of an electron is 'minus one' to the highest known degree of accuracy.

The French Scientist Coulomb established from his experiments that an electrical charge q_1 exerts a force on another charge q_2 separated by a distance r, which is given by

$$\mathbf{F} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}\hat{r} \tag{2.1}$$

where $\frac{1}{4\pi\varepsilon_0} = 10^{-7}\ c^2$ (by definition) = 9×10^9 (by experiment) $N.m^2/C^2$.

This equation shows that the force acts along the line which joins the two charges. For two electrons, for instance, we have (-q) (-q) = q^2 where q is the charge of an electron, which means that the force is positive and acts in the direction of increasing distance r, measured from either of the point charges. This means that it is a repulsive force. Like charges repel, and unlike charges attract (If two charges of one coulomb each are separated by a distance 1 mm, the repulsive force between them will be enormously high, i.e., 9×10^{15} N). If we add more electrons to the original two, we find from the experiment that this does not affect the interaction between the original two electrons. We, therefore, supplement the Coulomb's law with the another wonderful fact of nature, that is, the force on any charge is the vector sum of the Coulomb forces from each of the other charges. This fact of nature is called 'the principle of superposition'.

Let us now look at what happens from an alternative viewpoint (action through a medium). The repulsion (or attraction) of charge q_2 reveals that q_1 affects its surroundings. In other words, the influence of q_1 is still there even if q_2 were taken away. We can, therefore, think that q_1 produces something which is called the electric field **E** in the space around it. When q_2 is introduced, the field in its vicinity acts on it so that it experiences a force $\mathbf{F} = q_2\mathbf{E}$. This can be reconciled with the Coulomb's law if

$$\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\frac{q_1}{r^2}\hat{r} \quad . \tag{2.2}$$

In this case, we see that an electric field is a representation of the way in which it varies with position. Note that the word 'field' transfers our attention from the electrical charges to the space around them.

The electric field **E** has direction and magnitude (intensity). Note, then, that electric field lines begin at positive charge (source), and end at negative charge (sink), unlike magnetic field lines which neither begin nor end (no source, no sink).

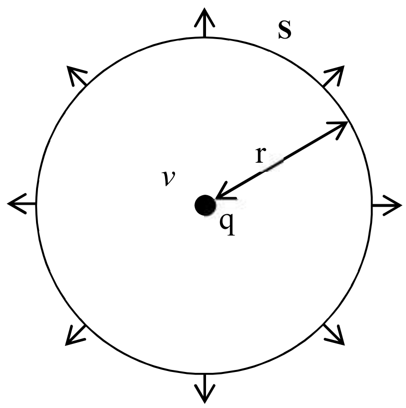


Figure 2.1. The charge q located at the origin is surrounded by a concentric spherical surface s of radius r.

The intensity of the field is crudely defined as the number of lines of force passing through a unit area at right angles to the direction of the lines. This simply means that the more intense the field, the higher the concentration of lines of force to represent it. If we consider an electrical charge q, the number of lines of force (hypothetical) it produces per unit area, by definition, will be $\mathbf{E} = \frac{1}{4\pi\varepsilon_0}\frac{q}{r^2}\hat{r}$. So, the intensity varies inversely as the square of the distance. Let us draw an imaginary sphere (as shown in Figure 2.2) with surface area S around this charge at a distance r and ask how many lines of force pass through it. The answer is $(1/4\pi\varepsilon_0)(q/r_1^2)\times 4\pi r_1^2 = q/\varepsilon_0$. This is independent of distance (Note that the electric flux at radius r per unit area, or electric flux density vector or electric displacement vector at radius r is $\mathbf{D} = \frac{1}{4\pi}\frac{q}{r^2}\hat{r}$, so that $\mathbf{D} = \varepsilon_0 \mathbf{E}$). So, we see that the total number of lines of force is constant from the charge to infinity. Note that this argument works only if the force falls off as the inverse square of the distance which is the case here. As already noted, we reiterate that the concept of lines of force cannot be used to interpret everything that happens in electromagnetism. For this, we must return to the basic equations. However, they often provide some insight into what happens at intuitive level.

Problem. Calculate the ratio of the electrical to gravitational forces acting on two particles (say electrons).

The ratio is constant, independent of the relative positions of the particles, and is given by

$$\frac{F_e}{F_g} = \frac{q_1}{m_1}\frac{q_2}{m_2}\cdot\frac{1}{4\pi\varepsilon_0 G}$$

$$= (1.60 \text{ x } 10^{-19}\text{ C}/9.1 \text{ x } 10^{-31}\text{ kg})^2 . (9 \text{ x } 10^9\text{ N.m}^2/\text{C}^2)/6.67 \text{ x } 10^{-11}\text{ N.m}^2/\text{kg}^2)$$

$= (3.09 \times 10^{22}) \,.\, (1.35 \times 10^{20}) = 4.17 \times 10^{42}$.

The gravitational force (attractive) is weaker than the electrostatic force (attractive or repulsive) by an enormous factor of $\approx 10^{42}$.

2.2. Electrostatics

We have seen in Section 2.1 that the number of lines of force, say N, leaving an electrical charge is given by $N = q/\varepsilon_0$. If we have an array of charges $q_1, q_2, \ldots, q_n$, we draw an imaginary closed surface around the volume of space they occupy. Recalling the principle of superposition, the net number of lines of force passing through the surface is $N = 1/\varepsilon_0 \sum_i^n q_i$. In performing the addition, we count field lines going outward through the surface as positive and those going inwards as negative. So, if there are equal number of positive and negative charges within the surface, the net number of lines of force passing through the surface will be zero. In general, we can write

$$N = \int_s \mathbf{E}.d\mathbf{s} = (1/\varepsilon_0)\sum_{i=1}^{n} q_i \,, \tag{2.3}$$

where d**s** is an element of area on the imaginary surface. This is known as Gauss' law, i.e., electric flux out of a closed surface is equal to the charge enclosed.

In general, the electrical charges are distributed throughout the volume contained by the surface. We can describe such a distribution in terms of an electrical charge density per unit volume, ρ. When we replace $\sum_{i=1}^{n} q_i$ by an integral $\int_V \rho dV$, Gauss' law becomes

$$\int_s \mathbf{E}.d\mathbf{s} = (1/\varepsilon_0)\int_V \rho dV \tag{2.4}$$

Using Divergence theorem on the left side of this equation, we have

$$\int_V \nabla.\mathbf{E} dV = \frac{1}{\varepsilon_0}\int_V \rho dV \,.$$

The integrals on both sides have to be equal for any arbitrary volume (or differentiating both sides), we get

$$\nabla.\mathbf{E} = \rho/\varepsilon_0 \,. \tag{2.5}$$

(Physically it means that, since dV appears on both sides of the equation, the relationship between $\nabla.\mathbf{E}$ and ρ must hold for any arbitrary point in space.)

It is an expression of the principle of conservation of charge, and it is true in every electromagnetic field at every point of space.

A simple exercise shows that the work done by a test charge around a closed path in an electrostatic field is zero. This means that electrostatic field is a conservative field, i.e.,

$$\oint \mathbf{E}.d\ell = 0 \tag{2.6}$$

Using Stokes' theorem, we have

$$\int_s \nabla \times \mathbf{E}.d\mathbf{s} = 0 \,.$$

Again the integrand has to be identically zero since the integral is zero for any arbitrary surface (or upon differentiation), we get

$$\nabla \times \mathbf{E} = 0 \quad (2.7)$$

Since the curl of gradient of a scalar function vanishes, we can write

$$\mathbf{E} = -\nabla V, \quad (2.8)$$

where V is electric scalar potential, and the minus sign implies that **E** points in the direction of decreasing V.

Note that electrostatics is a neat example of a vector field with zero curl ($\nabla \times \mathbf{E} = 0$) and a given divergence ($\nabla.\mathbf{E} = \rho/\varepsilon_0$).

2.3. Magnetostatics

We define the magnetic field intensity (or the magnetic induction) **B** in a similar way as that of electric field intensity **E**. Making use of representation of the magnetic field by lines of force, as in the case of the electric field, we can write for the number of lines of force N as:

$$N = \int_s \mathbf{B}.d\mathbf{s} = \int_V \nabla.\mathbf{B}\, dV \quad (2.9)$$

Since isolated magnetic charges do not exist, the number of outgoing lines of force through any imaginary surface around any distribution of magnets in space must be balanced by the number of incoming lines of force through the surface (i.e., outflow = inflow). This simply means that

$$N = \int_V \nabla.\mathbf{B}\, dV = 0 .$$

Again the integrand has to be identically zero since the integral is zero for any arbitrary volume (or upon differentiation), we get

$$\nabla.\mathbf{B} = 0 . \quad (2.10)$$

As we know that isolated electrical charges exist (positive charge as source, and negative charge as sink), but magnetic charges (magnetic monopoles) do not exist. The search for isolated magnetic poles has not yet been successful.

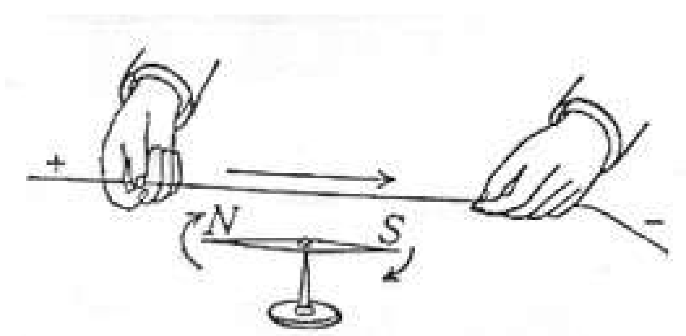


Figure 2.2 Experiment by Hans Christian Ørsted (1777-1851)

Biot-Savart found from their experiments a relationship between the magnetic field due to current i flowing in an element of the wire $d\ell$, which is known as Biot-Savart law, which is given by

$$d\mathbf{B} = \frac{\mu_0 i}{4\pi r^3}(d\ell \times \mathbf{r}). \tag{2.11}$$

This equation deals with an inverse square law (like Coulomb's law). The magnetic field is perpendicular to both $d\ell$ and **r**. The total field **B** due to the whole current system can be obtained by integrating the contributions from all the elements of the system.

A more fundamental approach to magnetic field uses a law that (like Gauss' law for electric fields) takes advantage of the symmetry present in certain problems to simplify the calculation of **B**. Let us, therefore, consider the magnetic field generated by a current in a long straight wire (see Figures 2.3a,b). From the symmetry of the problem, **B** depends on i and r so that (experimentally):

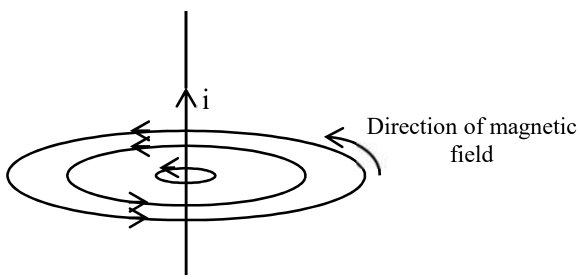


Figure 2.3a. Ampère's experiment consisting of a long straight wire carrying an electric current i.

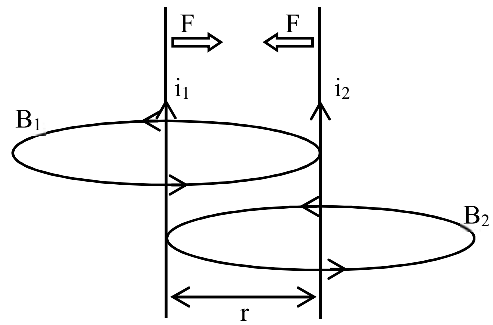


Figure 2.3b. Two parallel wires carrying currents in the same direction attract each other. The general rule is that parallel currents attract, and antiparallel currents repel.

$$B \propto i$$

$$B \propto 1/r$$

$$B = \frac{\mu_0}{2\pi}\frac{i}{r} \quad \text{or} \quad B(2\pi r) = \mu_0 i$$

Noting that $d\ell$ around the path is simply the length of the path, $2\pi r$, in the case of a circle, we can write

$$\oint \mathbf{B}.d\ell = \mu_0 \int_s \mathbf{J}.d\mathbf{s} \tag{2.12}$$

Using Stokes' theorem on the left side of this equation, we get

$$\int_s \nabla \times \mathbf{B}.d\mathbf{s} = \mu_0 \int_s \mathbf{J}.d\mathbf{s}$$

The integrals on the both sides have to be equal for any arbitrary surface (or differentiating both sides), we get

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \tag{2.13}$$

which is known as Ampère's law.

Note that magnetostatics is a neat example of a field with zero divergence ($\nabla.\mathbf{B} = 0$) and a given curl ($\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$).
If we take divergence of equation (2.13), we get

$$\nabla.(\nabla \times \mathbf{B}) = \mu_0 \nabla.\mathbf{J} .$$

Since divergence of curl of a vector vanishes, we have

$$\nabla.\mathbf{J} = 0 ,$$

which means that the current flows in closed loops.

In electrostatics, we saw that the curl of **E** is always zero, so we can represent **E** as the gradient of a scalar field or potential V ($\mathbf{E} = -\nabla V$). In magnetostatics, however, the curl of **B** is not always zero, so it is not possible to represent **B** as the gradient of a magnetic scalar potential. However, the divergence of **B** is always zero, so we can always represent **B** as the curl of a vector field, recalling the fact that the divergence of curl of a vector is always zero. We can, therefore, write **B** = curl **A**. The field **A** is called the magnetic vector potential (we will discuss about it in chapter 3).

2.4. Time varying magnetic field

Faraday's experiment showed that when the magnetic flux through a conducting circuit changed, a current flowed in it. This induced current flows in such a way as to oppose the original change (also known as Lenz's law). The induced current produces its own magnetic field and the direction of flow of the current is such that this secondary magnetic field acts in the opposite direction to the original magnetic field. The induced current is actually a secondary effect the value of which depends on the resistance of the circuit.

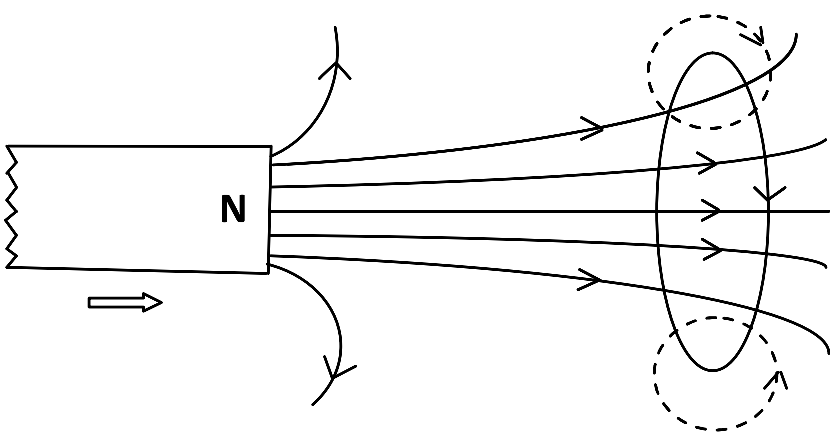


Figure 2.4 As the flux through the wire loop increases, current starts flowing in the loop. The magnetic field produced by this current opposes the flux changes that produce it.

The change in magnetic flux generates an emf in the circuit which then produces the observed current. We can write it as

$$\mathsf{E} = - dN/dt \tag{2.14}$$

where the magnetic flux is represented by the number of lines of force N and the minus sign reflects the opposition to a change (see Figure 2.4). We can write N as

$$N = \int_s \mathbf{B}.d\mathbf{s}$$

or
$$\frac{dN}{dt} = \frac{\partial}{\partial t}\int \mathbf{B}\cdot d\mathbf{s}$$

Note that the magnetic field **B** varies both with position and with time. Therefore, we use the partial derivative $\partial\mathbf{B}/\partial t$ rather than the total derivative $d\mathbf{B}/dt$.

The emf is the line integral of the electric field. We can, therefore, write equation (2.14) as

$$\int E.d\ell = -\frac{\partial}{\partial t}\int \mathbf{B}.d\mathbf{s} \tag{2.15}$$

Using Stokes' theorem on the left side of this equation, we rewrite it as

$$\int \nabla\times\mathbf{E}.d\mathbf{s} = -\frac{\partial}{\partial t}\int \mathbf{B}\cdot d\mathbf{s}.$$

The integrals on both sides have to be equal for any arbitrary surface (or differentiating both sides), we get

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \tag{2.16}$$

Since $\nabla.\mathbf{B} = 0$, and the divergence of a curl vanishes, we can write

$$\mathbf{B} = \nabla\times\mathbf{A},$$

where **A** is known as magnetic vector potential (we will discuss the physical meaning of magnetic vector potential **A** in chapter 3). Substituting this value of **B** in equation (2.16), we get

$$\nabla\times\mathbf{E} = -\frac{\partial}{\partial t}(\nabla\times\mathbf{A})$$

or
$$\nabla\times(\mathbf{E} + \frac{\partial\mathbf{A}}{\partial t}) = 0.$$

Compare it with the electrostatic field ($\nabla\times\mathbf{E} = 0$), and convince yourself that electromagnetic field is given by

$$\mathbf{E} = -\nabla V - \frac{\partial\mathbf{A}}{\partial t}. \tag{2.17}$$

Note also that time-varying electric field is a non-conservative field.

2.5. Time varying electric field

The net current flowing out of an imaginary closed surface will be given by

$$\int_s \mathbf{J}.d\mathbf{s}$$

If this is positive, then there is a net outflow of current. This means that there must be an accompanying reduction of charge within the volume enclosed by the surface. The total charge within the volume is $\int_V \rho\, dV$. Therefore, the rate of loss of this charge from the volume is $\frac{\partial}{\partial t}\left(\int_V \rho\, dV\right)$. Since this loss is responsible for the outflow of current, we have

$$\int_s \mathbf{J}.d\mathbf{s} = -\frac{\partial}{\partial t}\int_V \rho\, dV \ .$$

The minus sign reveals the fact that an outflow of current is equivalent to a decrease in the amount of charge within the volume enclosed by the surface.

Using divergence theorem on the left side of this equation, we get

$$\int_V \nabla.J dV = -\frac{\partial}{\partial t}\int_V \rho\, dV \ .$$

The integrals on both sides have to be equal for any arbitrary volume (or differentiating both sides with respect to the volume), we get

$$\nabla.\mathbf{J} = -\frac{\partial \rho}{\partial t} \quad \text{or} \quad \nabla.\mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \ .$$

(2.18)

This is called the continuity equation for electrical charge (or the conservation of charge).

2.6. The Lorentz force

Let us consider that a steady current carrying wire is immersed in a magnetic field **B**. Experiments show that a force d**F** is exerted by the field on a short length of the wire $d\ell$, i.e.,

$$d\mathbf{F} = i(d\ell \times \mathbf{B}) \ . \tag{2.19}$$

Note that the force is at right angles to the wire and to the magnetic field. Compare it with the Biot-Savart law (cf., Equation 2.11).

Now, we think of the current in terms of an electric charge q flowing along the wire, we have i = dq/dt, and if the charge is moving with velocity **v**, then $d\ell$ = **v** dt. Hence, i $d\ell$ = (dq/dt) **v** dt = dq **v**. We can then rewrite equation (2.11) as

$$d\mathbf{B} = \frac{\mu_0}{4\pi r^3}(dq\, \mathbf{v} \times \mathbf{r}) \ ,$$

or, integrating over the charge,

$$\mathbf{B} = \frac{\mu_0 q}{4\pi r^3}(\mathbf{v} \times \mathbf{r}) \ . \tag{2.20}$$

Similarly, we can rewrite equation (2.19) as d**F** = dq (**v** x **B**),
or, integrating

$$\mathbf{F} = q\ (\mathbf{v} \times \mathbf{B}). \tag{2.21}$$

Now, substituting equation (2.20) into equation (2.21), we get

$$\mathbf{F}_{mag} = \frac{\mu_0 q^2}{4\pi\, r^3}[\mathbf{v}\times(\mathbf{v}\times\mathbf{r})].$$

Compare it with the electric force (cf., equation 2.1), i.e.,

$$\mathbf{F}_{elect} = \frac{q^2}{4\pi\,\varepsilon_0\, r^3}\mathbf{r}\,.$$

The ratio of magnitudes of the electrical to magnetic forces will be

$$\frac{F_{mag}}{F_{elect}} = \mu_0\varepsilon_0 v^2 = \frac{v^2}{c^2}\quad . \tag{2.22}$$

The average speed of electrons in a wire is about 10^{-4} m/s, so $v^2/c^2 \approx 10^{-25}$. This means that the magnetic force is 10^{-25} weaker than the electrical force between the moving electrons. Number of electrons and protons being equal in wires, electrical forces would disappear because of balancing out. This balance is much more precise than one part in 10^{25}. Therefore, the small relativistic term (magnetic force) is the only term left which becomes the dominant term. It is the near-perfect cancellation of electrical effects which allowed relativity effects (i.e., magnetism) to be studied, and the correct equations to the order of v^2/c^2 to be discovered. That is why electromagnetic laws did not need any relativistic correction (unlike mechanics), because they were already correct to a precision of v^2/c^2.

In electromagnetic situations, electric and magnetic forces are present simultaneously, and we have

$$\mathbf{F}_{total} = \mathbf{F}_{elect} + \mathbf{F}_{mag} = q\,(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.23}$$

This is known as the Lorentz force. It should be noted here that $\mathbf{v} \times \mathbf{B}$ represents a kind of additional electric field which acts on any charged body in motion. Let us illustrate this point by looking at the Ohm's law:

$$I = V/R\,.$$

Considering an element of length $d\ell$ with cross sectional area $d\mathbf{s}$, we can write R in terms of the resistivity ρ of the wire as $R = \rho\, d\ell/d\mathbf{s}$. If we replace the resistivity of the material by its inverse, i.e., the conductivity $\sigma = 1/\rho$, we have $R = \frac{1}{\sigma}\frac{d\ell}{d\mathbf{s}}$. The current flowing through the small element of wire is

$$i = \mathbf{J}.d\mathbf{s} \;\text{ and }\; V = \mathbf{E}.d\ell\,.$$

Substituting all these in Ohm's law, we get

$$\mathbf{J}.d\mathbf{s} = \frac{\mathbf{E}.d\ell}{1/\sigma\, d\ell/d\mathbf{s}}\,.$$

or

$$\mathbf{J} = \sigma\,\mathbf{E}\,. \tag{2.24}$$

Comparing this with equation (2.23) indicates that Ohm's law needs modification in the presence of magnetic field. The modified form of Ohm's law, therefore, is

$$\mathbf{J} = \sigma\,(\mathbf{E} + \mathbf{v} \times \mathbf{B})\,. \tag{2.25}$$

Problem. Circulating charges: Suppose a beam of electrons is travelling in a uniform magnetic field. The magnetic deflecting force does not change the speed of the particles, and it always acts perpendicular to the velocity of the particle.

Since **B** is perpendicular to **v**, the magnitude of the magnetic force will be qvB, and Newton's second law with a centripetal acceleration v^2/r gives

$$qvB = mv^2/r,$$

$$\text{or } r = mv/qB = p/qB ,$$

which tells that smaller the speed, smaller will be the radius of charge circulation.

The angular frequency will be $\omega = v/r = qB/m$. Note that this frequency, which is known as cyclotron frequency, does not depend on the speed of the particle (if $v < c$).

Maxwell's Equations

3.1. Maxwell's equations

It is instructive to remind students of the state of the subject of electricity and magnetism at the time Maxwell published his first paper on Faraday's lines of force in 1856. The mathematical work of Laplace and Poisson had greatly contributed to the understanding of distribution of charges on bodies of different shapes. This led to the study of mechanical forces exerted by charges on the bodies and thereby to the formulation of Coulomb's law. Likewise, in magnetism, Ampère had developed his theory by studying the forces between current-carrying conductors. These developments viewed the subject from the *standpoint of action at a distance* (Students today are aware of relativity, and they are also aware that action at a distance is not instantaneous. With a source coil operating at 10 MHz and placed at a distance of one meter from the detector coil, one can demonstrate by observing the change in the phase lag by increasing the distance, thereby suggesting that an electromagnetic signal does take finite time to reach the detector.). As a result, the space outside an electric charge or magnetic dipole was not supposed to differ from empty space. Faraday's discovery of electromagnetic induction, however, showed that space occupied by the magnetic lines of force was not the same as the empty space. A circuit rotated in space through which no lines of magnetic force pass will not register an electric current.

Maxwell's objective obviously was to formulate Faraday's idea into a mathematical form. It is, therefore, important to emphasize that Maxwell had an intuition that action at a distance, as implied by Coulomb's law, for instance, was not true in physics. Having outlined Faraday's notion of magnetic lines of force and their representation by iron filings, he was dissatisfied with the *action-at-a-distance* theory of magnetism. While admitting that such theories are in fact 'in strict accordance' with the phenomena, he "(could not) help thinking that in every place where we find these lines of force, some physical state or action must exist in sufficient energy to produce the actual phenomena.". Action will always depend on what happens in the medium, and from this, Maxwell visualized that disturbances in the electromagnetic field could travel with speed of light.

While mathematically formulating the work of Faraday, Maxwell perceived that when a magnetic field was wrapped round a current, a current could equally be wrapped round a magnetic field. The relation between them was reciprocal in nature, and could be expressed mathematically by the curl of a vector (the word *curl* was first coined by Maxwell himself). When the two fields co-existed in space, the reaction between them could be expressed as the curl of a curl of a vector and this he simplified down to the well-known *wave equations*.

