# PREFACE

In a bid to standardize higher education in the country, the University Grants Commission (UGC) has introduced Choice Based Credit System (CBCS) based on five types of courses viz. *core, generic elective, discipline Specific, ability and skill enhancement* for graduate students of all programmes at Honours level. This brings in the semester pattern, which finds efficacy in sync with credit system, credit transfer, comprehensive continuous assessments and a graded pattern of evaluation. The objective is to offer learners ample flexibility to choose from a wide gamut of courses, as also to provide them lateral mobility between various educational institutions in the country where they can carry their acquired credits. I am happy to note that the university has been recently accredited by National Assesment and Accreditation Council of India (NAAC) with grade "A".

UGC (Open and Distance Learning Programmes and Online Programmes) Regulations, 2020 have mandated compliance with CBCS for U.G. programmes for all the HEIs in this mode. Welcoming this paradigm shift in higher education, Netaji Subhas Open University (NSOU) has resolved to adopt CBCS from the academic session 2021-22 at the Under Graduate Degree Programme level. The present syllabus, framed in the spirit of syllabi recommended by UGC, lays due stress on all aspects envisaged in the curricular framework of the apex body on higher education. It will be imparted to learners over the six semesters of the Programme.

Self Learning Materials (SLMs) are the mainstay of Student Support Services (SSS) of an Open University. From a logistic point of view, NSOU has embarked upon CBCS presently with SLMs in English/Bengali. Eventually, the English version SLMs will be translated into Bengali too, for the benefit of learners. As always, all of our teaching faculties contributed in this process. In addition to this we have also requisitioned the services of best academics in each domain in preparation of the new SLMs. I am sure they will be of commendable academic support. We look forward to proactive feedback from all stakeholders who will participate in the teaching-learning based on these study materials. It has been a very challenging task well executed, and I congratulate all concerned in the preparation of these SLMs.

I wish the venture a grand success.

**Professor (Dr.) Subha Sankar Sarkar**
Vice-Chancellor

Netaji Subhas Open University
Under Graduate Degree Programme
Choice Based Credit System (CBCS)
Subject : Honours in Physics

Course : Waves and Optics
Course Code : CC-PH-09

# Netaji Subhas Open University
## Under Graduate Degree Programme
## Choice Based Credit System (CBCS)
## Subject : Honours in Physics

## Course : Waves and Optics
## Course Code : CC-PH-09

### : Board of Studies :
### Members

**Professor Kajal De**
(*Chairperson*)
*Director, School of Sciences,*
*NSOU*

**Dr. Gautam Kumar Mallik**
*Associate Professor of Physics,*
*NSOU*

**Mr. Pranab Nath Mallik**
*Associate Professor of Physics,*
*NSOU*

**Dr. Gahul Amin**
*Assistant Professor of Physics,*
*NSOU*

**Dr. Gautam Gangopadhyay**
*Professor of Physics*
*University of Calcutta*

**Dr. Rupayan Bhattacharya**
*Retd. Principal, Gaurudas College*

**Dr. Amit Kumar Chakraborty**
*Associate Professor of Physics,*
*National Institute of Technology*

**Dr. Subhratanu Bhattacharya**
*Assistant Professor of Physics,*
*Kalyani University*

**Dr. Manik Sanyal**
*Associate Professor of Physics,*
*Barasat College*

### : Course Writer :

Unit : 1–7 (Waves)
**Dr. Asim Kr. Mukherjee**
*Associate Professor of Physics,*
*B B College, Asansol*

Unit : 8–14 (Optics)
**Dr. Gautam Kumar Mallik**
*Associate Professor of Physics, NSOU*

### : Course Editor :

Unit : 1–7 (Waves)
**Dr. Gautam Kumar Mallik**
*Associate Professor of Physics, NSOU*

Unit : 8–14 (Optics)
**Dr. Gahul Amin**
*Associate Professor of Physics, NSOU*

### : Format Editor :

**Dr. Gautam Kumar Mallik**
*Associate Professor of Physics, NSOU*

### Notification

**Dr. Ashit Baran Aich**
Registrar

**Netaji Subhas Open University**

**Course : Waves and Optics**
**Course Code : CC-PH-09**

# Unit-1 : Recapitulation of SHM

**Structure**

## 1.0   Objectives

After reading this unit, you will learn
- The necessary conditions for SHM.
- Formation and solution of differential equation of SHM.
- Calculation of kinetic and potential energy of a body executing SHM.
- Similarities between different type of SHM.

## 1.1   Introduction

We know that motion of a body is usually two types. (a) linear and (b) rotational. In these motions there is a particular type of motion wehre the body returns to its

initial state after a certain interval of time. These motions are called periodic motion. Thus of a body repeats its path after a regular interval of time, then it is called periodic motion. There is a proverb that "World is full of periodic motion". In fact in every nooks and corners of our daily life are get the example of periodic motion. The motion of a fan, motion of a pendulum clock, motion of leaves in trees—the motion of the earth round the sun etc. It a periodic motion fulfils some condition then it is calld simple harmonic motion. we can analyse and describe these motions by simple assumption.

In nature we get many examples of SHM and hence its importance and usefulness knew no bounds. We will show that these motions are the smallest building block of any kind of vibration.

## 1.2 Definition and characteristic

It a body in periodic motion has its acceleration always proportional to its instantaneous displacement and always in a direction opposite to displacement, then it is called simple harmonic motion. The main characteristic of this motion is that the force acting on it must be proportional to the displacement and always in a direction opposite to it.

### 1.2.1 Conditions

If the motion of a body is such that it satisfies the two following conditions—

(a) The acceleration $\infty$ instantaneous displacement ie $f \infty x$.

(b) The direction of acceleration is opposite to displacement.

Combining both we get $\vec{f} \infty - \vec{x}$

It is to be noted that in general we consider the motion of a particle is point mass but for an entended body are will consider the centre of mass of it.

### 1.2.2 Differential Equation

Let us consider a particle of mass m moving along x-axis. With its equilibrium point at 0. Let $\vec{f}$ is the acceleration of the mass when its displacement is $\vec{x}$ from 0. Then for SHM we must have

$$\vec{f} \infty - \vec{x} \qquad \text{... (1.2.1)}$$

or, $\quad m\vec{f} \infty - \vec{x}$ or, $m\vec{f} = -k\vec{x}$ or, $\vec{f} = \dfrac{k}{m}\vec{x}$
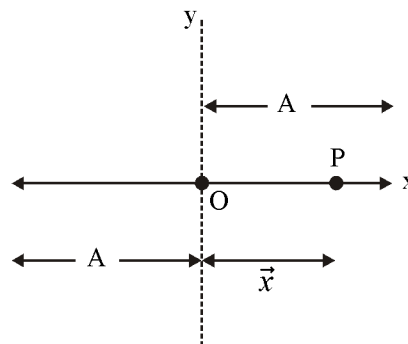
or, $\dfrac{d^2\vec{x}}{dt^2} = \omega^2 \vec{x}$      ... (1.2.2)

where $\omega = \sqrt{\dfrac{k}{m}}$ is the angular velocity of mass m.

### 1.2.3 Solution

The equation (1.2.2) is a linear second order homogeneous differential equation. It can have three types of solution—

1. trigonometric solution—      $x = A_1 \sin \omega t + B_1 \cos \omega t$

2. exponential solution—      $x = A_1 e^{i\omega t} + B_1 e^{-i\omega t}$

3. series solution—      $x = \displaystyle\sum_{n=0}^{\infty} a_n t^{k+n}$

Out of these the most familier is trigonometric solution.

If we now impose condition that of $t = 0$, $x = 0$, then we get $B = 0$, so $x = A \sin \omega t$ will be acceptable but if we put the condition that at $t = 0$, $x = A$, then we get $x = A \cos \omega t$ is the acceptable solution.

● If we consider exponential solution then $x = A_1 e^{i\omega t} + B_1 e^{-i\omega t}$.

But $e^{\pm i\omega t} = \cos \omega t \pm i \sin \omega t$, hence $x = \left(A_1 + B_1\right)\cos \omega t + i\left(A_1 - B_1\right)\sin \omega t$

It we put $A_1 + B_1 = A \cos \theta$ and $i\left(A_1 - B_1\right) = A \sin \theta$, then we get

$x = A \cos\left(\omega t - \theta\right)$ as the general solution.

● We can show that the series solution also leads to same solution is

$x = A \cos\left(\omega t - \theta\right)$, and putting $\theta = 0$ or $\theta = \pi/2$ we get the solution

$x = A \cos \omega t$ or $x = A \sin \omega t$.

### 1.2.4 Some parameters for description of SHM

In general, we see that SHM is described by $x = A \cos\left(\omega t - \theta\right)$. Here $x$ is the instantaneous displacement. $A$ is the maximum displacement called *amplitude*, $\omega$ is

the *angular velocity* which is $\frac{2\pi}{T} = 2\pi\upsilon$. Where $\upsilon$ is the *frequency* of oscillation is number of oscillation per unit time and T is the *time period* i.e. time required for one complete oscillation. The most important parameter related to SHM is *phase*, which is defined as the quantity (angle) that determines the state of motion of a oscillation. Here $(\omega t - \theta)$ is the phase of the SHM and $\theta$ is the initial phase of the motion called *epoch*.

The unit of displacement or amplitude is cm or meter.

The unit of frequency is cps or hertz.

The unit of time period is second.

The unit of phase is radian or degree.

**Exercise 1.** Which of the following describe SHM. Find the amplitude and frequency.

(a) $x = A\log(\omega t + \theta)$, (b) $x = A\cos^2(\omega t + \theta)$, (c) $x = e^{(\omega t + \theta)}$, (d) $x = A\sin^3 \omega t$.

2. A particle executes SHM with $x = 5\sin\frac{2\pi}{5}t$ cm. Find the distance the particle travels in 25.25 second.

3. The equation of an SHM is $x = 3\cos\left(25t + \frac{\pi}{3}\right)$. What is the amplitude, phase and epoch of this motion.

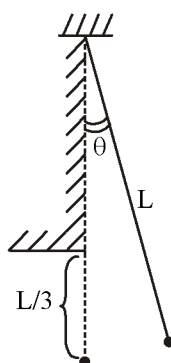4. Find the time period of oscillation of this restricted pendulum.



**Fig. 1.1**

5. Calculate the ratio of time taken by a particle execuiting SHM in describing the 1st half and next half of its amplitude.

## 1.3 Mechanical energy of a body executing SHM

A particle in SHM has both K.E and PE at any instant.

If $x = a\sin(\omega t + \phi)$ is the equation of SHM, then $v = \dfrac{dx}{dt} = \omega a\cos(\omega t + \phi)$ is the velocity and if m is the mass of the body, then.

KE at any instant $t$ is $KE(t) = \dfrac{1}{2}mv^2 = \dfrac{1}{2}m\omega^2 a^2 \cos^2(\omega t + \phi)$

$$= \frac{1}{2}m\omega^2\left(a^2 - x^2\right) \qquad\qquad \text{... (1.3.1)}$$

Similarly the PE is

$$\int_0^x Fdx = k\int_0^x Fdx = \frac{1}{2}kx^2$$

But $k = m\omega^2$ so $\text{PE} = \dfrac{1}{2}m\omega^2 a^2 \sin^2(\omega t + \phi) = \dfrac{1}{2}m\omega^2 x^2$   ... (1.3.2)

The Sum of PE and KE at any instant is ME = KE + PE

$$= \frac{1}{2}m\omega^2 a^2 = \text{Constant} \qquad\qquad \text{... (1.3.3)}$$

Now the time average of KE and PE are

$$\langle KE\rangle_T = \frac{1}{T}\int_0^T \frac{1}{2}m\omega^2 a^2 \cos^2(\omega t + \phi)\,dt = \frac{1}{4}m\omega^2 a^2$$

$$\langle PE\rangle_T = \frac{1}{T}\int_0^T \frac{1}{2}m\omega^2 a^2 \sin^2(\omega t + \phi)\,dt = \frac{1}{4}m\omega^2 a^2$$

Hence $\langle ME\rangle_T = \dfrac{1}{2}m\omega^2 a^2$      .... (1.3.4)

The average of KE over displacement is

$$\langle KE\rangle_x = \frac{1}{2a}\int_{-a}^a \frac{1}{2}m\omega^2\left(a^2 - x^2\right)dx$$

$$= \frac{1}{2a} m\omega^2 \cdot \frac{1}{2} \left[ a^2 x - \frac{x^3}{3} \right]_{-a}^{+a} = \frac{1}{2a} \frac{m\omega^2}{2} \left\{ a^3 - \frac{a^3}{3} - \left( -a^3 + \frac{a^3}{3} \right) \right\}$$

$$= \frac{1}{2a} \frac{m\omega^2}{2} \left\{ 2a^3 - \frac{2a^3}{3} \right\} = \frac{m\omega^2}{2a.2} .2a \left( \frac{2}{3} a^2 \right) = \frac{m\omega^2}{2} . \frac{2}{3} a^2$$

$$\langle PE \rangle_x = \frac{1}{2a} \int_{-a}^{+a} \frac{1}{2} m\omega^2 x^2 dx = \frac{1}{2a} . \frac{1}{2} m\omega^2 \left\{ \frac{a^3 + a^3}{3} \right\} = \frac{1}{2a} \frac{1}{2} m\omega^2 \frac{2a^3}{3}$$

$$= \frac{1}{2} m\omega^2 \frac{a^2}{3}$$

$$\therefore \quad \langle KE \rangle_x + \langle PE \rangle_x = \frac{m\omega^2}{2} \left\{ \frac{2}{3} a^2 + \frac{1}{3} a^2 \right\} = \frac{1}{2} m\omega^2 a^2 = \langle ME \rangle_T$$

Thus the average of mechnical energy both w.r. to time and displacement are same.

Average of KE and PE of a particle executing SHM :

$$KE = \frac{1}{2} m\omega^2 \left( a^2 - x^2 \right) \qquad\qquad PE = \frac{1}{2} m\omega^2 x^2$$

$$\langle KE \rangle_{disp} = 2. \frac{1}{4a} \int_{a}^{+a} \frac{1}{2} m\omega^2 \left( a^2 - x^2 \right).dx \qquad \langle PE \rangle_{disp.} = \frac{1}{2} m\omega^2 \frac{1}{2a} \int_{-a}^{+a} x^2 dx = \frac{1}{2} m\omega^2 \frac{1}{2a} \left\{ \frac{x^3}{3} \right\}_{-a}^{+a}$$

$$= \frac{1}{2a} . \frac{1}{2} m\omega^2 \left\{ a^2 x - \frac{x^3}{3} \right\}_{-a}^{+a} \qquad\qquad = \frac{1}{2} m\omega^2 . \frac{1}{2a} . \frac{2a^3}{3}$$

$$= \frac{1}{2a} . \frac{1}{2} m\omega^2 \left\{ a^3 - \frac{a^3}{3} - \left( -a^3 + \frac{a^3}{3} \right) \right\} \qquad = \frac{1}{2} m\omega^2 . \frac{a^2}{3}$$

Hence

$$= \frac{1}{2a} . \frac{1}{2} m\omega^2 \left\{ 2a^3 - \frac{2a^3}{3} \right\} \qquad\qquad \langle KE \rangle_{disp} + \langle PE \rangle_{disp} = \langle ME \rangle_{disp} = \frac{1}{2} m\omega^2 a^2$$

$$= \frac{1}{2} m\omega^2 \left\{ a^3 - \frac{a^2}{3} \right\}$$

$$= \frac{1}{2} m\omega^2 \frac{a^2}{3}$$

Time average of KE and PE :

$$\left\langle KE \right\rangle_T = \frac{1}{2} m\omega^2 a^2 \cos^2 \omega t = \frac{1}{4} m\omega^2 a^2$$

$$\left\langle PE \right\rangle_T = \left\langle \frac{1}{2} m\omega^2 a^2 \sin^2 \omega t \right\rangle = \frac{1}{4} m\omega^2 a^2$$

$$\therefore \quad \left\langle PE \right\rangle_T + \left\langle KE \right\rangle_T = \left\langle ME \right\rangle_T = \frac{1}{2} m\omega^2 a^2$$

## 1.4  Exampls of SHM

### 1.4.1 Mass spring system :

It $x$ is the instantaneous displacement of mass m, then the equation of motion is

$$m\frac{d^2x}{dt^2} = -kx \quad \text{or} \quad \frac{d^2x}{dt} = -\frac{k}{m}x$$

So if $\omega^2 = \frac{k}{m}$, then $\omega^2 = \sqrt{\frac{k}{m}}$ or, $T = 2\pi\sqrt{\frac{m}{k}}$ is the time period of oscillation.

In this example we have considered zero mass of the spring but if we assume the spring has mass M, then we can show that the time period becomes

$$T = 2\pi\sqrt{\frac{m+M/3}{k}} \ .$$
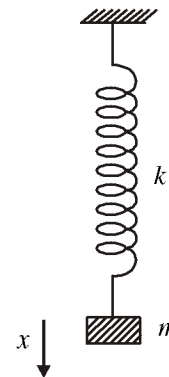


**Fig. 1.2**

### 1.4.2 Simple pendulum :

In case of a S.P. of bob mass $m$ and effective length L, the equation of motion is

$$m\frac{d^2x}{dt^2} = -mg\sin\theta \quad \text{and if } \theta < 10^\circ, \text{ then } \sin\theta \simeq \theta \text{ and as}$$

$\theta = \frac{x}{L}$ so the equation of motion becomes $m\frac{d^2x}{dt^2} = -mg\frac{x}{L}$

hence $\omega = \sqrt{\frac{g}{2}}$ or, $T = 2\pi\sqrt{\frac{L}{g}} = T_0$.



$mg \cos\theta$        $mg \cos\theta$

**Fig. 1.3**

But if $\theta > 10°$, then for large values of $\theta$ we have to take $\sin\theta = \theta - \dfrac{\theta^3}{3!} + \dfrac{\theta^5}{5!} + \ldots$

and detailed calculation shows that $T = 2\pi\sqrt{\dfrac{L}{g}} \times \left\{1 + \dfrac{\theta^2}{16} + \dfrac{90^4}{1024} + \ldots\right\}$

Thus with increase of angular displacement, the time period increases.

It is to be noted that when $\theta \simeq 7.25°$, the change in T is only 0.1%.

### 1.4.3 Compound pendulum :
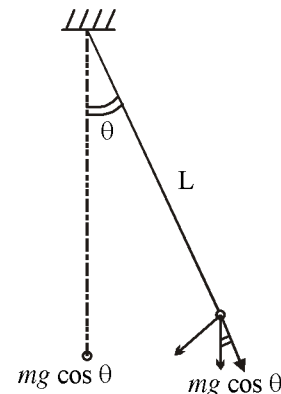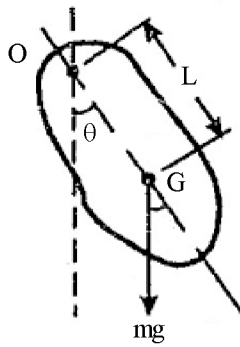


A rigid body of any arbitrary shape suspended from a smooth fixed point and allowed to oscillate in a vertical plane constitute a compound pendulum. If l is the distance between point of suspension and centre of gravity, then the equation of motion

$$I\dfrac{d^2\theta}{dt^2} = -mgL\sin\theta.$$ If $\theta < 4°$, then $\dfrac{d^2\theta}{dt^2} = -\dfrac{mgL}{I}\theta$

Hence $\omega = \sqrt{\dfrac{mgL}{I}}$ or, $T = 2\pi\sqrt{\dfrac{I}{mgL}}$

**Fig. 1.4**

Here I is the M.I of the compound pendulum about the point of suspension.

### 1.4.4 Torsional Pendulum :

When a disc or cylinder is suspended by a wire of torsional rigidity C, and if the disc or cylinder is slightly rotated by an angle $\theta$ and released, then its equation of motion will be

$$I\dfrac{d^2\theta}{dt^2} = -C\theta$$

or, $\dfrac{d^2\theta}{dt^2} = -\dfrac{C}{I}\theta$

which is the equation of motion of a SHM.



**Fig. 1.5**

Hence $\omega = \sqrt{\dfrac{C}{I}}$ or $T = 2\pi\sqrt{\dfrac{1}{C}}$. Here I is the moment of inertia of the disc or cylinder about its axis of rotation.

Here it is to be noted that in many SHM we put $\sin\theta = \theta$, i.e. it is valid only for

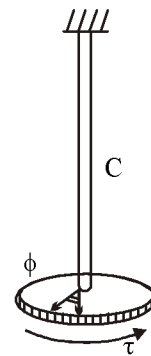$\theta < 4°$. But in this case of torsional pendulum $\dfrac{d^2\theta}{dt^2} \infty -\theta$ for any value of $\theta$. Here it is to be noted that for that value of $\theta$, we should not cross elastic limit of the wire.

### 1.4.5 Oscillation of charge and current :

We can show that if a charged capacitor is connected with an inductance then the charge and hence the current in the circuit starts executing SHM.

Let $q$ is the charge and i is the current in the LC circuit then from KVL we get

$-L\dfrac{di}{dt} = V_L$ and $V_c = \dfrac{q}{C}$ and hence

$$L\dfrac{di}{dt} = -\dfrac{q}{C} \text{ or, } L\dfrac{d^2q}{dt^2} + \dfrac{q}{C} = 0 \text{ as } i = \dfrac{dq}{dt}$$

or, $\dfrac{d^2q}{dt^2} = -\dfrac{1}{LC}q$ which is the equation of SHM.
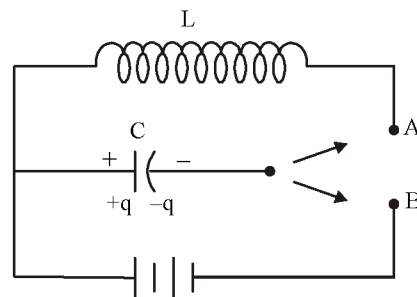


**Fig. 1.6**

Hence $\omega = \dfrac{1}{\sqrt{LC}}$ or, $T = 2\pi\sqrt{LC}$ is the time period of oscillation.

Here $q = q_0 \cos(\omega t + \delta)$ is the instantaneous charge and $i = \dfrac{dq}{dt} = -q_0\omega\sin(\omega t + \delta)$ is the instantaneous current.

We can make a one to one correspondence between the mass-spring system and LC oscillation.

| Mass spring System | LC Oscillation |
|---|---|
| 1. $x$ = displacement | 1. $q$ = charge in capacitor |
| 2. $v = \dfrac{dx}{dt}$ velocity | 2. $i = \dfrac{dq}{dt}$ = current in L |
| 3. m = mass corresponds to inertia | 3. L = self-inductance supplies inertia. |
| 4. K = force constant corresponds to elasticity | 4. C = capacitance supplies elasticity. |
| 5. $\dfrac{1}{2}mv^2 = KE$ of the system | 5. $\dfrac{1}{2}Li^2$ = magnetic energy with L |
| 6. $\dfrac{1}{2}kx^2 = PE$ associated with the spring | 6. $\dfrac{1}{2}CV^2 = \dfrac{1}{2}\dfrac{q^2}{C}$ = electrostatic energy stored in capacitor. |

**Exerise 6.** A 32 µF capacitor is charged by a 6V emf battery and then connected with a self-inductance of L = 500 mH. Find the frequency of oscillation of charge and also calculate the maximum value of current though the LC circuit.

### 1.4.6 Oscillation of two masses connected by a single spring.

Let us consider two masses $m_1$ and $m_2$ connected with a spring of force constant k.

If $x_1$ and $x_2$ are the instantaneous displant of $m_1$ and $m_2$, then the equation of motion are

$$m_1 \ddot{x}_1 = +k(x_2 - x_1)$$

$$m_2 \ddot{x}_2 = -k(x_2 - x_1) \text{ assuming } x_2 > x_1$$

**Fig. 1.7**

$$\therefore \quad m_1 m_2 \ddot{x}_1 = m_2 k(x_2 - x_1)$$

$$m_1 m_2 \ddot{x}_2 = -m_1 k(x_2 - x_1)$$

or, $\quad m_1 m_2 (\ddot{x}_2 - \ddot{x}_1) = -k(m_1 + m_2)(x_2 - x_1)$

If $x_2 - x_1 = z$, then above equation becomes

$$\frac{m_1 m_2}{m_1 + m_2} \ddot{x} = -kx \quad \text{or,} \quad \mu \ddot{x} = -kx \text{ which is the equation of SHM.}$$

or, $\quad \ddot{x} = -\dfrac{k}{\mu} x$, so $\omega = \sqrt{\dfrac{k}{\mu}}$

or, $\quad T = 2\pi \sqrt{\dfrac{\mu}{k}}$ is the time period of oscillation.

Here $\mu = \dfrac{m_1 m_2}{m_1 + m_2}$ = reduced mass of the system.

**Exercise 7.** If HCl molecule is supposed to be like two masses connected by a spring of force constant k, then assuming the frequency of vibration of the molecule as $8.57 \times 10^{13}$ Hz, find the value of k.
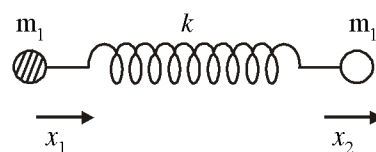
## 1.5  Importance and limitation of SHM :

From fourier series analysis we know that any periodic function that represents vibration may be expressed as a linear combination of many SHM as

$$f(t) = \sum_{n=0}^{\infty} a_n \cos n\omega t + \sum_{n=1}^{\infty} b_n \sin n\omega t$$

Thus any kind of vibrtion can be expressed in terms of many SHM. Thus SHM may be called the smallest building block of any kind of vibration.

If we assume a particle is at equilibrium at $x_0$. and if a force $F(x)$ acts on it, then using Taylor series expansion

$$F(x) = F(x_0) + \left(\frac{\partial F}{\partial x}\right)_{x=x_0}(x - x_0) + \frac{1}{2}(x - x_0)^2 \left(\frac{\partial^2 F}{\partial x^2}\right)_{n_0} + \dots$$

If $F(x_0) = 0$ and if $x - x_0$ is small, then the above expression becomes

$$F(x) = (x - x_0) \cdot \left(\frac{dF}{dx}\right)_{x_0} \quad \text{Thus} \quad F(x) \infty (x - x_0),$$

Now, if $\left(\frac{dF}{dx}\right)_{x_0} = -ve$, then the motion will be SHM.

But if $\left(\frac{dF}{dx}\right)_x = +ve$ or if it is zero, then the restoring force will vanish and the motion will no longer SHM. Also if the higher power of $(x - x_0)$ are not negligible then the motion will not be SHM. This motion may represent vibration but it will not be SHM. This is the limitation of SHM.

**Exercise 8.** A simple pendulum is suspended by a massless string of length L from a rigid support. If the rigid support start moving with an acceleration f in horizontal direction, then calculate the time period of oscillation of the pendulum.
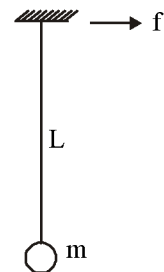
**Exercise 9.** A block of mass $m$ connected with a spring of force constant $k$ oscillate on a rough floor of coefficient of friction $\mu_s$. Calculate the frequency of oscillation and the instantaneous amplitude of the mass spring system.
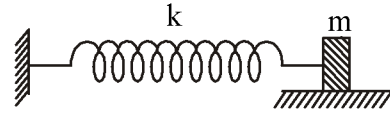
**Exercise 10.** A uniform rod of mass m and length $L$ is attatched with a disc of radius $r$, mass $M$ forms a pendulum. If it oscillates in a vertical plane, then calculate the time period of oscillation.
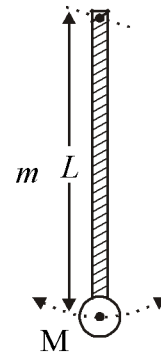
**Exercise 11.** The potential energy of a diatomic molecule is given by

$$V\left(r\right) = -2V_0\left(\frac{r_0}{r}\right)^6 + 2V_0\left(\frac{r_0}{r}\right)^6,$$ where $V_0$ & $r_0$ are constants and r is instantaneous

distance between the two molecules. Calculate the frequency of small vibration of the diatomic molecule.

**Exercise 12.** The PE of a diatomic molecule is represented by $V\left(r\right) = \frac{k_1}{r} + \frac{k_2}{r^8}$. If

$r_0$ is the equilibrium distance of the atoms in molecule, then find the expression for force constant of the system for small oscillation.

## 1.6 Summary

1. The most fundamental vibration associated with wave motion is the simple harmonic motion (SHM).

2. When a point rotates on the circumference of a circle with a uniform angular velocity, the foot of the perpendiculear on any one of its diameters will execute simple harmonic motion.

3. The differential equation for SHM is $\frac{d^2x}{dt^2} + \omega^2 x = 0$ where $\omega = \sqrt{\frac{k}{m}}$ is the

angular velocity of mass m and the exponential solution is $x = A_1 e^{i\omega t} + B_1 e^{-i\omega t}$

4.  The time average of KE and PE are $\langle KE \rangle = \frac{1}{4} m\omega^2 a^2 = \langle PE \rangle$ and mechanical

    energy $\langle ME \rangle = \frac{1}{2} m\omega^2 a^2$.

_____

## 1.7  Review Questions and Anwers

1.  Show that two rectangular SHMs of equal frequency but having a phase difference of $\pi/2$ can generate a circular motion.

2.  Show that a SHM can be considered as the superposition of two equal and opposite circular motions.

3.  A particle of mass m is under a unidimensional potential $V(n) = V_0 (1 - \cos Kx)$ where $V_0$ and $K$ are constants. Find the time period of small oscillations about the equilibrium position.

4.  A mass $M$ is suspended at the end of a spring of length $l$ and stiffness $s$. If the mass of spring is $m$ and the velocity of an element $dy$ of its length is proportional to its distance $y$ from the fixed end of the spring, show that the time period for small verticle oscillation is given by

    $$T = 2\pi \sqrt{\frac{M + m/3}{s}}$$

5.  A particle of mass m is attached to the midpoint of a vertical light string stretched between two rigid supports under a constant tension. Find the time period of small horizontal oscillations.

6.  A small pendulum consists of a rod of mass $m$ and length $l$ which is capable of oscillation about one end and the other end carries a mass $M$. Determine the natural frequency of oscillation.

### ANSWER

1.  Let the two SHMs be represented by

    $x = \cos \omega t$

    and $\quad y = a \cos(\omega t + \pi/2) = -a \sin \omega t$

    Now $\quad x^2 + y^2 = a^2(\cos^2 \omega t + \sin^2 \omega t) = a^2$

    It represents a circular motion.

2.  Circular motions can be represented by

$$\left.\begin{array}{l} x_1 = a\cos\omega t \\ y_1 = a\sin\omega t \end{array}\right\} \text{Anticlockwise} \qquad \left.\begin{array}{l} x_2 = a\cos\omega t \\ y_2 = a\sin\omega t \end{array}\right\} \text{Clockwise}$$

Adding thse equation we get $x = 2a\cos\omega t$, $y = 0$.

3.  For equilibrium, $\dfrac{dV}{dx} = 0$ or, $\sin Kx = 0$ or, $x = 0$

Now if x be the displacement from equilibrium at any instant then equation of motion is

$$m\frac{d^2x}{dt^2} = -\frac{dV}{dx} = -V_0 K\sin Kx$$

For small oscillation $\sin Kx \rightarrow Kx$

$$\frac{d^2x}{dt^2} = -\frac{V_0 K^2}{m}x = -\omega^2 x \quad \text{where} \quad \omega = \sqrt{\frac{V_0 K^2}{m}}$$

It represents a SHM. Its time period is

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{K}\sqrt{\frac{m}{V_0}}$$

4.  Let at any instant the displacement at the free end from equilibrium position be $x$. The velocity of the spring at the free end at that instant is $\dfrac{dx}{dt} = x$. Velocity of the spring at fixed end is zero. As the instantaneous velocity increases of an element $dy$ of the spring at a distance $y$ from the fixed end would be $\dfrac{x}{l}.y$. The kinetic energy of the element is $\dfrac{1}{2}(\rho dy)(xy/l)^2$ where r is the mass per unit length of the spring. Hence total kinetic energy of the mass-spring system is

$$\frac{1}{2}Mx^2 + \frac{1}{2}\frac{\rho x^2}{l^2}\int_0^l y^2 dy = \frac{1}{2}Mx^2 + \frac{1}{2}\left(\frac{\rho l}{3}\right)x^2$$

$$= \frac{1}{2}\left(M + \frac{m}{3}\right)x^2 \quad [\because \rho l = m]$$

Potential energy of the system is $\int\limits_{0}^{x} sx\,dx = \frac{1}{2}sx^2$

$\therefore$ Total energy of the system is

$$E = \frac{1}{2}\left(M + \frac{m}{3}\right)x^2 + \frac{1}{2}sx^2$$

Since total energy of the system remains constant, $dEidt = 0$

Thus, we get

$$x = \frac{s}{\left(M + \dfrac{m}{3}\right)}x$$

This shows that the motion is simple harmonic and the time period of oscillation is

$$T = 2\pi\sqrt{\frac{M + m/3}{s}}$$

5.  Let $x$ be the horizontal displacement of the mass $m$ at any instant of time t. The displacement is assumed to be small and hence the tension T in the string may be taken to remain constant. Referring to Fig. 1. P-1, the tensions in the two halves of the string can be resolved into vertical and horizontal components.
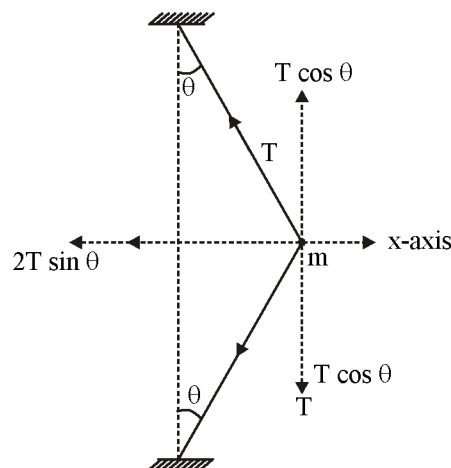


**Fig. 1. P–1**

The vertical components $T\cos\theta$ cancel out. Hence the restoring force on the mass is $2T\sin\theta$.

From Fig. 1.P–1

$$\sin\theta = \frac{x}{l/2} = \frac{2x}{l} \qquad \text{where } l \text{ is the length of the string.}$$

Therefore, the equation of motion is

$$m\frac{d^2x}{dt^2} = -2T\sin\theta = -2T\frac{2x}{l}$$

or, $\qquad \dfrac{d^2x}{dt^2} = -\omega^2 x \qquad\qquad$ where $\omega = \sqrt{\dfrac{4T}{ml}}$

This shows that the motion is simple harmonic and the time period of oscillation is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{ml}{4T}}$$

6. Let $\theta$ be the small angular displacement at any instant as shown in Fig.1. P–2 Now the kinetic energy of the system is

$$KE = \frac{1}{2}I_{rod}\dot\theta^2 + \frac{1}{2}I_{mass}\dot\theta^2$$

$$= \frac{1}{2}\left(\frac{1}{3}ml^2\right)\dot\theta^2 + \frac{1}{2}Ml^2\dot\theta^2$$

$$= \frac{1}{2}\left(M + \frac{m}{3}\right)l^2\dot\theta^2$$

Potential energy of the system is



**Fig. 1. P–2**

$$PE = Mgl(1-\cos\theta) + mg\frac{1}{2}(1-\cos\theta) = \left(M + \frac{m}{2}\right)gl(1-\cos\theta)$$

Since total energy (KE + PE) is constant,

$$\frac{d}{dt}(KE+PE) = 0$$

$$\left(M + \frac{m}{3}\right)l^2\ddot\theta + \left(M + \frac{m}{2}\right)gl\sin\theta = 0$$
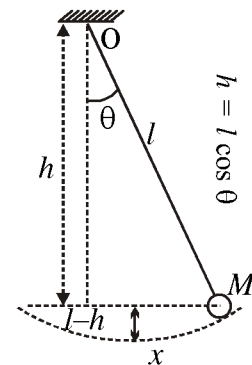
For small oscillation $\sin \theta \approx \theta$ and the equation of motion becomes,

$$\ddot{\theta} = -\omega^2 \theta \text{ where } \omega^2 = \frac{M + m/2}{M + m/3} \cdot \frac{g}{l}$$

∴ The time period of oscillation is

$$T = \frac{2\pi}{\omega} 2\pi \sqrt{\frac{M + m/3}{M + m/2} \cdot \frac{l}{g}}$$

# Unit-2 : Damped Harmonic Motion

**Structure**

## 2.0   Objectives

After reading this unit, you should be able to :
- Understand the various types of damping effects
- Write the differential equation of damped harmonic oscillator
- Analyze the weakly damped, critically damped and over damped motions
- Explain various parameters characterizing weak damping and
- Understand the damping in LCR circuit

## 2.1   Introduction

So far we have seen the SHM once starts, continues for infinite time with constant amplitude. But in practice it does not happen. The vibration of any system stops after a certain interval. This is because of the resistive force that acts on the system which

is not considered in previous section. This resistive force acting on a vibrating system is called damping force.

In general these damping forces are classified in three type.

1. **Viscous damping :** The resistive force offered by fluid medium on oscillating system. This force is found to be proportional to the instantaneous velocity of the body

i.e. $F_v = b\dfrac{dx}{dt}$, b = proportionality constant.

2. **Coulomb friction damping :** This damping appears due so direct contact of the vibrating body with any other body. This force may be represented by $F_c = \mu R$, where $\mu$ = coefficient of friction and R is the normal reaction.

3. **Structural damping :** This damping arises due to structural deformation or uneven shape of the body.

In this section we will discuss only the effect of viscous damping on oscillating

system. This damping may be expressed in general as $F = b\dfrac{dx}{dt} + c\left(\dfrac{dx}{dt}\right)^2 + d\left(\dfrac{dx}{dt}\right)^3 + \ldots$

but if $\dfrac{dx}{dt}$ is small, then it can be represented by $F_v = b\dfrac{dx}{dt}$.

## 2.2 Equation of damped vibration :

### Equation of damped vibration :

Let us consider a mass-spring system subject to a damping force. Then if the mass $m$ is displaced by $x$, then, it is acted upon by two forces

    (i)    $-kx$ restoring force

    (ii)    $-b\dfrac{dx}{dt}$ damping force

Hence the equation of motion

$$m\ddot{x} = -b\dot{x} - kx \qquad \ldots (2.2.1)$$

or, $\quad \ddot{x} + \dfrac{b}{m}\dot{x} + \dfrac{k}{m}x = 0$ and if we put $\dfrac{b}{m} = 2\gamma$ &

$\dfrac{k}{m} = \omega_0^2$ the equation is

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 = 0 \qquad \ldots (2.2.2)$$
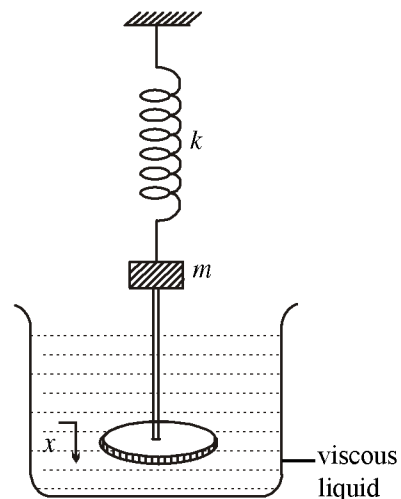


**Fig. 2.1**

Here $\gamma$ is called the damping factor and $\omega_0 = \sqrt{\dfrac{k}{m}}$ is the angular frequency of undamped oscillation.

Here it is to be noted that equation (2.2.2) is a second order linear homogeneous differential equation whose solution will give us the displacement of the oscillating system at any time t.

## 2.3 Solution of differential equatgion

To solve equation (2.2.2). Let $x = Ae^{\alpha t}$ where $\alpha$ is to be determined. Then we get $\dot{x} = \alpha Ae^{\alpha t}$ and $\ddot{x} = \alpha^2 Ae^{\alpha t}$

or, $\quad \left[\alpha^2 + 2\gamma\alpha + \omega_0^2\right]x = 0$

or, $\quad \alpha^2 + 2\gamma\alpha + \omega_0^2 = 0$

or, $\quad \alpha = \dfrac{-2\gamma \pm \sqrt{4\gamma^2 - 4\omega_0^2}}{2}$

So, $\quad \alpha_1 = -\gamma + \sqrt{\gamma^2 - \omega_0^2} \quad \& \quad \alpha_2 = -\gamma - \sqrt{\gamma^2 - \omega_0^2}$

hence $\quad x = A_1 e^{-\gamma t} e^{\left(\sqrt{\gamma^2 - \omega_0^2 \cdot}\right)t} + A_2 e^{-\gamma t} . e^{-\left(\sqrt{\gamma^2 - \omega_0^2 \cdot}\right)t}$

$$= e^{-\gamma t}\left\{A_1 e^{\sqrt{\gamma^2 - \omega_0^2} \cdot t} + A_2 e^{-\sqrt{\gamma^2 - \omega_0^2} \cdot t}\right\} \qquad \text{... (2.3.1)}$$

Here we may have three different cases.

### 2.3.1 Heavy damping

If $\gamma^2 > \omega_0^2$ i.e. damping is large, then the variation of (2.3.1) is depicted as shown in Fig. 2.2. This is called heavy damping and here if at $t = 0$, $x = x_0$, as $t$ increases $x(t)$ deceases exponentially but at a very slow rate due to presence of term $e^{-\sqrt{\gamma^2 - \omega_0^2} \cdot t}$
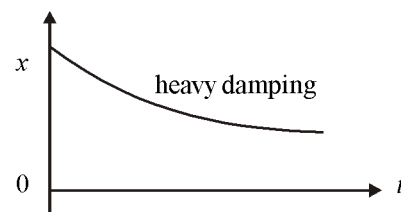


Fig. 2.2

### 2.3.2 Critical damping

If $\gamma^2 = \omega_0^2$, then the displacement is

$$x = \left(A_1 + A_2\right)e^{-\gamma t} \ \dots \ (2.3.2)$$

This is called critical damping. Here at $t = 0$, $x = x_0$, but as times goes on $x$ decreases at exponential fast rate which is compared with heavy damping in Fig. 2.3.
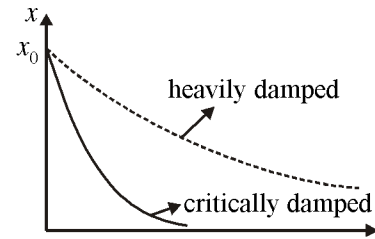


Fig. 2.3

### 2.3.2 Weak damping

If the damping is very weak, so that $\gamma^2 < \omega_0^2$, then the quantity $\sqrt{\gamma^2 - \omega_0^2} = i\omega$ is imginary. So, the solution becomes

$$x = e^{-\gamma t}\left\{A_1 e^{i\omega t} + A_2 e^{-i\omega t}\right\} = e^{-\gamma t}\left\{\left(A_1 + A_2\right)\cos\omega t + i\left(A_1 - A_2\right)\sin\omega t\right\}$$

$$= A_0 e^{-\gamma t}.\cos\left(\omega t - \delta\right) \text{ where } A_1 + A_2 = A_0 \cos\delta \ \left(A_1 - A_2\right)i = A_0 \sin\delta.$$

$$\dots (2.3.3)$$

From this expression we note that

(a) as time goes on the amplitude decreases but here we get an oscillation with decrease of amplitude. This is called damped vibration which is shown in Fig. 2.4.

(b) Here we also note that this vibration has frequency $\omega$ which is less than natural frequency $\omega_0$.

The time period of this oscillation is

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{\omega_0^2 - \gamma^2}} = \frac{2\pi}{\sqrt{\dfrac{k}{m} - \dfrac{b^2}{4m}}} \ \dots (2.3.4)$$
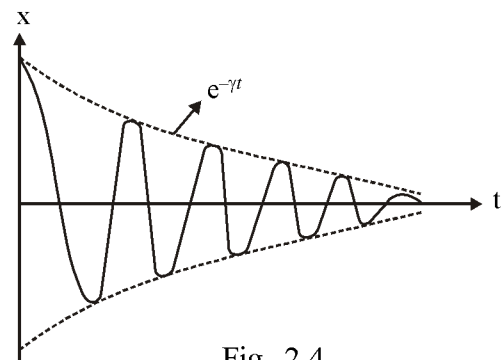


Fig. 2.4

**Exercise 1.** What is the unit of damping factor $\gamma$.

**Exercise 2.** In a weakly damped system the amplitude reduces from 8 cm to 2 cm in 160 sec. If during this time the system completes 80 vibrations, then find the value of damping factor $\gamma$ and the time period of vibration.

## 2.4  Engergy of a weakly damped oscillator

We have the instantaneous displacement of a weakly damped ocillator is $x = A_0 e^{-\gamma t} \cos(\omega t - \delta)$. So the KE. of the oscillator is

$$KE = \frac{1}{2}m\dot{x}^2 = \frac{1}{2}m\left[-\gamma A_0 e^{-\gamma t}\cos(\omega t - \delta) - A_0 e^{-\gamma t}.\omega \sin(\omega t - \delta)\right]^2$$

and

$$P.E. = \frac{1}{2}m\omega^2 x^2 = \frac{1}{2}m\omega^2 A_0^2 e^{-2\gamma t}\cos^2(\omega t - \delta)$$

$$\therefore \ T.E. = E = PE + KE = \frac{1}{2}m\omega^2 A_0^2 e^{-2\gamma t}\cos^2(\omega t - \delta) +$$

$$\frac{1}{2}m\left\{\gamma^2 A_0^2 e^{-2\gamma t}\cos^2(\omega t - \delta) + 2.\gamma A_0^2 \omega e^{-2\gamma t}\sin(\omega t - \delta)\cos(\omega t - \delta)\right.$$

$$\left. + A_0^2 \omega^2 e^{-2\gamma t}\sin^2(\omega t + \delta)\right.$$

Taking $\omega$ large i.e. T = small, we can take time average of this T.E. as

$$\langle E \rangle = \frac{1}{2}m\omega^2 A_0^2 e^{-2\gamma t} + \frac{\gamma^2 A_0^2 e^{-2\gamma t}}{2} \quad \text{as} \quad \left\langle \cos^2(\omega t - \delta)\right\rangle = \frac{1}{2}$$

and

$$\left\langle \sin^2(\omega t - \delta)\right\rangle = \frac{1}{2}$$

but

$$\left\langle \sin(\omega t - \delta)\cos(\omega t - \delta)\right\rangle = 0$$

We get

$$\langle E \rangle = \frac{1}{2}m\omega^2 A_0^2 e^{-2\gamma t} \qquad ... (2.4.1)$$
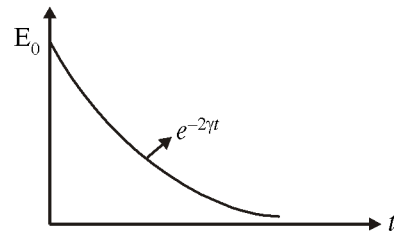
Fig. 2.5

Thus we see that the total energy of a weakly damped oscillator also decreases at exponentially fast rate.