Let us now present all the experimental laws of electricity and magnetism that we discussed so far (in SI units):

$\nabla \cdot \mathbf{E} = \rho/\varepsilon_0$	Gauss' law	(2.26)
$\nabla\times\mathbf{E} = -\dfrac{\partial \mathbf{E}}{\partial t}$	Faraday's law	(2.27)
$\nabla\cdot\mathbf{B} = \mathrm{C}$	No magnetic charges	(2.28)
$\nabla \times \mathbf{B} = \mu_0\, \mathbf{J}$	Ampère's law	(2.29)

This incomplete form of the equations (2.26 to 2.29) summarizes what was experimentally known at the time when Maxwell set them down. Equation (2.26) states that lines of force can be used to describe the electric field and that these lines of force always begin on positive charges and end on negative charges. Considering the electric field further, equation (2.27) states that there exists an electric potential in a static field so that, as long as only static fields are considered, the energy of a moving charged particle is conserved. The extra term in equation (2.27) represents the law of induction. These differential equations are completely equivalent to the integral form of the laws, namely, Coulomb's law, and Faraday's law of induction, as they were then known. Equation (2.28) expresses the fact that there are also lines of force for the magnetic field, but these neither begin nor end anywhere because there are no free magnetic poles. Finally, equation (2.29) explains that a magnetic field is generated near a current, i.e., it is equivalent to the Biot-Savart law. We will now examine these equations in some more details.

Now, suppose that we had access to all the knowledge about electromagnetic field that was available at the time of Maxwell. And also suppose that we had formed the view that action over a distance was not the basis of the phenomenon, but that *there should be local laws which must be expressed in terms of differential equations* rather than as integrals. Then we would have arrived at equations (2.26) to (2.29). Note that Maxwell's first great achievement was to realize that these laws could be expressed as a set of partial differential equations. And the second, of course, was to realize that these equations were quite inadequate to explain the electromagnetic phenomena.

One of the ways in which we teach our students today is to point out that these equations by themselves are not consistent if we assume them to remain valid even in circumstances in which the charges and currents change with time. Taking the divergence of equation (2.29) and remembering that *divergence of curl of a vector vanishes* (physically it means that $\nabla \times \mathbf{B}$ follows the lines of force for which the divergence is zero), we obtain the condition that $\nabla . \mathbf{J} = \mathbf{0}$, which means that all electric currents ought to flow in closed loops only. However, we know that $\nabla . \mathbf{J} = \mathbf{0}$ should not vanish in general because if it did, the charge conservation equation (any flow of electric charge must come from some supply), $\nabla . \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$ (cf., equation 2.18) implies that $\frac{\partial \rho}{\partial t}$, hence the electric charge density at each point in space is unchangeable. This is clearly false, since we cannot imagine to have a flow of discrete charged particles without changing the charge distribution with time. Thus the Ampère law cannot be a complete statement of the property of $\nabla \times \mathbf{B}$ This inconsistency is corrected by adding an extra term $\varepsilon_0\partial\mathbf{E}/\partial t$ to the current in equation (2.29). This would yield the correct equation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad . \tag{2.30}$$

Then from (2.26) and (2.30), we have the continuity equation (2.18).

The current density $\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$ in free space is Maxwell's displacement current. Combined with the Ampère current density **J**, it gives the closed flow represented by the expression $\mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$.

We do not necessarily regard this argument as a proof because this is probably not the only way of reconciling the inconsistent equations. One could have modified the equations in other ways to make them consistent. In fact, if we had stuck to the law of action at a distance, then we would just have thought, 'the derivation without the new term is not valid when the current is varying.' Therefore, we must modify equation (2.29) in such a case. So this is not meant as a rigorous proof. But it is the argument which makes the form of Maxwell's equations most plausible to the beginner. There is no evidence in Maxwell's papers that this was the path by which he arrived at his result or these arguments played any part in his reasoning. Maxwell arrived at the extra term by using a picture which we do not accept today. If we put an electric field on a capacitor in which there is a medium of very high dielectric constant, then most of the electric induction **D** is actually contributed by the separation of charges in the dielectric from one side to the other. It is then natural to expect that the motion of these charges would be accompanied by a current, i.e., the displacement current. But we would not postulate any such motion of charges in the vacuum. However, Maxwell, who in fact called **D** the electric displacement, had just such a picture in mind.

The need to add the 'displacement current' to the Ampère's law was seen by Maxwell as necessary so that the magnetic field should satisfy the laws of vector analysis when $\nabla.\mathbf{J} \neq 0$. We often give the impression that the famous Maxwell equation was deduced from the Ampère law. But this is not strictly true. Maxwell gave a brilliant, but intuitive solution from the purely mathematical consideration that the divergence of curl of any vector must vanish. There is no physical reason why a magnetic field must satisfy the laws of vector analysis. And during Maxwell's lifetime, there was no direct empirical evidence that his solution was physically meaningful.

The beginners find it difficult to appreciate the physical meaning of displacement current in free space where ρ and **J** are zero in which case there is no need to modify the Ampère's law. It is, therefore, important to emphasize that the crucial thing Maxwell did was to introduce the vacuum displacement current $\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$ which lead to the wave equation. It is of interest to note that one also needs the $-\frac{\partial \mathbf{B}}{\partial t}$ term in curl $\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$ in order to get the wave equation. From pedagogical viewpoint, we would clearly see that one could not form the wave equation in free space without the vacuum displacement current. Visualization of 'vacuum displacement current', i.e., $\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$, is Maxwell's outstanding discovery.

Taking the curl of $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$, we obtain

$$\nabla^2 E = \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} . \tag{2.31}$$

Similarly, taking the curl of $\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$, we obtain

$$\nabla^2 B = \mu_0 \varepsilon_0 \frac{\partial^2 B}{\partial t^2} . \tag{2.32}$$

Equations (2.31) and (2.32) established the possibility of electromagnetic waves in free space. The speed of these waves is given by $v = 1/\sqrt{\mu_0 \varepsilon_0} = 3 \times 10^8$ m/s = speed of light, revealing the electromagnetic nature of light. Maxwell writes, "We can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena." This is one of the great unifications of physics: what a prophetic prediction! The moment Maxwell conceived the idea of the displacement current, a new era started in the history of mankind.

Some important points:

(i) It is important to note that only after introducing the extra term $\varepsilon_0 \partial \mathbf{E}/\partial t$ in Ampère's law, the four equations yield profound consequences and rightly named after Maxwell. The equation $\nabla . \mathbf{E} = \rho/\varepsilon_0$ then expresses the fact that lines of **E** begin and end only on electric charges though under circumstances they may form closed loops. It should be stressed that in order to determine a vector uniquely from the given boundary conditions, one must know both its divergence and curl in the intervening space (cf., Helmholtz theorem). The equation $\nabla . \mathbf{E} = \rho/\varepsilon_0$ is not sufficient in a static or non-static case to determine the vector field **E** uniquely. In static case $\nabla \times \mathbf{E} = 0$, and in a non-static case, Faraday's law of induction, $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$, provides the extra equation required. It is to be noted that the latter equation is consistent with $\nabla . \mathbf{B} = 0$ everywhere, and hence the lines of force of **B** behave like streamlines in an incompressible fluid and do not end. They can only form closed loops.

(ii) Displacement current: In free space, the electric current can only be viewed in the form of a displacement current. But in a polarizable medium, both types of currents can coexist. It is important to illustrate the significance of the displacement current and associated physical implications. For instance, the fact that copper behaves like a good electric conductor for the propagation of electromagnetic waves well up to the frequency of the ultraviolet (UV) light, but it behaves like a dielectric at higher frequencies, is a penetrating example for the beginners.

For an electromagnetic wave propagating in an electric conductor, the conduction current density is $\mathbf{J}_c = \sigma \mathbf{E}_0 e^{j(k.r - \omega t)}$ and the displacement current density is
$\varepsilon \partial \mathbf{E}/\partial t = - \varepsilon\omega \mathbf{E}_0 e^{j(k.r - \omega t)}$. Thus the magnitude of the ratio of the displacement to the conduction current densities is equal to $\omega\varepsilon/\sigma$. Taking $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$ F/m and $\sigma = 5.8 \times 10^7$ mho/m for copper, this ratio becomes $\approx 10^{-18}\ \nu$ where ν is the frequency given by $\omega = 2\pi\nu$. Since $\nu = 10^{16}\ s^{-1}$

for the UV, the displacement current in a good electric conductor is small compared to the conduction current at frequencies below the UV. The displacement current overtakes the conduction current at higher frequencies, e.g., X-ray frequencies.

(iii) As a third example, it is seen that beginners find it difficult to answer the question: A 50 Hz alternating current is flowing in a wire. Is it a displacement current?

The answer is that it is almost entirely the Ampère conduction current but it does contain a very small amount of displacement current (At 50 Hz, the ratio of the displacement to conduction current densities is equal to $\approx 5\mathrm{x}10^{-17}$). Similar examples (e.g., At what frequencies the sea water, with σ = 4.3 mho/m and ε_r = 80, will behave like a conductor or dielectric?) of the ratio of **J** and $\varepsilon\,\partial\mathbf{E}/\partial t$ being frequency- and material-dependent are recommended for the beginners.
(iv) Why does vacuum behave like a polarizable medium both electrically and magnetically? The modern quantum field theories have answered this question satisfactorily. Maxwellian ether has transformed into a structured vacuum in the hands of Schrödinger and Feynman. But that is another story altogether.

3.2 Summary

All of electromagnetism is contained in a set of four equations known as Maxwell's equations:

$$\nabla\cdot\mathbf{E} = \rho/\varepsilon$$

$$\nabla\times\mathbf{E} = -\frac{\partial\mathbf{E}}{\partial t}$$

$$\nabla\cdot\mathbf{B} = C$$

$$\nabla\times\mathbf{B} = \mu\,\mathbf{J} + \mu\varepsilon\frac{\partial\mathbf{E}}{\partial t}$$

$\nabla\cdot\mathbf{E} = \rho/\varepsilon$ is true in general, in static (i.e., $\nabla\times\mathbf{E} = 0$) as well as in dynamic (i.e., $\nabla\times\mathbf{E} = -\frac{\partial\mathbf{E}}{\partial t}$) fields: Gauss' law is always valid. Since there are no magnetic charges, $\nabla\cdot\mathbf{B} = C$ is always true. Faraday's law $\nabla\times\mathbf{E} = -\frac{\partial\mathbf{E}}{\partial t}$ is again true in general. The last equation has something new, only the part of this equation (i.e. $\nabla\times\mathbf{B} = \mu\,\mathbf{J}$) holds for steady currents, but the correct general equation has a new part, i.e., $\mu\varepsilon\frac{\partial\mathbf{E}}{\partial t}$, that was discovered by Maxwell. We conclude this chapter by reiterating the crucial role of displacement current, $\varepsilon\frac{\partial\mathbf{E}}{\partial t}$, (Maxwell's real discovery), so that this set of four equations is justly known as Maxwell's equations.

Appendix: Some Unsolved Problems

(i) Show and demonstrate physically (i) $\nabla.(1/r^2) = 0$, (ii) $\nabla^2(1/r) = 0$, (iii) $\nabla(r^2) = 2r$, where $\mathbf{r = xi + yj + zk}$ is the position vector.

(ii) Prove that (i) $\nabla. \nabla \times V = 0$ (ii) $\nabla \times \nabla f = 0$; where V and f are respectively vector and scalar fields.

(iii) Prove that $\nabla. (\mathbf{r}/r^2) = 0$. Will it be valid at r=0? Explain why?

(iv) A test charge particle is moving from (0,4) to (0,-4) through a part of circle represented by $x^2 + y^2 = 16$.
The electric vector field is defined by a function $E = x\, i + (x^3 + 3\, x\, y^2)\, j$ in the space at each of its point.
Calculate the total work-done in this field while test charge is moving on the defined path.

(v) Find out the total electric flux from a cube $0 < x, y, z \leq 2$ m, containing volume charge density $\rho = x^3\, y^3\, z^3$ micro Coulomb m^{-3} and $\varepsilon_o = 8.86\times10^{-12}\ C^2/N.M^2$.

(vi) A 3-D slab is defined by (x,y,z) = [(0,0,0), (2m,2m,10m)]. The electric field vectors are passing through this slab, and this vector field is defined as E = (x 2 yz) i + (y 2 xz) j + (z 2 xy) k. Find the total flux enclosed with the slab.

(vii) A magnetic field vector is defined on a line as follows B = (x 2 +2 y) i+(3 x−y 2) j. The line is represented by y=5x+3 from (0,3) to (1,8). Estimate the line integral of this vector field over the given line segment.

(viii) A magnetic field is defined over a rectangular loop-path having the coordinates (x,y)=[(-2,-1); (+2,+1)] as B = (x 2 y) i + (y 2 x) j. What will be the curl of this field, and in which direction it will work? Estimate
its integration and thus magnitude over the enclosed surface.

(ix) In a cylinder of radius 'a', the current density is given by an expression $J = k\, r \cos\phi$ over (r, ϕ) surface (k is constant). Calculate the total current (I) in the cylinder. Can such cylinder produce any magnetostatic field at any given point outside? If yes then how? If no then why?

(x) Calculate the ratio of conduction current density to displacement current density (i) in a good conductor for which the conductivity is 3.8×10^7 S/m and the relative permittivity is 1.0, and (ii) a good dielectric for which the conductivity is 1.2×10^{-9} S/m and the relative permittivity is 5.0. The frequency of the time-varying field intensity that varies with time is 100 MHz.

References

1. Electricity and Magnetism: E. M. Purcell and D. J. Morin
2. Electricity and Magnetism: A S Mahajan and A A Rangwala

Module - V
Electromagnetic Waves

Lectures: 02

Introduction:

The module is made as an interactive session between the students and the instructor. There are two parts of it, the first part deals with the basic understanding of the Maxwell's equations, particularly the dynamical electric and magnetic fields. The second part deals with some concepts of the electromagnetic waves. We have kept some of the information to be filled by the students, such that they get involved in a bit of self-study and revise their concepts.

The instructor starts with asking a few conceptual questions. These questions can be displayed in the board or as a power point presentation and the instructor can pick up random students to try answering the questions. The discussion then follows based on these questions and the instructor explains concepts around them and provides a few facts also about the topic.

Questions to be asked before starting the module

The module for electromagnetic waves is designed with a set of questions in mind. These questions are not written as a separate set, but comes within the flow of the class and helps in having a proper interaction with the students as the class progresses. Here teachers ask these questions and give time to the students to work out the problems. If they are not able to work it out then the teachers help them do so. There are a total of such ***9 questions*** *in this modules*

1 PART – I

1.1 Introduction to Maxwell's equations

The theory of electrodynamics is a theory where we can find out the electric and magnetic fields in presence of charged particles or currents. Using the values of the electric and magnetic fields we then can calculate the trajectory of charged particles. The theory was devised over many years, initially for stationary charges and currents by Coulomb, Gauss and Ampere and later for moving charges and changing currents by Faraday and Maxwell. Let me first put in the board the equations of electrodynamics. These were written together by James Clark Maxwell and hence are called Maxwell's equations.

$$\oint \vec{E} \,.\, d\vec{S} = \frac{Q}{\epsilon_0},$$

$$\oint \vec{E} \,.\, d\vec{l} = -\frac{\partial \phi_B}{\partial t},$$

$$\oint \vec{B} \, . d\vec{S} = 0 \, ,$$

$$\oint \vec{E} \, . d\vec{l} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial \phi_E}{\partial t} \, , \qquad (1)$$

You are already familiar with the black parts of the equations from the discussions in other modules. These parts let us evaluate the electric and magnetic fields ($\vec{E}$ and $\vec{B}$) given the stationary charges and currents. Let us bring our attention to the parts of the equations where we have a time derivative. Notice that the time derivatives are written with a nabla sign (∂), this means that the quantities ϕ_B and ϕ_E are functions of time, but they are also functions of space. Here (the part of the equations in red color) only the time variations are differentiated over. The second equation is essentially Faraday's law, you have already learnt about it. We shall discuss it again. The quantity ϕ_B is the flux of the magnetic field $\vec{B}$ through an area. For example, if there is an uniform magnetic field $\vec{B}$ T (Tesla) along z direction and we want to calculate the flux through a surface of cross-section 5 square meter in the x-y plane, the value of ϕ_B will be 5 T m^2. The flux of electric field also can be calculated in exactly the similar way.

1.2 Walk through a Problem: The dynamical magnetic field

[1]This example is quite non trivial and has to do with our ill notion abut non conservative fields. Please go through it slowly and step by step.

Let us consider the following example of a circuit. We have a battery of e.m.f 10 V and two resistances of R_1 = 1 kΩ and R_2 = 9 kΩ arranged in series as shown in figure (1). We have attached two identical voltmeters one between A and B and other between E and F in the circuit. We shall call the rst one as V_1 and the other as V_2 from now on.

- **Question:** What is the magnitude of the current going through the circuit?

  Wait for the answers from the students. Since this is quite trivial, they will mostly come up with right answers. But still explain to the class the answer once.

Total e.m.f. in the circuit is from the battery of V = 10 V and total resistance is R = R_1 + R_2 = 10 kΩ: Hence, the current is I = V/R = 1 mA.

[1]This example is borrowed from the lectures of Prof. Walter Lwein, MIT

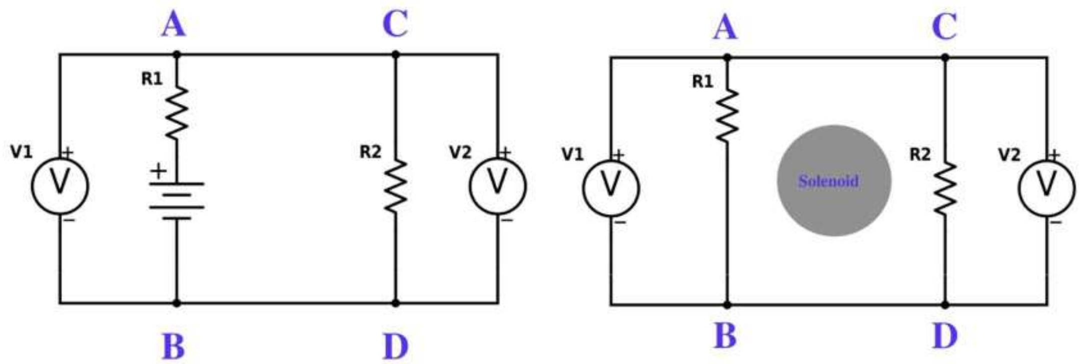


Figure 1: Circuit diagrams for the problem discussed in text. The left one has a battery and the right one uses a varying magnetic field for the required e.m.f. using a solenoid. For both the circuits the voltages are measured between A and B and C and D terminals using the voltmeters V_1 and V_2.

- **Question:** What would be the readings in the voltmeters V_1 and V_2?

  Again wait for the students to answer this. Check if they notice that they will have same reading and emphasize that fact. This is where the next example will not be trivial. You should also ask the students to actually calculate the voltage reading expected from the two voltmeters. After some discussion, explain the answer step by step as follows.

  It is rather obvious that both the voltmeters will read same voltage values, is not it? They are apparently connected to the same parts of the circuit, that is, A and C are the same points in the circuit, so are B and D. The voltmeter V_2 has resistance R_2 across the points B and D. The current in the circuit is 1 mA, which flows through this resistance. Hence, the voltage drop is $V_2 = 9$ V. On the other hand in between A and B there is the battery of 10 V and the resistance $R_1 = 1$ kΩ. The voltage drop across R_1 is 1 V, hence, $V_1 = 9$ V. They match exactly.

  If you are wondering what this example has to do with the Maxwell's equations, then here is the answer to that. We do not have any magnetic field in our problem (you can even have a constant magnetic field and the following discussion will work) and the e.m.f. provided by the battery is also constant. The left hand side of the second Maxwell's equation is what gives the e.m.f and as there is no changing magnetic field, the right hand side is zero. The close loop integral in the left hand side of this equation is to be evaluated along the circuit and the fact that it is zero is just the Kerchief's law of voltage. That ensures that the total voltage drops calculated by two voltmeters (which are apparently connected between same two points in the circuit) are exactly the same.

  Let us change the example a little bit. We know, there is a different way of producing e.m.f. in a circuit other than using a battery. That is by the application of Faraday's law, which says that a time varying magnetic flux creates an e.m.f.. So what we may do is that we remove the battery in the circuit and place a solenoid as shown in the right hand side of figure 1. We pass a time varying current through the solenoid. It creates a time varying magnetic field (see Ampere's law is in action here). This varying magnetic field produces a varying magnetic flux ϕ_B in the area given by the wire loop ABCD, which produces an e.m.f. in the circuit, just as the second Maxwell's equation say. At some point, the e.m.f. produced this way matches 10 V.

- **Question:** What is the magnitude of the current going in the circuit?

  Wait for the answers from the students. Since this is quite trivial, they will mostly come up with right answers. But still explain to the class the answer once. Emphasize that the current is not constant but varying, this is the value of the current at a moment when the produced e.m.f is 10 V.
  The e.m.f. in the circuit changes with time. When the e.m.f. in the circuit is V = 10 V and total resistance is $R = R_1 + R_2 = 10\ k\Omega$: Hence, the current is $I = V/R = 1$ mA.

- **Question:** What would be the reading for the voltmeters V_1 and V_2?

  We expect many students will not be able to answer this properly. Make them use the same steps as before to see that the voltmeters now have different readings. Then explain why they have different reading. Also show that the total voltage measured by them adds up to 10 V, which is what it should be.

  For the voltmeter V_1, there is the resistance R_1 and a current of 1 mA is passing through it. So the voltage drop it show is $V_{left} = 1$ V. For the voltmeter V_2, the resistance is R_2 and a same current of 1 mA is passing through it. So the voltage drop it show is $V_2 = 9$ V. Wait a second, the voltmeters are still connected through the terminals A, B, C and D. But now they read differently! This is because now Kerchief's law of voltage no longer holds, the closed loop integral in the left hand side of equation two is no longer zero! This is where you step into the dynamics.

1.3 Walk through a Problem: The dynamical electric field

[2]Let us first rewrite the fourth Maxwell's equation the following way

$$\oint \vec{B} \,.\, d\vec{l} = \mu_0 I + \mu_0 \frac{\partial \phi_D}{\partial t}\ , \tag{2}$$

Where $\vec{D} = \epsilon_0 \vec{E}$ is called the electric displacement vector and $\phi_D = \epsilon_0 \phi_E$. The equations we are considering here are the Maxwell's equations in free space, if we have to write equivalent equations inside a material medium, the concept of electric displacement vector becomes important. Here, we introduce this to tell you how certain concepts gets its name from this.

[2] This example is borrowed from NCERT Text book

To start with, let us again consider a circuit. It has a battery of e.m.f V volt, a resistance of R ohm. We have two circular coper plates P and Q attached in the circuit in a way that their planes stay parallel. Both of these has a cross sectional area A and the distance between the plates d can be adjusted. There is a probe at a distance r from the wire at position R. This probe can measure the magnetic field B at that point. [An example of such a probe is a Hall probe, which measures the magnetic field making use of the Hall effect.] We have kept d to zero initially. See left part of figure 2 for reference.

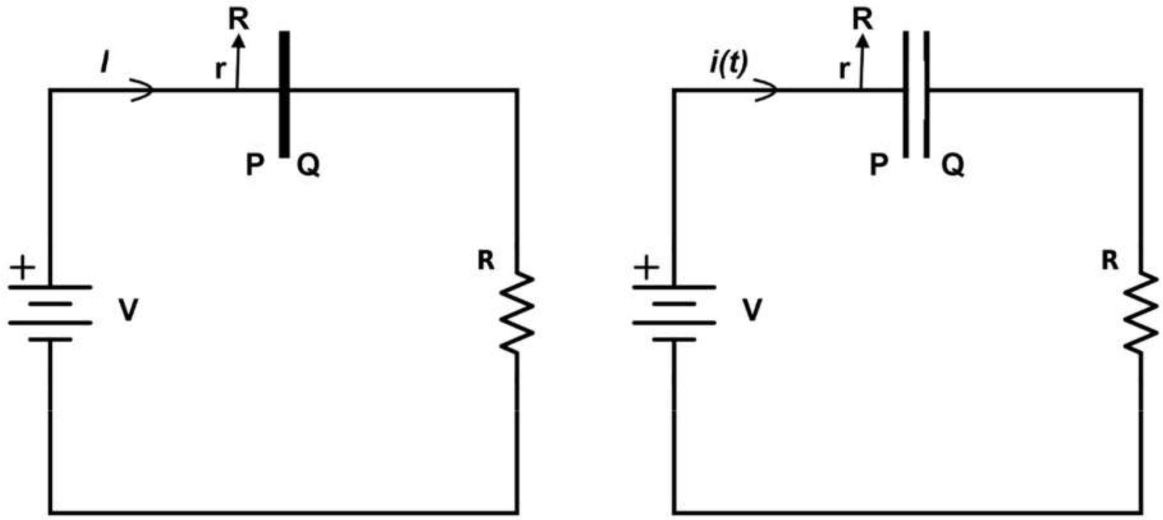


Figure 2: Circuit diagrams for the problem discussed in text. The left one has the plates P and Q kept without any distance between them while in the right figure they are kept at a distance d apart. Magnetic field measurements are done using a probe at the position R, a distance r away from the wire.

- **Question:** What is the magnitude of the current going in the circuit?

  This is straight forward. Since d = 0 the plates are touching each other. The current t will be V/R.

- **Question:** What is the magnitude of the magnetic field read by the probe at R?

  If the students have understood an earlier module where the Ampere's law is discussed, then they will be able to answer this. Still we shall go through this part once more here.