**Exercise 3.** In a mass-spring sysem the mass is m = 2kg and force constant k = 2N/m. If it is subject to a damping force of 0.2 Ns/m, then

(i) What is the frequency of damped oscillator.

(ii) How much time will be required to reduce the energy to $\frac{1}{10}$th of its maximum value.

## 2.5  Estimation of damping

The damping of a weakly damped oscillator can be estimated in three ways. These are—

    (a) Logarithmic decrement or log decrement ($\lambda$)

    (b) Relaxating time ($\tau$)

    (c) Quality factor or Q-factor.

### 2.5.1 Logarithmic decrement ($\lambda$) :

It is one of the easiest way of estimating damping of a oscillating system. To explain it let us consider the 1st few amplitude of a weakly damped oscillator.

We have    $x = Ae^{-\gamma t}\cos(\omega t - \delta)$

Putting $\delta = 0$, we see that

at $t = T = \dfrac{2\pi}{\omega}$, $x = Ae^{-\gamma T} = A_1$

at $t = 2T$, $x = Ae^{-\gamma 2T} = A_2$

at $t = 3T$, $x = Ae^{-\gamma.3T} etc. = A_3$
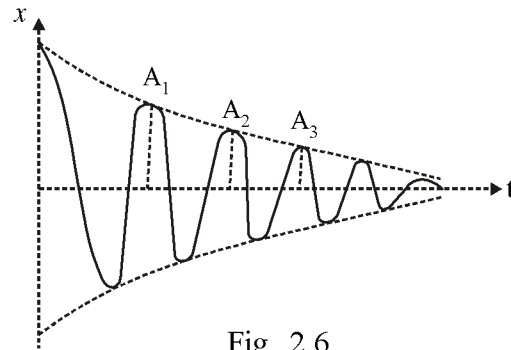


Fig. 2.6

$\therefore$ We see $\dfrac{A_1}{A_2} = \dfrac{A_2}{A_3} = \dfrac{A_3}{A_4}\ldots\ldots\ldots = e^{\gamma T}$ = constant = d (called decrement)

$\therefore \log d = \gamma T = \gamma$ = called logarithmic decrement.

Now to determine $\lambda$, we see $\dfrac{A_1}{A_2} \times \dfrac{A_2}{A_3} \times \dfrac{A_3}{A_4} \times \ldots\ldots\ldots \dfrac{A_{n-1}}{A_n} = \lambda^{n-1}$

or,    $\lambda^{n-1} = \dfrac{A_1}{A_n}$

or,    $\lambda = \dfrac{1}{n-1}\log\left(\dfrac{A_1}{A_n}\right)$        ... (2.5.1)

Thus from (2.5.1) we can determine $\lambda$ and hence $\gamma = \lambda/T$, the damping factor of the weakly damped oscillator.

### 2.5.2 Relaxation time ($\tau$)

We have seen that the amplitude of a weakly damped oscillator varies as $A = A_0$ $e^{-\gamma t}$. Now we can consider a time $t = \tau$, in which the amplitude reduces to $\dfrac{1}{e}$ times of its maximum value. This time $t = \tau$, is called relaxation time, and hence for $t = \tau$, we have $\gamma\tau = 1$ or $t = \dfrac{1}{\gamma}$. Thus the relaxation time is the reciprocal of damping factor.

Here it is to be noted that this parameter is usually used to estimate the damping of subatomic systems. On the other hand the log decrement is widely used to estimate the damping of ballistic galvanometer, mechanical oscillators, damping of a simple pendulum etc.

### 2.5.3 Quality factor (Q)

It is a parameter which measures the damping of a weakly damped oscillator by measuing the rate of decrease of energy. It is defined as

$$Q = 2\pi \times \frac{\text{Energy stored per unit time period}}{\text{Energy lost per unit time period}} = 2\pi \frac{\langle E \rangle}{T.\dfrac{d\langle E \rangle}{dt}}$$

Now $\langle E \rangle = E_0 e^{-\gamma t}$ $\therefore$ $\dfrac{d\langle E \rangle}{dt} = -\gamma \langle E \rangle$

$\therefore Q = 2\pi.\dfrac{1}{\gamma T} = \dfrac{\omega}{\gamma}$ , Hence Q-factor is the ratio of the angular frequency of damped oscillator and the damping factor.

It is to be noted that Q-factor is extensively used to estimate damping of forced vibrating oscillator, LCR circuit, etc.

## 2.6 Summary

In this section we have discussed—
1. Effect of viscous damping on real oscillating system.
2. The equation of motion of a damped oscillator.
3. It's solution and their nature in different condition.
4. The process of estimation of damping in different types of damped oscillator.

5. The final expression of the three parameters logarithmic decrement ($\lambda$), relaxation time ($\tau$) and Q-factor.

## 2.7 Review Questions and Answers

1. The angular amplitude of a simple pendulum decreases from 4° to 2° in making 50 oscillations. If $T = 2$ sec is the time period, then calculate the $\lambda$ and $\tau$.

2. Obtain a relation between $\lambda$ and $\tau$, $\tau$ and $Q$ factor and $\lambda$ and $Q$-factor.

3. State the unit and dimension of $\lambda$, $\tau$ and $Q$-factor.

4. A mass $m$ connected with a spring of force constant $k$ vibrates in a horizontal rough floor of coefficient of friction $\mu$. Set up the equation of motion and discuss the vibration of it.

5. The length of a simple pendulum is 10 cm. When it is allowed to oscillate in a viscous medium its frequency of oscillation becomes 9.0 rad/sec. Calculate the damping factor ($\gamma$) of this damped vibration.

6. A mass of 0.01 kg is acted upon by a restoring force of 0.01 $N.m^{-1}$ and a resisting force of $2 \times 10^{-3}$ $N.s.m^{-1}$. Find out whether the motion is oscillatory or non-oscillatory. Also, find the value of resistive force for the motion to be critically damped.

7. A mass $m = 2.5$ kg is suspended by a spring of stiffness constant $s = 40$ $N.m^{-1}$. the amplitude of vertical oscillations of the system is observed to decrease to 20% of the initial value after five consecutive cycles of oscillations. Determine the damping coefficient $R_m$ of the system.

8. A mass $m = 0.2$ kg is suspended by a light spring of stiffness constant $20N.m^{-1}$. The energy associated with the vertical oscillation of the system is observed to decay to $1/e$ of initial value in 50s. Assuming the damping force ($R_m v$) proportional to the velocity ($v$), find the damping force constant $R_m$. Also find the Q-value of the oscillator and the percentage change is frequency due to damping.

9. A body of mass $m = 10$ kg is subjected to a restoring force of stiffness constant $s = 10$ $N.m^{-1}$. The sysem is subjected to a damping force proportional to its velocity with the constant of proportionality as $R_m = 10$ $N.s.m^{-1}$. The mass is given an initial empulse of 0.68 $kg.m.s^{-1}$ when at rest. Find the subsequent displacement of the mass as a function of time.

10. Show that if the damping force is of constant magnitude the frequency of vibration of a damped oscillator is not affected by the magnitude of damping.

11. In a damped harmonic oscillator the amplitude of vibration decreases from 10 cm to 1 cm after 100 complete cycles of oscillation. The time period of oscillation is 2.3 sees. Estimate the first undamped amplitude.

## ANSWER

1. We have $\lambda = \dfrac{1}{n-1} \ln\left(\dfrac{A_1}{A_n}\right) = \dfrac{1}{n-1} \ln\left(\dfrac{\theta_1}{\theta_n}\right)$, here $\theta_1 = 4°$, $\theta_n = 2°$ & $n = 50$.

$$\therefore \quad \lambda = \frac{1}{49} \ln\left(\frac{4}{2}\right) = \frac{\ln 2}{49} = \frac{0.693}{49}.$$

But $\lambda = \lambda T$ $\quad \therefore \dfrac{0.693}{49} = \gamma.2$ $\quad$ or, $\lambda = \dfrac{0.693}{98}$ and hence $\tau = \dfrac{1}{\gamma} = \dfrac{98}{0.693}$ sec.

2. We have $\lambda = \gamma.T$, $\tau = \dfrac{1}{\gamma}$ and $Q = \dfrac{\omega}{\gamma} = \dfrac{2\pi}{\gamma T}$

$$\therefore \quad \lambda = \frac{T}{\tau}; \quad Q = \frac{2\pi\tau}{T} \text{ and } Q = \frac{2\pi}{\lambda}.$$

3. $\lambda$ and $Q$ are unit and dimensionless as they are ratio of same quantity. But $\tau$ has unit sec and diension $[T]$.

4. The equation of motion of the mass is

$$\frac{md^2 x}{dt^2} = -kx - \mu mg$$

or, $\quad \dfrac{d^2 x}{dt^2} + \left(\dfrac{kx}{m} + \mu g\right) = 0$

or, $\quad \dfrac{d^2 x}{dt^2} + \dfrac{k}{m}\left(x + \dfrac{m\mu g}{k}\right) = 0$ $\quad$ putting



Fig. 2.P–1

$z = x + \dfrac{\mu mg}{k}$, we get

$$\frac{d^2 z}{dt^2} + \omega_0^2 z = 0 \quad \therefore \quad z = A \sin \omega_0 t, \quad \omega_0 = \sqrt{\frac{k}{m}} \text{ and } x = A \sin \omega_0 t - \frac{m\mu g}{k}$$
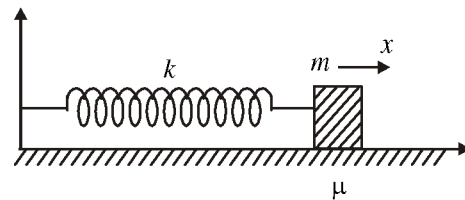
5. $T = 2\pi\sqrt{\dfrac{L}{g}}$ taking $g = 10$ $m$/sec$^2$, when $L = 10$cm, $T = 2\pi\sqrt{\dfrac{10}{1000}}$

$\therefore$ $\omega_0 = \dfrac{2\pi}{T} = \sqrt{100}$ $= 10$ rad/sec, here $w = 9$ rad/sec.

$\therefore$ $\omega = \sqrt{\omega_0^2 - \gamma^2}$ or $\gamma^2 = \omega_0^2 - \omega^2 = 10^2 - 9^2 = 19$ $\therefore$ $\gamma = \sqrt{19}\, s^{-1}$

6. Here $= 0.01$ kg, s $= 0.01$ $N.m^{-1}$ and $R_m = 2 \times 10^{-8}$ $N.s.m^{-1}$.

So, $b = \dfrac{R_m}{2m} = \dfrac{2 \times 10^{-3}}{2 \times 0.01} = 0.1$ N.s.m$^{-1}$ kg$^{-1}$

$\omega_0 = \sqrt{\dfrac{s}{m}} = \sqrt{\dfrac{0.01}{0.01}} = 1$

As $b < \omega_0$, the motion is oscillatory.

For the motion to be critically damped $b = \omega_0$

or, $\dfrac{R_m}{2m} = \sqrt{\dfrac{s}{m}}$

or, $R_m = 2\sqrt{sm} = 2\sqrt{0.01 \times 0.01} = 0.02$ N.S.m$^{-1}$

7. The damped vibration of the system is given by

$x = Ae^{-bt}\cos(\omega t - \phi)$

where $b = R_m/2m$, $\omega = \sqrt{\omega_0^2 - b^2}$ and $\omega_0 = \sqrt{s/m}$

The maximum amplitude in a cycle occurs when $\cos(\omega t - \phi) = 1$. Therefore the maximum amplitudes on the same side of the mean position are,

$A_1' = Ae^{-bt_1}, A_2' = Ae^{-b(t_1+T)}, \ldots\ldots, A_6' = Ae^{-b(t_1+5T)}$

where $T = 2\pi/\omega$ is the time period.

Now in the problem,

$\dfrac{A_1'}{A_6'} = \dfrac{1}{0.20} = 5$

or, $e^{b.5T} = 5$ or, $bT = \dfrac{\ln 5}{5} = 0.322$ or, $b.\dfrac{2\pi}{\sqrt{\omega_0^2 - b^2}} = 0.322$

Putting $\omega_0 = \sqrt{s/m} = \sqrt{40/25} = 4$ rad s$^{-1}$, we get b = 0.20.

Damping coefficient $R_m = 2bm = 2 \times 0.20 \times 2.5 = 1$ N.S.m$^{-1}$.

8.  The energy decays according to the law

$$\langle E \rangle = E_0 e^{-2bt} \qquad \text{where } 2b = R_m/m$$

$\therefore$ At $t = 50$ sec, $\dfrac{E_0}{e} = E_0 e^{-2b.50}$

or, $2b = (1/50)\text{s}^{-1}$

Hence $R_m = 2bm = \dfrac{1}{50} \times 0.2 = 4 \times 10^{-3}$ N.s.m$^{-1}$

The undamped natural frequency

$$\omega_0 = \sqrt{\frac{s}{m}} = \sqrt{\frac{20}{0.2}} = 10 \text{ rad s}^{-1}$$

$$Q = \frac{\omega}{2b} = \frac{1}{2b}\sqrt{\omega_0^2 - b^2} = 50\sqrt{10^2 - \frac{1}{10^4}} = 500$$

Now, $\omega = \sqrt{\omega_0^2 - b^2} = \omega_0\left(1 - \dfrac{b^2}{\omega_0^2}\right)^{1/2} = \omega_0\left(1 - \dfrac{b^2}{2\omega_0^2}\right)$

$$= \omega_0\left(1 - \frac{1}{8Q^2}\right) \qquad \left[\because Q = \frac{\omega_0}{2b}\right]$$

or, $\dfrac{\omega_0 - \omega}{\omega_0} = \dfrac{1}{8Q^2}$

$\therefore$ The percentage change in frequency due to damping is

$$\frac{\omega_0 - \omega}{\omega_0} \times 100 = \frac{100}{8Q^2} = \frac{100}{8 \times (500)^2} = 5 \times 10^{-4} \%$$

9.  The equation of motion of the mass is

$$mx = -R_m x - sx$$

and its solution is of the form

$$x = e^{-bt} \left( A \cos \omega t + b \sin \omega t \right)$$

where $b = \dfrac{R_m}{2m}$, $\omega = \sqrt{\omega_0^2 - b^2}$ and $\omega_0 = \sqrt{s/m}$ ; A and B are constants to be determined from given initial conditions.

At t = 0, x = 0, This indicates that A = 0.

$$x = Be^{-bt} \sin \omega t$$

Now the impulse is equal to change of momentum. Therefore

$$mx_0 - 0 = 0.68$$

∴ Initial velocity $\quad x_0 = \dfrac{0.68}{10} = 0.068 \ m.s^{-1}$

$$b = \frac{R_m}{2m} = \frac{10}{2 \times 10} = 0.5 \, s^{-1}, \ \omega_0 = \sqrt{s/m} = 1 \ \text{rad.} \ s^{-1}, \ \omega = \sqrt{\omega_0^2 - b^2} = 0.866$$

$$x = -bBe^{-bt} \sin \omega t + \omega Be^{-bt} \cos \omega t$$

At $t$ = 0, 0.068 = $\omega$B.

$$B = \frac{0.068}{0.866} = 0.078 \, m$$

Thus $\quad x = 0.078e^{-0.5t} \sin 0.866t$

This equation describes the subsequent motion.

10. The equation of motion in this case is

$$m\ddot{x} = -F_0 - sx$$

or, $\quad m\ddot{x} = s\left(x + F_0/s\right) = 0$

Introducing a new variable $\quad y = x + F_0/s$ we get

$$m\ddot{y} + sy = 0$$

or, $\quad \ddot{y} + \omega_0^2 y = 0$

where $\omega_0 = \sqrt{s/m}$ is the angular frequency for undamped oscillation. Obviously the damped natural frequency is equal to the undamped natural frequency and does not depand on the magnitude of the damping force.

11. Log decrement

$$\lambda = \frac{1}{2n} \log_e \frac{A_1'}{A_{n+1}'} = \frac{1}{2 \times 100} \log_e \frac{10}{1} = 0.0115$$

∴ Undamped amplitude

$$A \approx A_1 \left(1 + \lambda / 2\right) = 10 \left(1 + \frac{0.0115}{2}\right) = 10.06 \text{ cm.}$$

# Unit-3 : Forced Vibration and Resonance

**Structure**

## 3.0  Objectives

After studying this unit, you should be able to :

- define forced harmonic opcillations
- write the differential equation for weakly damped forced harmonic oscillator
- solve the differential equation for weakly damped forced hamonic oscillator
- describe the phenomenon of resonance
- apply the solution of wearkly damped forced harmonic oscillator to explain resonance
- obtain an expression for the power absorbed by a forced oscillator and
- define the quality factor of a forced oscillator and sharpness of resonance.

## 3.1  Introduction

We have seen that the amplitude of vibration of a system decreases with due to damping force on it. But if the vibrating system is subject to the action of an extenal periodic force, it is said to be in a state of forced vibration. In this condition initially the body tends to vibrate with its own natural frequency ($\omega_0$), while the applied force

tries to make the body vibrate with the frequency of forcing system. Soon the natural vibration dies away due to damping force and finally the body continue to vibrate with the frequency of periodic force. This vibration is called forced vibration. As an example a bridge vibrates under the influence of marching soldiers, the housing of a motor vibrate owing to periodic impulses from an irregularity in the shaft, a tuning fork vibrate when exposed to periodic force of a sound wave, etc.

## 3.2  Equation of motion of a forced system

Let us consider a mass-spring system connected with a simple pendulum, which exert a periodic force $F = F_0 \cos \omega t$ on the mass-spring system. Let the damping force on it is $-p\dfrac{dx}{dt}$. Then the equation of motion of the forced system is



Fig. 3.1

$$m\frac{d^2x}{dt^2} = -p\frac{dx}{dt} - kx + F_0 \cos \omega t$$

or,  $\quad \dfrac{d^2x}{dt^2} + \gamma\dfrac{dx}{dt} + \omega_0^2 x = f_0 \cos \omega t \qquad\qquad$ ... (3.2.1)

where  $\gamma = \dfrac{p}{m}$,  $\omega_0 = \sqrt{\dfrac{k}{m}}$  and  $f_0 = \dfrac{F_0}{m} \qquad\qquad$ ... (3.2.2)

This is a second order linear inhomogeneous equation.
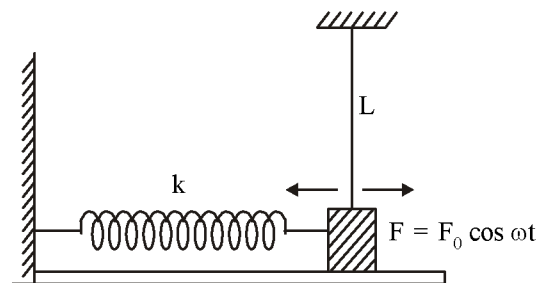
## 3.3  Solution of the differential equation

The solution of this equation will have two parts. One $x_1$ called complementary function obtained by putting R.H.S. equal to zero, i.e. $x_1 = Ae^{-\gamma t/2}\cos(\omega't - \phi)$ ...(3.3.1) and the other called particular integral $x_2$ keeping R.H.S. intact. The solution $x_1$ gradually decreases due to exponential term, but the solution $x_2$ remains also the time. So after a sufficiently long time the forced vibrator settles down to vibrate with the frequency of periodic force i.e.

$$x_2 = B_0 \left(\cos \omega t - \delta\right) \qquad\qquad \text{... (3.3.2)}$$

This just after the exernal driving force is put on the solution is $x = x_1 + x_2$. This is called transient state. But after some time $x_1 \to 0$ and we get $x = x_2$ This is called steady state.

Now to find the amplitude and phase of steady state solution let us take $F = F_0$ cos $\omega t = F_0\, e^{i\omega t}$ whose real part respresents the periodic force.

Thus we get $\dfrac{d^2x}{dt^2} + \gamma\dfrac{dx}{dt} + \omega_0^2 x = f_0 e^{i\omega t}$ ... (3.3.3)

hence assuming $x = B_0 e^{i(\omega t - \delta)}$, we get

$$\left[\left(-\omega^2 + \omega_0^2\right) + i\omega\gamma\right]x = f_0 e^{i\omega t}$$

or, $\quad B_0 = \dfrac{f_0 e^{i\delta}}{\left(\omega_0^2 - \omega^2\right) + i\omega\gamma}$ $\quad$ or, $\quad B_0 e^{-i\delta} = \dfrac{f_0}{\left(\omega_0^2 - \omega^2\right) + i\omega\gamma}$

or, $\quad B_0\cos\delta - iB_0\sin\delta = \dfrac{f_0\left\{\left(\omega_0^2 - \omega^2\right) - i\omega\gamma\right\}}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$

Thus $\quad B_0\cos\delta = \dfrac{f_0\left(\omega_0^2 - \omega^2\right)}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$ and $B_0\sin\delta = \dfrac{\omega\gamma\, f_0}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$

Thus $\quad B_0 = \dfrac{f_0}{\left\{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2\right\}^{1/2}}$ ... (3.3.4)

and $\quad \delta = \tan^{-1}\left(\dfrac{\omega\gamma}{\omega_0^2 - \omega^2}\right)$ ... (3.3.5)

Thus we get the steady state solution

$x = B_0\cos(\omega t - \delta)$ ... (3.3.6)

## 3.4 Variation of displacement with frequency ($\omega$)

Thus we see that the amplitude and phase of displacement varies with the frequency of periodic force.

Here we note that

When $\omega \to 0$, $\quad B_0 \to \dfrac{f_0}{\omega_0^2} = \dfrac{F_0}{k}$

,, $\quad \omega \to \omega_0$, $\quad B_0 \simeq \dfrac{f_0}{\omega\gamma}$

,, $\quad \omega \to large$, $\quad B_0 \simeq \dfrac{f_0}{\omega^2}$

Moreover we see that when $\dfrac{d}{d\omega}\left\{\sqrt{\left(\omega_0^2-\omega^2\right)^2+\omega^2\gamma^2}\right\}=0$, then the amplitude

will be maximum, and this gives $\omega_\gamma=\sqrt{\omega_0^2-\gamma^2\!/\!2}$. Thus the amplitude will be maximum when the frequency of forcing system is

$$\omega_\gamma=\sqrt{\omega_0^2-\gamma^2\!/\!2} \qquad\qquad \text{... (3.4.1)}$$

This is known as amplitude reasonance and the maximum value of this displacement

is given by $\qquad B_{\max}=\dfrac{f_0}{\gamma\sqrt{\omega_0^2-\gamma^2\!/\!4}}$ $\qquad\qquad$ ... (3.4.2)

Fig. 3.2 shows the variation of amplitude of forced oscillator with the frequency of the periodic force. We see for small damping $\omega_\gamma\simeq\omega_0$. but for large damping amplitude resonance frequency differes from the natural freqency of oscillator.
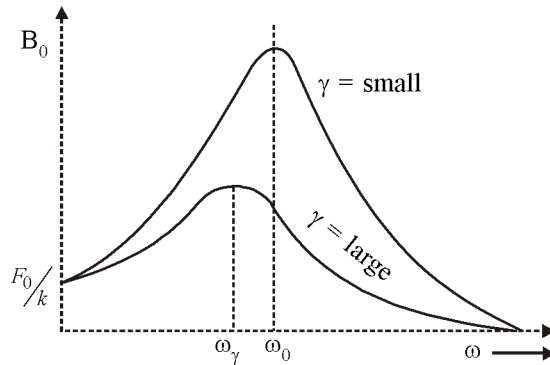


Fig. 3.2

## 3.5 Resonance

Now we see that the instantaneous velocity of the forced vibration is

$$v=\frac{dx}{dt}=\frac{d}{dt}\left[B_0\cos\left(\omega t-\delta\right)\right]=-\omega V_0\sin\left(\omega t-\delta\right)$$

or, $\qquad v=\omega B_0\cos\left\{\omega t-\left(\omega_2-\delta\right)\right\}=\omega B_0\cos\left\{\omega t-\phi\right\}$ $\qquad$ ... (3.5.1)

So the velocity amplitude is $\omega B_0=\dfrac{\omega f_0}{\sqrt{\left(\omega_0^2-\omega^2\right)^2+\omega^2\gamma^2}}$

and we see this $\qquad \omega B_0=\dfrac{f_0}{\sqrt{\dfrac{\left(\omega_0^2-\omega^2\right)^2}{\omega^2}+\gamma^2}}=f_0\!/\!\gamma=\text{max.}$

when $\omega = \omega_0$ exactly. This is known as velocity reasonance.