  We know how to answer this. There is a constant current in the circuit and no electric field (or at least time varying electric field) so we can use Ampere's law. We take a loop around the wire at R with a radius r. By symmetry the only component of the magnetic field is along the direction of the loop. This allow us to calculate the integral in the left hand side of the third equation easily, it is B2πr, where B is the magnitude of the magnetic field. In the first term of the right hand side of equation 2 (the second term is zero here as there is no electric field anywhere here that is changing with time), we have the total current that passes through an area of which our loop is a boundary. Take this area in any way you like, the total current pass through it is I. Hence, clearly

$$B2\pi r = \mu_0 I, \qquad i.e\ B = \frac{\mu_0 I}{2\pi r}. \tag{3}$$

Let us now move the two plates apart and keep d nonzero. The plates P and Q start to act like a parallel plate capacitor. As time progresses, the capacitor charges up. That reduces the current in the circuit and eventually the current goes to zero. As the capacitor charges up an electric field is developed across the plates. So the first question is following:

- **Question:** At a given time if the capacitors have a charge of q, what is the magnitude of the electric field developed across them?

Again we would expect a significant part of the student will answer this. Instructors need to help the students to understand that the plates are charged and they can use Gauss's law (and symmetry in the system) to find the electric field.

To solve this problem you have to make use of the first equation of Maxwell, that is the Gauss's law. I am sure you have already done it in an earlier module. To solve this, we consider a cylinder with its axis perpendicular to one of the plate, say P. Radius of the cylinder is smaller than the radius of the plate (see figure 3).

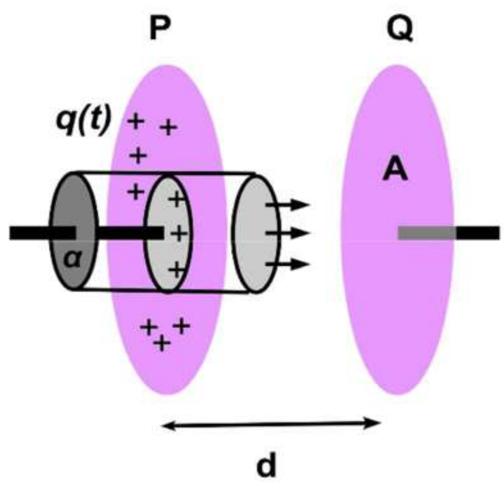


Figure 3: Schematic for calculation of the electric field between plates P and Q using Gauss's law. The plate P is charged to q(t) at a given time leaving on it a surface charge of q(t)=A. A gaussian pillbox is considered with area.

If the cross section of the cylinder have an area, then I leave it to you to show that

$$E(t)\alpha = \frac{q(t)\alpha}{\epsilon_0\, A}. \tag{4}$$

See I have written E(t) and q(t), since they both changes with time and above result is valid for a given time. A more advanced reader will note that the above expression may not be exactly correct (we have to consider retarded fields), but for the present purpose this expression is ne. Riding on this expression we can also calculate the total flux of the electric field through an area same or bigger than the plates P and Q as

$$\phi_E = \frac{q(t)}{\epsilon_0} \tag{5}$$

How much is the second term of equation 2?

$$\mu_0 \frac{\partial \phi_D(t)}{\partial t} = \mu_{0i}\ \ (t). \tag{6}$$

Here we have used the fact that the time derivative of the charge is a current. But notice carefully, this is not a current owing in the circuit! This is the time change of the charge accumulated in the capacitor. As this current is a direct consequence of the change in the displacement vector, we have called this as i_D(t), the displacement current. Let us trace back once and see how much is this current.

- **Question:** What is the magnitude of the magnetic field read by the probe at R?

This is where the main concept comes in. Convince the students that what method they may adopt to calculate the magnetic field at probe R, the probe will read a value irrespective of how we calculate the field there. This means if we get to two different results using two different methods, the values we get must be equal.

We see now answering this not as straight forward as before (when d was zero). That is because now there is a changing electric field in between the plates P and Q and that also may give rise to a contribution. let us work this through slowly. We know how to solve this, we consider a loop around the wire at R with a radius of r (just as before). The left hand side of the equation 2 still reads then $B2\pi r$. To calculate the right hand side, we need to attach a surface to this loop. Previously whatever we choose this surface as, it will be such that the current carrying wire will go through it once (or odd number of times) and the total current that goes through it will be I. Now if we choose the surface as in the left hand side of the figure 4, then it is still like that. Through this surface there is no electric field and hence there is no contribution from the second term in the right hand side. The result for us remains the same, i.e:

$$B(t) = \frac{\mu_0 i(t)}{2\pi r}, \tag{7}$$

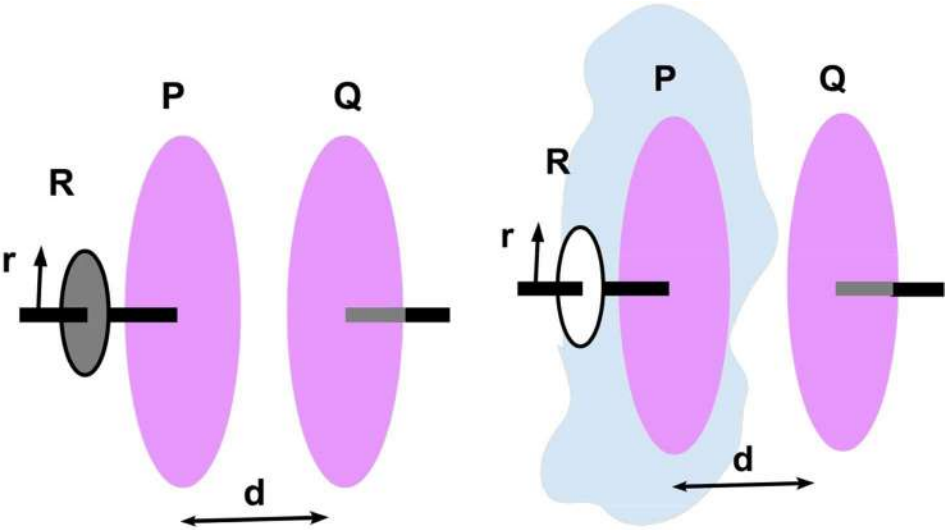


Figure 4: Two ways of calculating the magnetic field at R. We have considered a circular loop around the wire with radius r. Line integral of the magnetic field is calculated for this loop. In left figure a surface is attached to this loop (the grey circular area), through which a current of *i(t)* passes. In the right figure the surface is different, it encloses the plate P. No current pass through this, but there is a changing ϕ_D through this surface.

Only is the fact that the current is now changing with time and we need to take the current at the exact same time as when we calculate the field. A more advanced reader will note that the above expression may not be exactly correct (we have to consider retarded fields), but for the present purpose this expression is ne.

But wait a moment. We might have chosen a surface as in the right hand side of the figure 5. It is completely valid to choose such one. Now there is no current through this surface and the first term gives zero contribution. But there is a changing electric field and flux. That will contribute through the displacement current. Now we can use our result in equation 6 to calculate the magnetic field as

$$B(t) = \frac{\mu_0 i_D(t)}{2\pi r}. \tag{8}$$

If we do an experiment, the probe at R should read a number for the strength of the magnetic field at a given instance of time. It is independent of what surface we choose to calculate! That means, whatever we choose

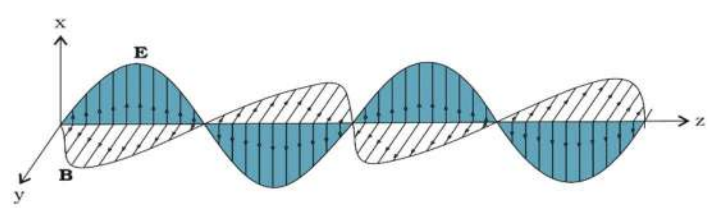


Figure 5: Schematic of a snapshot of the electric and magnetic field of an electromagnetic wave propagating along z direction. The picture is adopted from NCERT book.

as our surface, we should get the same value for the magnetic field strength! Clearly, if the magnetic fields in equation 7 and equation 8 are same then the current through the circuit $i(t)$ must be same as the displacement current $i_D(t)$. Physically this means that the rate of charge accumulation in the capacitor must be same as the rate of ow of charge in the circuit. That has to be right? The charges that accumulates in the capacitor comes from the circuit only! The charges cannot be created or destroyed at any point in space. This is called the principle of local conservation of electric charge.

The term with the displacement current was added to the Ampere's law by Maxwell. This term, as discussed above is the consequence of the conservation of the charge. And consequence of this term is the path towards the electromagnetic waves, which we shall discuss next.

2 PART -II

2.1 Electromagnetic Waves

Question: Here is a question for you straight away. You asked your parents to give you a FM radio set, but they denied. You are clever enough to make one yourself! So you start to design the electronics. A radio detects the electric field of the electromagnetic waves. In your receiver circuit you found you need some inductors and capacitors. Question is, what is the maximum time constant your circuit can have to detect the FM signal from the nearby station? What should be the typical size of the antenna that you will use?

We all know the FM stations broadcast at frequencies near to 100 MHz. Your receiver need to be as fast as to detect this signal. Hence, the time constant of the circuit has to be less than 1/100 MHz = 10^{-8} sec. The antenna used need to be smaller than one wavelength, the wavelength of 100 MHz signal is $\lambda\ c/v = 3m$.

Reason for the circuit need to work as fast as above is the fact that the electric field of the wave oscillates rapidly in time. If we measure its amplitude at a given point in space (like by the antenna) the amplitude changes as time progresses. So you need to measure it before it change much. Interestingly, if you move with the radio, you will again see a change in the amplitude of the electric field as you move. That apparently is the reason why you need an antenna of similar size of the wavelength or half or quarter of one wavelength). Let us discuss a few things about the electromagnetic waves here.

- Electromagnetic wave is disturbances in electric and magnetic fields connected in a very special way. These disturbances change with time and spread out in space. One easiest way we can produce an electromagnetic wave is by making a charge move back and forth in a wire. As the charge moves it accelerates and decelerates. That produce a time varying current. We have seen already that a time varying current produces a time varying magnetic field. And then Faraday's law tells us that a time varying magnetic field will create a time varying electric field and e.m.f.. This e.m.f. in turn creates a time varying current again. This procedure repeats. As a result, the information of the oscillation of the charge gets embedded in the electric and magnetic fields that lies to large distances and oscillates. This oscillating electric and magnetic fields constitutes the electromagnetic wave. A part of the energy used to make the charge oscillate is used up to make the fields oscillate and the energy (as well as a momentum) propagates with these oscillations in different directions.

- A wave (electromagnetic, sound or anything else) can be mathematically represented as (for plane wave going along z direction)

$$W(x,t) = A_1 f(z - ct) + A_2 g(z + ct) \tag{9}$$

where c is the speed at which the information propagates, A_1 and A_2 are the amplitudes. The first term in the right hand side represents a wave propagating in the direction of the positive z axis and the second part is the wave propagating in the direction of the negative z axis. It is easier to understand the wave propagation for waves with only one frequency. For a monochromatic and plane electromagnetic wave propagating along the positive z direction, the electric and magnetic fields would be

$$\vec{E}(z,t) = E_o cos(kz - \omega t)\hat{x} \qquad B(z,t) = B_0 cos\, A_2 g(kz - \omega t)\hat{y}. \tag{10}$$

Since the electric and magnetic field in this example does not depend on the x and y coordinates, the electric field vector measured at any value of z and t is same for all x - y. That is the value of the fields are same over a plane. This is why this is an example of a plane monochromatic wave propagating along z direction. It is polarized along the x direction (definition of the polarization direction varies in literature, here we have adopted the convention that the wave is polarized along the direction of oscillation of the electric field). The quantity k is known as the wave vector and is related to the angular frequency! $\omega = 2\pi v$ (where v is the frequency) and the speed of the wave c and the wavelength λ in the following way:

$$k = \frac{2\pi}{\lambda} = \frac{2\pi v}{c} = \frac{\omega}{c}. \tag{11}$$

The electric and the magnetic fields both changes with time and as it propagates (in space). This is captured in the fact that they are function of both *t* and *z*. To see the effects individually, let us assume we are observing at a given point in space, say at $z = 0$. Then we see the electric (and magnetic) field oscillates in time. It oscillates at a pace determined by the quantity ω or v. The frequency v tells us the number of times the field oscillates in one sec. Think about the physical meaning of the wavelength in the similar way. Figure 5 shows the schematic of an electromagnetic wave captured at a particular time.

- Electromagnetic wave is a transverse wave. That means the disturbance in the electric and magnetic fields are perpendicular to the direction of propagation of the wave. In the above example the wave is propagating along the *z* direction, whereas the electric and magnetic fields are along *x* and *y* directions.
- Unlike any other waves, electromagnetic wave can travel through free space. It does not need a material medium. Speed of the wave can be written in terms of the constants ϵ_0 and μ_0. In SI

unit the speed of the electromagnetic wave in free space is taken to be 299,792,458 m s^{-1}. From our everyday experiences of waves, like that of sound or in the river, the speed of the wave depends on the speed of the observer or the source of the wave relative to the medium it propagates. But this is not true for the electromagnetic wave in free space, it does not even need a medium to propagate. One of the postulates of the Special Relativity by Einstein in 1905 was that the speed of the electromagnetic wave in free space is a constant and does not depend on the relative motion of the source and the observer. This had a far reaching consequence in physics. Speed of the electromagnetic wave in free space c is a fundamental constant of nature. Maxwell's equation gives $C = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$. The value of μ_0 is chosen to be $4\pi \times 10^{-7}\mathrm{H\,m^{-1}}$ and then ϵ_0 comes to be $8.854 \times 10^{-12}\mathrm{F\,m^{-1}}$. In a medium, the speed of the electromagnetic wave is slower by a factor of the refractive index of the medium. The refractive index is a consequence of the relative permittivity and permeability of the medium, in fact the refractive index $n\ is\ n = \frac{1}{\sqrt{\epsilon_r \mu_r}}$. Since most of the medium is not magnetic or μ_r is very close to one ew may approximately write $n = \frac{1}{\sqrt{\epsilon_r}}$.

2.2 Flavors of Electromagnetic Waves: Mental Aptitude

Frequency of the electromagnetic wave can be from a few Hz (oscillations per second) to a few 10^{20} Hz (in principle it can be anything!). At different frequencies/wavelengths the production and detection procedures are completely different. This is why we also give separate names to different frequency ranges. The part of electromagnetic spectrum at which our eye is sensitive to have wavelengths around 400 700 nm. This is called the visible range. Electromagnetic waves of different frequencies/wavelengths are used for different purposes in modern day life. Here is a table to summarize it.

Name	Wavelengths [m]	Frequency [Hz]	Used in
radio	> 0.1		Communication
microwave	$0.1 \sim 1. \times 10^{-3}$		Cooking
infrared	$1 \times 10^{-3} \sim 700 \times 10^{-9}$		Heat detection
visible	$700 \times 10^{-9} \sim 400 \times 10^{-9}$		To see
ultravviolet	$400 \times 10^{-9} \sim 1 \times 10^{-9}$		Sterilisation from bacteria
X-ray	$1 \times 10^{-9} \sim 1 \times 10^{-10}$		Scanning through skins and security purposes
gamma rays	< 10 10		Medical purposes

Table 1: Different flavors of electromagnetic waves.

I have kept the columns with frequencies empty. Try to fill it up. Which of these have more frequencies than the modern day computer? The values of the frequency has a lot to do with the detection of the wave. Think about it. Also add more points in the application of different flavors.

Exercise:

1. Write the static limit of the Maxwell's equations.
2. Interchange the Electric field and the Magnetic field in the Maxwell's equation for the static case. What is the difference between the new sets of equations with the original Maxwell's equations?
3. Write Down the Maxwell's equations in a case where there is no current or charges. Now repeat the exercise in the first two questions.
4. Can you think of an experiment by which you can measure the free space permittivity and permeability?
5. We have used Ohm's law in one of the example. When do you think current in the circuit will not be directly proportional to the applied electromotive force/voltage (even for conductors)?
6. Fill up the frequency column in Table 1. Can you convert the frequencies to energy units? Write down the energies in terms of electron volts and Joules.
7. Suppose that you want to make a dipole antenna to receive X-Ray radiation, what should be size of the antenna? Do you think you can make such a device? Can you suggest an alternate way of measuring X-radiation?

References

1. Introduction to Electrodynamics: D J Griffith
2. Classical Electricity and Magnetism: Panofsky and Phillips

Module - VI
Wave Optics

Lectures: 02

Overview:

The module on the physical optics is designed to understand the wave nature of light in interference, diffraction and polarization phenomenon. It discusses that how the corpuscular model of light developed by Isaac Newton could not explain the phenomenon of reflection and refraction. In this module, the classification of interference phenomenon and their conditions are discussed. The Huygens's principle is discussed to describe interference and diffraction. It also discusses about the various applications of interference, diffraction and polarization in our daily life.

Outcome of the module:

After studying this module, the students would be able to understand the conditions needed to produce interferences; Identify interferences by wave front division or by amplitude division. They can explain the effects of superposition of waves in interference and diffraction. It provides the insight of the functioning of various optical systems based on interference, diffraction and polarization and applications in our daily life.

Pretest Questions.

Q.1 Is it possible to explain interference and diffraction by particle nature of light?

Q.2 What are coherent sources and can we produce them from an independent source?

Q.3 Explain whether the interference or diffraction or both take place in the diffraction phenomenon.

Q4. What are the differences in the intensity pattern on changing the number of slits on the grating surface?

Q5. What is polarization of electromagnetic waves and how it is useful in our daily life?

Introduction

Descartes gave the corpuscular model of light in 1637 and explained the laws of reflection and refraction of light at an interface. In the corpuscular model, the speed of light would be greater if the ray of light on refraction bends towards the normal. Isaac Newton further developed this corpuscular model of light. Later, the wave theory of light was developed by Dutch physicist Christian Huygens. The wave theory of light could explain the phenomenon of reflection and refraction; however, it predicted that on refraction if the wave bends towards the normal then the speed of light would be less in the second medium. This in the contradiction to the prediction made by using the corpuscular model of light. It was confirmed much later by experiment that the speed of light in water is less than the speed of light air confirming the prediction of the wave model. The wave theory of light was not easily accepted because of Newton's authority and also because light could travel through vacuum and it was felt that a wave would always require a medium to propagate from one point to the other. However, it was firmly established that light is a wave phenomenon after performing famous interference experiment by Thomas Young in 1801.After the interference experiment of Young in 1801, various experiments were carried out involving interference and diffraction of light waves. The only major difficulty was that since it was thought that a wave required a medium for its propagation, how light waves could propagate through vacuum. This was explained by Maxwell famous electromagnetic theory of light. Using wave equation, Maxwell calculated the speed of electromagnetic waves in free space and he found that the theoretical value was very close to the measured value of speed of light. Maxwell explained that light waves are associated with changing electric and magnetic fields and changing electric field produces a time and space varying magnetic field and a changing magnetic field produces a time and space varying electric field. The changing electric and magnetic fields result in the propagation of electromagnetic waves even in vacuum.

Interference of light

Today, we produce interference effects with little difficulty. In the days of Sir Isaac Newton and Christian Huygens, however, light interference was not easily demonstrated. There were several reasons for this. One was based on the extremely short wavelength h of visible light around 20 millionths of an inch and the obvious difficulty associated with seeing or detecting interference patterns formed by overlapping waves of so short a wavelength, and so rapid a vibration around a million billion cycles per second! Another reason was based on the difficulty before the laser came along of creating coherent waves, that is, waves with a phase relationship with each other that remained fixed during the time when interference was observed.

Classification of interference phenomenon

In general, the interference phenomenon is classified in two classes depending on the method of obtaining the coherent sources.

(i) Division of wavefront: The wavefront originating from a common source is divided in to two parts by employing mirrors, prism or lenses and the two wave fronts thus separated travel unequal distances and are finally brought together to produce interference. The devices used to obtain the interference by division of wavefront are a Fresnel biprism, Lloyd's mirror, Billet's split lens and laser.

(ii)Division of amplitude: The amplitude of the incoming beam is divided into two or more parts either by partial reflection or refraction. These divide parts travel different paths and are finally brought together or produce interference. Examples like interference in thin films, Newton's rings, Michelson's interferometer and Fabry-Perot interferometer etc.

Conditions for sustained interference of light

The intensity must be maximum and zero at the points corresponding to constructive and destructive interference to obtain perfect interference. The following conditions must be fulfilled to obtain well defined and observable interference:

(i) The frequency of light emanating from the two interfering sources must be same. If their frequency is not same, the phase difference between the interfering waves will vary continuously. Consequently, the resultant intensity at any point will vary with time alternately maximum and minimum.

(ii)The amplitude of interfering waves must be equal. If this condition is not satisfied, the minimum intensity will not be zero and so the contrast of the fringes formed will not be good.

(iii) The distance between the two coherent sources should be small as much as possible. The distance between the consecutive dark or bright fringes is inversely proportional to the distance between the two sources. Hence, the fringes of maximum and minimum intensity will lie so close together that the fringes will not be separately visible.

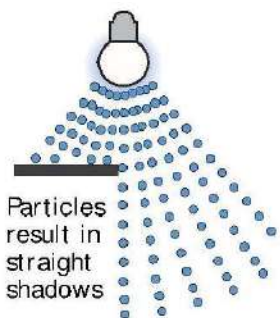
(iv)The two sources should be narrow. A broad source is equivalent to a large number of fine sources lying side by side. Each pair of fine sources will give its own interference pattern. The fringes of different interference patterns will overlap causing a general illumination.

(v) The two interfering waves must be propagated along the same line. If this condition is not satisfied, the vibrations will not be along a common line.

<u>Are Photons Particles or Waves</u>?

Newton believed that light was particles:

Light travels in straight lines!



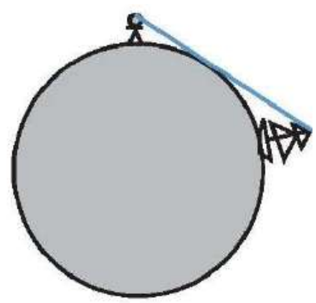


Image is in the public domain

Image by Jeffrey Jose

A wave is a vibration of some medium through which it propagates, e.g., water waves, waves propagating on a string

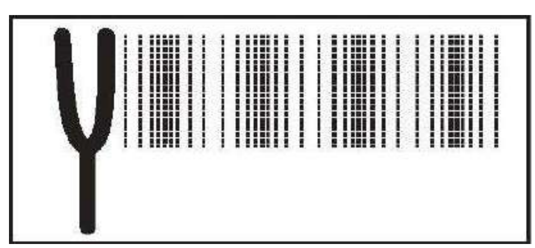


Image by May Wong http://www.flickr.com/photos/maywong_photos/5569932571/ on flickr

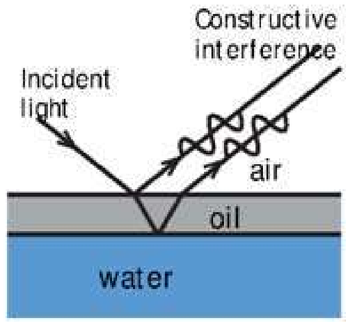
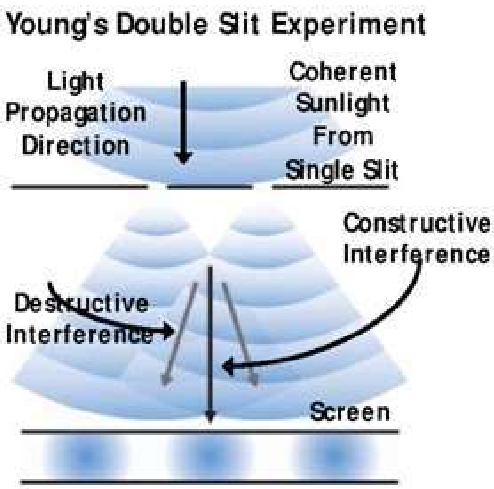


Image by Yoko Nekonomania http://www.flickr.com/photos/nekonomania/4827035737/ on flickr



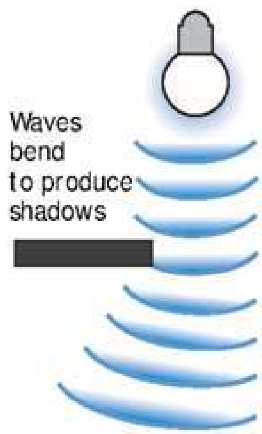


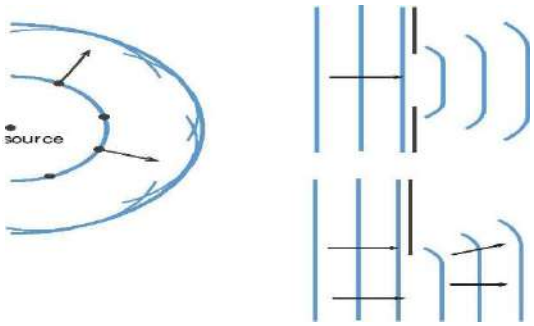
Image by Pieter Kuiper http://commons.wikimedia.org/wiki/File:Compact-Disc-spectrum-Mercury.jpg on wikimedia commons

- Huygens asumed that light is a form of wave motion rather than a stream of particles.
- Huygens's Principle is a geometric construction for determining the position of a new wave at some point based on the knowledge of the wave front that preceded it.
- Al points on a given wave front are taken as point sources for the production of spherical secondary waves, called wavelets, which propagate in the forward direction with speeds characteristic of waves in that medium

– After some time has elapsed, the new position of the wave front is the surface tangent to the wavelets

As you might expect, the heuristic idea of Huygens can be fully justified through various derivations associated with the Maxwell equations.

Huygens' Principle Explains Diffraction

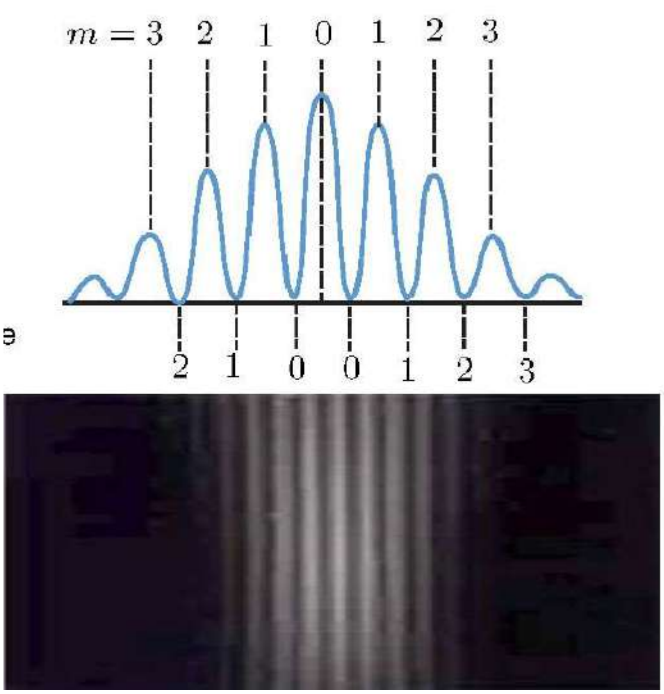


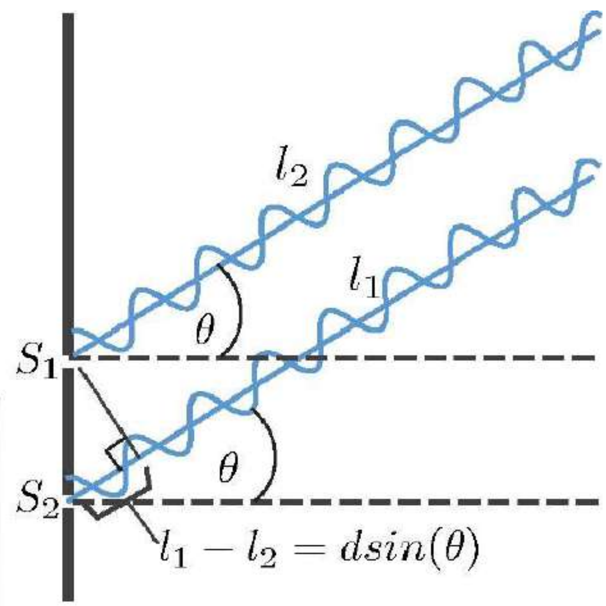
Interference

$d\sin(\theta) = m$, $\quad m = 0, \pm 1, \pm 2, \dots\dots\dots \quad$ *Constructive*

$d\sin(\theta) = (m + 1/2)\,\lambda$, $\quad m = 0, \pm 1, \pm 2, \dots\dots\dots \quad$ *Destructive*

Where m is the order of an interference fringe





Diffraction

Diffraction is why we can hear sound even though we are not in a straight line from the source — sound waves will diffract around doors, corners, and other barriers.

The amount of diffraction depends on the wavelength, which is why we can hear around corners but not see around them.

Single-Slit Diffraction

•According to Huygens's principle, each portion of the slit acts as a source of light waves.

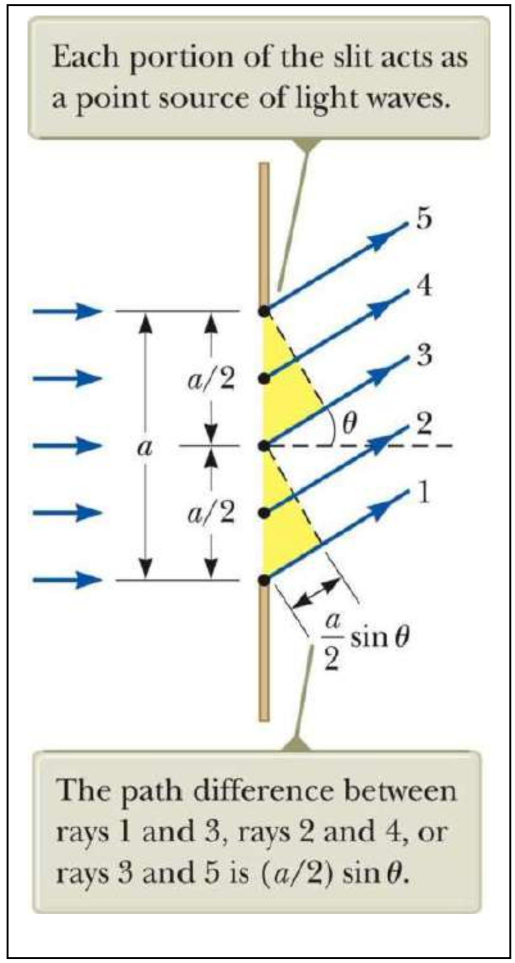
•Therefore, light from one portion of the slit can interfere with light from another portion.

•The resultant light intensity on a viewing screen depends on the direction θ

•The diffraction pattern is actually an interference pattern.

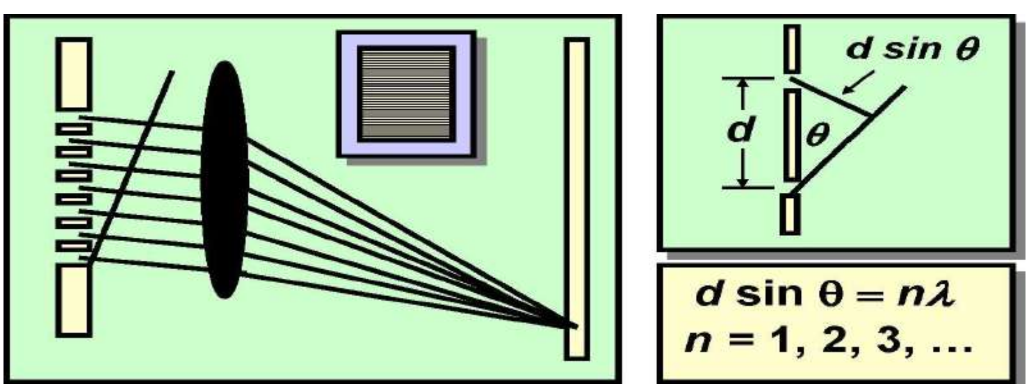
– The different sources of light are different portions of the single slit.

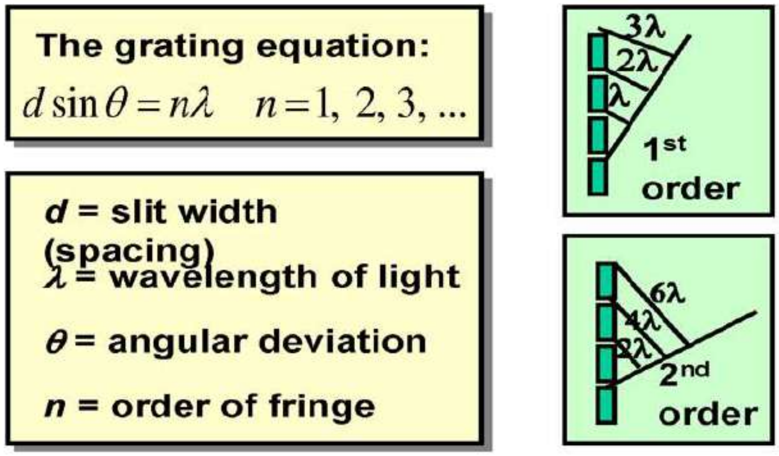
The results of the single slit cannot be explained by geometrical optics



The Diffraction Grating

A diffraction grating consists of thousands of parallel slits etched on glass so that brighter and sharper patterns can be observed than with Young's experiment. Equation is similar.





Resolution

Diffraction through a small circular aperture results in a circular pattern of fringes. This limits our ability to distinguish one object from another when they are very close.

The location of the first dark fringe determines the size of the central spot:

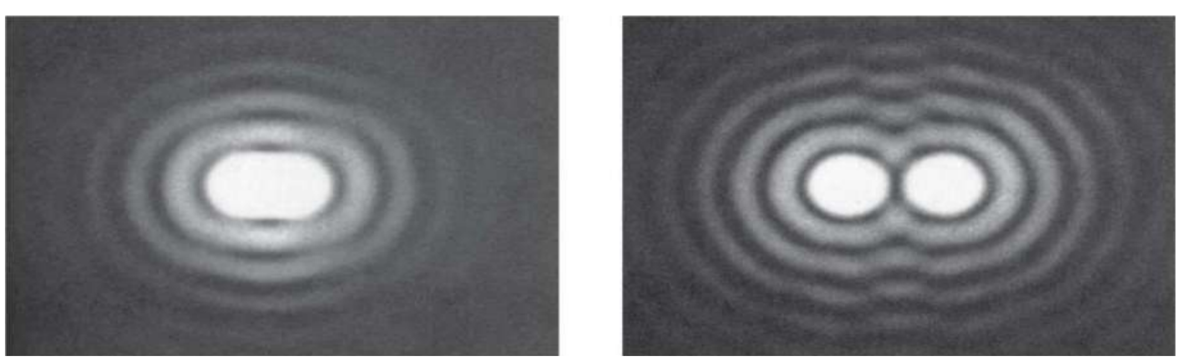
$$\sin\theta = 1.22\frac{\lambda}{D}$$

Rayleigh's criterion relates the size of the central spot to the limit at which two objects can be distinguished:

If the first dark fringe of one circular diffraction pattern passes through the center of a second diffraction pattern, the two sources responsible for the patterns will appear to be a single source.

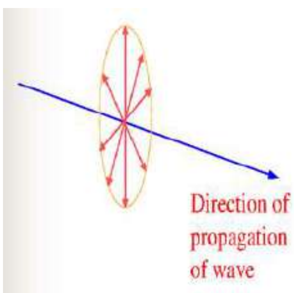
The size of the spot increases with wavelength, and decreases with the size of the aperture.

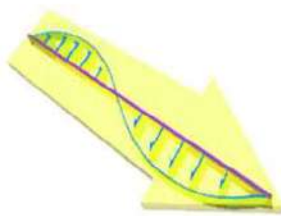
On the left, there appears to be a single source; on the right, two sources can be clearly resolved,

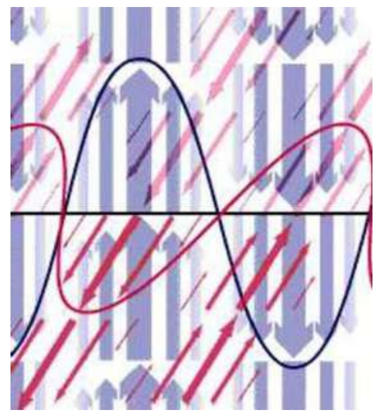


Polarization

- Polarization is a characteristic of all transverse waves.
- Oscillation which take places in a transverse wave in many different directions is said to be unpolarized.
- In an unpolarized transverse wave oscillations may take place in any direction at right angles to the direction in which the wave travels.





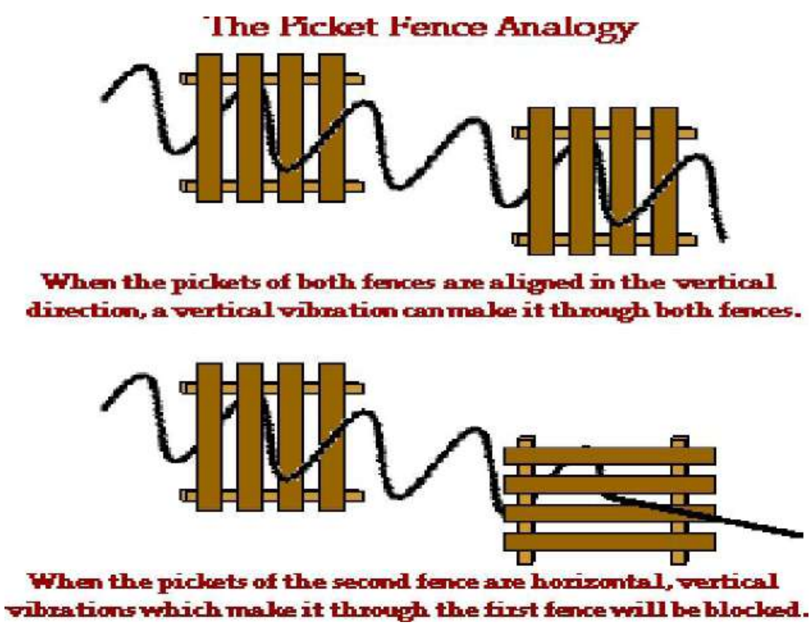


- **Polarization of Electromagnetic Waves**

- Any electromagnetic wave consists of an electric field component and a magnetic field component.

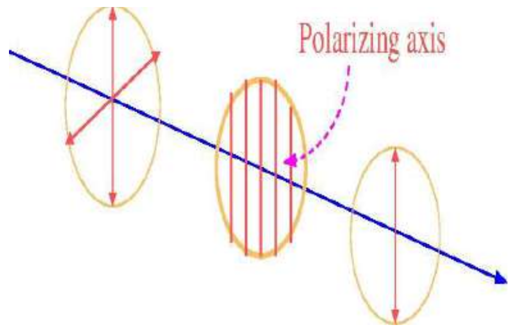
Polarization by Selective Absorption

• Polarization of light by selective absorption is analogous to that shown in the diagrams.



Polaroid

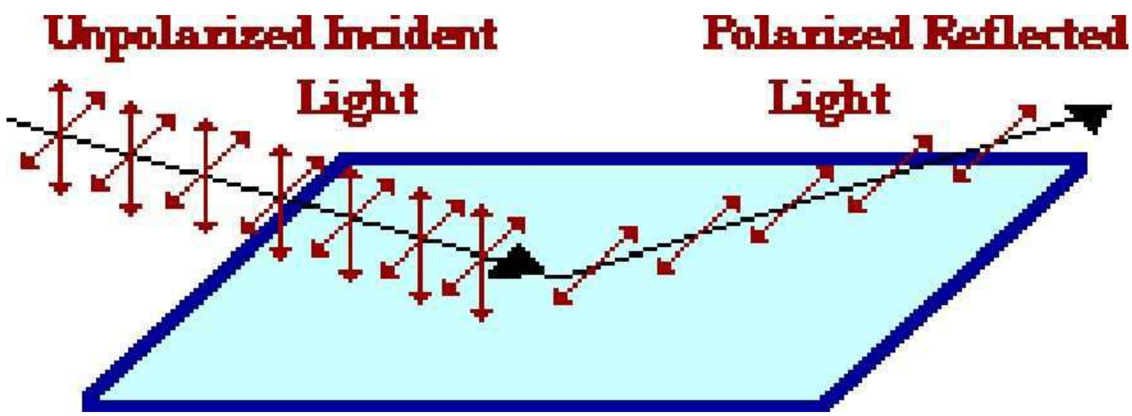
- A Polaroid filter transmits 80% or more of the intensity of a wave that is polarized parallel to a certain axis in the materials called the polarizing axis.



- Polaroid is made from long chain molecules oriented with their axis perpendicular to the polarizing axis; these molecules preferentially absorb light that is polarized along their length

<u>Polarization by Reflection</u>

- Unpolarizes light can be polarized, either partially or completely, by reflection.
- Amount of polarization in the reflected beam depends on the angle of incidence



Reflection of light off of non-metallic surface results in some in degree of polarization parallel to the surface

Post test Questions

Q1. Is the interference phenomenon inconsistent with the energy conservation? Explain.

Q2. What is Huygens's principle and how it is useful in explaining the interference and diffraction phenomenon?

Q3. What do mean by resolving power of an optical instrument?

Q4. Is it possible to explain the transverse nature of light by interference or diffraction phenomenon?

Q5. How polarized light are produced?

Q6. Is Brewster's angle same for all the medium to produce plane polarized light?

Q7. How one can have two refracted rays when a ray of light is incident on certain crystals?

Q8. What is the principle of working a Polaroid?

References

1. Optics: B K Mathur
2. Optics: Ajoy Ghatak

Module - VII

Semiconductor electronics: Materials, Devices and Simple Circuits

Lectures: 03

Objective:

In this module we will show the evolution of electronics from semiconductors. The students will learn regarding different semiconducting materials, electronic devices and simple circuits.

Pre requisite:

The students may be asked following questions before starting this module.

1. What are the differences between semiconductors and metals?
2. How the resistance changes with temperature in semiconductor?
3. How the resistance changes with temperature in metals?
4. What are the majority and minority charge carriers in *n*-type and *p*-type superconductors and why?
5. How are energy bands formed in solids?
6. How do you classify metals, semiconductors and insulators? How does one can understand them via electron energy band diagram?
7. Define an intrinsic and extrinsic semiconductor. What role does donor and acceptor impurities play in forming extrinsic semiconductor?
8. Describe how a *p-n* junction is formed and its characteristics in forward and reverse biased mode. How one can exploit these properties to use it for rectification of ac voltages?
9. Describe in brief the concept of filtering a time varying voltage signal to obtain a steady dc output voltage.

1. Introduction:

Controlled flow of electrons is the feature required in any device forming the building blocks of the electronic circuits. Prior to the discovery of transistor in 1948 such devices mostly were vacuum tubes containing many electrodes like cathode, anode, grids etc. The applied voltages at each electrode gave a large degree of control over the flow of electrons and hence the desired operation of the device can be ensured. Some examples of such devices are valves, triodes, tetrodes etc. each one with its specific function. These devices mostly require a vacuum inside a tube which houses the various electrodes hence they were bulky in size and difficult to operate comfortably. Moreover they consume high power, need high voltages for their operation (~100 eV) and have a limited life. With the discovery of modern solid state based semiconductors in 1930's and its applications in device manufacturing offered significant advantages over vacuum tubes in many of the areas mentioned above. Hence gradually the electronic industry replaced vacuum tubes with semiconductor devices extensively.

The figures and examples in this presentation have been obtained from NCERT Physics textbook for class 12th course of the central board of secondary education (CBSE), New Delhi.

2. Classification of metals, conductors and semiconductors:

2.1 On the basis of conductivity:

Based on values of resistivity ρ and conductivity σ (=1/ρ);

(i) *Metals: V*ery low resistivity (or high conductivity).
$\rho \sim 10^{-2} - 10^{-8}\ \Omega$ m
$\sigma \sim 10^{2} - 10^{8}$ S m^{-1}
(ii) *Semiconductors: R*resistivity or conductivity intermediate to metals and insulators.
$\rho \sim 10^{-5} - 10^{6}\ \Omega$ m
$\sigma \sim 10^{5} - 10^{-6}$ S m^{-1}
(iii) *Insulators:* High resistivity (or low conductivity).
$\rho \sim 10^{11} - 10^{19}\ \Omega$ m
$\sigma \sim 10^{-11} - 10^{-19}$ S m^{-1}

Semiconductors are mainly of two types:
(i) Elemental Semiconductors: e.g. Si and Ge.
(ii) Compound Semiconuductors: CdS, GaAs, CdSe, InP, anthracene, doped pthalocyanines, polypyrrole, polyaniline, polythiophene etc.

In this module we will restrict ourselves to elementary semiconductors particularly. The general concepts introduced here for discussing the elemental semiconductors, by-and-large, apply to most of the compound semiconductors as well.

2.2 On the basis of energy bands:

The overlap of electronic orbitals of atoms with each other leads to a continuous distribution of their energies in solids giving rise to band formation. Thus every atomic orbital of the solid becomes a part of some band inside the solid. The valence electrons of atoms form a band known as the valence band of the solid. Then by definition we have the valence band as being completely occupied at 0 K. The lowest unoccupied energy band is called the conduction band. The relative position of valence band and conduction band varies for distinguishing metals, semiconductors and insulators. An electron in a conduction band must be considered as free contributing to electrical conduction.

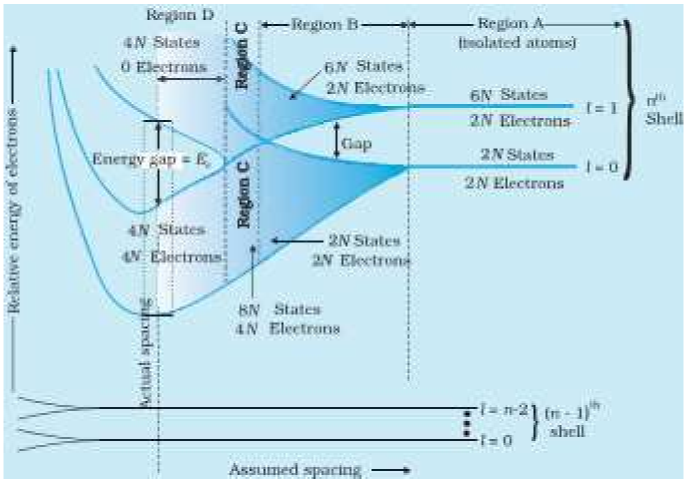


FIG. 1 *Illustrating the band formation in solids from atoms coming close to each other.*

Consider the band formation in Si or Ge crystal having N atoms. Each atom has a $2s^2 2p^2$ valence electronic configuration hence we have 4N valence electrons in the crystal in total. However every atom has 4 *p*-electronic states unoccupied hence the crystal has 4N unoccupied electronic states too. When the atoms come together during the formation of a solid the valence electronic states of neighbouring atoms overlap with each other giving rise to slight shift of their energies with respect to that inside the atom. It causes both downward and upward shift of their mutual energies. *When the number of atoms N is very large, ideally infinity, then the energy distribution for those electrons becomes continuous leading to the formation of energy bands.* These energy bands are most prominent for the valence electrons since they occupy outermost orbits and hence they strongly interact with each other inside the solid. The core electrons are well shielded from the electrostatic potential of the neighbouring atoms by the outermost electrons which keeps their wave function confined close to the nuclei of the atoms and hence their wave functions' overlap with those of the neighbouring ones is negligible; hence the band formation among them is insignificant. *The valence electrons form valence band and the unoccupied electronic states form the conduction band.*

The formation of the band is nicely illustrated in fig. 1. At large values of the separation between the atoms (region A) its electronic states are mostly influenced by the electrostatic potential of the nucleus of that atom itself and hence the overlap of the valence states is weak giving rise to well defined valence electronic energies corresponding to that atom. As one slowly decrease the separation the overlap of the valence states increases gradually leading to the spread of the electronic energy of the atom (region B). This amounts to the formation of a band from those states. Further decrease in the separation leads to the further downward (upward) shift of the lower (upper) end of the conduction (valence) band such that they both cross each other (region C). With continuous decrease in the separation leads to the reappearance of the energy gap E_g between valence and conduction band (region D).

It must be noted that the no. of occupied (unoccupied) states in a band are same as the no. of occupied (unoccupied) energy states of the atoms from where they were created.

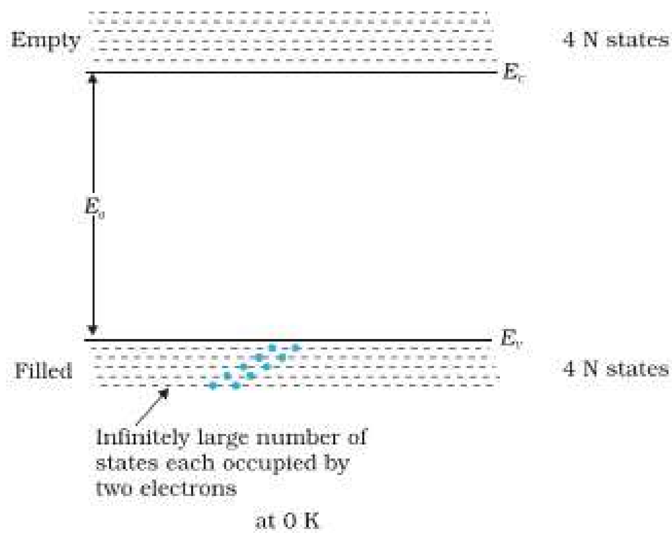


FIG.2 *The energy band position in a semiconductor at 0 K. The upper band, called the conduction band, consists of infinitely large number of closely spaced energy states. The lower band, called the valence band, consists of closely spaced completely filled energy states.*

The lowest energy level in the conduction band is denoted as E_C and the highest energy level in the valence band is denoted as E_V. Above E_C and below E_V there are a large number of closely spaced energy levels, as shown in fig. 2. The gap between E_V and E_C is called the energy band gap (Energy gap E_g).

It may be large, small, or zero, depending upon the material. These cases are discussed in the following figure.