Thus the velocity of the oscillator will be maximum when the frequency of the forcing system is exactly equal to the natural frequency of vibration of the system.
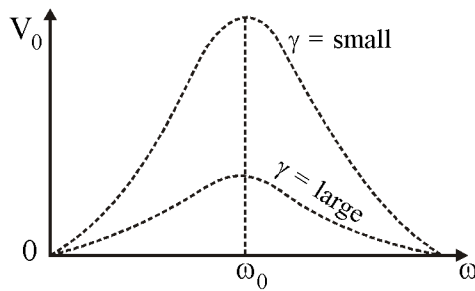


Fig. 3.3

Here we see     $\omega B_0 = v_0 \to 0$ as $\omega \to 0$

$v_0 \to \dfrac{f_0}{\gamma}$   when $\omega = \omega_0$

$v_0 \to 0$ when $\omega \to$ large

The variation of $V_0$ with $\omega$ is as shown beside.

## 3.6  Rate of eneregy dissipation

We know that the energy of a damped oscillator decreases exponentially as $E = E_0 e^{-\gamma t}$. So in order to maintain steady state oscilation of the oscillator, the driving force must supply the power to the oscillator to compensate the loss of energy due to damping force.

Now this rate of dissipation of energy by the oscillator is

$$P(t) = F(t) \times \frac{dx}{dt} = \text{force} \times \text{velocity}$$

But we have  $\dfrac{dx}{dt} = \omega \{ B_0 \cos\delta\cos\omega t - B_0 \sin\delta\sin\omega t \}$

$$= \omega \{ B_1 \cos\omega t - B_2 \sin\omega t \}$$

Where     $B_1 = B_0 \sin\delta = \dfrac{f_0 \omega\gamma}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}}$

and     $B_2 = B_0 \cos\delta = \dfrac{f_0 \left(\omega_0^2 - \omega^2\right)}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}}$

$\therefore$     $P(t) = \omega F_0 \cos\omega t \{ B_1 \cos\omega t - B_2 \sin\omega t \}$

∴ The average power dissipation is

$$\langle P(t) \rangle = \omega F_0 \langle \cos^2 \omega t \rangle B_1 - \frac{\omega F_0 B_2}{2} \langle \sin 2\omega t \rangle$$

But $\qquad \langle \cos^2 \omega t \rangle = \frac{1}{2}$ & $\langle \sin 2\omega t \rangle = 0$

So $\qquad \langle P \rangle = \frac{1}{2} \omega F_0 B_1 = \frac{1}{2} \dfrac{\omega F_0 \cdot f_0 \omega \gamma}{\sqrt{\left(\left(\omega_0^2 - \omega^2\right)^2 + \omega^2 \gamma^2\right)}}$

$$\langle P \rangle = \frac{1}{2} F_0^2 p \dfrac{\omega^2}{\sqrt{\left(\left(\omega_0^2 - \omega^2\right)^2 + \omega^2 \gamma^2\right)}} \qquad \text{... (3.6.1)}$$

This is the average power dissipated by the damping force and this must be equal to the power supplied by the forcing system in order to maintain constant amplitude oscillation.

## 3.7 Sharpness of resonance and Q-factor

From the expression of average power dissipation we see that it is maximum when $\omega = \omega_0$ and this maximum value is

$$\langle P \rangle_{\text{max}} = \frac{1}{2} F_0^2 \frac{p}{\gamma^2}$$

∴ $\qquad \langle P \rangle = \langle P \rangle_{\text{max}} \dfrac{\omega^2 \gamma^2}{\left\{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2 \gamma^2\right\}} \qquad \text{... (3.7.1)}$

From the variation of average power dissipation with the frequency of forcing system we see that there are two frequencies $\omega_1$ and $\omega_2$ at which this dissipative power reduces to half of its maximum value and the difference of this two frequencies is called bandwidth. Thus Bandwidth

$(BW) = \omega_2 - \omega_1$ ... (3.7.2)
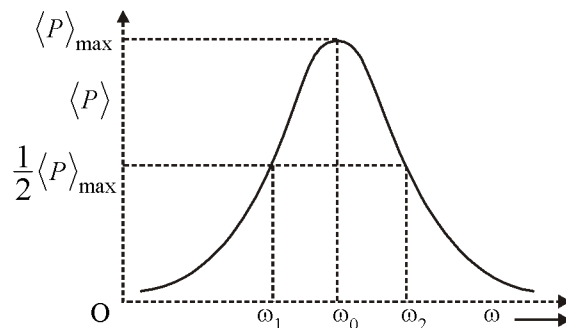


Fig. 3.4

These frequencies are called half power frequencies.

Now at $\omega_1$ and $\omega_2$ we get

$$\frac{1}{2} = \frac{\omega^2\gamma^2}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$$

or,  $\left(\omega_0^2 - \omega^2\right)^2 = \omega^2\gamma^2$  or,  $\omega_0^2 - \omega^2 = \pm\omega\gamma$

Thus we get two equations $\omega^2 - \gamma\omega - \omega_0^2 = 0$ and $\omega^2 + \omega\gamma - \omega_0^2 = 0$ and the positive roots of these two equations gives us

$$\omega_1 = -\frac{\gamma}{2} + \sqrt{\omega_0^2 + \gamma^2/4} \quad \text{and} \quad \omega_2 = +\frac{\gamma}{2} + \sqrt{\omega_0^2 + \gamma^2/4}$$

Thus B.W. $= \omega_2 - \omega_1 = \gamma$                                    ... (3.7.3)

Since Q-value is equal to $\dfrac{\omega_0}{\gamma}$; hence we can write $Q = \dfrac{\omega_0}{\gamma} = \dfrac{\omega_0}{\Delta\omega}$    ... (3.7.4)

Thus lower the bandwidth greater is the Q-value.

Thus from the variation of $\langle P \rangle$ with $\omega$ we see that if Q is large then the resonance is sharp which means the resonance takes place only for a small range of values of $\omega$ near $\omega_0$. But if Q is small then the resonance is not sharp or blurred which means that the resonance takes place over a wide range of frequencies on both side of $\omega_0$. Thus the sharpness of resonance of a forced vibrator is determined by its Q-factor.
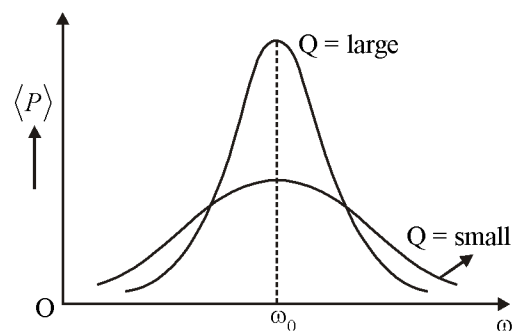


Fig. 3.5

## 3.8 LCR Resonance circuit
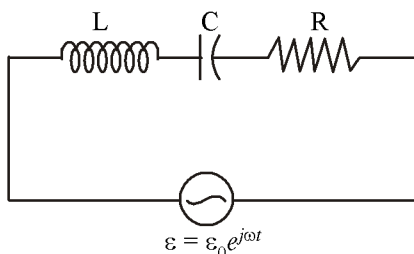


Fig. 3.6

So far we have discussed the forced vibration of mechanical systems but in a series LCR circuit with a a.c. source we can observe this forced vibration also.

Let us consider an inductance L, capacitance C and resistance R in series with an ac source of emf $\varepsilon = \varepsilon_0 e^{j\omega t}$, $j = \sqrt{-1}$.

Then if $i$ is the instantaneous current in the circuit, then using KVL (kirchhoff's voltage law) we get

$$\varepsilon - L\frac{di}{dt} = iR + \frac{q}{C}$$

or, $$L\frac{di}{dt} + iR + \frac{q}{C} = \varepsilon_0 e^{j\omega t}$$

or, $$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i = j\omega\varepsilon_0 e^{j\omega t}$$

Thus if we put $i = i_0 e^{j\omega t}$, then we get

$$\left[-\omega^2 L + j\omega R + \frac{1}{C}\right]i_0 = j\omega\varepsilon_0$$

or, $$\left[R + j\left(\omega L - \frac{1}{\omega C}\right)\right]i_0 = \varepsilon_0$$

or, $$i_0 = \frac{\varepsilon_0}{R + j\left(\omega L - \frac{1}{\omega C}\right)}$$

Hence $$\left|i_0\right| = \frac{\varepsilon_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)}}$$ and $i_0 = \left|i_0\right| e^{-j\delta}$

where $$\delta = \tan^{-1}\frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}$$

Thus the current in the ckt is $$i(t) = \frac{\varepsilon_0 . e^{\left[j\left[\omega t - \tan^{-1}\frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}\right]\right]}}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}$$

Thus we see that the current not only depends on L, C & R, but also on the frequency $\omega$ of the ac source. This current is maximum when

$\omega_0 L - \dfrac{1}{\omega_0 C} = 0$ is $\omega_0 = \dfrac{1}{\sqrt{LC}}$. this is called the

resonance frequency of the circuit.



Fig. 3.7

Comparing this eqation

$\dfrac{d^2 i}{dt^2} + \dfrac{R}{L}\dfrac{di}{dt} + \dfrac{i}{LC} = \dfrac{j\omega\varepsilon_0}{L} e^{j\omega t}$ with the equation
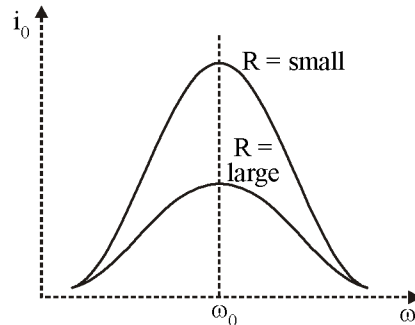
of forced vibration $\dfrac{d^2 x}{dt^2} + \gamma\dfrac{dx}{dt} + \omega_0^2 x = f_0 e^{i\omega t}$ we get damping factor $\gamma = \dfrac{R}{L}$ and

$\omega_0 = \dfrac{1}{LC}$.

So the Q-value of the ac circuit is given by

$$Q = \dfrac{\omega_0}{\gamma} = \dfrac{\omega_0}{R/L} = \dfrac{\omega_0 L}{R} = \dfrac{1}{\omega_0 C R}$$

Thus we see that smaller the value of R grdater the value of Q and this Q-factor also measures the sharpness of resonance.

## 3.9 Summary

1. Due to damping force the amplitude of a vibration body decreases. But this amplitude can be retained constant by imposing a periodic force on it. This is called forced vibration.

2. The general differential eqation of a forced vibration is

$$m\dfrac{d^2 x}{dt^2} + p\dfrac{dx}{dt} + kx = F_0 \cos\omega t \text{ or } F_0 e^{i\omega t}.$$

3. $\omega_0 = \sqrt{\dfrac{k}{m}}$ is the natural frequency of vibration of the body. $\omega$ is the frequency

of the forcing system and $\gamma = \dfrac{p}{m}$ is called damping factor.

4. The general solution of the above equation is

$$x = Ae^{-\gamma t/2}\cos\left(\omega't - \delta'\right) + B_0\cos\left(\omega t - \delta\right).$$ This is known as transient state but after a certain lapse of time, we get only $x = B_0\cos\left(\omega t - \delta\right)$. This is called steady state.

5. For a particular frequency $\omega_\gamma = \sqrt{\omega_0^2 - \gamma^2/4}$, the amplitude is maximum. This is called amplitude resonance and for $\omega = \omega_0$, the velocity is maximum. This is called velocity resonance.

6. For sustained oscillation in a forced vibration we must have average power dissipated by damping force must be equal to the power supplied by the forcing system.

   i.e $\quad \langle P \rangle_{supply} = \langle P \rangle_{diss}$.

7. The sharpness is a very important characteristics of resonance. And this sharpness of resonance is determined by Q-factor of the forced oscillator and it is $Q = \dfrac{\omega_0}{\Delta\omega}$.

8. Like mechanical system, the forced vibration is equally observed in electrical circuit also. LCR series resonance circuit is a good example of it.

## 3.10  Review Questions and Answers

1. A mass-spring system has m = 500 gm, force constant k = 200N/m. It is subjected to a damping force 0.5 N/ms$^{-1}$. If this mass is subject to a force of amplitude 5N & frequency $\omega$ = 30 rad/sec, then calculate the frequency of osillation at amplitude resonance and velocity resonance. What is the Q-value.

2. Discuss the variation of phase of a forced oscillator with frequency of forcing system.

3. Show that in a forced vibration the avergae power dissipated against damping force is equal to the power supplied by the driving force.

4. In a series LCR circuit R = 10$\Omega$ and C = 50$\mu$f. The emf of ac source is $\varepsilon$ = 20 sin 100 $\pi t$ volt. For what value of L, the current in the circuit will be maximum. And what is that maximum current?

5. If $x = B_0 e^{i\left(\omega t - \delta\right)}$ is the displacement of a forced oscillator and the value of $B_0$ at resonance is $B_{max}$ then show that $\dfrac{B_{max}}{B_0} = Q$, where Q is the Q-factor of the circuit.

6. Consider a weakly damped harmonic oscillator driven by a force $F = F_0 \cos \omega t$. If $m$ is the mass and $R$ is the damping force per unit velocity show that the width of the amplitude resonance curve is approximately $\sqrt{3}\,R/m$, where the width is measured between two frequencies where the amplitude equals half the maximum value.

7. In case of a forced vibration under a simple harmonic force the half-power frequencies correspond to the phase angle range given by $\tan\phi = \pm 1$ where $\phi$ is the phase difference between the displacement and the driving force.

8. Show that in forced vibration the total energy of the vibrating system is not constant. Also show that

$$\frac{Average\ PE}{Average\ KE} = \frac{\omega_0^2}{\omega^2}$$

where $\omega_0$ is the undamped natural (angular) frequency and $\omega$ is the angular frequency of the driving force.

9. Consider the following equation which represents the equation of motion of a weakly damped harmonic oscillator driven by a constant external force $F_0$.

$$m\ddot{x} + R_m\dot{x} + sx = F_0$$

Solve the equation for small damping and sketch graphically the nature of variation of $x$ with $t$.

## ANSWER

1. Here $\quad \omega_0 = \sqrt{\dfrac{k}{m}} = \sqrt{\dfrac{200}{0.500}} = \sqrt{400} = 20$ rad/sec

here $\quad \gamma = \dfrac{0.5}{0.5} = 1$ rad/sec

$\therefore \quad \omega_\gamma = \sqrt{\omega_0^2 - r^2/4} = \sqrt{400 - \dfrac{1}{4}} = \sqrt{\dfrac{1599}{4}}$ rad/sec $\simeq 200$ rad/sec

$\omega_0 = 20$ rad/sec is the frequency of velocity reasonance;

$Q = \dfrac{\omega_0}{\gamma} = \dfrac{20}{1} = 20$

2. If $x = B_0 \cos(\omega t - \delta)$ is the displacement of the forced oscillator then

$$\delta = \tan^{-1}\left[\frac{\omega\gamma}{\omega_0^2 - \omega^2}\right]$$

Now    $\delta \to 0$    when    $\omega \to 0$

$\quad\quad \delta \to \dfrac{\pi}{2}$    ,,    $\omega = \omega_0$

$\quad\quad \delta \to \pi$    ,,    $\omega \to$ v.v.large

Thus the variation of $\delta$ with $\omega$ is as shown in Fig. 3.P-1. $\delta$ is almost $\pi = 180°$ when $\omega$ is very very large.
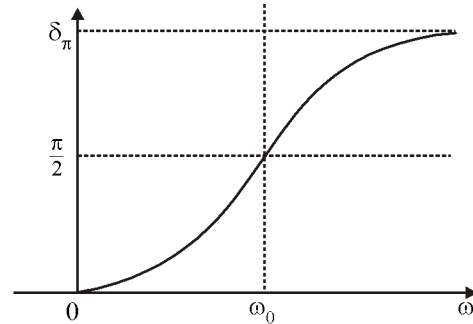


Fig. 3.P-1

3. Here $F = F_0 \cos \omega t$ is dring force $x = B_0 \cos(\omega t - \delta)$ displacement and

$$v = \frac{dx}{dt} = -\omega B \sin(\omega t - \delta) = \omega B \cos(\omega t - \phi) \text{ where } \phi = \frac{\pi}{2} - \delta .$$

$$\therefore \quad v = \omega B \cos\phi \cos\omega t + \omega B \sin\phi \sin\omega t = \omega\left[B_1 \cos\omega t + B_2 \sin\omega t\right]$$

Damping force $F = p\dfrac{dx}{dt}$        $B_1 = B\cos\phi, \quad B_2 = B\sin\phi$

$\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad = B\sin\delta \quad\quad = B\cos\delta$

Now power dissipated is

$$P = p\frac{dx}{dt} \times \frac{dx}{dt} = p\left(\frac{dx}{dt}\right)^2$$

$$\therefore \quad P_{diss} = p\omega^2 \left\{B_1 \cos\omega t + B_2 \sin\omega t\right\}^2$$

$$\left\langle P_{diss} \right\rangle = p\omega^2 \frac{1}{2}\left(B_1^2 + B_2^2\right) = \frac{1}{2}p\omega^2 \frac{f_0^2}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$$

and    $P_{supp} = F(t).\dfrac{dx}{dt} = F_0 \cos\omega t.\omega\left\{B_1 \cos\omega t + B_2 \sin\omega t\right\}$

$$\left\langle P_{supp} \right\rangle = F_0 \omega B_1 \left\langle \cos^2 \omega t \right\rangle = \frac{1}{2}F_0\omega B_1 = \frac{1}{2}F_0\omega \frac{f_0\omega\gamma}{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}$$

as $f_0 = \dfrac{F_0}{m}$, hence $\left\langle P_{supp} \right\rangle = \left\langle P_{diss} \right\rangle$

4.  Here $\omega_0^2 = \dfrac{1}{LC}$ and it is 100 radians, so $(100)^2 \dfrac{1}{L \times 50 \times 10^{-6}} = \dfrac{10^6}{50L}$

    or, 50L = 100, L = 2 Henry.

    The current is maximum when resonance takes place and this maximum current is

    $$i_{\max} = \frac{\varepsilon_0}{R} = \frac{20}{10} = 2 \text{ amp.}$$

5.  From Eqn. (3.3.4), we have $B_0 = \dfrac{f_0}{\left\{ \left(\omega_0^2 - \omega^2\right)^2 + \omega^2 d^2 \right\}^{1/2}}$ ... (1)

    Again, from (3.4.1), we have $\omega_\gamma = \sqrt{\omega_0^2 - d^2/2}$ ... (2)

    Substituting $\omega_\gamma$ for $\omega_0$ in Eqn. (1), we have

    $$B_{\max} = \frac{f_0}{\gamma \left( \omega_0^2 - \dfrac{d^2}{4} \right)^{1/2}}$$

    $$\therefore \quad \frac{B_{\max}}{B_0} = \frac{\omega_0^2}{\gamma \left( \omega_0^2 - \dfrac{d^2}{4} \right)^{1/2}} = \frac{\omega_0}{\gamma \left( 1 - \dfrac{\gamma^2}{4\omega_0^2} \right)^{1/2}} = \frac{Q}{\left( 1 - \dfrac{1}{4Q^2} \right)^{1/2}}$$

    For low damping, Q is very large, hence

    $$= \frac{B_{\max}}{B_0} = Q \left( 1 - \frac{1}{4Q^2} \right)^{-\frac{1}{2}} \simeq Q \left( 1 + \frac{1}{8Q^2} \right) \simeq Q$$

    Thus $B_{\max} = Q B_0$ (Proved).

6.  Amplitude $A = \dfrac{f_0}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4b^2 \omega^2}}$

    where $f_0 = F_0/m$, $b = R/2m$ and $\omega_0$ is the undamped frequency.

$$\omega_{res} = \sqrt{\omega_0^2 - 2b^2}$$

$$A_{\max} = \frac{f_0}{2b\omega_0}$$

Now at $A = \dfrac{A_{\max}}{2}$

$$\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4b^2\omega^2} = 2 \times 2b\omega_0$$

or, $\left(\omega_0^2 - \omega^2\right)^2 + 4b^2\omega^2 = 16b^2\omega_0^2$

As $b^2 < \omega_0^2$ we can write $4b^2\omega^2 = 4b^2\omega_0^2$

$$\left(\omega_0^2 - \omega^2\right) = 12b^2\omega_0^2$$

or, $\omega_0^2 - \omega^2 = \pm 2\sqrt{3}b\omega_0$

or, $\omega^2 = \omega_0^2 \left(1 \pm 2\sqrt{3}b / \omega_0\right)$

or, $\omega = \omega_0 \left(1 \pm \sqrt{3}b / \omega_0\right)$

Therefore, $A = \dfrac{A_{\max}}{2}$, at frequencies

$$\omega_1 = \omega_0 - \sqrt{3}b \text{ and } \omega_2 = \omega_0 + \sqrt{3}b$$

$\therefore$ Width $\omega_2 - \omega_1 = 2\sqrt{3}b = 2\sqrt{3}, \quad \dfrac{R}{2m} = \sqrt{3}R/m$

7. Average power [from Eqn. (3.6.1)

$$\langle P \rangle = \frac{mbf_0^2}{\left(\dfrac{\omega_0^2}{\omega} - \omega\right)^2 + 4b^2}$$

At half power points,

$$\left(\frac{\omega_0^2}{\omega}-\omega\right)^2 = 4b^2 \quad \text{or,} \quad \omega_0^2 - \omega^2 = \pm 2b\omega$$

Therefore the phase angle $\phi$ is given by

$$\tan\phi = \frac{2b\omega}{\omega_0^2 - \omega^2} = \pm 1$$

8.  For forced vibration under a force $F = F_0 \cos \omega t$ the displacement is given by

$$x = A\cos(\omega t - \phi)$$

Now total energy E = K.E. + P.E.

$$= \frac{1}{2}m\dot{x}^2 + \frac{1}{2}sx^2$$

$$= \frac{1}{2}mA^2\omega^2 \sin^2(\omega t - \phi) + \frac{1}{2}sA^2 \cos^2(\omega t - \phi)$$

where m = mass, s = stiffness and $\omega_0 = \sqrt{s/m}$

Clearly E is not constant and varies with time t.