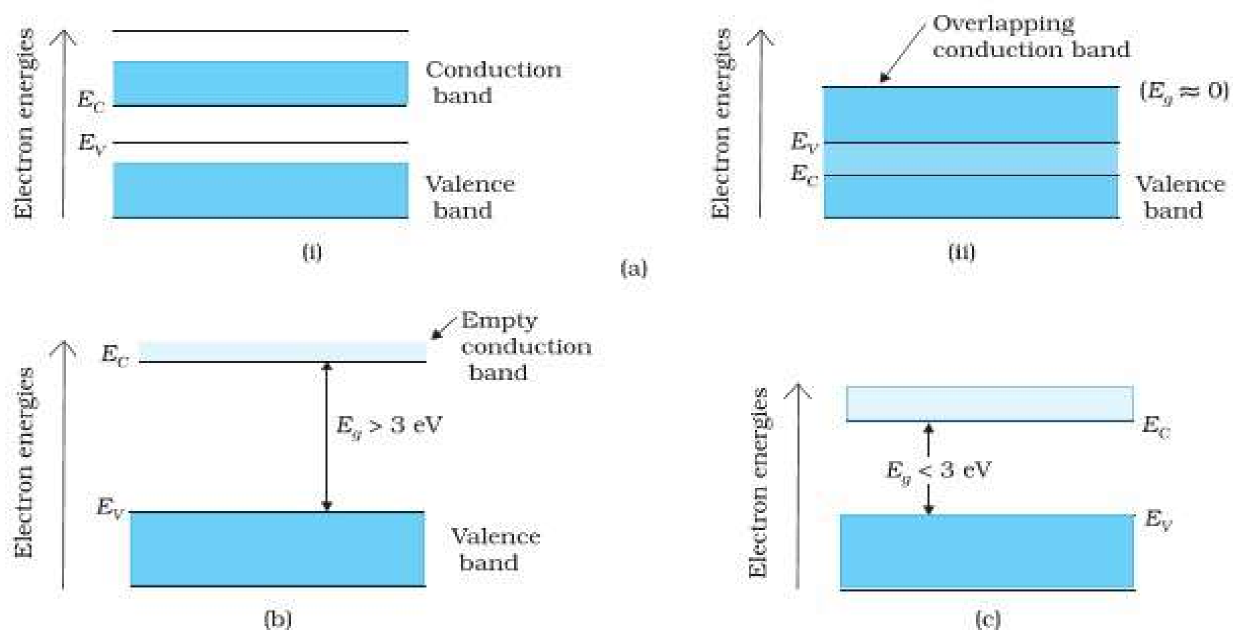


FIG. 3 *Difference between energy bands of (a) metals, (b) insulators and (c) semiconductors.*

Case I: This refers to a situation, as shown in (a) of fig. 3. Metals form either when the conduction band is partially filled and the valence band is partially empty or when the conduction and valence bands overlap. When one has either of the above situations an extremely small amount of energy given to the sample can easily excite large no. of electrons into the unoccupied states hence such materials easily conduct electricity and they are called as metals. The metals have high conductivity and low resistivity.

Case II: In this case, as shown in (b), a large band gap E_g exists ($E_g > 3$ eV). The valence band is completely filled and the conduction band is completely empty. One has to supply energy of the order of 3 eV in order to excite an electron from the valence band to the conduction band in order to generate a free electron. Such a high value of energy cannot be supplied under ambient conditions hence these materials do not conduct electricity and are called as insulators. A natural source of energy for causing excitation of the electrons in solids is the atmospheric temperature which can cause thermal excitations of the electrons, however, at practical temperatures e.g. room temperature $T = 300$ K the available energy to cause thermal excitations is ~ 25 meV $<< E_g$. Thus thermal excitation is inadequate in causing electrical conduction in insulators. The insulators have low conductivity and high resistivity.

Case III: This situation is shown in (c). Here a finite but small band gap ($E_g < 3$ eV) exists. Because of the small band gap, at room temperature some electrons from valence band can acquire enough energy to cross the energy gap and enter the conduction band. These electrons (though small in numbers) can move in the conduction band and contribute to electrical conduction. Hence, the resistance of semiconductors is not as high as that of the insulators but also not low as that of metals. Semiconductors sometimes demonstrate metallic like behavior at high temperatures due to that existence of free electrons at high temperatures.

***Fermi Level**

Top of the collection of electron energy levels at absolute zero temperature. This concept comes from Fermi-Dirac statistics. Electrons are fermions and by the Pauli exclusion principle cannot exist in identical energy states.

Intrinsic Semiconductor:

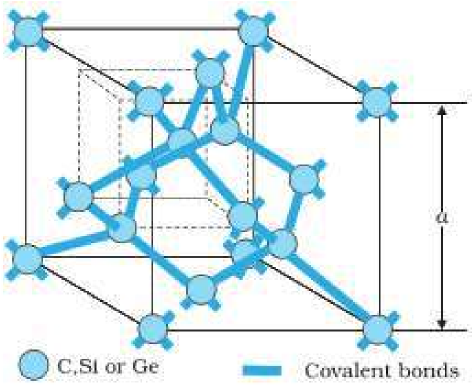


FIG. 4 *Three dimensional diamond like crystal structure for Carbon, Silicon or Germanium with respective lattice spacing 'a' equal to 3.56, 5.43 and 5.66 Å.*

Common examples of intrinsic semiconductors are elements Si and Ge which have 4 valence electrons in the outermost shell in $2s^2 2p^2$ electronic configuration. From fig. 4 we observe that each atom is coordinated by 4 nearest neighbours forming covalent bonds with each other. In each covalent bond a valence electron from one atom is shared with the corresponding one of the neighbouring atom thus the 4 neighbouring atoms will 'add' 4 additional electrons to the valence shell of the central atom completing its inert shell configuration.

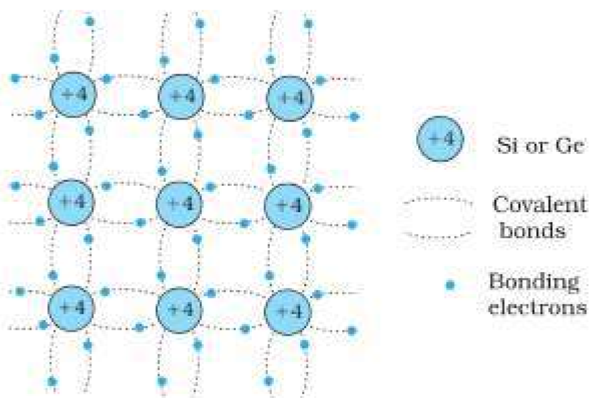


FIG. 5 *Schematic two-dimensional representation of Si or Ge structure showing covalent bonds at low temperature (all bonds intact). +4 symbol indicates inner cores of Si or Ge.*

A two dimensional schematic representation of the covalent bonding is represented in fig. 5. The covalently bonded electrons are depicted as shuttling between the sharing atoms. The figure shows an idealized scenario of all covalent bonds being intact which is expected only at low temperatures and in an ideal crystal without lattice defects. In real crystals there will be imperfections in the crystalline lattice during crystal growth. Similarly at high temperatures there will be thermal excitations of the electrons which will force some of the electrons to break away from the bonds. However in order to study the effect of intrinsic electronic structure in giving rise to semi-conductivity we imagine an idealized case when we do not have any lattice defects and assume that we have a perfect crystal produced for us. For such a case at low temperatures all the bonds are intact and we do not have any thermal excitation induced breaking of bonds.

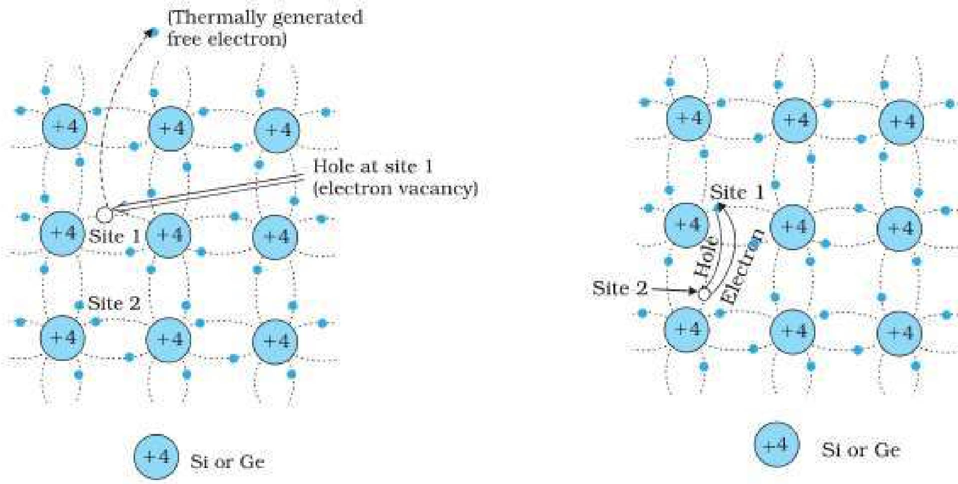


FIG. 6 *(a) Schematic model of generation of hole at site 1 and conduction electron due to thermal energy at moderate temperatures. (b) Simplified representation of*

possible thermal motion of a hole. The electron from the lower left hand covalent bond (site 2) goes to the earlier hole site1, leaving a hole at its site indicating an

apparent movement of the hole from site 1 to site 2.

At high temperatures the electrons which have broken these bonds become free and contribute to electrical conduction (fig. 6). When the electron is thermally excited it leaves a vacancy at the site of the bond from which it got excited which is called a 'hole'. A hole is an *effective* positive charge with the magnitude of the charge same as that of the outgoing electron. Under the influence of temperature the hole generated at one site can 'migrate' to another site thus holes inside the semiconductors are mobile. In semiconductors both these charge carriers i.e. the free electrons and the 'holes' contribute to the electrical conduction independently and hence the net electrical conduction in semiconductors is the sum of these individual contributions.

In intrinsic semiconductors, the number of free electrons, n_e is equal to the number of holes, n_h. T hat is $n_e = n_h = n_i$, where n_i is called intrinsic carrier concentration.

The thermally excited free electron moves completely independently as conduction electron and thus gives rise to an electron current, I_e under an applied electric field. Remember that the motion of hole is only a convenient way of describing the actual motion of *bound* electrons, whenever there is an empty bond anywhere in the crystal. Under the action of an electric field, these holes move towards negative potential giving the hole current, I_h. The total current, I is thus the sum of the electron current I_e and the hole current I_h: i.e. $I = I_e + I_h$.

Example: *C, Si and Ge have same lattice structure. Why is C insulator while Si and Ge intrinsic semiconductors?*

Solution: *The 4 bonding electrons of C, Si or Ge lie, respectively, in the second, third and fourth orbit. Hence, energy required to take out an electron from these atoms (i.e., ionisation energy E) will be least for Ge, followed by Si and highest for C. Hence, number of free electrons for conduction in Ge and Si are significant but negligibly small for C. Note that the band gap of diamond (carbon) is 5.5 eV, of Si is 1.11 eV and of Ge is 0.67eV*

10. **Extrinsic Semiconductor:**

The intrinsic semiconductors require finite temperatures for its operation. Only then one can get electron-holes available for conduction and semiconducting operation. However technologically it is much more desirable to have the semi conductivity to be realized at all temperatures. Moreover the electrical conductivity of intrinsic semiconductors is quite low for the sake of their technological use. Hence it is desirable to improve the properties of intrinsic semiconductors in order to tailor them for better technological use. One idea in this direction is to use foreign atoms added to intrinsic semiconductor in very small quantities in parts per million (ppm) level (called impurities) which would enhance their physical properties for technological exploitation e.g. one can enhance the electrical conductivity manifold upon such impurity addition. Such materials are known as extrinsic semiconductors (also called doped semiconductors) since the semiconducting properties are primarily due to the impurity atoms than from the parent compound itself. The impurity atoms added are technically called as *dopants*. An important consideration which must go into while choosing a dopant for doping is that the size of the dopant atom should be close to the size of the parent atom so as not to distort the crystal lattice significantly. This will ensure that the dopant merely adds its promised physical properties to the parent compound without generating spurious features arising due to the distortion of the lattice.

Mainly two types of dopants are used in the doping tetravalent Si or Ge:

(i) **Pentavalent (valency 5):** like Arsenic (As), Antimony (Sb), Phosphorous (P), etc.

(ii) **Trivalent (valency 3):** like Indium (In), Boron (B), Aluminium (Al), etc.

You will see from above that the dopant atoms lie in either side of Si and Ge in the periodic table of elements. Thus they are expected to have atomic sizes close to Si and Ge hence their substitution in place of Si or Ge is not expected to create significant crystal structure modifications in the parent lattice. Both these substitutions give rise to entirely different types of semi conductivity.

(i) n-type semiconductor:

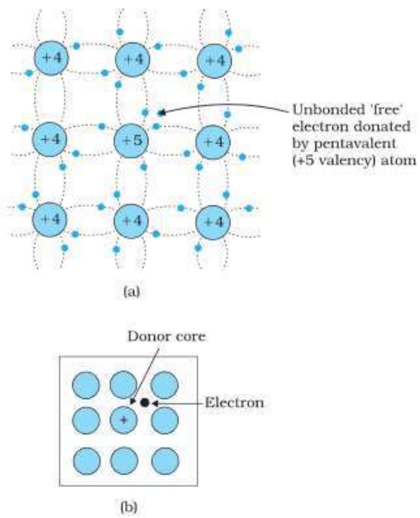


FIG. *7(a) Pentavalent donor atom (As, Sb, P, etc.) doped for tetravalent Si or Ge giving n-type semiconductor, and (b) Commonly used schematic representation of n-type material which shows only the fixed cores of the substituent donors with one additional effective positive charge and its associated extra electron.*

If we dope a pentavalent impurity (As, Ab, P) into Si or Ge, then the 4 valence electrons of the pentavalent atom will be involved in the formation of 4 covalent bonds with its neighboring Si or Ge atoms and the fifth valence electron of the dopant will find itself loosely attached to its parent nucleus (see fig.7). This is because the 4 covalent bonds will apparently complete the inert shell configuration for the dopant atom which will screen the positive nuclear charge from attracting the fifth valence electron. Hence that electron will be very loosely bound to the dopant atom and with very small amount of energy it will get excited to the conduction band and contribute to the conduction. E.g for pentavalent impurity in Si this energy ~ 0.05 eV and for Ge ~ 0.01 eV. Since this extra electron is donated by the pentavalent dopant hence it is also called as the donor impurity. Since the extra electron lies very close to the conduction band the free electron concentration in such cases is very weakly temperature dependent.

In such semiconductors the number of free electrons n_e is contributed largely by the no. of dopant atoms and weakly by the intrinsically generated free electrons by thermal excitation. Moreover the since the free electron concentration in doped semiconductors is enhanced in comparison with intrinsic semiconductors the rate of hole recombination also enhanced which further reduces the no. of available holes n_h at any temperature. Hence the electrical conduction in such compounds occurs primarily by the free conduction electrons and less by the holes in the valence band. Thus, with proper level of doping the number of conduction electrons can be made much larger than the number of holes. Hence in an extrinsic semiconductor doped with pentavalent impurity, electrons become the *majority* carriers and holes the *minority* carriers. These semiconductors are, therefore, known as n-type semiconductors. For n-type semiconductors, we have,

$n_e >> n_h$.

(ii) p-type semiconductor:

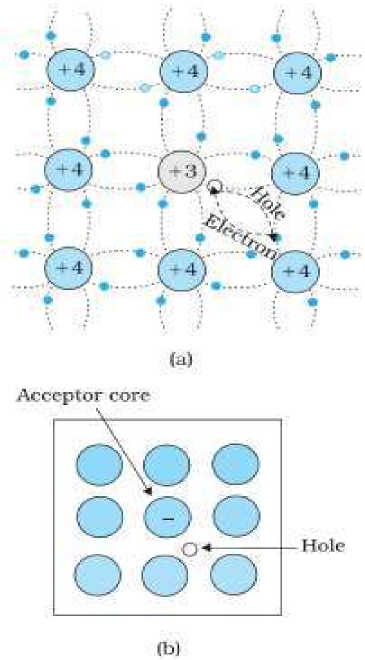


FIG. 8 *(a) Trivalent acceptor atom (In, Al, B etc.) doped in tetravalent Si or Ge lattice giving p-type Semiconductor. (b) Commonly used schematic representation of p-type material which shows only the fixed core of the substituent acceptor with one effective additional negative charge and its associated hole.*

This is obtained when we dope trivalent impurity like (Al, B, In etc.) into Si or Ge. The trivalent impurity will form 3 covalent bonds with neighbouring Si or Ge atoms while the last neighbour will not have any bonding with the impurity (see fig.8). When the impurity tries to attract an additional electron from the surroundings it will generate a hole at that place. This hole itself is not stable with time but will undergo annihilation with the electron coming from its neighbouring Si or Ge atom. This process continues over the whole lattice hence the hole is delocalized over the whole lattice and hence is available for electrical conduction. These holes are in addition to the intrinsically generated hole due to thermal excitation of electrons in Si or Ge atoms. Hence the holes are in excess to electrons in such semiconductors and these are called *p*-type semiconductors and the impurities are called acceptor impurities. The excess of holes further deplete the thermally generated free electrons via recombination. Hence we have for *p*-type semiconductors $n_h >> n_e$.

Thus we see that the effect of dopant in extrinsic semiconductors is to enhance the majority charge carriers and reduce the minority charge carriers thus leading to the conductivity provided essentially by majority charge carriers.

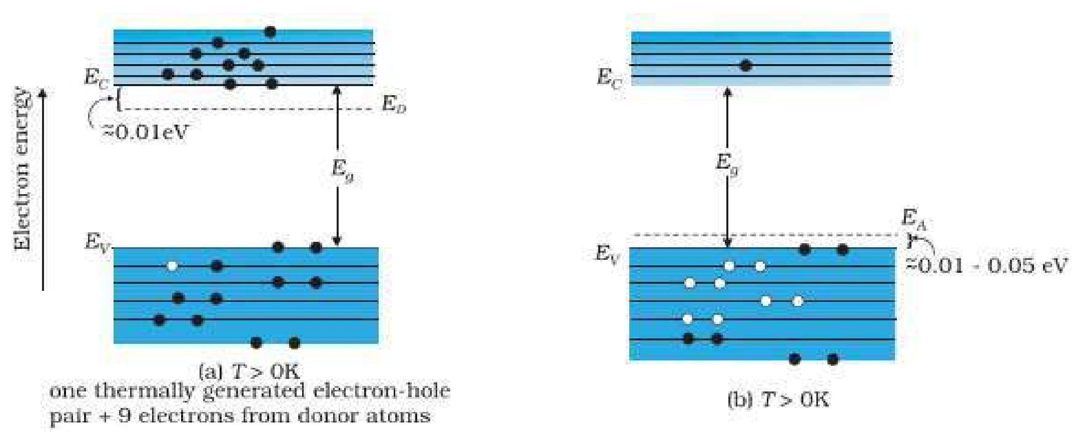


FIG. 9 *Energy bands of (a) n-type semiconductor at T > 0K, (b) p-type semiconductor at T > 0K.*

The energetics arising due to the donor and acceptor action of the impurity is schematically represented in fig. 9. The donor (E_D) and acceptor (E_A) impurity states are formed just below the bottom of the conduction band E_C and just above top of the valence band E_A respectively as shown in the figure. The energy separation between the impurity states and the corresponding edges of the bands are very small ~ 0.01-0.05 eV. As a result, at room temperature all the impurity states are populated and hence we have large no. of electrons and holes in the conduction and valence band respectively as shown in the figure. The electron and hole concentration in a semiconductor in thermal equilibrium is given by $n_h\, n_{e=}\, n_i^2$, where n_i is the intrinsic carrier concentration.

11. *p-n* junction:

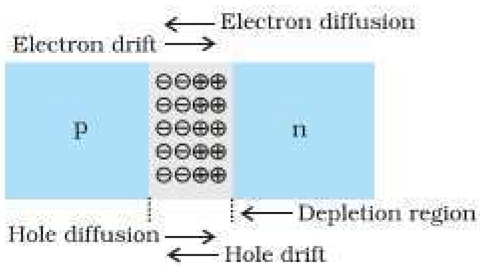


FIG.10 *p-n junction formation process.*

When we dope small amount of donor impurity into *p*-type semiconductor wafer on its side then we get a junction between *p*- and *n*-types of semiconductors called as a *p-n* junction. This device holds both types of majority charge carriers, electrons and holes together, but which occupy different regions in the device. Two main charge transport mechanisms operate within the junction namely diffusion and drift. Diffusion is the process in which the majority charge carriers on the *n*-side of the junction i.e. electrons migrate to the *p*-side of the junction as a result of the electron concentration gradient across the junction and vice versa for the majority charge carriers on the *p*-side of the junction i.e. holes. As a consequence we a small region (~tenths of µm thick) around the junction from which all the majority charge carriers have been removed due to diffusion called as a depletion region (see fig. 10). The diffusion of charge carriers creates local positive or negative charge at the donor or acceptor impurity site respectively. Hence the depletion region is actually a space-charge region with the *n*-side of it acquiring a positive charge and the *p*-side of it acquiring a negative charge. Such a space-charge development leads to the formation of electric field across the depletion region from the positive charge to the negative charge which attracts the minority charge carriers across both *n*- and *p*-sides towards opposite direction giving rise to a drift current. Thus the diffusion and drift currents are in opposite direction to each other. Initially as the junction is formed the diffusion dominates the drift and hence the depletion region starts developing. After it gets well developed the wide depletion region gives rise to significant amount of electric field to increase the drift current. This happens till the diffusion current equals the drift current at which an equilibrium condition is reached and no net current flows across the junction.

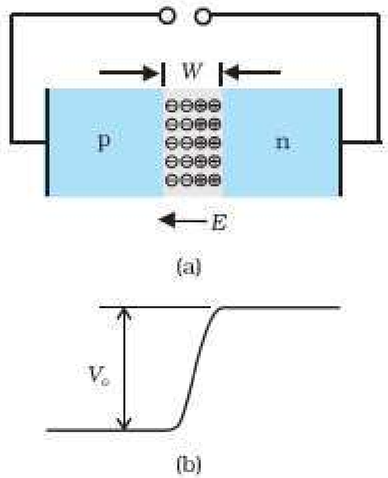


FIG. 11 (a) Diode under equilibrium (V = 0), (b) Barrier potential under no bias.

The formation of such an electric field across the junction gives rise a potential barrier across the junction which stops further flow of charge carriers across the junction. Such a potential barrier in represented in fig. 11 which describes a case where no external voltage has been applied to the junction.

Example: *Can we take one slab of p-type semiconductor and physically join it to another n-type semiconductor to get p-n junction?*

Solution: *No! Any slab, howsoever flat, will have roughness much larger than the inter-atomic crystal spacing (~2 to 3 Å) and hence continuous contact at the atomic level will not be possible. The junction will behave as a discontinuity for the flowing charge carriers.*

12. **Semiconductor Diode:**

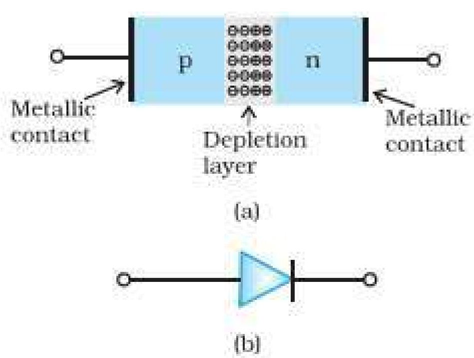


FIG. 12 (a) Semiconductor diode, (b) Symbol for p-n junction diode.

A semiconductor diode is merely a p-n junction with metallic leads at the ends for electrical connection to be made as shown in the fig. 12. The direction of arrow in the symbol used for diode represents the conventional direction in which the current through the diode will flow when it is operated in forward bias mode.

12.1 p-n junction diode under forward bias:

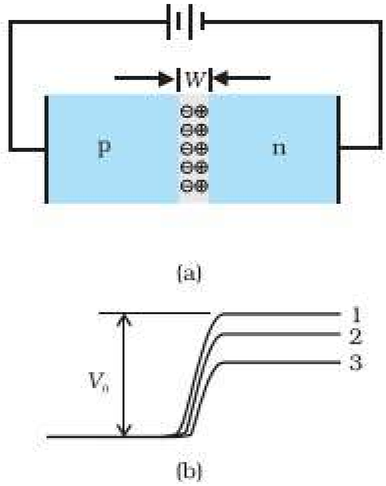


FIG. 13 *(a) p-n junction diode under forward bias, (b) Barrier potential (1) without battery, (2) Low battery voltage, and (3) High voltage battery.*

When we apply positive potential to the *p*-side and negative potential to *n*-side then the diode is said to operate under forward bias. The applied voltage (V) leads to the reduction in the height of the barrier potential since the applied voltage is in reverse direction to the barrier potential (see fig. 13). In that case the depletion region reduces in width since the holes are injected into the *n*-side from the *p*-side by the applied potential and vice versa for the electrons thus reducing the size of the depletion region. The injected holes into the *n*-side constitute minority charge carriers and the injected electrons into the *p*-side constitute minority charge carriers there hence the process is also called minority carrier injection. This leads to the establishment of net current in the direction of applied voltage the magnitude of which increases with the applied voltage. The effective barrier height under forward bias is ($V_o - V$). The magnitude of the net current flow depends on the height of the barrier. Higher the barrier height lesser will be the current since the resistance to the current flow will be higher in that case.

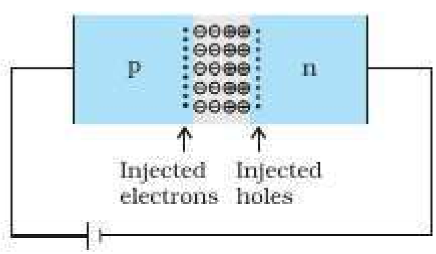


FIG. 14 *Forward bias minority carrier injection.*

At the junction boundary, on each side, the minority carrier concentration increases significantly compared to the locations far from the junction. Due to this concentration gradient, the injected electrons on *p*-side diffuse from the junction edge of *p*-side to the other end of *p*-side. Likewise, the injected holes on *n*- side diffuse from the junction edge of *n*-side to the other end of *n*-side (see fig. 14). This motion of charged carriers on either side gives rise to an electric current. The total diode forward current is sum of hole diffusion current and electron diffusion current. The total current is usually in mA.

12.2 p-n junction diode under reverse bias:

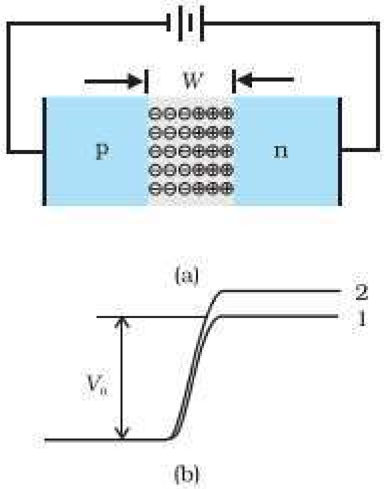


FIG. 15 (a) Diode under reverse bias, (b) Barrier potential under reverse bias.

When we apply external voltage (V) to the diode such that the positive potential is applied to the n-side and the negative potential is applied to p-side then the resulting arrangement is said to be in reverse biased operation of the diode (fig. 15). Here the applied voltage is in the same direction as the barrier potential and hence the barrier potential increases in height and concomitantly the depletion region widens. The effective barrier height under reverse bias is ($V_o + V$). This suppresses the diffusion of electrons from n-side to p-side and diffusion of holes from p-side to n-side thus reducing the overall diffusion current remarkably as compared to that in forward bias.

The electric field direction in case of reverse bias gives rise to an enhancement of the drift current which is of the order of μA. This is quite low as compared to that in forward bias (~ mA) since the drift current is a result of minority charge carrier motion whereas the diffusion current arises due to majority charge carriers.

The reverse bias current is dependent more on the minority carrier concentration on either side of the junction rather than on the magnitude of the applied voltage. The current under reverse bias is essentially voltage independent upto a critical reverse bias voltage, known as breakdown voltage (V_{br}). When $V = V_{br}$ the diode current increases sharply. Even a slight increase in the bias voltage thereafter causes large change in the current.

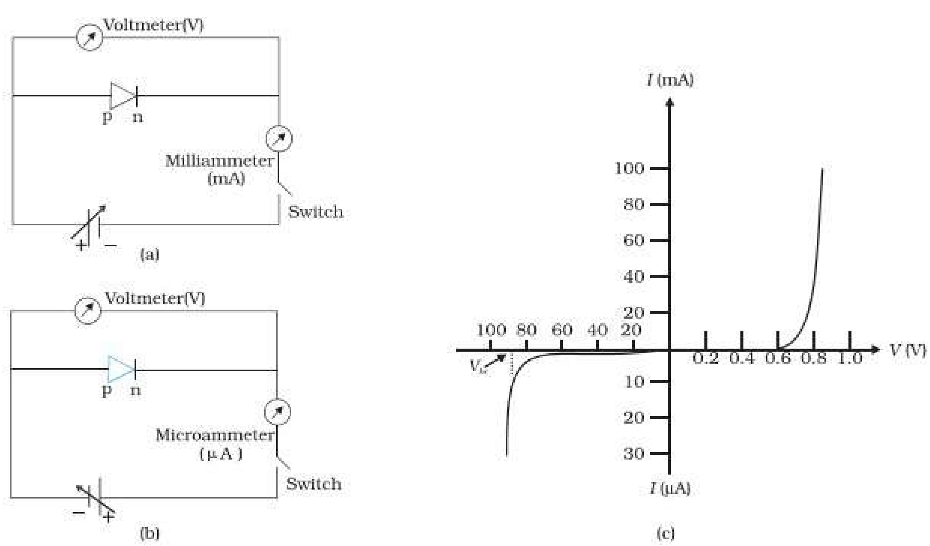


FIG. 16 Experimental circuit arrangement for studying V-I characteristics of a p-n junction diode (a) in forward bias, (b) in reverse bias. (c) Typical V-I characteristics of a silicon diode.

The circuit for measuring the *V-I* characteristic curve for a *p-n* junction diode is shown in fig. 16. The voltage applied across the diode is variable. A milliammeter and a microammeter is used to measure the current under forward and reverse bias conditions resectively. In forward bias, the current increase very slowly till it reaches a certain threshold or cut-in voltage (~0.2V for germanium diode and ~0.7 V for silicon diode) after there is very rapid increase of the current for a very small increase in the applied voltage. In reverse bias, the current remains practially unchanged for a large range of applied voltage. When the voltage reaches V_{br} (~100 eV), breakdown occurs and the reverse bias current increase rapidly.

Thus we see that the action of a *p-n* junction diode is to allow the current to flow only in one direction while blocking the flow in the other direction. This property allows it to be used the electrical circuits to allow the current flow only in one direction. For diodes, we define a quantity called *dynamic resistance* as the ratio of small change in voltage ΔV to a small change in current ΔI i.e. $r_d = \Delta V/\Delta I$.

7. Application of Junction Diode as a Rectifier:

A **rectifier** is an electrical device that converts alternating current (AC) to direct current (DC). The process is known as **rectification**.

From the V-I characteristic of a junction diode we see that it allows current to pass only when it is forward biased. So if an alternating voltage is applied across a diode the current flows only in that part of the cycle when the diode is forward biased. This property is used to rectify alternating voltages. The following two rectifier circuits can be used:

7.1 Half-Wave Rectifier

In half-wave rectification, the rectifier conducts current only during the positive half-cycles (positive-half cycles) of input a.c. supply. Figure 17a shows the circuit where a single crystal diode acts as a half-wave rectifier. The a.c supply to be rectified is applied in series with the diode and load resistance R_L. The a.c. voltage across the secondary winding AB changes polarities after every half-cycle. During the positive half-cycle of input a.c. voltage, end A becomes positive w.r.t. end B, and diode is forward biased, hence it conducts current. During the negative half-cycle, end A is negative w.r.t. B, and diode is reverse biased, hence no current conducts (Fig.17 b).

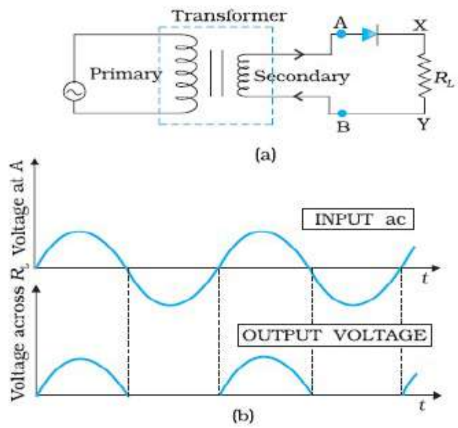


FIG. 17 (a) Half-wave rectifier circuit, (b) Input ac voltage and output voltage waveforms from the rectifier circuit.

7.2 Full-Wave Rectifier

In full-wave rectification, current flows through the load in the same direction for both the half-cycles of input a.c. voltage. This can be achieved with two diodes working alternately (Fig. 18 a). For the first half-cycle of input voltage, one diode D_1conducts and supplies current to the load (R_L), and for the other half-cycle, diode D_2does so. Current being always in the same direction through the load. Therefore, a full-wave rectifier utilizes both half-cycles of input a.c. voltage to produce the d.c. output (Fig. 18 b & c).

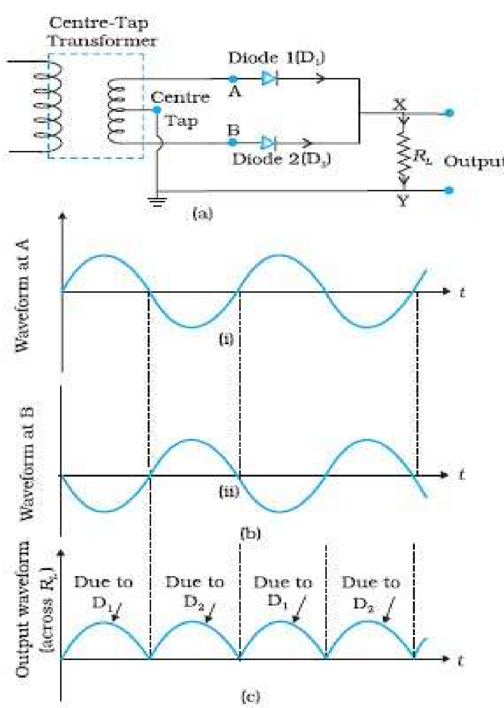


FIGURE 18 *a) A Full-wave rectifier circuit; (b) Input wave forms given to the diode D1 at A and to the diode D2 at B; (c) Output waveform across the load R_L connected in the full-wave rectifier circuit.*

8. Special Purpose p-n Junction Diode

8.1 Zener diode

This diode has heavily doped p- and n- sides of the junction, due to this depletion region formed is very thin (10^{-6}m) and the electric field at the junction is very high (~ 10^6 V/m) even for a very small reverse bias voltage. When reverse bias voltage V reaches V_z, a sudden rise in current occurs, this is called Breakdown Voltage (Fig.19b). Beyond V_z a large change in current can be produced even for a negligible change in reverse bias voltage. We can say that zener voltage remains constant even though the current through zener diode varies to very large vales. Hence, this property of zener diode is used to regulate the supply voltage.

When the reverse bias voltage $V = V_z$, then the electric field strength is so high that pulls the valence electrons from the host atoms on the p-side which are accelerated to n-side. These electrons account for high current observed at the breakdown. The electric field required for this phenomenon to occur is of the order of 10^6 V/m.

Fig. 19a shows the symbol of a zener diode. It may be seen that it is just like an ordinary diode except that the bar is in z-shape.

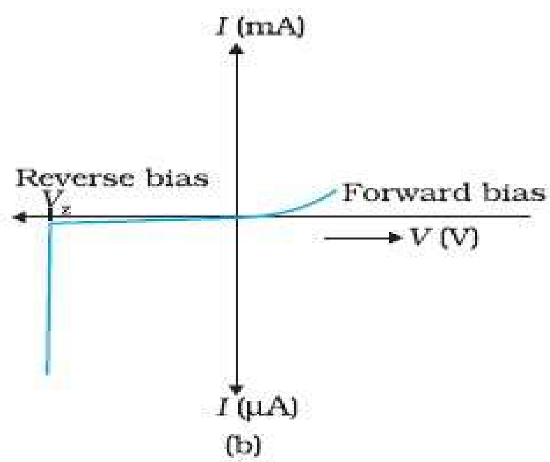


Fig. 19 *Zener diode, (a) symbol, (b) I-V characteristics.*

Zener diode as a voltage regulator

A zener diode can be used as a voltage regulator to provide a constant voltage from source whose voltage may vary over sufficient range.

The circuit arrangement is shown in Fig. 20. The unregulated dc voltage V_I is connected to the Zener diode through a series resistance R_s. When V_I increases, the current through Rs and Zener diode also increases causing the increase in voltage drop across Rs without any change in the voltage across the Zener diode. This is because of the specific feature of zener diode, that in the breakdown region, Zener voltage remains constant even though the current through the Zener diode changes.

In the same way if the input voltage decreases, the current through Rs and Zener diode also decreases causing the decrease in voltage drop across Rs without any change in the voltage across the Zener diode. Hence any increase or decrease in the input voltage results in increase or decrease of the voltage drop across Rs without any change in voltage across the Zener diode, and we get the constant (regulated) voltage in the output (measured across R_L). We can select the Zener diode as per the required output voltage and R_s.

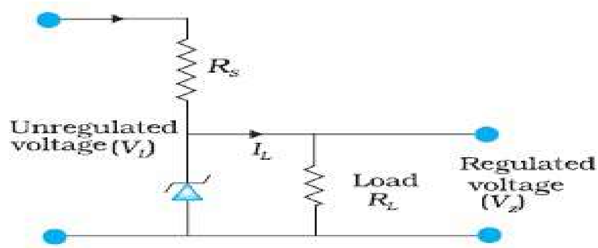


Fig. 20 *Zener diode as DC voltage regulator.*

8.2 Optoelectronic junction devices

(i) Photodiode

A photodiode is a reverse-biases silicon or germanium p-n junction in which reverse current increases when the junction is exposed to light. The value of reverse current depends on the incident light intensity. When light (photons) falls on the p-n junction, the energy is imparted by the photons to the atoms in the junction. This will knock out the electrons (more free electrons and more holes). These additional free electrons will increase the reverse current.

(ii) Light emitting diode (LED)

This is heavily doped p-n junction that emits spontaneous radiation when forward biased. It is enclosed in a transparent cover, so that emitted light comes out. Hence, named Light Emitting Diode.

When LED is forward biased, the electrons move from n to p (where they are minority carrier), and holes from p to n (where they are minority carrier). Therefore, at the junction boundary the minority carrier concentration increases from their original value (unbiased). Near the junction on either side, the excess minority carriers recombine with majority carriers. Thus, on recombination the energy in the form of photons is released. When the forward current of the diode is small, the intensity of light emitted is small, and as the forward current increases, intensity of light increases and reaches a maximum.

Ex. An LED is made of GaAsP having a band gap of 1.9 eV. Determine the wavelength and colour of radiation emitted.

Ans. We have $\lambda = hc/E_g = (6.63 \times 10^{-34} \times 3 \times 10^{8}) / (1.9 \times 1.6 \times 10^{-19})$

$= 6543$ angstrom. The emitted radiation is red.

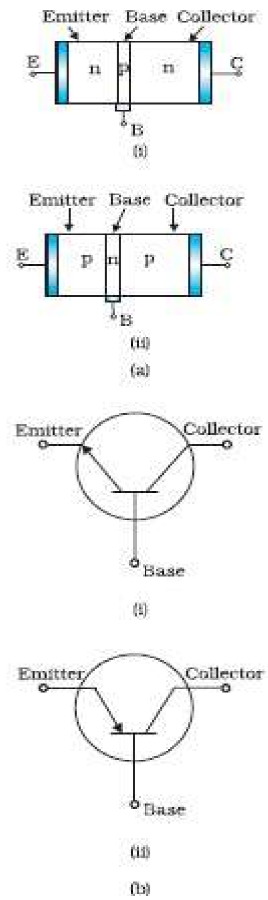


Fig. 21(a) & (b) are schematic representations and symbols of a n-p-n transistor and p-n-p transistors.

9. Junction Transistor

A transistor has three doped regions forming two p-n junctions. Two p-n junctions are formed by sandwiching either p-type or n-type semiconductor between a pair of opposite types. There are two types of transistors as shown in Fig.21, i.e. n-p-n and p-n-p transistor. The transistor is entirely a new type of electronic device, and is capable of achieving amplification of weak signals.

Emitter- The section on one side of the transistor that supplies chare carriers is called emitter. The emitter is always forward biased w.r.t base so it can supply a large number of majority carriers.

Collector- The section on the other side that collects the charges is called the collector. The collector is always reverse biased. Its function is to remove charges from its junction with the base.

Base- The middle section which forms two p-n junctions between the emitter and collector is called the base.

9.1 Basic transistor circuit configurations and transistor Characteristics

Transistor can be connected in either of the following three configurations namely Common Emitter (CE), Common Base (CB) and Common Collector (CC).

In a circuit the input/output connections have to be such that one of these (E, B or C) is common to both the input and the output.

Common emitter transistor characteristics

Transistor is most widely used in the CE configuration and more commonly used transistors are n-p-n Si transistors. With p-n-p transistors the polarities of the external power supplies are to be inverted. Therefore, it is easy to understand the transistor circuits using n-p-n transistors.

In CE arrangement, input is between base and emitter and the output is between collector and emitter. Here, emitter of the transistor is common to both input and output circuits and hence the name common emitter connection (Fig.22). The variation of base current I_B with the base-emitter voltage V_{BE} at constant collector-emitter voltage V_{CE}. is called the input characteristic, while the variation of collector current I_C with the collector-emitter voltage V_{CE} at constant base current I_B is called the output characteristic (Fig. 23 a and b) . We see that the output characteristics are controlled by the input characteristics. This implies that the collector current changes with the base current.

Input resistance-It is the ratio of change in base-emitter voltage to the change in base current at constant V_{CE} i.e.

$r_i = (\Delta V_{BE} / \Delta I_B)_{VCE}$

Output resistance -It is the ratio of change in collector-emitter voltage to the change in collector current at constant I_B i.e.

$r_o = (\Delta V_{CE} / \Delta I_C)_{IB}$

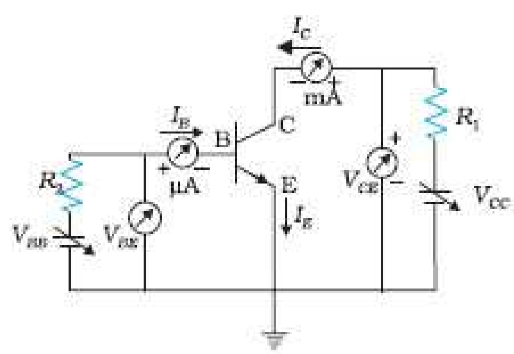


Fig. 22 *Circuit arrangement for studying the input and output characteristics of n-p-n transistor in CE configuration.*

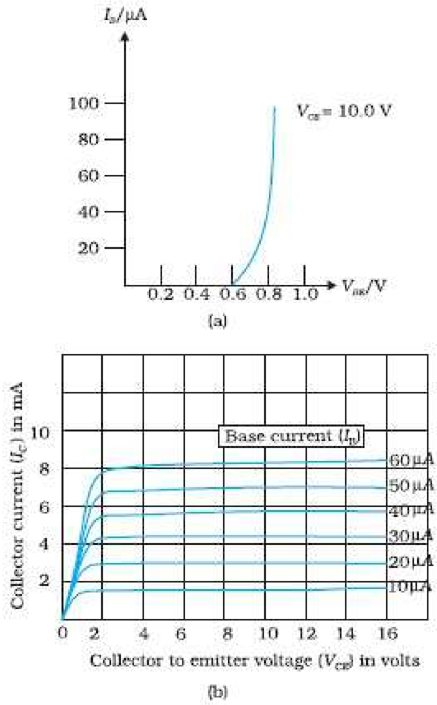


Fig. 23 *(a) Typical input characteristics, and (b) Typical output characteristics.*

9.2 Transistor as an Amplifier (CE-Configuration)

Fig. 24 shows the common emitter n-p-n amplifier circuit. In general, amplifiers are used to amplify alternating signals. Now let an ac input signal V_i (to be amplified) is superimposed on the bias V_{BB}(dc) as shown in Fig. 24, then the base current will have sinusoidal variations superimposed on the value of IB. As a consequence the collector current will also have sinusoidal variations superimposed on the value of IC, producing in turn corresponding change in the value of VO. We can measure the ac variations across the input and output terminals by blocking the dc voltages by large capacitors.

Change in I_C due to the change in I_B is known as *ac current gain*, which is represented as *βac*.

$\beta ac = \Delta I_C / \Delta I_B = i_C / i_B$

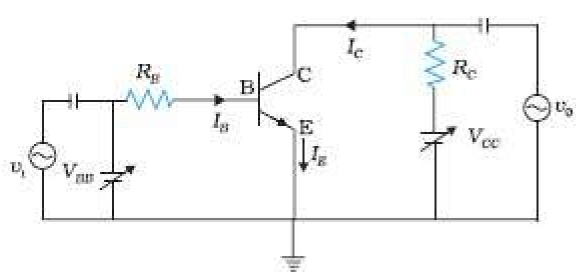


Fig. 24 *Circuit of a CE-transistor amplifier.*

Final Assessment

Question 1: The density of silicon atoms is 6×10^{28} per m^3. After doping it with 5×10^{22} atoms per m^3 of Arsenic and 4×10^{20} per m^3 atoms of Indium calculate the number of electrons and holes produced as a result. Given that $n_i = 1.5 \times 10^{16}$ m^{-3}. Is the material *n*-type or *p*-type?

Question 2: In a *p-n* junction diode, the current I can be expressed as

$$I = I_0 \exp\left(\frac{eV}{2k_B T} - 1\right)$$

where I_o is called the reverse saturation current, V is the voltage across the diode and is positive for forward bias and negative for reverse bias, and I is the current through the diode, k_B is the Boltzmann constant (8.6×10^{-5} eV/K) and T is the absolute temperature. Given that $I_o = 6 \times 10^{-12}$ A and $T =$ 300 K, then

(a) What will be the forward current at a forward voltage of 0.7 V?

(b) What will be the increase in the current if the voltage across the diode is increased to 0.8 V?

(c) What is the dynamic resistance?

(d) What will be the current if reverse bias voltage changes from 2 V to 3 V?

Question 3: Describe the origin of diffusion and drift currents in a *p-n* junction. How does a competition between them help us to achieve the valve action for electric current flow through the junction?

References

1. Solid State Physics: A. J. Dekkar
2. Solid State Physics: C. Kittel

Module – VIII
Modern Physics
(Dual Nature of Radiation and Matter)

Lectures: 02

Objective

The main objective of this module is to introduce the concepts of duality of radiation and matter. This module will simplify the concepts of particle aspects of radiation and wave aspects of particle. The concepts will be supported by experimental findings, hypothesis and theoretical framework.

Prerequisites

This module requires knowledge of electromagnetic wave and Newtonian mechanics.
The instructors may access this knowledge by asking following questions:

- Which wave is more energetic: Infrared or ultraviolet?
- How will you a characterize a wave?
- How will you characterize a moving particle?
- Why do we need two separate frameworks for wave and particle in classical physics?
- Does our institution permits to consider a particle manifesting wave like proper-ties in a wavelike experiments and particle-like properties in particle like experiments?
- Name at least two experiments which demonstrates wave characteristics.
- Name at least two experiments which demonstrates particle characteristics.
- Do you think of any experiment where the electromagnetic waves are interacting with matter?
- Consider the electromagnetic waves emanating on an isolated atom. What would be your anticipation of possible outcome of this interaction?
- Can you think of a moving electron as a wave or X-ray waves as a beam of particles?

Introduction

By now the students must be knowing about a classical particle which is characterized by definite position, momentum, velocity, momentum and can be sensed by naked eyes with a definite structure. The motion of these classical particles are governed by Newton's laws. In case of a wave one can characterize by wavelength, frequency and amplitude. The wave is extended over the space and contrast to the particles, cannot be localized. The motion of an electromagnetic wave is governed by Maxwell's equations. Classical Physics deals with these two objects separately. The Maxwell's equations together with Hertz experiments on electromagnetic waves in 1887 established the wave nature of light. The end of 19th century and beginning of 20th century witnessed many experiments on conduction of electric discharge through various gases filled at low pressure in a vaccum tube. These experiments lead to many discoveries of the atomic world. In 1895 W. Roentgen discovered X-rays by Roentgen and in 1897 J. J. Thom-son discovered the electrons (a fundamental particle). These discoveries unraveled lots of mysteries in the atomic world and exposed the inadequacy of the classical formalism. The term wave-particle duality refers to the behavior where both wave-like and particle-like properties are exhibited under different conditions by the same entity. The particle aspect of wave put forward a strong belief of wave aspects of particle by De Broglie. In this module we will focus mainly the wave aspect of light and particle aspect of wave by introducing some experiments and theoretical framework.

Wave nature of light

Which of the following support the wave nature of light?

1. Maxwell's equations of electromagnetism
2. Hertz experiments on generation and detection of electromagnetic (EM) waves
3. Young's double slit experiment
4. Phenomena of diffraction

Electron emission

- The conductivity of a metal is due to the presence of free electrons.
- Discuss: If a piece of metal is isolated why do these free electrons not leave the metal spontaneously?
- If one of these electrons is given a certain minimum amount of energy, it can indeed leave the metal.
- This minimum energy depends on the metal and is called its *work function*.
- This minimum energy can be supplied by any of the following processes-
    1. Thermionic emission: using heat.
    2. Field emission: by applying very high electric field, e.g. in a spark plug
    3. Photo-electric emission: by illuminating with suitable light

Photoelectric effect

Hertz's observations

- Discovered in 1887 during EM wave experiments.

- *Phenomena:* high voltage sparks across the detector enhanced by ultraviolet (UV) light.
- *Explanation:* electrons near the surface absorb enough energy from the radiation and escape.

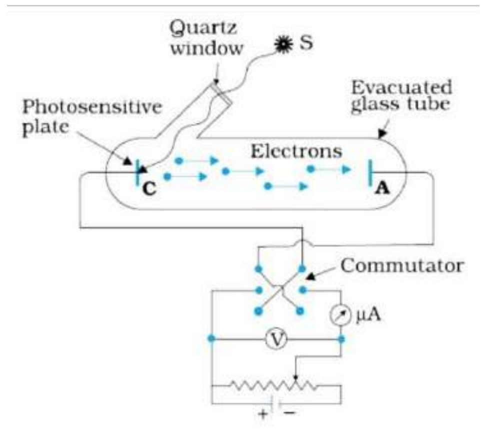


Figure 1: *Experimental setup for photoelectric effect. Source: NCERT*

Hallwachs' and Lenard's observations (1886-1902)

- *Experimental setup:* Figure 1
- *Observations:*

  – UV light on emitter plate → current in circuit
  – No UV light → no current

- *Exercise:* Describe a possible physical process that can explain these two observations. (Answer: UV light → electron ejection → electrons attracted to the collector plate A by electric field).
- *Hallwachs' experimental observations (1888):*

  – Negatively charged zinc plate in an electroscope $\xrightarrow{\text{UV light}}$ lost its charge
  – Uncharged or positively charged zinc plate in an electroscope $\xrightarrow{\text{UV light}}$ more positively charged

- Which of the following explanations is more plausible for above observations

  – UV light is positively charged

  – UV light is neutral but causes the metal to lose negatively charged particles

Discovery of Electron (1897)

- Was identified as the particle emitted from the metal surface in the above experiments.

- Emitted particles were called photoelectrons and the phenomena was called photoelectric effect.
- *Observation:* No electron emitted if the frequency of UV light is below certain minimum value called threshold frequency.
- *Exercise:* What can be concluded about the threshold frequency of the mentioned metals from the following observation:
  "Certain metals such as zinc, cadmium, magnesium, etc. respond only to UV light while some alkali metals such as lithium, sodium, potassium, cesium and rubidium are sensitive even to visible light."
  (*Hint:* UV light has higher frequency than visible light.)