Now average values of K.E. and P.E. are given by,

$$\langle K.E.\rangle = \frac{1}{4}mA^2\omega^2$$

and $\quad \langle P.E.\rangle = \frac{1}{4}sA^2 = \frac{1}{4}mA^2\omega_0^2$

$$\left[\because \; <\sin^2(\omega t - \phi>) = <\cos^2(\omega t - \phi)> = \frac{1}{2}\right]$$

9.  The given equation can be rewritten as

$$\ddot{x} + 2b\dot{x} + \omega_0^2 x = F_0/m \quad [\text{where } \omega_0^2 = s/m \text{ and } 2b = R_m/m ]$$

or, $\quad \ddot{x} + 2b\dot{x} + \omega_0^2\left(x - \frac{F_0}{s}\right) = 0$

Introducing a new variable $y = x - F_0/x$ we get

$$\ddot{y} + 2b\dot{y} + \omega_0^2 y = 0$$

For $b < \omega_0$ we have the solution

$$y = Ae^{-bt} \cos(\omega t - \phi)$$

or, $\quad x = \dfrac{F_0}{s} + Ae^{-bt} \cos(\omega t - \phi)$

where $\omega = \sqrt{\omega_0^2 - b^2}$ and A, $\phi$ are constants determined by the initial conditions
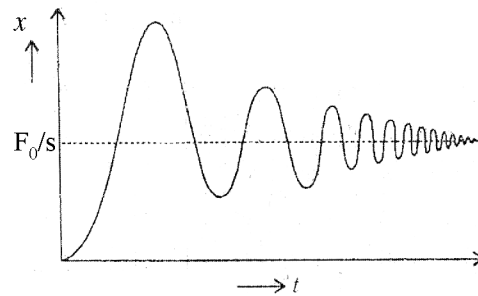


Fig. 3.P-2 : Forced vibration with aconstant force

The variation of x with t is as shown in Fig. 3.P-2.

# Unit-4 : Superposition of SHM

**Structure**

## 4.0 Objectives

After reading this unit, you will learn :
- principle of superposition
- applications of superposition principle in different cases
- formation of beats
- Lissaous figures and its uses

## 4.1 Introduction

In the previous unit we have discussed the condition of SHM and its differential equation and its solution. We have also shown same inportant applications. But in nature many times we see a number of SHM. acts simultaneously on a single system. For example in a sitar a point on a wire, a number of SHM of different frequency superpose on each other. In such condition the resultant motion can be obtained by using a principle called superposition principle. And the resultant vibration may be of different types and form depending upon the number of frequencies, their magnitude, phase and direction of oscillation. In this unit we will learn the resultant of two, three or more SHM of same or different frequencies in either same or perpendicular direction.

At the end we will discuss some methods to observe these resultant vibration by mechanical, optical or electronic method.

## 4.2 Superposition Principle

This principle states that "if two SHMs of same frequency but different amplitude and phase are superposed, then the resultant will also be a SHM of same frequency but of different amplitude and phase". In fact this principle is a consequence of the fact that for a homogeneous 2nd order differential equation. If $y_1$ is a solution and $y_2$ is another solution, then $y = y_1 + y_2$ must be another solution of the same equation.

The differential equation of SHM is $\dfrac{d^2x}{dt^2} = -\omega^2 x$ is a 2nd order linear homogeneous

differential equation, hence if $x_1 = A\sin\left(\omega t + \phi_1\right)$ is a solution and $x_2 = B\sin(\omega t + \phi_2)$ is another solution, then their linear combination $x = A\sin(\omega t + \phi_1) + B\sin(\omega t + \phi_2)$ will also be a solution i.e. is it will represent another solution.

It is to be noted that if the differential equation contains terms like $\left(\dfrac{dx}{dt}\right)^2$ or $x^2$

terms then this principle will not hold.

## 4.3 Superposition of two SHMs of same frequency acting along the same direction

1. **Trigonometric method** : Let $x_1 = a_1\cos\left(\omega t + \theta_1\right)$ and $x_2 = a_2\cos\left(\omega t + \theta_2\right)$ represent two SHMs of same frequency in same direction (along x-axis). Then if these two are supsposed, then the resultant will be

$$x = x_1 + x_2 = a_1\cos\left(\omega t + \theta_1\right) + a_2\cos\left(\omega t + \theta_2\right)$$

$$= a_1\cos\omega t\cos\theta_1 - a_1\sin\omega t\sin\theta_1 + a_2\cos\omega t\cos\theta_2 - a_2\sin\omega t\sin\theta_2$$

$$= \left(a_1\cos\theta_1 + a_2\cos\theta_2\right)\cos\omega t - \left(a_1\sin\theta_1 + a_2\sin\theta_2\right)\sin\omega t$$

$$= a_1\cos\delta\cos\omega t - a\sin\delta\sin\omega t$$

$$= a\cos\left(\omega t + \delta\right)$$

where the resultant amplitude is

$$a^2 = \left(a_1\cos\theta_1 + a_2\cos\theta_2\right)^2 + \left(a_1\sin\theta_1 + a_2\sin\theta_2\right)^2$$

or,     $a = \sqrt{a_1^2 + a_2^2 + 2a_1 a_2 \cos(\theta_1 - \theta_2)}$                    ... (4.3.1)

and     $\tan\delta = \dfrac{a_1 \sin\theta_1 + a_2 \sin\theta_2}{a_1 \cos\theta_1 + a_2 \cos\theta_2}$                    ... (4.3.2)

**Vector addition method :** We know that a SHM can be represented by a rotating vector method.

Let $\overrightarrow{OP} = \vec{a}_1$ and $\overrightarrow{OQ} = \vec{a}_2$ are two rotating vectors rotating at angular velocity $\omega$. If $\omega t + \theta_1$ and $\omega t + \theta_2$ are the angles made by $\vec{a}_1$ and $\vec{a}_2$ at any time $t$, then their resultant will be given by
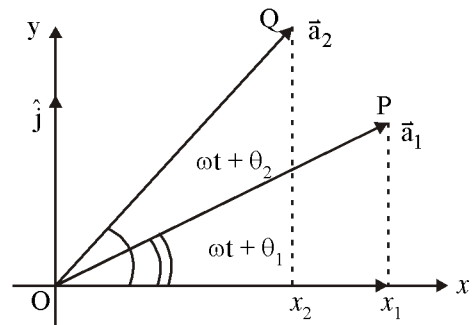
$$\vec{a} = \vec{a}_1 + \vec{a}_2 = \hat{i}a_1 \cos(\omega t + \theta_1) + \hat{j}a_1 \sin(\omega t + \theta_1) +$$

$$\hat{i}a_2 \cos(\omega t + \theta_2) + \hat{j}\sin(\omega t + \theta_2)$$

If we consider only the $x$ componenet of this vector then we see that the resultant will also represent a SHM of same frequency $\omega$. But of amplitude equal to

$$A\cos\delta = a_1 \cos\theta_1 + a_2 \cos\theta_2 \text{ and } A\sin\delta = a_1 \sin\theta_1 + a_2 \sin\theta_2$$

Hence the resultant will be $x = A\cos(\omega t + \delta)$

In this process are can add any number of SHM of same frequency but of different amplitude and initial plase. If we have $n$ no. of SHM given by

$$x_1 = a_2 \cos(\omega t + \theta_1), \quad x_2 = a_2 \cos(\omega t + \theta_2) \dots x_n = a_n \cos(\omega t + \theta_n)$$

then the resultant SHM will have amplitude 'a',

$$a\cos\delta = a_1 \cos\theta_1 + a_2 \cos\theta_2 + \dots a_n \cos\theta_n = \sum_{i=1}^{n} a_i \cos\theta_i$$

$$a\sin\delta = a_1 \sin\theta_1 + a_2 \sin\theta_2 + \dots a_n \sin\theta_n = \sum_{i=1}^{n} a_i \sin\theta_i$$

So     $a = \sqrt{\left(\sum_{i=1}^{n} a_i \cos\theta_i\right)^2 + \left(\sum_{i=1}^{n} a_i \sin\theta_i\right)^2}$                    ... (4.3.3)

and resultant phase $\tan\delta = \dfrac{\sum\limits_{i=1}^{n} a_i \sin\theta_i}{\sum\limits_{i=1}^{n} a_i \cos\theta_i}$ ... (4.3.4)

3. **Using complex form :** We know that $e^{i\theta} = \cos\theta + i\sin\theta$ hence a SHM. $a\cos(\omega t + \theta)$ may be expressed as R.P. of $a\,e^{i(\omega t+\theta)}$. Similarly $y = a\sin(\omega t + \theta)$ may be expressed as $y = ImP.$ of $a\,e^{i(\omega t+\theta)}$.

Thus if we have $x_1 = a_1\cos(\omega t + \theta_1)$ and $x_2 = a_2\cos(\omega t + \theta_2)$ then their resultant will be given by

$$x = x_1 + x_2 = R.P.\,a_1 e^{i(\omega t+\theta_1)} + R.P.\,a_2 e^{i(\omega t+\theta_2)}$$

$$= R.P.\,e^{i\omega t}.\left[a_1 e^{i\theta_1} + a_2 e^{i\theta_2}\right]$$

$$= R.P.\,Ae^{i\omega t}.e^{i\delta} \quad \text{(say)}$$

So $Ae^{i\delta} = a_1 e^{i\theta_1} + a_2 e^{i\theta_2}$

Thus $A^*A = |A|^2 = \left(a_1 e^{i\theta_1} + a_2 e^{i\theta_2}\right).\left(a_1 e^{-i\theta_1} + a_2 e^{-i\theta_2}\right)$

$$= a_1^2 + a_2^2 + a_1 a_2 e^{i(\theta_1-\theta_2)} + a_1 a_2 e^{-i(\theta_1-\theta_2)}$$

$$= a_1^2 + a_2^2 + a_1 a_2 \left\{e^{i(\theta_1-\theta_2)} + e^{-i(\theta_1-\theta_2)}\right\}$$

$$= a_1^2 + a_2^2 + 2a_1 a_2 \cos(\theta_1 - \theta_2)$$

Thus $A = \sqrt{a_1^2 + a_2^2 + 2a_1 a_2 \cos(\theta_1 - \theta_2)}$ ... (4.3.5)

and $\tan\delta = \dfrac{a_1 \sin\theta_1 + a_2 \sin\theta_2}{a_1 \cos\theta_1 + a_2 \cos\theta_2}$ ... (4.3.6)

**Exercise 1.** Let we have n number of SHM of same angular frequency and same amplitude but each has a constant phase difference from other, then find the resultant of these due to superposition.

Let $x_1 = ae^{i\omega t}$, $x_2 = ae^{i(\omega t + \phi)}$, $x_3 = ae^{i(\omega t + 2\phi)}$ ..... $x_n = ae^{i\{\omega t + (n-1)\phi\}}$

Then when they are superposed.

The resultant will be $x = x_1 + x_2 + ... + x_n$

$$x = ae^{i\omega t}\left\{1 + e^{i\phi} + e^{2i\phi} ... + e^{(n-1)i\phi}\right\}$$

$$= ae^{i\omega t} \cdot \frac{\left(1 - e^{in\phi}\right)}{\left(1 - e^{i\phi}\right)} = ae^{i\omega t} \cdot \frac{e^{\frac{in\phi}{2}}}{e^{\frac{i\phi}{2}}} \cdot \frac{\left\{e^{-\frac{in\phi}{2}} - e^{-\frac{in\phi}{2}}\right\}}{\left\{e^{-\frac{i\phi}{2}} - e^{-\frac{i\phi}{2}}\right\}}$$

$$= ae^{i\left(\omega t + \left(\frac{n-1}{2}\right)\phi\right)} \cdot \frac{\sin\frac{n\phi}{2}}{\sin\frac{\phi}{2}} = Ae^{i\left(\omega t + \left(\frac{n-1}{2}\right)\phi\right)}$$

where $A = a\dfrac{\sin\frac{n\phi}{2}}{\sin\frac{\phi}{2}}$, and $\delta = \left(\dfrac{n-1}{2}\right)\phi$

## 4.4 Superposition of two SHM in same direction but of different frequency

During superpostion of two SHMs in same direction if the frequencies are different, then a new phenomena is developed which is called beat phenomena.

To explain this let $x_1 = a_1 \cos(\omega_1 t + \phi_1)$ and $x_2 = a_2 \cos(\omega_2 t + \phi_2)$ are two SHMs in same direction but of different frequencies $\omega_1$ and $\omega_2$. Then thier superposition gives

$$x = x_1 + x_2 = a_1 \cos(\omega_1 t + \phi_1) + a_2 \cos(\omega_2 t + \phi_2)$$

Let $\omega_1 = \omega$ and $\omega_2 + \Delta\omega$

$\therefore \quad x = a_1 \cos(\omega t + \phi_1) + a_2 \cos\{\omega t + \Delta\omega t + \phi_2\}$

$= a_1 \cos\omega t \cos\phi_1 - a_1 \sin\omega t \sin\phi_1 + a_2 \cos\omega t \cos\phi_2' - a_2 \sin\omega t \sin\phi_2'$

where $\quad \phi_2' = \Delta\omega t + \phi_2$

$$= \left(a_1 \cos\phi_1 + a_2 \cos\phi_2'\right)\cos\omega t - \left(a_1 \sin\phi_1 + a_2 \sin\phi_2'\right)\sin\omega t$$

$$= A\cos\delta\cos\omega t - A\sin\delta\sin\omega t$$

$$= A\cos\left(\omega t + \delta\right) \qquad\qquad ... (4.4.1)$$

where $\quad A\cos\delta = a_1 \cos\phi_1 + a_2 \cos\phi_2'$

$$A\sin\delta = a_1 \sin\phi_1 + a_2 \sin\phi_2'$$

and $\quad A = \left\{\left(a_1 \cos\phi_1 + a_2 \cos\phi_2'\right)^2 + \left(a_1 \sin\phi_1 + a_2 \sin\phi_2'\right)^2\right\}^{\frac{1}{2}}$

$$= \left\{a_1^2 + a_2^2 + 2a_1 a_2 \left(\cos\phi_1 \cos\phi_2' + \sin\phi_1 \sin\phi_2'\right)\right\}^{\frac{1}{2}}$$

$$= \left\{a_1^2 + a_2^2 + 2a_1 a_2 \cos\left(\phi_2' - \phi_1\right)\right\}^{\frac{1}{2}} \qquad ... (4.4.2)$$

Here in equation (4.4.1) amplitude A is a function of time t and we see when

$$\phi_2' - \phi_1' = \Delta\omega t + \phi_2 - \phi_1 = 0,\ 2\pi,\ 4\pi,\ 6\pi \text{ etc. } A = (a_1 + a_2) = \text{maximum.}$$

but when $\quad \phi_2' - \phi_1' = \pi,\ 3\pi,\ 5\pi$ etc. $A = (a_1 - a_2) = $ minimum.

Thus equation (4.4.1) shows a vibration whose amplitude will vary periodically with time and the time difference between these two successive maximum and minimum

is $\Delta\omega\left(t_2 - t_1\right) = 2\pi$ or $t_2 - t_1 = t_b = \dfrac{2\pi}{\Delta\omega} = \dfrac{2\pi}{\omega_2 - \omega_1}$ $\qquad ... (4.4.3)$

Fig. 4.2 displays graphically the result of superpostion of two SHMs of slightly different frequencies. As an examiple, we show in Fig 4.2 superposition of two SHMs having frequencies 4Hz and 5Hz. They are further assumed to have same amplitudes and same initial phases.

It is  clear from Fig 4.2 that the time interval between two successive maxima or minima in the resultant pattern in this case is 1 sec. The resultant amplitude thus varies with a frequency of 1 Hz which is equal  to the difference of frequencies (5 Hz – 4 Hz) of the component vibrations.

**Beats :**

In caee of sound waves the superpostion of two waves having slightly different frequencies causes the intensity of the resulting sound to increase and decrease

periodically with time. This phenomenon is known as beats. The numnber of beats per second equals the difference of frequencies of the component waves.
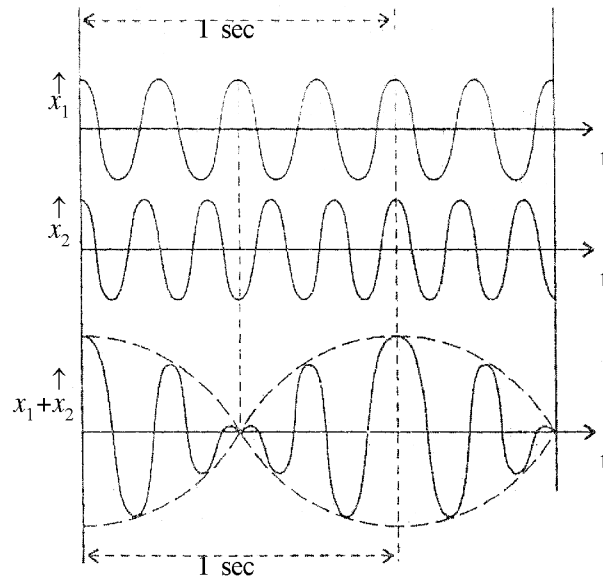


**Fig. 4.2 : Superposition of SHMs with slightly different frequencies**

**Exercise 2.** Two colinear SHM are $x_1 = 20\sin\left(52\pi t + \dfrac{\pi}{3}\right)$ and

$x_2 = 24\sin\left(48\pi t + \dfrac{\pi}{6}\right)$ cm. Find the equation of resultant vibration.

## 4.5 Superposition of a large number of SHMs

Let us consider $N$ number of SHM of same amplitude in same direction but whose frequencies differ from each other by same amount.

Let $\qquad x_1 = ae^{i\omega t}$, $\quad x_2 = ae^{i(\omega + \Delta\omega)t}$, $\quad x_3 = ae^{i(\omega + 2\Delta\omega)t}$ $\quad\ldots\ldots\quad x_N = ae^{i[\omega + (N-1)\Delta\omega]t}$

$\therefore \quad x = x_1 + x_2 + \ldots\ldots + x_N = ae^{i\omega t}\left\{1 + e^{i\Delta\omega t} + e^{i\Delta\omega t} + \ldots\ldots + e^{(N-1)i\Delta\omega t}\right\}$

$$= ae^{i\omega t}\left\{\frac{1-e^{iN\Delta\omega t}}{1-e^{i\Delta\omega t}}\right\}$$

$$= ae^{i\omega t}\frac{e^{\frac{iN\Delta\omega t}{2}}\left\{e^{\frac{iN\Delta\omega t}{2}}-e^{-\frac{iN\Delta\omega t}{2}}\right\}}{e^{\frac{i\Delta\omega t}{2}}\left\{e^{\frac{i\Delta\omega t}{2}}-e^{-\frac{i\Delta\omega t}{2}}\right\}}$$

$$= ae^{i\left\{\omega+\frac{N-1}{2}\Delta\omega\right\}t}\cdot\frac{\sin\frac{N\Delta\omega t}{2}}{\sin\frac{\Delta\omega t}{2}}$$

$$= Ae^{i\left\{\omega+\frac{N-1}{2}\Delta\omega\right\}t} \qquad\qquad ...(4.5.1)$$

Thus the amplitude of resultant vibration of N number of SHM will be

$$A = a\frac{\sin\frac{N\Delta\omega t}{2}}{\sin\frac{\Delta\omega t}{2}} \qquad\qquad ...(4.5.2)$$

Thus the intensity of resultant vibrtion due to superpotion of N number of such SHM will be proportional to

$$\frac{\sin^2\frac{N\Delta\omega t}{2}}{\sin^2\frac{\Delta\omega t}{2}}$$

The variation of intensity due superposition of such $N$ number of waves of equally separated frequency is shown in Fig. 4.3.
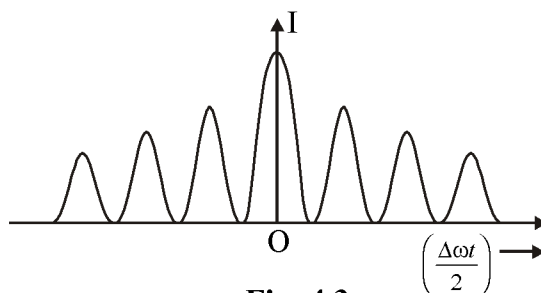


**Fig. 4.3**

The above analysis is helpful in finding the diffraction pattern due to a plane diffraction grating in the study of physical optics.

**Exercise 3.** Three SHMs are of same frequncy $\omega = 100\pi$ but their phase differ form one another by $\frac{\pi}{3}$ . If each have same amplitude a = 5 cm then find their resultant amplitude.

**Exercise 4.** You have N = 100 number of SHMs each of same amplitude a = 2cm but frequencies starting from $\omega$ = 100$\pi$ differ from each other by $\Delta\omega$ = 2$\pi$. Find the amplitude of resultant vibration of these.

## 4.6 Superpoisition of two SHMs in mutually perpendicular direction

### 4.6.1. Same frequency

Let us us consider $x = a\cos\omega t$ and $y = b\cos(\omega t + \phi)$ are two SHMs in mutually perpendicular direction. If they superpose on a particle simultaneously, then their resultant vibration may be obtained as

$$\frac{y}{b} = \cos\omega t\cos\phi - \sin\omega t\sin\phi = \frac{x}{a}\cos\phi - \sqrt{1 - \frac{x^2}{a^2}}\sin\phi$$

or, $\qquad \left[\frac{y}{b} - \frac{x}{a}\cos\phi\right]^2 = \left(1 - \frac{x^2}{a^2}\right)\sin^2\phi$

or, $\qquad \frac{y^2}{b^2} + \frac{x^2}{a^2} - \frac{2xy}{ab}\cos\phi = \sin^2\phi \qquad\qquad$ ... (4.6.1)

This shows that the resultant motion will depend on the phase difference $\phi$.

(a)    If $\phi$ = 0, then (4.6.1) gives $y = \frac{b}{a}x$            ... (4.6.2)

which represents a straight line with slope $\frac{b}{a}$.

(b)    If $\phi = \frac{\pi}{2}$, then (4.6.1) reduces to $\frac{y^2}{b^2} + \frac{x^2}{a^2} = 1$      ... (4.6.3)

which is the equation of an ellipse and it reduces to a circle if $a$ = $b$, i.e.
$\qquad x^2 + y^2 = a^2$                     ... (4.6.4)

(c)    If $\phi$ = $\pi$, then (4.6.1) gives $y = -\frac{b}{a}x$        ... (4.6.5)

which is again a straight line with slope $-\frac{b}{a}x$

(d)    If $\phi$ has any other value excluding these three, then (4.6.1) will represent an inclined ellipse



(a)  $\phi = 0$

(b)  $\phi = \dfrac{\pi}{2}$

(c)  $\phi = \pi$

(d)  $\phi$ = arbitrary $\pi > \phi > 0$

(e)  $\phi$ = arbitrary $-\pi > \phi > 0$
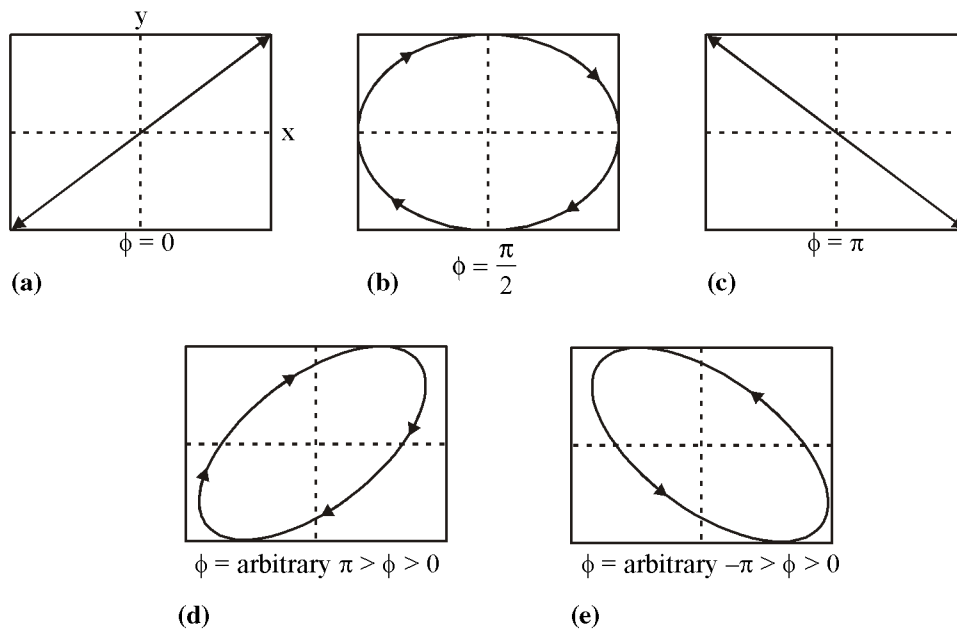
**Fig. 4.4**

Here it is to be noted that where $\pi > \phi > 0$, the motion will be clockwise but when $-\pi < \phi < 0$, the motion will be anticlockwise.