Experimental study of photoelectric effect

- *Experimental setup:* Fig. 1
- *Exercise:* If the tube is made vacuum, can there be a current in the circuit (other than due to photoelectric effect that is)?
- *Exercise:* In the presence of suitable illumination, i.e. condition for photoelectric effect is satisfied, will there be a current if the battery is reversed?
- Things that can be varied in the experiment —
  – potential of the collector plate A with respect to the emitter plate C, both in magnitude and sign
  – Intensity and frequency of incident light
  – nature of material of the emitter plate C

Summary of observations

- *Exercise:* Discuss which of the following observations can have a plausible explanation using wave nature of light.

Effect of intensity of light on photocurrent

- *Observation:* linear dependence
- *Plausible wave explanation:* With higher intensity more electrons will be emitted in shorter time resulting in higher current

Effect of potential on photoelectric current

- *Observation with positive potential:* increase with increasing potential before saturating
- Maximum current is called saturation current
- *Exercise:* Give an explanation for saturation. (Answer: All emitted electrons reach the collector plate at high enough potential and hence no further increase in current.)

- *Observation with negative (retarding) potential:* photocurrent decreases rapidly to zero.
- There is a certain sharply defined, critical value of negative potential V_0 called cut-off or stopping potential.
- *Exercise:* Explain the cut-off or stopping potential. (*Answer:* Emitted electrons have different kinetic energy. The potential required to stop even the most energetic electrons from reaching the collector plate is the cut-off or stopping potential.)
- *Observation:* Cut-off or stopping potential does not depend on intensity
- *Plausible wave explanation:* Higher intensity means more photoelectrons but it does not necessarily mean higher kinetic energy for individual electrons.

Effect of frequency of incident radiation on stopping potential

Discussion below is for same "frequency weighted" intensity

- *Observations:*
  - Different stopping potential but same saturation current
  - Higher frequency) stopping potential more negative) greater maximum kinetic energy
- *Stopping potential versus frequency.*
  - Straight line
  - Minimum cut-off frequency v_0 ↔ stopping potential is zero
    - v_0 is called threshold frequency.
    - Maximum kinetic energy is linear with frequency
    - $v < v_0$ → no photoelectric emission is observed.
- Wave explanation is not possible as there is no practical difference between light just below and just above the threshold frequency.
- *Observation:* photoelectric emission is an instantaneous process (10^{9} sec or less) even when the incident radiation is exceedingly dim
- Again wave explanation is not possible as lower intensity would mean delayed emission.

Photoelectric effect versus wave theory of light

- Striking contradiction with the frequency dependence especially the existence of threshold frequency,
- quantitative contradiction in time scale.

Einstein's photoelectric equation: energy quantum of radiation (1905)

- No continuous absorption of energy from radiation
- Radiation → discrete units "quanta" of energy of radiation

$$\text{energy of each quantum} = h\nu$$

where h is the Planck's constant.

- Photoelectric effect → electron absorbs one quantum

$$K_{max} = h\nu - \phi_0 \qquad \text{(Einstein's photoelectric equation)}$$

Where ϕ_0 is work function

- Higher intensity ⇒ more number of quanta
- Explains the photoelectric effect including all the observations in contradiction with wave theory of light
- Further experimental verification by Millikan who tried to disprove the theory

Particle nature of light: the photon

- Photoelectric effect ⇒ light made up of quanta or packets of energy ($h\nu$)
- Another important contribution of Einstein on photoelectric effect: the light quantum has momentum $h\nu/c$
- Well-defined energy and momentum ⇒ light quantum is like a particle → "photon"
- Further confirmation → Compton (1924)

Summary of photon picture

- Radiation and matter → radiation behaves like particle called photon
- Photon: energy = $h\nu$, momentum = $h\nu/c$
- Each photon of given frequency has same energy and momentum
  higher intensity ⇒ more number of photons
- Photons are electrically neutral ⇒ not deflected by electric and magnetic fields
- Photon-particle collision

  – Total energy and momentum conserved

  – Number of photons may not be conserved $ a photon may get absorbed or a new one created

Exercise: Examples 11.1, 11.2, 11.3 from NCERT.

De Broglie Hypothesis

Light behaves as particle and this has been evident from the Einstein's explanation of photoelectric effect in 1905. Now "converse also holds true?" In 1923 de Broglie postulated that if waves can behave like particles, then particles should be able to behave like waves. According to de Broglie, a moving material particle can be associated with a wave or in other words a wave can guide the motion of the particle. Such waves associated with the moving material particles are known as de Broglie waves or matter waves.

Suppose a particle of mass m is moving with the velocity v, the de-Broglie wave-length associated with the particle is given by

$$\lambda = \frac{h}{mv} \tag{1}$$

Proof: According to Planck's hypothesis, the energy associated with a photon of frequency ν is $E = h\nu$. Further, relativistic mass-energy relation given by Einstein for the particle of rest mass m_0 and momentum p is $E = \sqrt{p^2c^2 + m_0^2c^4}$ Since, the rest mass m_0 of the photon is zero, the energy of photon became $E = pc$. Therefore, we can write,

$$h\nu = pc \text{ or } p = \frac{h\nu}{c} = \frac{h}{\lambda} \text{ or } \lambda = \frac{h}{p} \tag{2}$$

As, de Broglie have stated that matter particle behaves as photon so the same expression can be used for the matter particles moving with velocity v and momentum $p = mv$ as

$$\lambda = \frac{h}{mv} \tag{3}$$

The above relation is known as de Broglie relation that connects the momentum/velocity of the particle to the wavelength of a wave associated with the particle.

Important points:

i. In everyday life, de Broglie wavelength comes out to be very small for massive particles.

ii. De Broglie wavelength is charge independent.

iii. It was found that velocity of de Broglie waves is always greater than the velocity with which the particle travels.

De-Broglie wavelength for an accelerated electron:

If an electron is supposed to accelerate through the potential difference of V volt, then electrical potential energy of the electron gets converted into its kinetic energy i.e. kinetic energy of electron, $K = eV$. We Know that,

$$K = \frac{p^2}{2m} \; or \; p = \sqrt{2mk} \; \Rightarrow \; p = \sqrt{2meV} \tag{4}$$

Thus, de Broglie's wavelength associated with an accelerated electron is expressed as:

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2meV}} \tag{5}$$

or
$$\lambda = \frac{6.63 \times 10^{-34}}{\sqrt{2 \times 9.1 \times 10^{-31} \times 1.6 \times 10^{-19}}} \times \frac{1}{\sqrt{V}} \mathrm{\mathring{A}} \tag{6}$$

or
$$\lambda = \frac{12.27}{\sqrt{V}} \mathrm{\mathring{A}} \tag{7}$$

De-Broglie wavelength and temperature:

The average kinetic energy of a particle at absolute temperature T is given by:

$$K = \frac{3}{2} k_B T \tag{8}$$

where k_B is the Boltzmann's constant. Suppose *m* is the mass of particle travelling with velocity *v*, then we have

$$K = \frac{1}{2} m v^2 \tag{9}$$

The momentum of the particle is

$$p = mv = \sqrt{2mK} = \sqrt{2m \times \frac{3}{2} k_B T} = \sqrt{3mk_B T}, \tag{10}$$

and de Broglie wavelength

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{3mk_B T}} \tag{11}$$

Experimental confirmation of de Broglie hypothesis (Davisson and Germer's Ex-periment):

C.J. Davisson and L.H. Germer in the year 1927 carried out an experiment, popularly known as Davisson Germer experiment to explain the wave nature of electrons through electron diffraction. In their experiment Davisson and Germer scattered a 54 eV mono-energetic beam of electrons from a Nickel crystal. The electron source and detector were symmetrically located with respect to the crystal's normal as indicated in the figure 2.

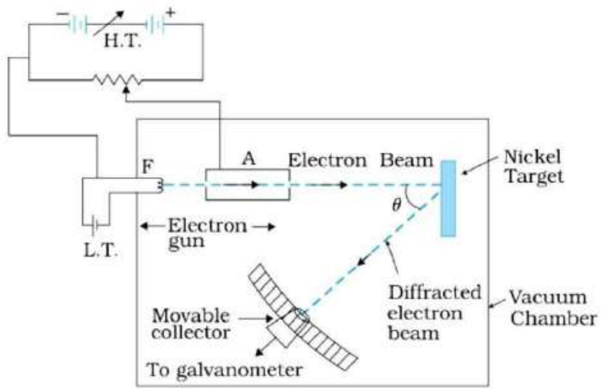


Figure 2: *Experimental setup for Davisson-Germer experiment. Source: NCERT*

Experimental set-up and working:

It consists of an electron gun, which comprised of a tungsten filament F, coated with barium oxide and heated by a low voltage power supply. Electrons emitted from this electron gun were accelerated to a desired velocity by applying suitable potential difference from a high voltage power supply. These emitted electrons were made to pass through a cylinder perforated with fine holes along its axis, thus producing a fine collimated beam. This beam produced from the cylinder is made to fall on the surface of a nickel crystal. This leads to scattering of electrons in various directions. The intensity of the beam of electrons is measured by the electron detector which is connected to a sensitive galvanometer (to record the current) and can be moved on a circular scale. The intensity of the scattered electron beam is measured for different values of angle of scattering, θ (angle between the incident and the scattered electron beams) by moving the detector on the circular scale at different positions. By varying accelerating potential difference, we finally obtained the variation of the intensity (I) of the scattered electrons with the angle of scattering, θ. The accelerated voltage was varied from 44V to 68V. A strong peak was noticed in the intensity (I) of the scattered electron for an accelerating voltage of 54 V at a scattering angle $\theta = 50^o$. This peak can be explained as a result of the constructive interference of electrons scattered from different layers of the regularly spaced atoms of the crystals. The wavelength of matter waves was calculated with the help of electron diffraction, which measured to be 1:65Å.

Proof: The glancing angle ϕ (i.e. the angle between the direction of scattered electron and crystal atomic plane) is given by:

$$\phi + \theta + \phi = 180^0 \tag{12}$$

Or

$$\phi = \frac{1}{2}(180^0 - \theta) \tag{13}$$

For maximum intensity

$$\phi = \frac{1}{2}(180^0 - 50^0) = 65^0 \tag{14}$$

Now according to Bragg's Law:

$$2d \ \sin\phi = n\lambda \tag{15}$$

For first order diffraction n = 1 and inter atomic separation for Ni crystal, *d = 0.91*Å. Thus,

$$\lambda = 2 \times 0.91 \sin 65^0 = 1.65\text{Å} \tag{16}$$

Now according to de-Broglie hypothesis, the wavelength of the wave associated with the electron accelerated to 54 volt is given by,

$$\lambda = \frac{12.27}{\sqrt{V}}\text{Å} = \frac{12.27}{\sqrt{54}}\text{Å} = 1.66\text{Å} \tag{17}$$

Thus, there is a close agreement between the estimated value of de-Broglie wavelength and the experimental value determined by Davisson and Germer. We have seen that the scattered electrons in the Davisson-Germer experiment produced interference fringes that were identical to those of Bragg's X-ray diffraction. Thus this experiment gave a strong evidence for the de-Broglie hypothesis or the wave nature of matter. The de Broglie hypothesis has been basic to the development of modern quantum mechanics. The wave properties of electrons have been utilized in the design of electron microscope which is more efficient than optical microscope.

0.0.1 Questions:

- An electron is accelerated through a potential difference of 180 V. Calculate its associated wavelength.
- A 100 grams ball rolls along a field with a speed of 20 cm/s. How large is its associated wavelength? Given $h = 6.63 \times 10^{-34} Js$
- Can we observe de Broglie wavelength associated with a football? Why not?

Outcome of this module

- Emission of electrons from a metal when light of proper frequency incident on its surface is called photoelectric emission.
- In photoelectric emission, electrons gain energy from light.
- The stopping potential increases with increase in frequency of incident light.
- There exists a frequency v_0 for every material below which no photoelectric effect takes place.

- The maximum velocity of photoelectrons increases with increasing frequency of incident light but is independent of the intensity of incident light.
- The number of photoelectrons emitted from each square centimeter of the emit-ting surface for any particular frequency is proportional to the intensity of incident light.
- Einstein assumed light to consist of photons, each having energy hv, where v is frequency and h is Planck's constant.
- Photoemissive type of phototube is based on the photoelectric effect.
- The saturation current of a phototube increases with increasing intensity of the incident light.
- Particles in motion have waves associated with them. The wavelength is given by = h=p, where, p is the momentum.

Exercise

- In photoelectric emission, what happens to the incident photons?
- Photons in a red light have a wavelength of 660 nm. ($1\text{nm} = 10^{-9}\text{m}$) What is the energy of this photon?
- What is the difference between a photon and a matter particle?
- Why is the wave nature of matter not apparent in daily life?
- How is velocity of photoelectrons affected if the wavelength of incident light is increased?
- The threshold frequency of a metal is 5×10^{-14} Hz. Can a photon of wavelength 6000Å emit an energetic photoelectron?
- Does the threshold frequency for a metal depend on the incident radiations?
- What are the various uses of photocell?
- What was the aim of Davisson and Germer's experiment? On what principle does it depend?
- Describe the experiment used for studying the photoelectric effect.
- Explain the terms (a) Saturation voltage and (b) Stopping potential.
- State the laws of photoelectric emission.
- Describe the salient features of Einstein's theory of photoelectric effect.

References

1. Quantum Mechanics: Ajoy Ghatak
2. Quantum Mechanics: Powell crasemann

Module - IX
Basics of Atomic and Nuclear Physics

Lectures: 02

Objectives

The basic objective of this module is to bridge the gap between what the students need to know before they can start taking the advanced courses in the college level and what they are actually aware of from the intermediate level. To serve this purpose this module has been prepared so that the students after carrying out this course can answer and explain following questions:

What is Matter?
What is atom?
What is the structure of an atom?
Discovery of electron, proton & neutron
Rutherford α particle scattering experiment
Classical atomic models: J.J. Thomson’s plum pudding & Rutherford’s planetary model
Shortcomings of Classical models: an unstable system
Bohr’s semi-classical model: Stationary orbits and quantization of angular momentum

Curiosity driven questions to start the subject

1. What is an atom? What are atoms made of? How do atoms form? What is the modern view of the structure of the atom?
2. What are the differences among protons, neutrons and electrons? What kind of charges they have? What are the exact relative masses of protons, neutrons and electrons? How many times bigger is a proton than an electron?
3. Why are electrons so far away from the nucleus of an atom?
4. Why do protons and neutrons stay together in the nucleus?
5. What holds an electron revolving around the nucleus? Why don't they just go zooming around everywhere?
6. How much of an atom is empty space?
7. How the emission and absorption of radiation takes place in an atom?

Introduction

Experiments on electric discharge through gases revealed that atoms of different elements contain negatively charged constituents (electrons) that are identical for all atoms. However, atoms on a whole are electrically neutral. Therefore, an atom must also contain some positive charge to neutralize the negative charge of the electrons. But what is the arrangement of the positive charge and the electrons inside the atom? Or in other words, what is the structure of an atom?

Answer to this question laid the discovery of electron, proton and neutron. In this context, the famous Rutherford α particle scattering experiment for the discovery of nucleus, different atomic models (e.g. plum pudding model proposed by J. J. Thomson & Rutherford's planetary model, also called the nuclear model of the atom), etc. play an important step towards how we see the atom. However, Both the Thomson's as well as the Rutherford's models constitute an unstable system. Thomson's model is unstable electrostatically, while Rutherford's model is unstable because of electromagnetic radiation of orbiting electrons.

The model proposed by Neils Bohr laid the foundation of the quantum theory by postulating specific orbits in which electrons do not radiate. This model explained many facets of the atomic structure and spectra of the Hydrogen atom. However, Bohr's semi-classical model based on some aspects of classical physics and some aspects of modern physics also does not provide a true picture of the simplest hydrogenic atoms.

1. **Matters**

- Anything that has mass and takes up space (volume)
    - Examples:
        - A brick has mass and takes up space
        - A desk has mass and takes up space
        - A pencil has mass and takes up space
        - Air has mass and takes up space

All of the above examples are considered matter because they have mass and take up space. Can you think of anything that would not be considered matter?

2. ATOMS

- Matter is anything that takes up space and has mass. All matter is made of atoms.
- Atoms are the basic building blocks of matter. They make up everything around us; Your desk, the board, your body, everything is made of atoms!
- Atoms are too small to see without powerful microscopes.

Atoms are so small that…

- It would take a stack of about 50,000 aluminum atoms to equal the thickness of a sheet of aluminum foil from your kitchen.

- If you could enlarge a penny until it was as wide as the US, each of its atoms would be only about 3 cm in diameter – about the size of a ping-pong ball.

- A human hair is about 1 million carbon atoms wide.

C-C-C-C-C-… + 999,995 more

- A typical human cell contains roughly 1 trillion atoms.

1 trillion atoms →

- A speck of dust might contain 3×10^{12} (3 trillion) atoms.

Is made of approximately 3 trillion atoms

- it would take you around 500 years to count the number of atoms in a grain of salt.

Just one of these grains

What is An Atom Made Of?
(First Approximation)

Electron(s)

Nucleus

Proton(s) **Neutron(s)**

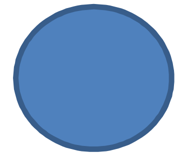
Atomic Theory

- Because we cannot see atoms, we use models to teach and learn about atoms.
- The atomic theory has changed over time as new technologies have become available.

Remember: Scientific knowledge builds on past research and experimentation.

Atomic Theory by John Dolton:

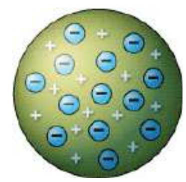
All matter is made of atoms. Atoms are too small to see, indivisible and indestructible. All atoms of a given element are identical.



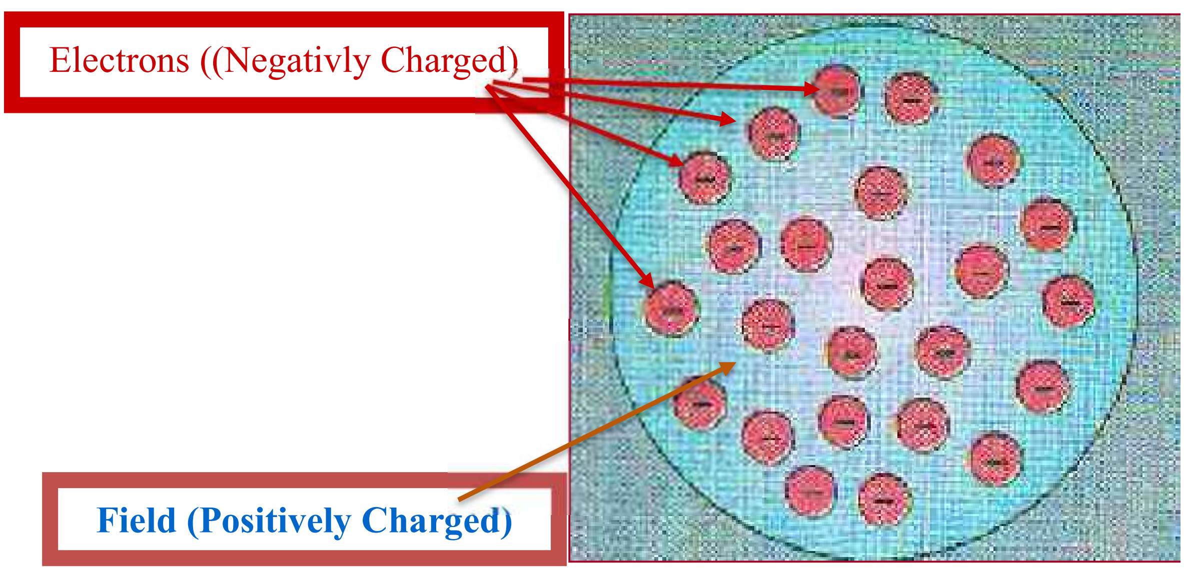
In 1803: John Dalton (England) formulated the modern version of the atomic theory. In his model: Matter is made up of small indivisible particles, called atoms. All atoms in a given chemical element are exactly alike, while the atoms of different elements differ by atomic weight. Atoms can neither be created nor destroyed. A chemical reaction is just a simple rearrangement of atoms and the same number of atoms must be present before and after the reaction.

Atomic Theory by J. J Thompson:

Discovered the negative electron, and predicted that there also must be a positive particle to hold the electrons in place.



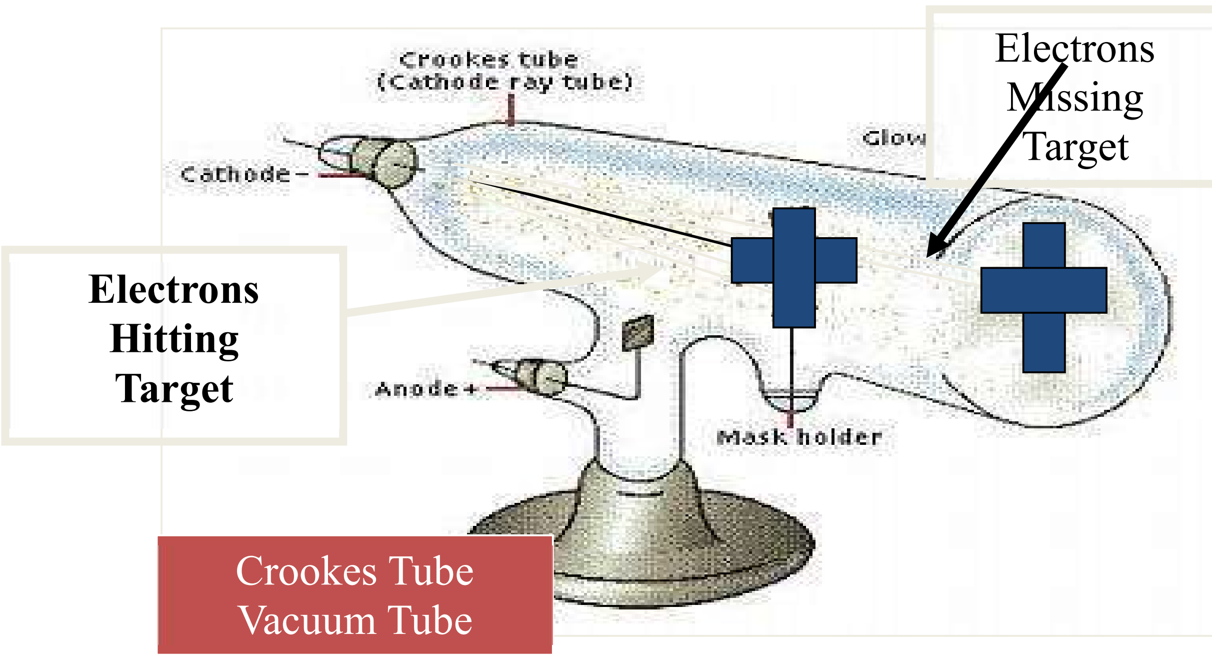
The Plum Pudding Model J.J. Thompson- 1904



In Today's Vernacular- The Chocolate Chip Model

In Thomson's model, the atom is composed of electrons (which Thomson still called "corpuscles," though G. J. Stoney had proposed that atoms of electricity be called electrons in 1894) surrounded by a soup of positive charge to balance the electrons' negative charges, like negatively charged "plums" surrounded by positively charged "pudding". The electrons (as we know them today) were thought to be positioned throughout the atom in rotating rings. In this model the atom was also sometimes described to have a "cloud" of positive charge.

Discovery of Subatomic Particle: Electron



In 1832: Michael Faradey showed that chemical changes occur when electricity is passed through an electrolyte. He stated that electricity is made up of particles called Atoms of Electricity.

In 1891: Later, G. J. Stoney suggested the name “Electron”, fundamental unit of electric charge, to describe the atoms of electricity, Michael Faraday’s 1832 experiment.

In 1897: In 1897, J. J. Thomson determined the charge/mass, as given in Equation below, ratio of these cathode ray particles “electrons” and found the value of 11 1.76 10 e Coulomb m Kg

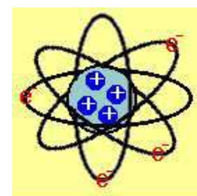
$$e/m = 1.76 \times 10^{11}\ Coulomb/Kg$$

Electrons, Electricity???

Electricity was thought of as a stream, (analogous to water flowing as a pipe)

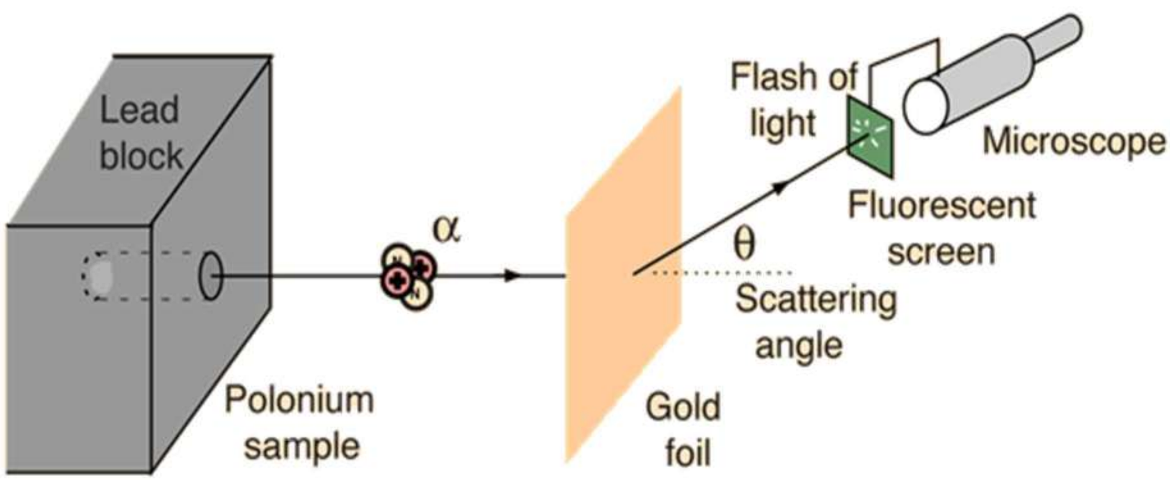
Now thought of as the motion of ***Discrete particles*** having ***mass*** (Electrons) Carrying the smallest possible unit of ***charge.***

Atomic Theory by Ernest Rutherford:



Discovered the nucleus of an atom and named the positive particles in the nucleus "protons". Concluded that electrons are scattered in empty space around the nucleus.

Discovery of Subatomic Particle: Nucleus



Rutherford called it a "Central Charge (1920 called it a Proton)

Rutherford Scattering

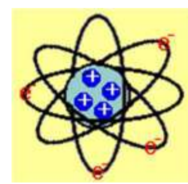
Alpha particles from a radioactive source were allowed to strike a thin gold foil. Alpha particles produce a tiny, but visible flash of light when they strike a fluorescent screen. Surprisingly, alpha particles were found at large deflection angles and some were even found to be back-scattered.

This experiment showed that the positive matter in atoms was concentrated in an incredibly small volume and gave birth to the idea of the nuclear atom.

In so doing, it represented one of the great turning points in our understanding of nature.

Atomic Theory by James Chadwick:

Discovered that neutrons were also located in the nucleus of an atoms and that they contain no charge.



Electrons could never exist in nucleolus
(a way towards the discovery of the Neutron)

Since the time of Rutherford it had been known that the atomic mass number A of nuclei is a bit more than twice the atomic number Z for most atoms and that essentially all the mass of the atom is concentrated in the relatively tiny nucleus.

As of about 1930 it was presumed that the fundamental particles were protons and electrons, but that required that somehow a number of electrons were bound in the nucleus to partially cancel the charge of A protons.

But by this time it was known from the modern physics (Uncertainty and from particle in a box type confinement) calculations that there just wasn't enough energy available to contain electrons in the nucleus.

Discovery of the Neutron James Chadwick – 1935

An experimental breakthrough came in 1930 with the observation by Bothe and Becker that bombardment of beryllium with alpha particles from a radioactive source produced neutral radiation which was penetrating but non-ionizing. They presumed it was gamma rays.

Curie and Joliot showed that when you bombarded a paraffin target with this radiation, it ejected protons with energy about 5.3 MeV. This proved to be inconsistent with gamma rays. Because, the necessary energy for the gamma ray explanation should be much greater (nearly ten times or even more).

Chadwick was able to prove that the neutral particle could not be a gama-photon rather they are a unique particle neutron.

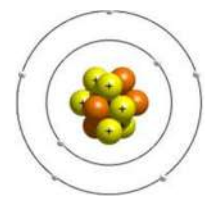
Essentially the same mass as an Proton

Has no Charge

With the Proton, Makes up the Nucleus

Atomic Theory by Neils Bohr:

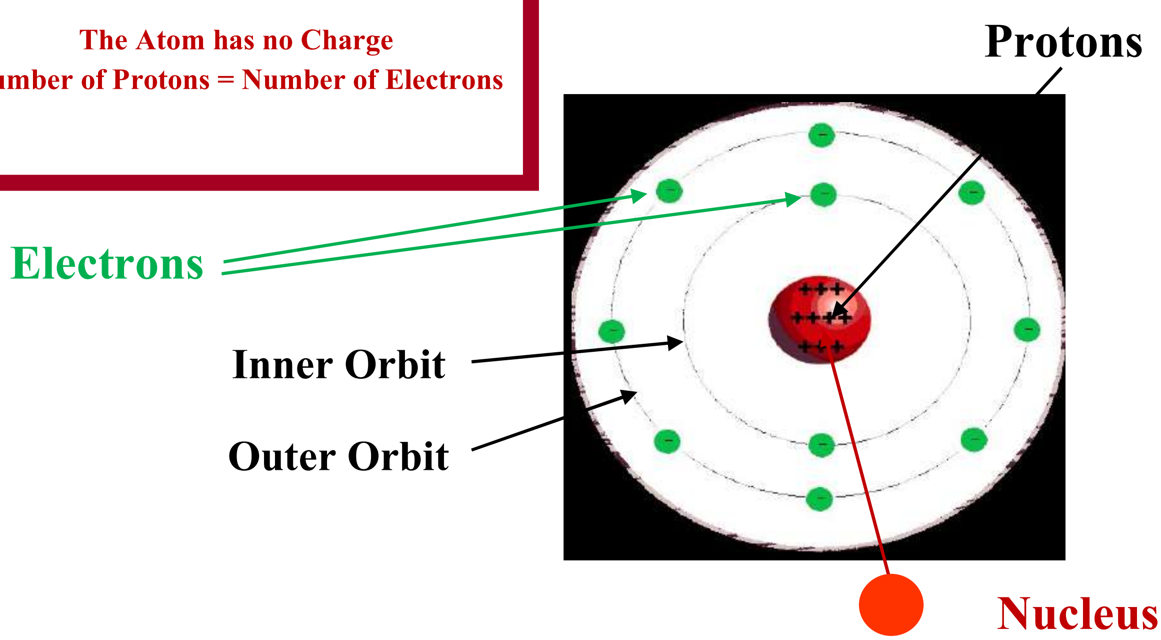
Concluded that electrons are located in planet-like orbits around the nucleus in certain energy levels.



Bohr's Model: 1913 – Atomic Model with Fixed Orbits proposed –

Modeled after the Solar System

The Atom has no Charge
Number of Protons = Number of Electrons



Bohr’s Postulates

- Stationary States: An atomic system possesses a number of states in which no emission of radiation takes place. Such states are called stationary states.

In case of Hydrogen atom, each stationary state corresponds to circularly orbiting electron and nucleus at the centre.

- Allowed stationary states: Only those orbits are possible for which the relation $p=n(h/2\pi)$ holds good, where p is the angular momentum, h is the Planck’s constant, n is the principal quantum number (1,2,3,….)

- Emission of emitted radiation: Emission of radiation occurs only when the electron jumps from one of the allowed states E_{n1} to another of lower energy of E_{n2}, in accordance with the relation

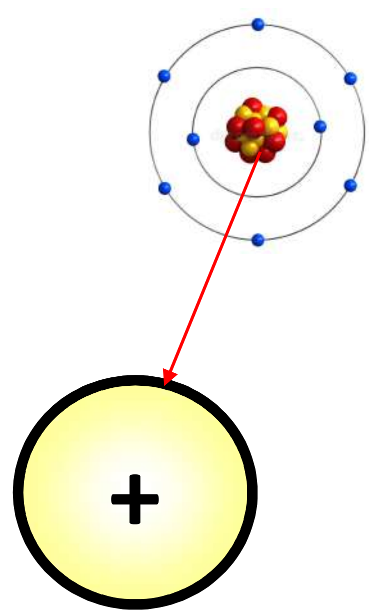
$$h\nu=E_{n1}-E_{n2}$$

Bohr radius: $r=n^2h^2/4\pi^2me^2Z$

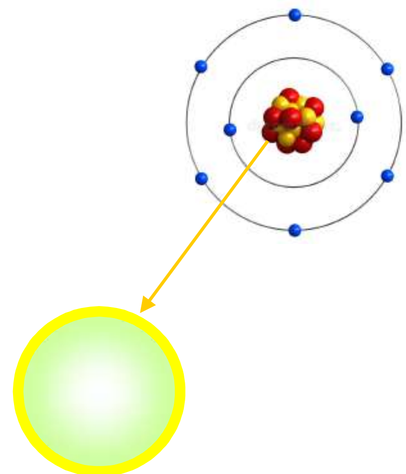
Emitted Frequency: $\nu=RZ^2(1/n_2^2 - 1/n_1^2)$, R, the Rydberg constant = $2\pi^2me^4/ch^3=109677\ cm^{-1}$.

Protons (+)

- Positively charged particles
- Help make up the nucleus of the atom
- Help identify the atom (could be considered an atom’s DNA)
- Equal to the atomic number of the atom
- Contribute to the atomic mass
- Equal to the number of electrons

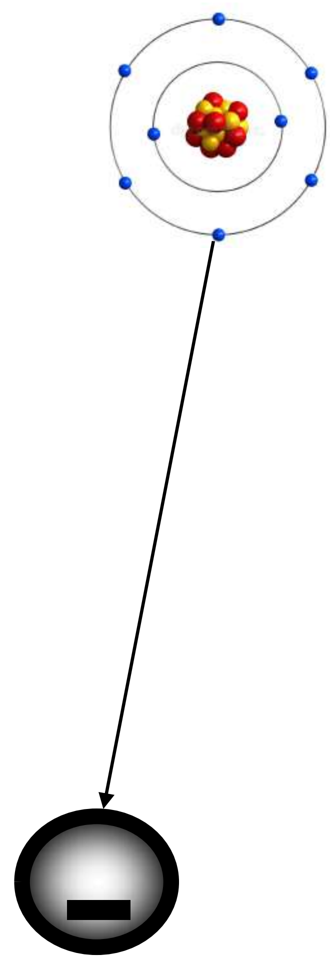


Neutrons



- Neutral particles; have no electric charge
- Help make up the nucleus of the atom
- Contribute to the atomic mass

Electrons (-)

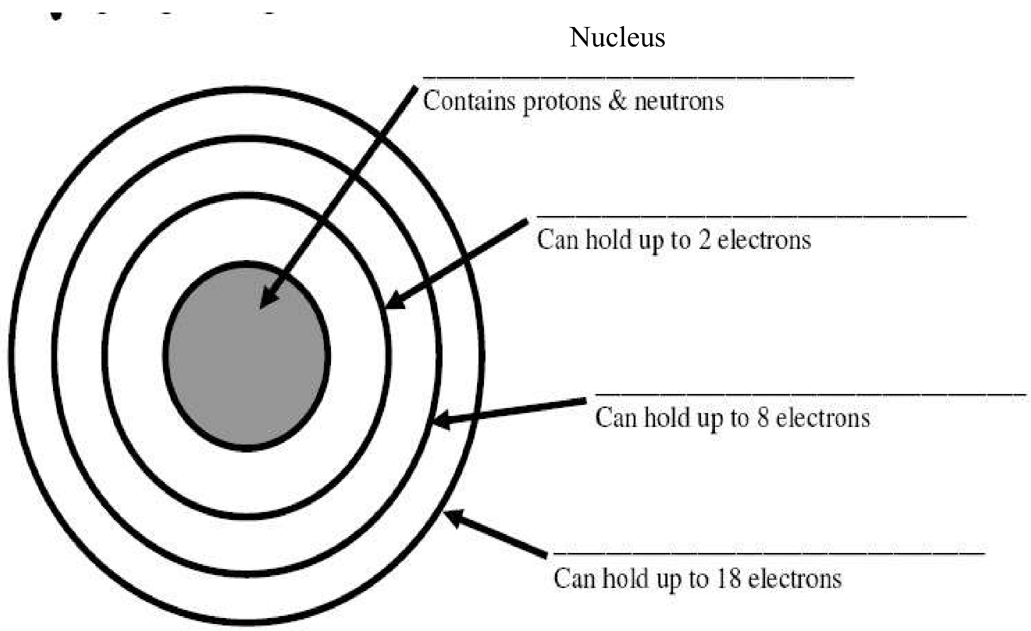


- Negatively charged particles
- Found outside the nucleus of the atom, in the electron orbits/levels; each orbit/level can hold a maximum number of electrons ($1^{st} = 2$, $2^{nd} = 8$, $3^{rd} = 8$ or 18, etc…)
- Move so rapidly around the nucleus that they create an electron cloud
- Mass is insignificant when compared to protons and neutrons
- Equal to the number of protons
- Involved in the formation of chemical bonds

Electrons have special rules….

- You can't just shove all of the electrons into the first orbit of an electron.
- Electrons live in something called **<u>shells or energy levels</u>**.
- Only so many can be in any certain shell.

Nucleus



What Did Bohr's Model Accomplish?

Stability

Electrical Forces balanced by Centrifugal Forces

Identity

Change in number of Protons/Electrons Changes the Element

Regeneration

Once Pulled Apart, the Atom Reforms as before

How Big is It, (Are They)?

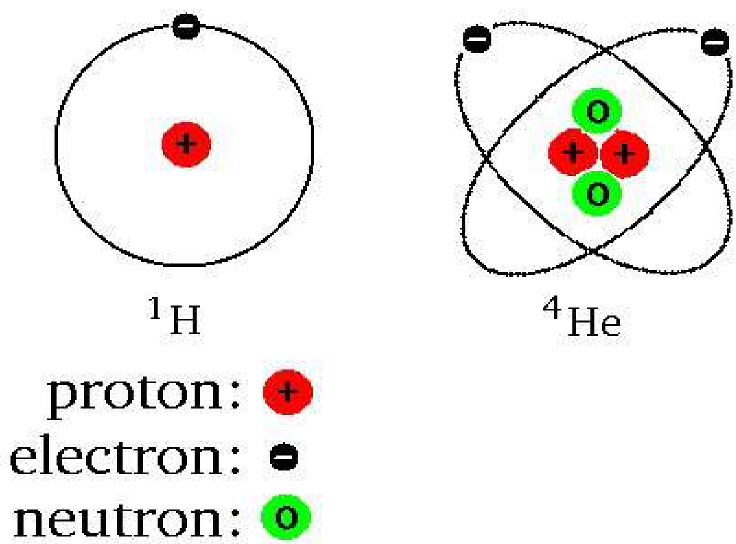
1 Proton (or Neutron)----1.67 x10^{-24} Grams
=0.00000000000000000000000167 Grams

1 electron-----9.1x10^{-28} Grams
=0.0000000000000000000000000091Grams

The Proton weighs 1800 x the Electron

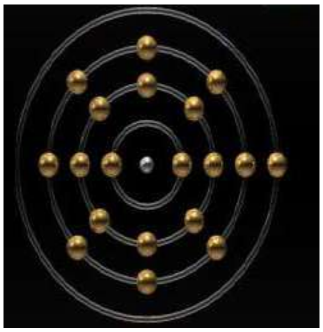
The Atom is about 5 X 10^{-8} cm

Bohr Model of Hydrogen and Helium



Bohr Model of Potassium(K) and Uranium (U)

$39P_{19}$



19 Protons
20 Neutrons
Atomic Mass=39

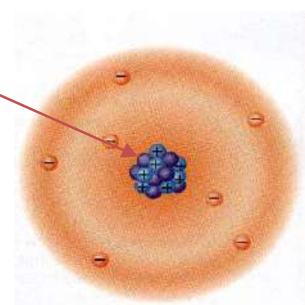
$235U_{92}$ $238U_{92}$

92 Protons
143 Neutrons
Atomic Mass = 235

92 Protons
146 Neutrons
Atomic Mass=238

Therefore Atoms….

Protons and neutrons "hang out" together at the core of the atom called the nucleus.



Protons + neutrons = atomic mass

Protons = atomic number

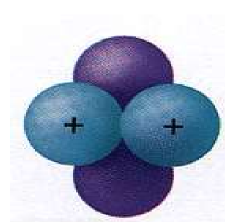
Different elements = Different number of protons.

and…

What is the atomic number of this element?

What is the atomic mass?

What is the name of this element?



Helium

and…

If you added 3 protons, what element would you have?

Boron

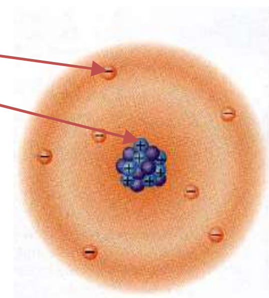
and…

"Charge" of atoms

Protons and electrons attract each other.

If you find the same number of protons and electrons, the element has NO CHARGE.

Nitrogen



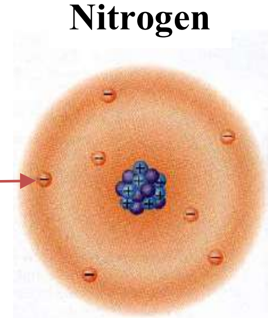
N

and…

"Charge" of atoms

If the atom "gets" an electron, it becomes "negatively charged"

Nitrogen



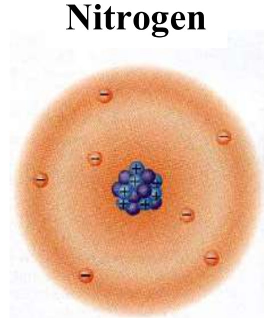
$\mathbf{N^{-1}}$

and…

"Charge" of atoms

If the atom "loses" an electron, it becomes "positively charged"

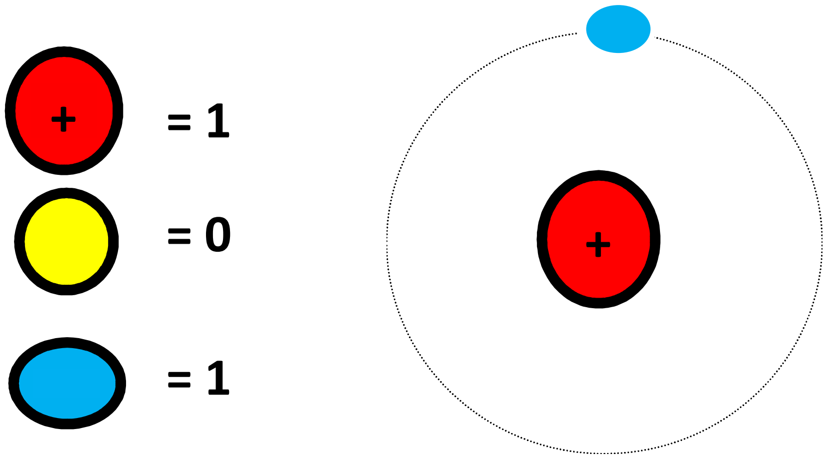
Nitrogen



$\mathbf{N^{+1}}$

Furthermore, Hydrogen (H) Atom

- **Notice the one electron in the first orbital**

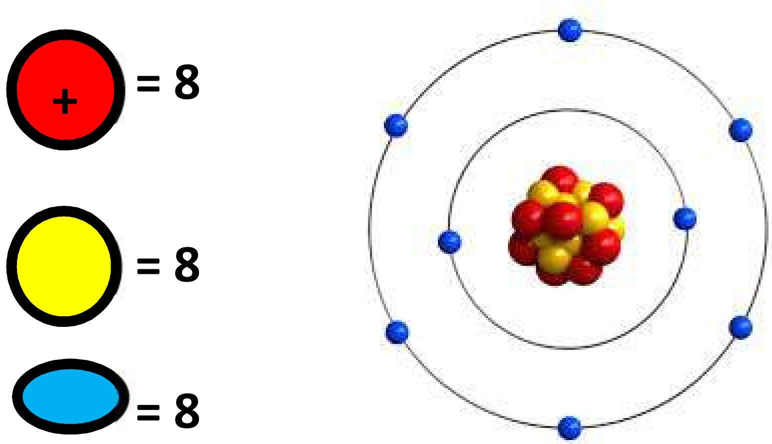


How many more electrons can fit in the 1st orbital/ level?

Even though there are no neutrons present, Hydrogen is still considered an atom

Oxygen (O) Atom

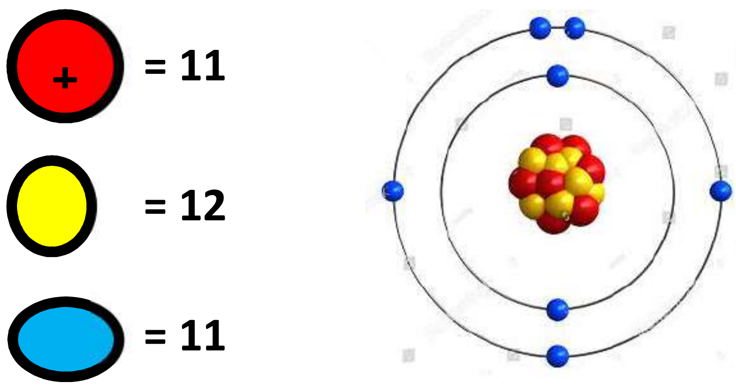
- **Notice the two electrons in the first orbital/level and the six in the second**



How many more electrons can fit in the 2nd orbital/ level?

Sodium (Na) Atom

- **Notice the two electrons in the first orbital/level, eight in the second, and one in the third**



How many more electrons can fit in the 3rd orbital/ level?

Limitations of Bohr's Theory

1. It fails to explain the spectra of multi-electron atoms (atoms having more than one electron).
2. According to Bohr, the circular orbits of the electrons are planar. However, this is not true as the electrons move around the nucleus in three-dimensional space.
3. It fails to explain the cause of chemical combination and shapes of the molecules arising out of it.
4. One of the major drawbacks of Bohr's theory is that it assumes a definite knowledge about the position and momentum of electrons at the time of measurement. However, Heisenberg (1927) suggested that it is impossible to measure simultaneously both the exact position and momentum of a subatomic particle such as an electron. This statement is known as Heisenberg's uncertainty principle.
5. It cannot explain the relative intensities of spectral lines.
6. Bohr's theory failed to account for the splitting of spectral lines on the application of magnetic field (Zeeman effect) and also the application of the electric field (stark effect)

Nevertheless, Bohr's theory has its own importance, as it was the first theory to introduce the concept of quantization in the behaviour of subatomic particles. Bohr's theory was abandoned 12 years after its formulation in favour of the present quantum theory of atomic structures.

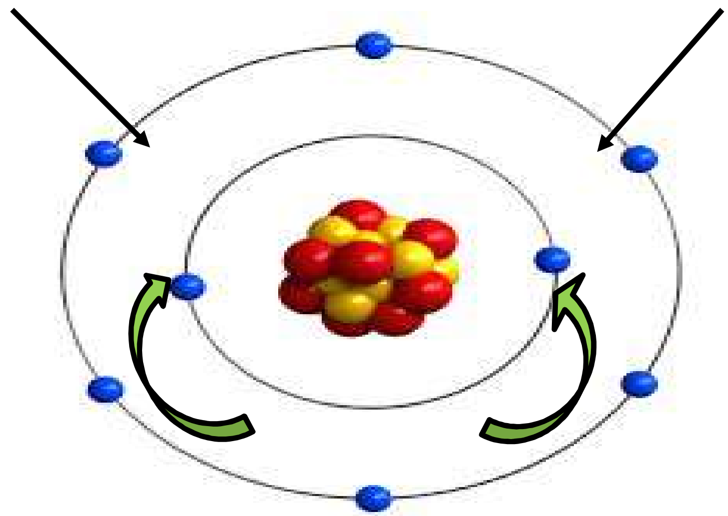
Nucleus: The Atom's "Center"

The nucleus is the center of an atom. It is made up of nucleons (protons and neutrons) and is surrounded by the electron cloud.

Protons and neutrons are grouped together to form the "center" or **nucleus** of an atom.

The size (diameter) of the nucleus is between 1.6 fm (10−15 m) (for a proton in light hydrogen) to about 15 fm (for the heaviest atoms, such as uranium). These sizes are much smaller than the size of the atom itself by a factor of about 23,000 (uranium) to about 145,000 (hydrogen).

Notice that the electrons are not apart of the nucleus



The nucleus has most of the mass of an atom, though it is only a very small part of it. Almost all of the mass in an atom is made up from the protons and neutrons in the nucleus with a very small contribution from the orbiting electrons.

Atomic Number and Mass number: isotopes and isobars

The number of protons is called the atomic number and determines the chemical element.

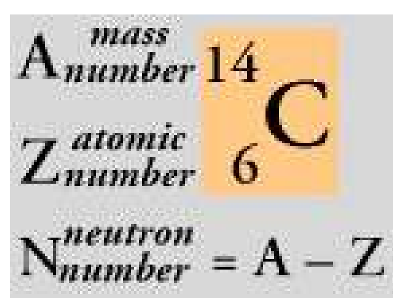
The mass number of an element is defined as the total number of protons and neutrons present in the nucleus of an atom of the element.

The mass number is the atomic weight of the element, corrected to the nearest integer.

Mass number (A) = number of protons + number of neutrons.

Nuclide symbol:

This consists of the symbol for the element E with the atomic number Z written as a subscript at the lower left and the mass number A at the upper right, as a superscript.



Thus in the isotope of carbon represented as ${}_6C^{14}$, there are six protons and eight neutrons in the carbon nucleus. Therefore, its mass number is 6 + 6 =12.

Isotopes and isobars

Isobars are atoms of different elements having the same mass number but different atomic numbers.

For example: $_1H^1$ and $_1H^2$

Nuclei of a given element (same atomic number) may have different numbers of neutrons and are then said to be different isotopes of the element.

For example: $_1H^3$ and $_2He^3$

Questions to test the concept of students after the course is finished

1. Which of the following will not show deflection from the path on passing through an electric field? Proton, electron, neutron.
2. An atom having atomic mass number 13 has 7 neutrons. What is the atomic number of the atom?
3. Nickel atom can lose two electrons to form Ni ion. The atomic number of nickel is 28. From which orbital will nickel lose two electrons.
4. What did Rutherford's foil experiment reveal?
5. If Rutherford's planetary model were correct, atoms would be extremely unstable. Explain why?
6. How is Bohr's model of hydrogen similar to Rutherford's planetary model? How are the two models different?
7. How does Bohr's model account for the atomic spectra?
8. The Balmer series in the hydrogen spectrum corresponds to the transition from n = 2 to n = 3,4,………. This series lies in the visible region. Calculate the wave number of line associated with the transition in Balmer series when the electron moves to n = 4 orbit. (R = 1 09677 cm)
9. Calculate the energy and frequency of the radiation emitted when an electron jumps from n = 3 to n = 2 in a hydrogen atom.

References

1. Atomic Physics: H. E. White
2. Physics of Atoms and Molecules: B. Brandsden