In this section we have discussed the resultant of two vibrations.

### 4.6.2 Frequency ratio 1:2

Let two rectangular SHMs of frrequencies in the ratio 1:2 differing in phase by $\delta$ be represented by

$$x = a\cos\omega t \qquad\qquad ... (4.6.6)$$

$$y = b\cos(2\omega t + \delta) \qquad\qquad ... (4.6.7)$$

$$\frac{y}{b} = \cos 2\omega t.\cos\delta - \sin 2\omega t.\sin\delta$$

$$= \left(2\cos^2\omega t - 1\right)\cos\delta - 2\sin\omega t.\cos\omega t.\sin\delta$$

Using Equation (4.6.6) we get

$$\frac{y}{b} = \left(2.\frac{x^2}{a^2} - 1\right)\cos\delta - 2\frac{x}{a}\sqrt{1 - \frac{x^2}{a^2}}\sin\delta$$

or, $\left[\dfrac{y}{b} + \cos\delta - \dfrac{2x^2}{a^2}\cos\delta\right]^2 = \dfrac{4x^2}{a^2}\left(1 - \dfrac{x^2}{a^2}\right)\sin^2\delta$　　　　... (4.6.8)

This is an equation of fourth degree in $x$ and, in general, represents closed curve having two loops. for a given value of $\delta$ the exact nature of the curve can be traced.
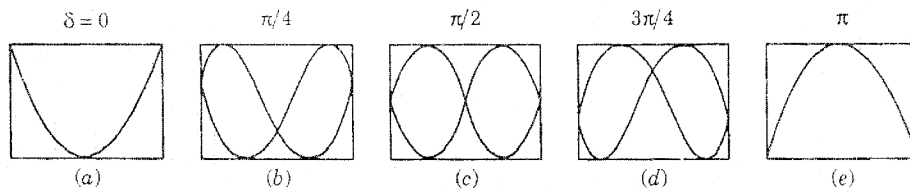


Fig 4.5 Resultant pattern of two rectangular SHMs having frequency ratio 1 : 2

For example, if $\delta = 0$ then Eqn. (4.6.8) reduces to

$$\left(\dfrac{y}{b} + 1 - \dfrac{2x^2}{a^2}\right)^2 = 0$$

It represents two coincident parabolas [Fig 4.5.(a)] given by

$$x^2 = \dfrac{a^2}{2b}(y + b)$$　　　　...(4.6.9)

It $\delta = \pi/2$ Eqn. (4.6.8) reduces to

$$\dfrac{4x^2}{a^2}\left(\dfrac{x^2}{a^2} - 1\right) + \dfrac{y^2}{b^2} = 0$$　　　　... (4.6.10)

This equation represents a curve containing two loops as shown in Fig. 4.5.(c)

### 4.6.3 Frequencies in any commensurate ratio

If the frequency ratio of the rectangular SHMs becomes large then analytical method of obtaining the resultant pattern becomes very combersome. In that case one can use graphical method using the concept of rotating vectors.

## 4.7　Lissajous figures

The figures or curves formed by the superposition of two simple harmonic motions at right angles to each other are known as *Lissajous figures*. The shape of these curves depends on the ratio of frequencies as well as on the initial phase relationship of the

component simple harmonic motions. The curves shown in Fig. 4.4 and Fig. 4.5 are examples of Lissajous figures.

**Demonstration of Lissajous figures :**

There are a number of experimental methods for obtaining Lissajous figures. These may be classified as mechanical optical and electrical methods. We only discuss electrical method here.

**Electrical method**

Most suitable method of demonstrating Lissajous figures is to use a modern cathode ray oscilloscope (CRO) Here a narrow beam of electrons iron an electron gun is passed through two pairs of parallel metal plates arranged at right angles to each other. When electric field is set up one pair of plates deflects the electrons in horizontal direction and the other pair deflects them in vertical direction. The electrons finally impinge on a fluorescent screen and produce visible spot.
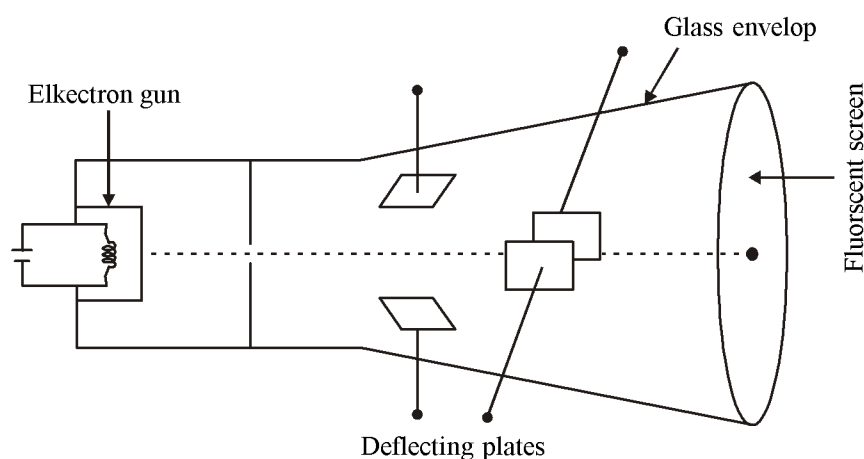


**Fig. 4.6 Cathode-ray oscilloscope**

To display a Lissajous figure on the CRO screen two simple harmonic vibrations are at first converted into sinusoidal voltages with the help of two microphones. These voltages, after proper magnification, are applied to the two pairs of deflecting plates of the CRO. As the electrons move under the simultaneous action of two sinusoidal electric fields at right angles to each other they traces out Lissajous figure on the CRO screen.

**Use of Lissajous figures :**

Lissajous figures have many important uses. For example, these figures may be

used to find the ratio of two exactly commensurate frequencies of the component vibrations. For this purpose draw a horizontal and a vertical line to intersect the curve and find the number of intersections each line makes with curve. Now the ratio of the number of intersections of the horizontal line to that on the vertical line will give the ratio of the vertical to the horiozontal frequency i.e.,

$$\frac{\upsilon_{vertical}}{\upsilon_{horizonal}} = \frac{\text{No. of cuts on horozontal line}}{\text{No. of cuts on vertical line}}$$

Lissajous figures can be used to compare to two nearly equal frequency. If the frequencies differe slightly Lissajous figures change gradually and passes through a complete cycle of changes in a time T given by

$$T = \frac{1}{\Delta\upsilon}$$

So by measuring T we can find the difference $\Delta\upsilon$ of frequencies. Again, if one of the frequencies is knowsn then the other can be found out.

Lissajous figures can also be used to measure the phase difference δ between two signals of same frequency. Depending on the value of δ the Lissajous figure will be an ellipse or one of its degenerate forms. Refgerring to Fig. 4.7 let the distance of the point of intersection of the ellipse with y-axis be A and maximum vertical displacement be B. Now from Eq. (4.6.1) A = b sin δ and B = b. Hence



Fig. 4.7 Measurement of phase differences

$$\delta = \sin^{-1}(A/B)$$

Thus measuring A and B we can determine δ.

## 4.8  Summary

1.  Superposition principle and its applications to (a) two collinear SHMs oscillations with equal and unequal frequencies and (b) two SHMs perpendicular oscillations have been discussed.

2.  Formation of Lissajous figures and its uses have been mentioned.

## 4.9   Review Questions and Answers

1.  Let $x_1 = x_{01} \sin \omega t$ and $x_2 = x_{02} \sin \omega t$ represent solutions of a SHM equation. Then show that $x_1 + x_2$ will also be solution of that equation. If $\omega$ of $x_1$ and $x_2$ where different then will they satisfy the same condition?

2.  The equation of motion of a pendulum is $\dfrac{d^2\theta}{dt^2} + \dfrac{g}{l} \sin \theta = 0$. Does it satisfy the superpostion principle. If not then why?

3.  The equation of two SHMs are $x_1 = A \cos (\omega t - 30°)$ and $x_2 = A \cos (\omega t + 45°)$. Find their resultant using vector method and complex form method.

4.  Two SHMs are vibrating with same frequency and amplitude in mutually perpendicular direction. The phase of $x$ directed SHM is 60° ahead that of $y$ directed motion. Discuss the trajectory of resultant motion.

5.  the earth moves round the sun in an elliptical orbit under the action of gravitational force. Can the motion of earth be regarded as the superposition of two SHMs in mutually perpendicular direction.

6.  Two SHM of same amplitude but different frequency superpose on a particle simultaneously in mutually perpendicular direction. Will the resultant force on the particle be propostional to the displacement and in direction opposite to it.

7.  Lissajous figure obtained by using two tuning forks A and B having slightly different frequencies undergoes complete cycle of changes in 20 sec. When $B$ is loaded with a little wax, the time reduces to 10 sec. If the frequency of A is 288 Hz, calculate the frequency of $B$ before loading it with wax.

8.  Suppose a particle is simultaneously subjected to three collinear SHMs all of the same frequency and having amplitudes 1 cm, 0.5 cm and 0.25 cm respectively. the phase of the second relative to the first is 30° and that of the third relative to the second is 60°, find the resultant motion.

<div align="center"><u><b>ANSWER</b></u></div>

1.  Let $\dfrac{d^2 x}{dt^2} + \omega^2 x = 0$ is the equation of SHM. Then if $x_1$ and $x_2$ are two solutions

    then $\dfrac{d^2 x_1}{dt^2} + \omega^2 x_1 = 0$ and $\dfrac{d^2 x_2}{dt^2} + \omega^2 x_1 = 0$ so adding or subtrating we get

    $\dfrac{d^2}{dt^2}\left(x_1 \pm x_2\right) + \omega^2 \left(x_1 \pm x_2\right) = 0$ hence $x_1 \pm x_2$ is also a solution of the equation.

But if $\omega_1$ and $\omega_2$ are different then we get $\dfrac{d^2x_1}{dt^2} + \omega_1^2 x_1 = 0$ and $\dfrac{d^2x}{dt^2} + \omega_2^2 x_2 = 0$,

hence $\dfrac{d^2}{dt^2} + \left(x_1 + x_2\right) + \left(\omega_1^2 x_1 + \omega_2^2 x_2\right) = 0$ and hence if $\omega_1 \neq \omega_2$ so it does not represent the solution of same equation.

2. There if $\theta_1$ and $\theta_2$ are two solutions then $\dfrac{d^2\theta_1}{dt^2} \dfrac{g}{l} \sin\theta_1 = 0$ and $\dfrac{d^2\theta_2}{dt^2} \dfrac{g}{l} \sin\theta_2 = 0$

but adding these we get $\dfrac{d^2}{dt^2}\left(\theta_1 + \theta_2\right) + \dfrac{g}{l}\left(\sin\theta_1 + \sin\theta_2\right) = 0$ and as

$\sin\theta_1 + \sin\theta_1 \neq \sin\left(\theta_1 + \theta_2\right)$, hence it does not satisfy superposition theorem.

3. Using vector method we get the resultant R is

$$R = \left(x_1^2 + x_2^2 + 2x_1x_2 \cos 75°\right)^{1/2} = \left(2a^2 + 2a^2 \cos 75°\right)^{1/2}$$

$$= \sqrt{2}a\left\{1 + \cos 75°\right\}^{1/2} = 1.567a \ \ ... \ (1)$$

In complex form $x_1 = ae^{i\left(\omega t - 30°\right)}$, $x_2 = ae^{i\left(\omega t + 45°\right)}$

So $x = x_1 + x_2 = ae^{i\omega t}\left\{e^{-i\pi/6} + e^{i\pi/4}\right\}$

So the amplitude of resultant

$$R = a\left|\left\{\cos\dfrac{\pi}{6} + \cos\dfrac{\pi}{4} + i\left(\sin\dfrac{\pi}{4} - \sin\dfrac{\pi}{6}\right)\right\}\right|$$



Fig. 4.P-1

$$= a\left|\left(\dfrac{\sqrt{3}}{2} + \dfrac{1}{\sqrt{2}}\right) + i\left(\dfrac{1}{\sqrt{2}} - \dfrac{1}{2}\right)\right| = 1.568a \ \ ... \ (2)$$
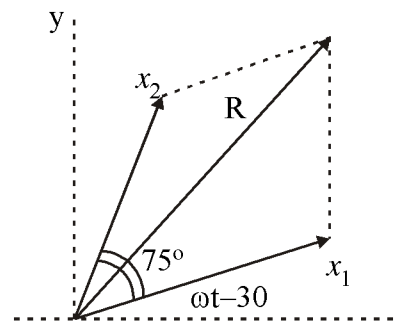
4. Let $x = a\cos\left(\omega t + \dfrac{\pi}{3}\right)$ and $y = a\cos\omega t$ are the two SHMs in mutually

perpendicular direction. Then $\dfrac{x}{a} = \cos 60° - \sin\omega t \sin 60°$

or, $\dfrac{x}{a} = \dfrac{1}{2} \cdot \dfrac{y}{a} - \sqrt{1 - \left(\dfrac{b}{a}\right)^2} \dfrac{\sqrt{3}}{2}$

or, $\left(\dfrac{x}{a} - \dfrac{y}{2a}\right)^2 = \dfrac{3}{4}\left(1 - \dfrac{y^2}{a^2}\right)$

Hence $\dfrac{x^2}{a^2} + \dfrac{y^2}{4a^2} - \dfrac{xy}{a^2} = \dfrac{3}{4} - \dfrac{3y^2}{4a^2}$

or, $\dfrac{x^2}{a^2} + \dfrac{y^2}{a^2} - \dfrac{xy}{a^2} = \dfrac{3}{4}$

or, $x^2 + y^2 - xy = \dfrac{3}{4}a^2$        ... (1)



**Fig. 4.P-2**

This is the equation of an oblique ellipse. Hence the trajectory will be elliptcal.

5. Try yourself.

6. Try yourself.

7. The frequency difference before loading is

$$\Delta \upsilon = \dfrac{1}{20} \text{ or, } \upsilon_A - \upsilon_B = 0.05 \text{ Hz}$$

After loading $\upsilon_B$ and the difference $\upsilon_A - \upsilon_B = \dfrac{1}{10} = 0.1\,\text{Hz}$ increases. This indicates that $\upsilon_A > \upsilon_B$. Thus

$$\upsilon_A - \upsilon_B = 0.05 \text{ Hz}$$

or, $\upsilon_B = \upsilon_A - 0.05$ Hz

$$= (288 - 0.05) \text{ Hz}$$
$$= 287.95 \text{ Hz.}$$

8. The resultant is

$$x = \cos \omega t + 0.5 \cos\left(\omega t + 30°\right) + 0.25 \cos\left(\omega t + 90°\right)$$

$$= \cos \omega t + 0.5\left[\cos \omega t. \dfrac{\sqrt{3}}{2} - \sin \omega t. \dfrac{1}{2}\right] - 0.25 \sin \omega t$$
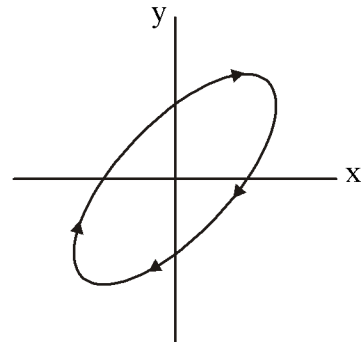
$$= \left(1 + \frac{\sqrt{3}}{4}\right) \cos\omega t - \frac{1}{2}\sin\omega t$$

$$= A\cos(\omega t + \phi)$$

where $= A\cos\phi = 1 + \dfrac{\sqrt{3}}{4} = 1.433$, $A\sin\phi = \dfrac{1}{2} = 0.5$

$\therefore$ For the resultant motion, amplitude $A = \sqrt{(1.433)^2 + (0.5)^2} = 1.52$ cm and the

phase relative to the first is $\phi = \tan^{-1}\dfrac{0.5}{1.433} = 19.23°$.

# Unit-5 : Wave Motion

**Structure**

## 5.0   Objectives

From this unit, you will learn
- What is wave and how it is produced
- characteristic of mechanical wave
- The mathematical formulation of a sinusiodal wave
- The equation of one dimensional, and 3-dimensional progressive wave
- Plane wave and spherical wave solution
- Relation among volumetric strain, particle velocity and group velocity
- Energy distribution in plane progressive sound wave
- Sound intensity presure and loudness levels
- Ripple measurement of water waves

## 5.1  Introduction

A wave may be defined as the fluctuation of some physical quantity with position and time. In case of acoustic waves propagating through any material medium the particles of the medium vibrate about their mean position and the distubance thus created is transmitted from one layer to the next due to elasticity of the medium. It is the disturbance which is transmitted but there is no bodily moment of the medium. For the propagation of an acoustic wave the medium must have inertia and elasticity distributed throughout the medium. If an element of the medium is displaced it also displaces the neighbouring elements due to elasticity. In this way the disturbance is transmitted. The disturbance at a particular position varies with time and at any given instant it varies with position. Thus the disturbance associated with a wave is a function of both position and time.

An acoustic wave way be generated when a vibrating source is coupled to an open elastic medium. The waves travel away from the vibrating source that created them. Such waves are called *travelling* or *progressive waves.* These waves can transport energy and momentum. If the vibrating source oscillates with harmonic motion the progressive waves it produces are called harmonic progressive waves. However, it should be kept in mind that a wave need not be necessarily periodic. In general, there are two types of wave motion.

**Longitudinal waves—** Wave in which the direction of disturbance is parallel to the direction of wave propagation are called *longitudinal waves.*

**Transverse waves—** The waves in which the direction of disturbance in the medium is transverse to the direction of wave propagation are called *transverse waves.*

Sound waves are longitudinal elastic waves. A gaseous medium can sustain only longitudinal waves because transverse waves require a shear force to maintain them. However, both the transverse and longitudinal waves can travel in a solid.

The most common example of a wave is that produced by dropping a stone into the water of a quiet pond. Such waves are not truly periodic and change their shape during propagation. However, for the mathematical description of wave motion we shall start with a periodic wave of *constant type* in which the disturbance varies periodically both in space and time and the wave does not change shape during propagation. The curve showing the variation of displacement of a particle with time *t* at a fixed space-point is called *time-displacement* curve. In the simplest case of harmonic variation the time-displacment curve is sinusoidal (Fig. 5.1). For a wave of constant type all the particles in the path of propagation of the wave will have identical time-displacement curve. Here it should be noted that the word 'particle' to which we

refer is not necessarily an individual molecule but may be very small volume of the medium over which the conditions of displacment, velocity, acceleration, pressure, density etc. may be taken to be the same all over it. A particle away from the source goes through the same motion as that of a particle nearer to the source but at some later time because the wave takes some time in going from one point to another. In a homogenous medium this time or phase lag is proportional to the distance between the partilce.
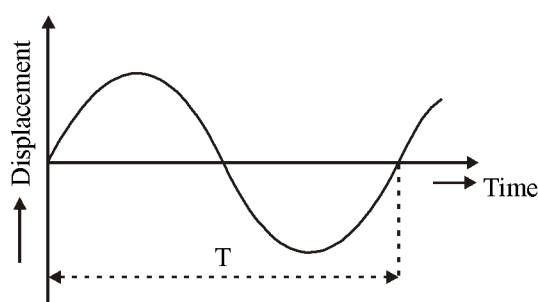


**Fig. 5.1 Time-displacement curve
at a fixed position**
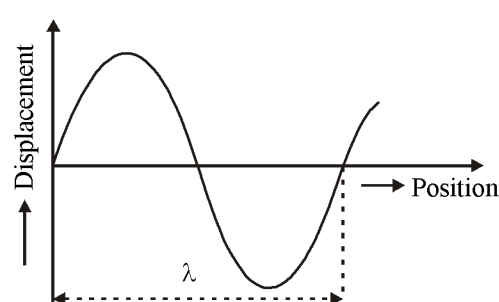


**Fig. 5.2 Spece-displacement curve
at a particular instant of time**

The curve showing the displacement as a function of distance from some suitably chosen point on the line of propagation, at a particular instant of time is called a *space-displacement curve*. It gives the shape or the form of the wave and therefore, is also called *waveform* or *wave profile*. For the simplest case of a harmonic wave it is sinusoidal (Fig. 5.2)

A study of the time-displacement curve shows that a given particle on the line of propagation undergees a complete cycle of vibration at equal intervals of time, called *time period (T)*. This is the time required for the generation of a complete wave. The *frequency* (v) of a wave is defined as the number of complete waves generated per second and is given by $v = 1/T$. The unit of frequency is SI system is hertz (Hz). The space displacement curve shows that particles separated by a certain distance or an integral multiple of this distance are in the same phase i.e., in the same state of motion. The minimum distance between two particles along the line of propagation having the same phase of motion is called the *wavelength* ($\lambda$).

The particles of the medium do not move bodily. They merely vibrate about their mean position whereas the disturbances created are transmitted through the medium with a velocity determined by the properties of the medium. When the waves are harmonic progressive waves, this velocity is called the *phase velocity (c)*.

If we imagine a surface at all points of which the particle of the medium at any given instant are in the same phase of motion, then the surface is called a *wavefront*.

For a point source in a homogeneous and isotropic medium the wavefronts are concentric speherical surfaces. Such waves are called *spherical waves.* For a long line source wavefronts are cylindrical. If the wavefronts are parallel planes we call it *plane waves.* In this case the waves move only in a fixed direction without spreading laterally. A small part of a wavefront at a large distance from the source may be considered to be plane Students shouold go through Article 8.4, Unit-8 for definition of wavefront and their properties.

In the mathematical description of a wave motion we are to relate the phase difference between particles to their physical separation in space. Let us consider one dimensional plane progressive harmonic waves travelling in a medium with a phase velocity $c$. Consider two planes $A_1$ and $A_2$ separated by a distance $x$ as shown in Fig. 5.3. Let $\psi(x, t)$ be the *field parameter* which may the displacement, velocity, presure or

**Fig. 5.3  Propagation of disturbance $\psi(x, t)$**

any other physical paramenter that varies with space and time like a wave. Suppose the harmonic oscillations of the particles at the plane $A_1(x = 0)$ at any time instant $t$ is given by

$$\psi (0, t) = a \sin \omega t$$

where $a$ is the amplitude and $\omega$ is the angular frequency. This disturbance propagates throught the medium with a phase velocity $c$. So the disturbance at $x = 0$ takes a time $x/c$ to reach the point $x$ on the plane $A_2$. Now the displacement $\psi(x,t)$ of a particle at the position $x$ on the plane $A_2$ at any time $t$ will be the same which another particle at $x = 0$ had at same earlier time $t' = t - x/c$. Therefore, for a *plane progressive harmonic wave* we have

$$\psi (x, t) = \psi (0, t')$$
$$= a \sin \omega t'$$
$$= a \sin \omega (t - x/c)$$
$$= a \sin (\omega t - kx) \qquad \qquad ...(5.1.1)$$
$$= a \sin 2\pi / \lambda (ct - x) \qquad \qquad ...(5.1.2)$$

where $k = \omega c = 2\pi / T.v.\lambda = 2\pi / \lambda$ is known as wave number.

Note that at a fixed position $x$, $y$ varies harmonically in time $t$ with a period T. At a fixed instant of time, $\psi$ varies harmonically in space with a period $\lambda$, For a wave

travelling in the negative x-direction the corresponding equation is

$$\psi(x,\ t) = a\sin(\omega t + kx) \qquad\qquad ...(5.1.3)$$

## 5.2 Wave Equation

We now develope a differential equation which governs the propagation of one dimensional waves. Let us consider the equation representing a plane progressive harmonic wave

$$\psi(x,\ t) = a\sin(\omega t - kx) \qquad\qquad ...(5.2.1)$$

Differentiating twice with respect to $t$ we get

$$\frac{\partial^2\psi}{\partial t^2} = -\omega^2 a\sin(\omega t - kx) \qquad\qquad ...(5.2.2)$$

Differentiating Eqn. (5.2.1) twice with respect of $x$ we get

$$\frac{\partial^2\psi}{\partial x^2} = -k^2 a\sin(\omega t - kx) \qquad\qquad ...(5.2.3)$$

Combining Eqs. (5.2.2) and (5.2.3) we get

$$\frac{\partial^2\psi}{\partial t^2} = c^2\frac{\partial^2\psi}{\partial x^2} \qquad\qquad ...(5.2.4)$$

where $c = \omega/k$ is the velocity of the wave. Eqn. (5.2.4) is the well-known one dimensional differential *wave equation*.

## 5.3 Plane wave solution

The Eqn. (5.2.4) can be solved by the method of separation of variables.

Let $\qquad\quad \psi(x,\ t) = F_1(x).F_2(t) \qquad\qquad ...(5.3.1)$

where $F_1$ is a function of $x$ only $F_2$ is a function of $t$ only. Now substituting (5.3.1) in Eqn. (5.2.4) and dividing throughout by $F_1.\ F_2$ we get

$$\frac{1}{F_2}\cdot\frac{d^2F_2}{dt^2} = \frac{c^2}{F_1}\cdot\frac{d^2F_1}{dx^2} \qquad\qquad ...(5.3.2)$$

Left hand side of this equation is a function of $t$ only and right hand side is a function of $x$ only. So for the validity of Eqn. (5.3.2) each side must be equal to some

constant $- \omega^2$ (say). We choose the negative sign in order to have a periodic solution. thus we get

$$\frac{d^2F_1}{dx^2} + k^2F_1 = 0 \qquad (5.3.3)$$

$$\frac{d^2F_2}{dt^2} + \omega^2F_2 = 0 \qquad \ldots (5.3.4)$$

where $k = \omega/c$.

Solutions to these equations are of the form

$$F_1(x) = A_1e^{ikx} + A_2e^{-ikx}$$

$$F_2(t) = B_1e^{i\omega t} + B_2e^{-i\omega t}$$

Thus

$$y(x, t) = F_1(x).F_2(t)$$

$$= \left(A_1e^{ikx} + A_2e^{-ikx}\right)\left(B_1e^{i\omega x} + B_2e^{-i\omega t}\right)$$

$$= A_1B_1e^{ik(ct+x)} + A_2B_2e^{-ik(ct+x)} + A_1B_2e^{-ik(ct-x)} + A_2B_1e^{ik(ct-x)} \quad \ldots (5.3.5)$$

Thus the solution can be put in the form

$$\psi(x,t) = f_1(ct - x) + f_2(ct + x) \qquad \ldots (5.3.6)$$

Theoretically, the functions $f_1$ and $f_2$ can have any form but for correct representation of a wave we must choose a proper form. For example, a plane progressive harmonic wave travelling along +ve x-axis can be represented by

$$\psi(x,t) = a\sin k(ct - x) = a\sin(\omega t - kx)$$

or, $\qquad \psi(x,t) = a\cos k(ct - x) = a\cos(\omega t - kx)$

Similarly a plane progressive harmonic wave travelling along –ve x-axis can be represented by

$$\psi(x,t) = a\sin(\omega t + kx)$$

or, $\qquad \psi(x,t) = a\cos(\omega t + kx)$

## 5.4  Spherical wave solution

The three dimensional wave equation is

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \nabla^2 \psi \qquad\qquad \ldots (5.4.1)$$

where, in general, the field paramenter $\psi = \psi$ $(r,\ \theta,\ \phi,\ t)$. if the waves have spherical symmetry then $\psi$ will be independent of the angular coordinates $\theta$ and $\phi$ of a spherical coordinate system. In this case Eqn. (5.4.1) reduces to

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \cdot \frac{1}{r^2}\frac{\partial}{\partial r}\left( r^2 \frac{\partial \psi}{\partial r}\right)$$

Putting $\psi = \psi'/r$

$$\frac{\partial^2}{\partial t^2}\left(\frac{\psi'}{r}\right) = c^2 \cdot \frac{1}{r^2}\frac{\partial}{\partial r}\left[ r^2 \frac{\partial}{\partial r}\left(\frac{\psi'}{r}\right)\right]$$

or, $\quad \dfrac{1}{r}\dfrac{\partial^2 \psi'}{\partial t^2} = \dfrac{c^2}{r^2}\dfrac{\partial}{\partial r}\left[ r\dfrac{\partial \psi'}{\partial r} - \psi'\right] = c^2 \cdot \dfrac{1}{r}\dfrac{\partial^2 \psi'}{\partial r^2}$

or, $\quad \dfrac{\partial^2 \psi'}{\partial r^2} = c^2 \dfrac{\partial^2 \psi'}{\partial r^2}$

This equation is of the same form as the one dimensonal wave equation (5.2.4). Hence its general solution is, therefore,

$$\psi' = f_1(ct - r) + f_1(ct + r)$$

or, $\quad \psi = \dfrac{1}{r}f_1(ct - r) + \dfrac{1}{r}f_2(ct + r) \qquad\qquad \ldots (5.4.2)$

The first term of Eqn. (5.4.2) represents a spherical wave diverging from the origin of coordinates with a velocity c. the second terms represents a similar wave converging to the origin. A diverging harmonic spherical wave can be represented by

$$\psi(r.t) = A.\frac{1}{r}\cos(\omega t - kr) + B.\frac{1}{2}\sin(\omega t - kr) \qquad\qquad \ldots (5.4.3)$$

In exponential form,

$$\psi(r.t) = \frac{A}{r}\exp\left[ i(kr \pm \omega t)\right]$$

## 5.5  Volumetric strain, particle velocity and wave velocity

**Volumetric strain :** Let a wave be set up in a cylinder of cross-section unity the axis of the cylinder coinciding with the diredtion of propagation ACE of the

Wave (Fig. 5.4) Let *ABCDE* represent the displacement curve of the wave–the waveform.

Consider *P*, *Q* to be two close points *dx* apart. The displacement at *P* = *PR* and that at *Q* = *QS*. So the volume of the gas limited by *P* and *Q* is *PQ* × 1 = *PQ*. When the wave is in the position as shown in Fig. 5.4 the change in volume between *P* and *Q* is

$$QS - PR = ST = d\psi$$

∴    Volume strain $= \dfrac{\text{change in volume}}{\text{original volume}}$

$$= \frac{ST}{PQ} = \frac{ST}{RT} = \frac{d\psi}{dx}$$
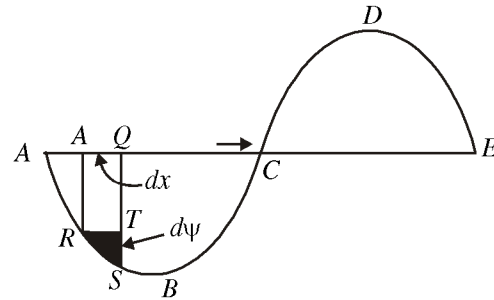
**Fig. 5.4**

Volumetric strain = Slope of the displacement curve.

Thus the volumetric strain at a given point is measured by the slope of the displacement curve at that point.

**Particle velocity—** Let the equation of the wave be given as

$$\psi = a\sin\frac{2\pi}{\lambda}(ct - x) \qquad\qquad ...(5.5.1)$$

Differentiating (5.5.1) with respect to time. We obtain

Particle Velocity $\quad = \dfrac{d\psi}{dt} = a\dfrac{2\pi c}{\lambda}\cos\dfrac{2\pi}{\lambda}(ct - a) \qquad\qquad ...(5.5.2)$

Differentiating (5.5.1) with respect to *x*, we get

Volume strain $\quad = \dfrac{d\psi}{dx} = -a\dfrac{2\pi}{\lambda}\cos\dfrac{2\pi}{\lambda}(ct - x) \qquad\qquad (5.5.3)$

If *d*ψ/*dx* is positive a rarefaction occurs; if however *d*ψ/*dx* is *negative, a compression* takes place.

Comparing (5.5.2) and (5.5.3), we obtain

$$= \frac{d\psi}{dt} = -c\frac{d\psi}{dx}$$

∴  Particle velocity = – wave velocity × volume strain.

This is an important relation between volume strain, particle velocity and the wave velocity.

## 5.6 Energy distribution in plane progressive sound waves

When sound waves pass through any medium the particles of the medium are set into oscillation, so these particles now possess extra amount of energy which is partly kinetic and partly potential in nature. The total average energy per unit volume of the medium is called *energy density*.

Suppose a plane, harmonic longitudinal wave is moving through a medium along +*ve* x-direction. The displacement of a particle at the position *x* at any time *t* may be represented by

$$\psi(x,t) = a\cos(\omega t - kx) \qquad \dots (5.6.1)$$

where $\omega$ is the angular frequency, $k$ the propagation constant and a is the amplitude of vibration. Let us now consider an infinitesimally thin layer of thickness $\delta x$ and and of unit cross-section. The layer is considered so thin that all the particles within the layer may be assumed to possess the same velocity $\partial\psi/\partial t$. The mass of this layer is $\rho_0\delta x$ where $\rho_0$ is the equilibrium density. Therefore, kinetic energy of the layer is

$$\Delta E_k = \frac{1}{2}(\rho_0\delta x)\left(\frac{\partial\psi}{\partial t}\right)^2 \qquad \dots (5.6.2)$$

Using Eqn. (5.6.1) we get the kinetic energy per unit volume as

$$E_k = \frac{1}{2}\rho_0 a^2\omega^2 \sin^2(\omega t - kx) \qquad \dots (5.6.3)$$

Now the time average value of $\sin^2(\omega t - kx)$ over complete time period (or its space average value over a complete wavelength $\lambda$) is 1/2. Thus average kinetic energy per unit volume becomes

$$\bar{E}_k = \frac{1}{2}\rho_0 a^2\omega^2 \qquad \dots (5.6.4)$$

During the transmission of the acoustic wave the voluem element is periodically compressed and rarefied. So there is change in potential energy associated with this volume change. Suppose an increase of pressure from $P_0$ to $P_0 + p$ compresses unit volume of the medium to a volume (1 – s) where $s = -\frac{\partial\psi}{\partial x}$ is the condensation. The potential energy associated with this volume is equal to the work done in the process of compression. Thus potentrial energy per unit volume is given by

$$E_p = \int_0^s pds \qquad \dots (5.6.5)$$

Now by difinition bulk modulus

$$B = \frac{p}{s} \quad \text{or, } p = Bs$$

Therefore, $\quad E_p = B\int_0^s s\,ds = \frac{1}{2}Bs^2$

$$= \frac{1}{2}B\left(-\frac{\partial\psi}{\partial x}\right)^2 \qquad \qquad \dots (5.6.6)$$

Using Eqn. (5.6.1) we get

$$E_p = \frac{1}{2}Ba^2k^2\sin^2\left(\omega t - kx\right) \qquad \qquad \dots (5.6.7)$$

Now the time or space avewrage value of $\sin^2(\omega t - kx)$ is 1/2. Therefore the average value of potential energy per unit volume becomes

$$\overline{E}_p = \frac{1}{2}Ba^2k^2$$

$$= \frac{1}{4}\rho_0 a^2\omega^2 \quad \left[\because c = \sqrt{B/\rho_0} = \omega/k\right] \qquad \dots (5.6.8)$$

$\therefore$ Energy density is

$$\overline{E} = \overline{E}_k + \overline{E}_p = \frac{1}{2}\rho_0 a^2\omega^2 \qquad \qquad \dots (5.6.9)$$

Thus we find that the average values of the kinetic energy and potential energy densities in the sound wave are equal. Another interesting feature is that unlike the case of simple vibrations here the instantaneous values of the kinetic and potential energies are in phase [Eqs. (5.6.3) and (5.6.7)] i.e., both of them reach maximum value or the zero value at the same instant of time. A compression or rarefaction produces a maximum in the energy of any element of volume. So the energy in the wave is distributed in



Fig. 5.5 Energy distribution in space for a sound wave in a fluid

the wave system with distance as shown in Fig. 5.5. The regions of maximum energy alternate with regions of little or no energy.

## 5.7  Intensity of sound waves

The intensity of sound or acoustic intensity at a point is defined as the average rate of flow of energy through a unit area surrounding the point held normal to the direction of wave propagation. If $\delta E$ is the average energy crossing an area $\delta S$ nornally per unit time then intensity at a point is given by

$$I = \lim_{\delta S \to 0} \frac{\delta E}{\delta S}$$

For a plane progressive wave through a homogenoues isotropic medium the energy that crosses unit area in unit time in the direction of propagation of the wave equation of propagation of the wave equals the amount of energy contained in a cylinder of unit cross-section and length c. The volume of this cylinder is c. Therefore,
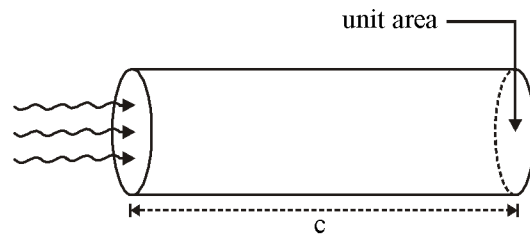


Fig. 5.6 : To calculate intensity

$$1 = \text{average energy density} \times c$$

$$= \left( \overline{E}_k + \overline{E}_p \right) \times c$$

$$= \frac{1}{2} \rho_0 c a^2 \omega^2 \qquad \qquad \text{... (5.7.1)}$$

where we have used the Eqn. (5.6.4) and (5.6.8). This relation shows that the intensity is proportional to the square of the amplitude a.

Most of the modern acoustic detectors respond to the variation of acoustic pressure rather than to the intensity. So it is useful to express $I$ in terms of acoustic pressure p which is given by

$$p = \left( -\partial \psi / \partial x \right).B \qquad \qquad \text{where } \psi = a \cos (\omega t - kx)$$

$$\therefore \quad p = -Bak \sin \left( \omega t - kx \right) \qquad B = \text{Bulk modulus of the gaseous medium}$$

$$= -\rho_0 a \omega c \sin \left( \omega t - kx \right) \qquad \left[ \because c = \sqrt{B / \rho_0} = \omega / k \right]$$

$$= \rho_0 \cos \left( \omega t - kx + \pi / 2 \right) \qquad \qquad \text{... (5.7.2)}$$

where peak acoustic pressure $p_0 = \rho_0.a\omega c$

The r.m.s accustic pressure $p_0$ is given by

$$p^2_{r.m.s} = \frac{1}{T}\int_0^T p^2 dt$$

A simple integration using Eqn. (5.7.2) shows that

$$p_{r.m.s} = \frac{p_0}{\sqrt{2}} = \frac{\rho_0 a\omega c}{\sqrt{2}} \qquad \qquad ... (5.7.3)$$

Hence Eqn. (5.7.1) can be expressed as

$$I = \frac{p^2_{r.m.s}}{\rho_0 c} \qquad \qquad ... (5.7.4)$$

## 5.8  Specific acoustic impedance

While discussing the propagation of plane waves through a modium certain analogies may be drawn between some acoustic quantities and some electrical quantities in the electrical circuit theory. If the displacement is represented by $\psi = a\cos(\omega t - kx)$ then particle velocity $u = \partial\psi / \partial t = -a\omega\sin(\omega t - kx)$

and acoustic pressure

$$p = -B\frac{\partial\psi}{dx} = -Bak\sin(\omega t - kx)$$

Therefore, $\dfrac{\text{acoustic pressure}}{\text{particle velocity}} = \dfrac{p}{u} = \dfrac{Bk}{\omega} = \rho_0 c \quad \left[\because c = \sqrt{B/\rho_0} = \omega/k\right].$

If we take acoustic pressure to be analogous to e.m.f. and particle velocity to electric current then their ratio will be analogous to electrical impedance. Thus the ratio $p/u$ is the *acoustic impedance* presented by the medium to the acoustic waves. The product $\rho_0 c$ is found to have greater significance as a characteristic property of the medium than $\rho_0$ or $c$ alone. For this $\rho_0 c$ is called *characteristic impedance* of the medium. the ratio $p/u$ is also called *specific accoustic impedance.* It is to be noted that we speak of impedance between two points in an electrical circuit, but the acoustic impedance is a point property. For plane progressive waves the specific acoustic impedance is a real quantity but for speherical waves or standing plane waves it has also a reactive component. In electrical circuit the product e.m.f $\times$ current represents instantaneous power in ac curcuits. Similarly we can show that the product $pu$ represents the instantaneous acoustical power,

$$pu = Ba^2\omega k\sin^2(\omega t - kx) = \rho_0 ca^2\omega^2\sin^2(\omega t - kx)$$

*Its average value is*

$$\overrightarrow{pu} = \frac{1}{2}\rho_0 ca^2\omega^2$$

$$= I \text{ (acoustic intensity)}$$

$$= \text{Energy flow per unit area per second}$$

## 5.9 Sound intensity, pressure and loudness levels

In theoretical study absolute intensity of sound is expressed in units of $W/m^2$. But in practical applications it is customary to express sound intensities in terms of some standard intensity by using a longarithmic scale. One reason for doint this is the variation of sound intensities over a wide range (approximately $10^{-12}$ to $10$ W/m$^2$) in our audible acoustical environment. The other reason is that our ear does not respond linearly to sound intensities. The sensation of hearing depends on the logarithim of intensities. The commonly used standard intensity for air borne sounds is $10^{-12}$ W/m$^2$ which approximately corresponds to the lower limit to intensity for audibility in air at a frequency of 1 kHz. The *intensity level (IL)* of a sound of intensity $I$ is defined by

$$\text{IL} = 10 \log_{10}(I/I_0) \qquad \qquad \text{... (5.9.1)}$$

where IL is expresed in decibets (dB) and $I_0$ is the standard reference intensity. Bel is a larger unit of IL. 1 Bel = 10 decibels. Nowadays the noise pollution level is expressed in units of decibels.

Most of the modern sound detectors like headphones, loudspeakers respond to the changes in acoustic pressure rather than to the intensity. Hence it would be more useful to exprese Eq. (5.9.1) in tems of sound pressures. Since $I = p_{rms}^2 / \rho_0 c$, we may define *sound pressure level (SPL)* in decibels by the equation.

$$SPL = 20\log_{10}\frac{P_{rms}}{p_0} \qquad \qquad \text{... (5.9.2)}$$

where $p_0$ is effective pressure of a standard sound. The commonly used reference pressure for specifying SPL in air is $2 \times 10^{-5}$ N/m$^2$. It corresponds almost exactly to $I_0 = 10^{-12}$ W/m$^2$. Hence almost identical numerical values are obtained by using either of the above two equations for sound waves in air.

For underwater measurements it is the SPL and not the IL which is most commonly used. Two standard pressure levels ($p_0$) are used for specifying SPL in water. $p_0$ is

taken as $2 \times 10^{-5}$ N/m$^2$ to express underwater noise levels and as 0.1 N/m$^2$ for calibration and investigation of the acoustic characteristics of sonar transducers and hydrophones.

Loudness of sound is the degree of auditory sensation produced by sound waves reaching the ear. It depends not only on the intensity of the sound but also on the frequency and sensitiveness of the ear. Loudness is a physiological quantity rather than a physical quantity Loudness depends on mental interpretation and hence cannot be measured exactly with an instrument. There is no absolute scale for the measurement of loudness. Our ear cannot make a quantitative assessment of the loudness. However, it has been found that there is fair agreement between observers as to when a pure tone of one frequency appears as loud as a tone of another frequency. Therefore, it is possible to make a quantitative measurement of loudness by comparing the sound under test with some standard sound. The standard is chosen to be a tone of frequency 1 kHz and of acoustic presure $2 \times 10^{-5}$ N/m$^2$. The unit of *loudness level (LL)* is *phon*. The intensity level of the standard tone is raised until it appears equally loud with the sound under test, then the rise of the intensity level of the standard tone in DB is said to be the equivalent loudness of the sound in number of phons.

## 5.10 Water waves : Ripple meaurement

A ripple tank can be used to measure and calculate frequency, wavelength, and the speed of waves on the surface of the water. A ripple tank is a transparent shallow tray of water with a light shining down through it onto a white card below in order to clearly see the motion of the ripples created on the water's surface. Ripples can be made by hand but to generate regular ripples it is better to use a motor.

**Aim of the experiment**

To measure the frequency, wavelength, and speed of waves in a ripple tank.

**Method**

1. Set up the ripple tank as shown in Fig. 5.7 with about 5 cm depth of water.
2. Adjust the height of the wooden rod so that it just touches the surface of the water.
3. Switch on the lamp and motor and adjust until low frequency waves can be clearly observed.
4. Measure the length of a numebr of waves then divide by the number of waves to record wavelength. It may be more practical to take a photograph of the card with the ruler and take measurements from the still picture.
5. Count the numebr of waves passing a point in tens seconds then divide by ten to record frequency.

6. Calculate the speed of the waves using : wave speed = frequency × wavelength.
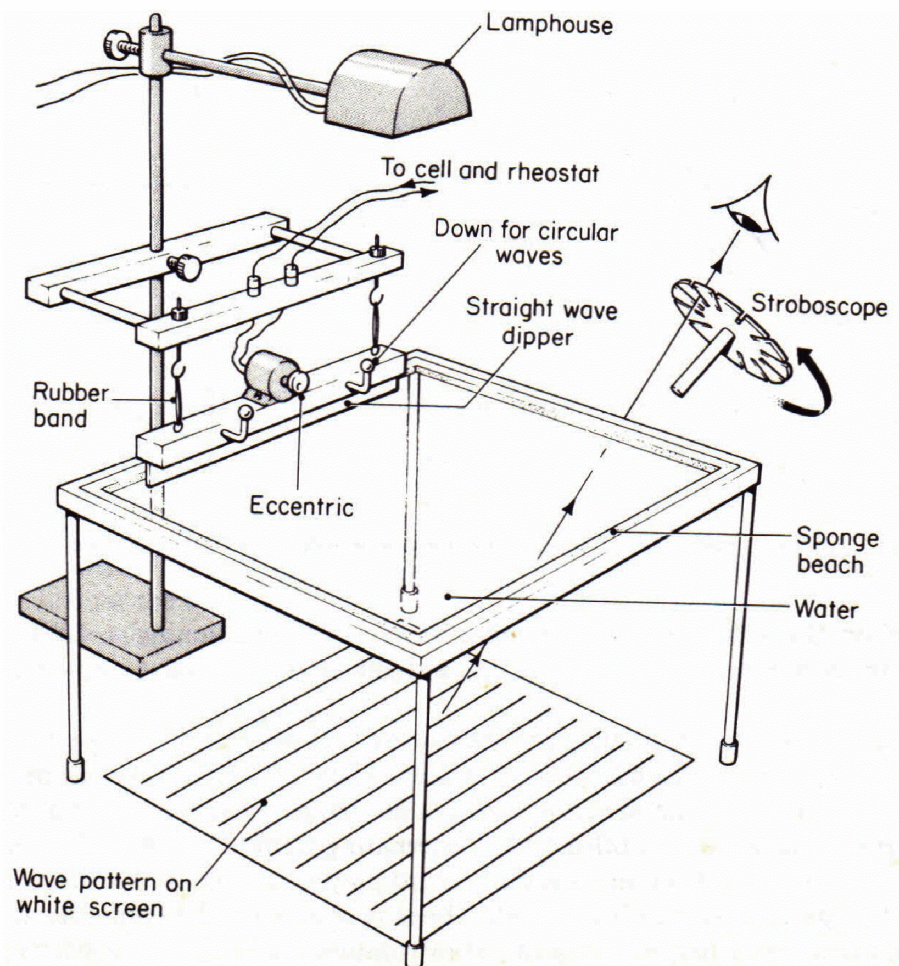


**Fig. 5.7 : The ripple tank**

## 5.11 Summary

1. For a sinusoidal wave, the displacement is given by

$$\psi = a \cos\left[kx \pm \omega t + \phi\right]$$

where $a$ represents the amplitude of the wave. $\omega(= 2\pi v)$ the angular frequency of the wave, $k (= 2\pi/\lambda)$ the wave number and $\lambda$ represnets the wavelength associated with the wave. The upper and lower signs correspond to waves propagating in the $-x$ and $+x$ directions respectively. Such a displacement is indeed produced in

a long stretched string at the end of which a continuously vibrating tuning fork is placed. The quantity $\phi$ is known as the phase of the wave.

2. The most general solution of the wave equation

$$\frac{\partial \psi}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}$$

is of the form

$$\psi = f(x - vt) + g(x + vt)$$

where $f$ and $g$ are arbitrary functions of their argument. The first term on the RHS of the above equation represents a disturbance propagating in the $+x$ direction with speed $v$ and similarly, the second term represents a disturbance propagating in the $-x$ direction with speed $v$. Thus if we can derive the wave equation from physical considerations, then we can be sure that waves will result and $\psi$ will represent the displacement associated with the wave.

3. For a sperical wave the displacment is given by

$$\psi = \frac{A}{r} e^{i(kr \pm \omega t)}$$

where the $+$ and $-$ signs correspond to incoming and outgoing waves respectively. Notice that the factor $1/r$ term implies that the amplitude of a spherical wave decreases inversely with r, and therefor the intensity will fall off as $1/r^2$.

4. For a 3–D wave we have $\psi(r,t) = A\sin\left(\omega t - \vec{k}.\vec{r}\right)$ and wave equation is

$$\frac{\partial^2 \psi(r,t)}{\partial t^2} = v^2 \nabla^2 \psi(r,t).$$

5. Particle velocity = – wave velocity × volume strain.

6. Energy density $= \frac{1}{2}\rho_0 a^2 \omega^2$

7. Intensity of sound wave $(I) = \frac{1}{2}\rho_0 v a^2 \omega^2$

## 5.12 Review Questions and Answers

1. State two disfference between a plane wave and a spherical wave.

2. Solve the equation $\dfrac{\partial^2 \psi}{\partial x^2} = \dfrac{1}{c^2} \dfrac{\partial^2 \psi}{\partial t^2}$ to obtain an expression of $\psi(x, t)$.

3. Prove that for one dimensional progressive wave the average kinetic energy and potential energy is always same.

4. If the velocity of sound in air is 340 m/sec. and the frequency of a progressive wave is 255 Hertz, then how many complete waves will be in a distance of 60 meter.

5. Show that the function $\psi(x,t) = f_1(ct - x) + f_2(ct + x)$ satisfies the one dimensional differential wave equation.

6. Consider the one dimensional wave equation $\dfrac{\partial^2 \psi}{\partial t^2} = c^2 \dfrac{\partial^2 \psi}{\partial t}$. Find the travelling wave solution satisfying the initial conditions $y(x, 0) = g(x)$ and $\psi(x, 0) = 0$.

7. Show that the acoustic pressure, condensation, dilatation and particle velocity all satisfy the differential wave equation.

8. Compare the intensities of sound in air and water at N.T.P. for the same acoustic pressure. Given the density of air $\rho_a = 1.21$ kg/m$^3$, density of water $\rho_\omega = 998$ kg/m$^3$, the speed of sound in air $c_a = 343$ m/s and the speed of sound in water $c_\omega = 1480$ m/s.

9. A source of sound emits energy isotropically in all directions at the rate of 1 J/s. Determine the intensity level at a distance of 10m from the source. Take the reference as intensity $10^{-12}$ W/m$^2$.

10. An air-conditioning unit operates with a sound intensity level of 73 dB. If it is operated in an office with surrounding intensity level of 68 dB, what will be the resultant intensity level?

11. Find the sound energy density in air of a free plane progressive acoustic wave having an IL of 80 dB. Given the velocity of sound in air = 343 m/s and reference intensity $I_0 = 10^{-12}$ W/m$^2$.

12. The intensity of a sound is $10^{-2}$ W/m$^2$ at a frequency of 1 kHz. What is its loudness level? Take the reference intensity level = $10^{-12}$ W/m$^2$.

13. An intense plane harmonic wave in air at STP has an r.m.s. excess pressure of 20 Nm$^2$. The frequency is 1 kHz and the phase velocity 350 m/s. Calculate the maximum value of particle displacement and particle velocity. Given $\gamma_{air} = 1.40$.

## ANSWER

1.  $y_1 = A\sin(\omega t - kx)$ represents plane wave but $y_2 = \dfrac{A}{r}\sin(\omega t - kr)$ represents spherical wave.

    ● The wavefront is a plane surface for plane wave but the wavefront is a spherical surface.

    ● The amplitude remains constant at all points for plane wave but the amplitude varies inversely with distance from source.

2.  $\dfrac{\partial^2 \psi}{\partial x^2} = \dfrac{1}{c^2}\dfrac{\partial^2 \psi}{\partial t^2}$ ....(1), Let $\psi(x,t) = X(x)f(t)$, where $X(x)$ is a function of $x$ only, $f(t)$ is a function of $t$ only.

    Hence $\psi = X(x)f(t)$ and from equation (1) we get

    $$f(t)\frac{d^2 X}{dx^2} = \frac{1}{c^2}X(x)\frac{d^2 f}{dt^2}$$

    or, $\quad \dfrac{1}{X(x)}\dfrac{d^2 X}{dx^2} = \dfrac{1}{c^2}\dfrac{1}{f(t)}\dfrac{d^2 f}{dt^2} = -k^2$ say.

    $\therefore$ We get $\quad \dfrac{d^2 X}{dx^2} + k^2 x = 0 \qquad$ ... (2)

    $$\frac{d^2 f}{dt^2} + k^2 c^2 f(t) = 0 \quad ... (3)$$

    Hence $X(x) = A_1 e^{\pm ikx}$ and $f(t) = A_2 e^{\pm i\omega t}$, where $\omega = kc$.

    This $\psi(x,t) = Ae^{i(\omega t \pm kx)}$ or, $\psi = A\cos(\omega t \pm kx)$ etc.

3.  Consider one dimensional progressive wave $y = a\sin(\omega t - kx)$

    This $\quad v_p = \dfrac{dy}{dt} = \omega a\cos(\omega t - kx)$ and $KE = \dfrac{1}{2}mv_p^2$

    $$KE = \frac{1}{2}m\omega^2 a^2 \cos^2(\omega t - kx)$$

∴    Average KE is $\langle KE \rangle = \frac{1}{2} m\omega^2 a^2 \left\langle \cos^2 (\omega t - kx) \right\rangle$

$$= \frac{1}{4} m\omega^2 a^2 \quad ...(1)$$

Now Force on particle is  $F = mf = m\omega^2 y$. as  $f = \frac{d^2 y}{dt^2}$  = acceleration

∴ PE $= \int_0^y F.dy = \int_0^y m\omega^2 y \, dy = \frac{1}{2} m\omega^2 y^2 = \frac{1}{2} m\omega^2 a^2 \sin^2 (\omega t - kx)$

Hence average PE is   $\langle PE \rangle \quad = \frac{1}{2} m\omega^2 a^2 \left\langle \sin^2 (\omega t - kx) \right\rangle$

$$= \frac{1}{4} m\omega^2 a^2 \quad ...(2)$$

Hence $\langle KE \rangle = \langle PE \rangle$.

4.   The number of complete waves is $\frac{L}{\lambda} = \frac{L}{\vartheta/\lambda} = \frac{L\lambda}{\vartheta}$.

5.   We have       $\frac{\partial \psi}{\partial x} = -f_1'(ct - x) + f_2'(ct + x)$

$$\frac{\partial^2 \psi}{\partial x^2} = f_1''(ct - x) + f_2''(ct + x) \qquad\qquad ...(1)$$

where  $f'(ct \pm x)$  indicate differentiation with respect the argument.

Again       $\frac{\partial \psi}{\partial t} = cf_1'(ct - x) + cf_2'(ct + x)$

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \left[ f_1''(ct - x) + f_2''(ct + x) \right] \qquad\qquad ...(2)$$

Combining (1) and (2) we get

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2}$$

Thus $\psi$ satisfies the differential wave equation.

6.  Let the solution be

$$\psi(x,t) = f_1(x-ct) + f_2(x-ct)$$

where $f_1$ and $f_2$ are two arbitrary functions.
Now using the initial conditions we get

$$\psi(x,0) = f_1(x) + f_2(x) = g(x) \qquad \qquad \text{... (1)}$$

and $\qquad \qquad \psi'(x,0) = -cf_1'(x) + cf_2'(x) = 0 \qquad \qquad \text{... (2)}$

From Equation (2),

$$f_1'(x) = f_2'(x)$$

Which on integration gives,

$$f_1(x) = f_2(x) + c_1 \qquad \qquad \text{... (3)}$$

Solving Equation (1) and (3) we get

$$f_1(x) = \frac{1}{2}\big[g(x) + c_1\big], \quad f_2(x) = \frac{1}{2}\big[g(x) - c_1\big]$$

Therefore the required travelling wave solution is

$$\psi(x,t) = \frac{1}{2}\big[g(x-ct) + c_1\big] + \frac{1}{2}\big[g(x+ct) - c_1\big]$$

$$= \frac{1}{2}g(x-ct) + \frac{1}{2}g(x+ct)$$

7.  If $\psi(x,t)$ be the displacement at any position $x$ at any time $t$ then acoustic pressure is given by $p = -B\left(\dfrac{\partial \psi}{\partial x}\right)$ where B is the bulk modulus.
Now $\psi$ satisfies the wave equation

$$\frac{\partial^2 \psi}{\partial t^2} = c^2 \frac{\partial^2 \psi}{\partial x^2} \qquad \qquad \text{... (1)}$$

We assume $\psi$ to be a well-behaved function with differentiable derivatives of higher order. Now differentiating (1) with respect to $x$ we can write

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial \psi}{\partial x}\right) = c^2 \frac{\partial^2}{\partial x^2}\left(\frac{\partial \psi}{\partial x}\right)$$

Multiplying both sides by $-B$ and using $p = -B\left(\dfrac{\partial \psi}{\partial x}\right)$, we get

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial t}$$

Similarly we can show that condensation $s = \dfrac{\partial \psi}{\partial x}$, dilatation $\Delta = -\dfrac{\partial \psi}{\partial x}$ and particle

velocity $u = \dfrac{\partial \psi}{\partial t}$ all satisfy the wave equation.

8. Intensity $I = \dfrac{p_{rms}^2}{\rho c}$

$$\therefore \qquad \frac{I_{air}}{I_{water}} = \frac{p_{rms}^2 / (\rho c)_{air}}{p_{rms}^2 / (\rho c)_{water}} = \frac{\rho_\omega}{\rho_a} \times \frac{c_\omega}{c_a} = \frac{998}{1.21} \times \frac{1480}{343} = 3559$$

9. The sound intensity at 10 m from the source is

$$I = \frac{\text{power radiated}}{\text{area}} = \frac{1}{4\pi(10)} \quad \text{W/m}^2$$

Therefore $\quad IL = 10\log_{10}\dfrac{I}{I_0} = 10\log_{10}\dfrac{1}{4\pi(10)^2 \times 10^{-12}}$ dB = 89 dB

10. $\qquad\qquad (IL)_1 = 10\log_{10}\dfrac{I_1}{I_0} = 73$ dB $\quad$ or, $I_1 = 10^{7.3} I_0$

$$(IL)_2 = 10\log_{10}\frac{I_2}{I_0} = 68 \text{ dB} \quad \text{or, } I_2 = 10^{6.3} I_0$$

$\therefore$ The total absolute intensity $I = I_1 + I_2 = \left(10^{7.3} + 10^{6.8}\right) I_0$

$\therefore$ Resultant $IL = 10\log_{10}\dfrac{I}{I_0} = 10\log_{10}\left(10^{7.3} + 10^{6.8}\right)$ dB = 74.19 dB

11. $IL = 10\log_{10}\dfrac{I}{I_0} = 80$ dB $\quad$ or, $I = 10^8 I_0 = 10^8 \times 10^{-12}$ W/m$^2$ = $10^{-4}$ W/m$^2$.

Now energy density $= \dfrac{I}{c} = \dfrac{10^{-4}}{343}$ J/m$^3$ = $2.9 \times 10^{-7}$ J/m$^3$.

12. Loudness level in number of phons = IL in dB

$$= 10 \log_{10} \frac{I}{I_0} = 10 \log_{10} \frac{10^{-2}}{10^{-12}} = 100$$

13. Let displacement $\psi = a \cos(\omega t - kx)$

Excess presure $p = -B \frac{\partial \psi}{\partial x} = Bak \sin(\omega t - kx)$

where bulk mudulus $B = \gamma . P_0 = 1.40 \times 1.013 \times 10^5 \text{ N/m}^2$

$p_{\max} = Bak$  or,  Maximum particle displacement

$$\alpha = \frac{p_{\max}}{Bk} = \frac{\sqrt{2} P_{rms} \times c}{\gamma P_0 \times \omega} \qquad [\because c = \omega/k]$$

$$= \frac{\sqrt{2} \times 20 \times 350}{1.40 \times 1.013 \times 10^5 \times 2\pi \times 10^3} \text{ m} = 1.11 \times 10^{-5} \text{ m}$$

Particle velocity $u = \frac{\partial \psi}{\partial t} = -a\omega \sin(\omega t - kx) = a\omega \cos(\omega t - kx + 90°)$

$\therefore$ Maximum particle velocity $= a\omega$

$$= 1.11 \times 10^{-5} \times 2\pi \times 10^3 \text{ m/s} = 6.97 \times 10^{-2} \text{ m/s}.$$

# Unit-6 : Velocity of Waves

**Structure**

## 6.0  Objectives

After reading this unit, you will learn :

- how to set up the expression for velocity of transverse vibrations of a stretched string
- how to find the expression for velocity of longitudinal waves in a fluid (Newton's formula) and Laplace correction.

## 6.1  Introduction

The transverse vibration of a stretched string is a simple example of a vibrator with distributed characteristics of masses throughout the medium. It is also a simple example of a medium of wave propagation. There is, therefore, much theoretical interest in the study of the vibration of strings. Here we consider the transverse vibration of an **ideal string**. An ideal string is a hypothetical one and cannot be realized in practice.

A sound wave propagates, through an elastic fluid medium as an alteration in pressure, particle displacement or particle velocity. To drrive an expression for velocity of a plane longitudinal wave in fluid medium we need some approximations.

## 6.2  Velocity of Transverse Vibrations of a Strectched String

A string is a *perfectly uniform* and *flexible filament* of solid matter whose length is large compared to its radius.

When a string is stretched between two fixed points (Fig. 6.1) and a portion of it is plucked perpendicular to its length, *transverse vibrations* are set up in it. The particles vibrate at right angles to its length and the movement is handed over from particle to particle. This results in the propagation of a transverse wave along the string with a velocity determined by some constants of the string.
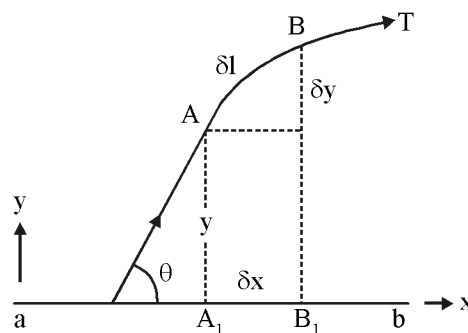
Fig. 6.1

**Expression for velocity :** Let a string be stretched between two fixed points and let ab be its rest position along x-axis (Fig. 6.1). Let a portion of the string be now drawn along y-axis, i.e., at right angles to its length such that the *amplitude* of vibration is *small*. Since the string is perfectly flexible, we can take the *tension T* of the string to remain essentially *constant* along the string acting tangentially at each point as the waves travel along it.

Consider a small segment $AB$ $(= \delta l)$ of the string in the displaced position. If $\theta$ be the angle made by the tangent at $A$ with $ab$, the transverse component of the tension

at A is  $T \sin \theta \simeq T \tan \theta = T\left(\dfrac{\partial y}{\partial x}\right)$,  since $\theta$ is small.

Transverse component of tension at  $B = T\dfrac{\partial}{\partial x}\left(y + \dfrac{\partial y}{\partial x}\delta x\right) = T\dfrac{\partial y}{\partial x} + T\dfrac{\partial^2 y}{\partial x^2}\delta x$.

∴ Resultant force along y-direction on the element  $AB = T\dfrac{\partial^2 y}{\partial x^2}\delta x$.

If $\rho$ be the mass per unit length of the string $\rho \delta x$ is the mass of the element AB.

∴     $\pi \delta x \times$ acceleration of  $AB = T\dfrac{\partial^2 y}{\partial x^2}\delta x$

or,     $\rho \delta x \dfrac{\partial^2 y}{\partial t^2} = T\dfrac{\partial^2 y}{\partial x^2}\delta x$

or,     $\dfrac{\partial^2 y}{\partial t^2} = \dfrac{T}{\rho}\dfrac{\partial^2 y}{\partial x^2} = v^2\dfrac{\partial^2 y}{\partial x^2}$                    ... (6.2.1)

where $v = \sqrt{\dfrac{T}{\rho}}$ .

This is the differential equation for transverse wave motion alsong the stretched string.

$\therefore$   Wave velocity,   $v = \sqrt{\dfrac{T}{\rho}}$                           ... (6.2.2)

## 6.3  Velocity of Longitudinal Waves in a Fluid

A sound wave propagates throguh an elastic fluid medium as an alternation in pressure, particle displacement or particle velocity. To derive an expression for the velocity of a plane longitudinal wave in fluid medium we start with the following assumptions :

(i) The medium is homogeneous isotropic and has no dissipative forces.

(ii) The medium is continuous i.e., the wavelength ($\lambda$) of the wave is much greater than the mean free path.

(iii) The medium is of infinite extent so that we can neglect the boundary effects.

(iv) The equilibrium pressure $P_0$ and density $\rho_0$ are the same everywhere i.e., we neglect the effect of gravity.

(v) The wave is of small amplitude, the strain developed in the medium is so small that the Hooke' law is obeyed.

Let us consider a tube of fluid of unit cross-section with its axis in the direction of propagation of the wave and two plane sections $A$ and $B$ of the tube at $x$ and $x + \delta x$ as shown in Fig. 6.2. Here $\delta x \gg \lambda$.
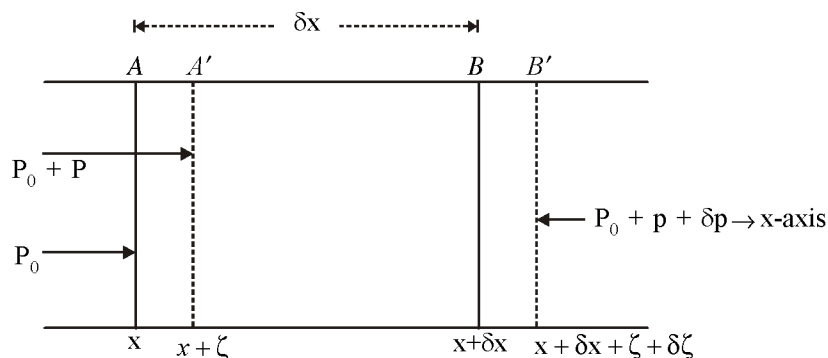


Fig. 6.2  A tube of unit cross-section and of thickness $\delta x$

The Volume of the slice $AB$ is $\delta x$. Under the influence of the sound wave let the particles in the layer $A$ are displaced parallel to the axis of the tube by a distance $\xi$ $A$ to $A'$ and those at $B$ are displaced by $\xi + \delta\xi$ to $B'$ where $\xi \ll \delta x$. Therefore, the final volume of the slice $A'B'$ is

$$\left(x + \delta x + \xi + \delta\xi\right) - \left(x + \xi\right) = \delta x + \delta\xi = \delta x + \frac{\partial\xi}{\partial x}.\delta x$$

Now the fractional change in volume, called *dilataztion* ($\Delta$) is given by

$$\Delta = \frac{\text{change in volume}}{\text{original volume}} = \frac{\left(\delta x + \frac{\partial\xi}{\partial x}\cdot\delta x\right) - \delta x}{\delta x} = \frac{\partial\xi}{\partial x} \qquad ... (6.3.1)$$

When an acoustic wave passes through any medium the medium is deformed due to change in pressure from layer to layer. The pressure $p$ in excess of normal equilibrium pressure $P_0$ is known as *excess* or *acoustic pressure*. Let $p$ be the exces pressure on

$A'$ and that on $B'$ be be $p + \delta p = p + \frac{\partial p}{\partial x}\cdot\delta x$. Therefore, the resulting force on the slice between $A'$ and $B'$ along x-axis is

$$p - \left(p + \frac{\partial p}{\partial x}\delta x\right) = -\frac{\partial p}{\partial x}\cdot\delta x$$

From Newton's law this force must be equal to mass $(\rho_0.\delta x)$ multiplied by the acceleration $\frac{\partial^2\xi}{\partial t^2}$, where $\rho_0$ is the equlibrium density. Thus

$$\rho_0\delta x\cdot\frac{\partial^2\xi}{\partial t^2} = -\frac{\partial p}{\partial x}\cdot\delta x$$

or, $\qquad \rho_0\cdot\frac{\partial^2\xi}{\partial t^2} = -\frac{\partial p}{\partial x} \qquad\qquad ... (6.3.2)$

Now by definition bulk modulus of the medium is

$$B = \frac{\text{volume stress}}{\text{volume strain}} = \frac{p}{-\Delta}$$

$$p = -B\frac{\partial\xi}{\partial x} \qquad\qquad ... (6.3.3)$$

Substituting Equation (6.3.3) in Equation (6.3.2) we get

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \frac{\partial^2 \xi}{\partial x^2} \qquad \qquad \dots (6.3.4)$$

where $v = \sqrt{B/\rho_0}$

Thus the displacement $\xi$ satisfies the well-known differential wave equation. Hence we can identify $v = \sqrt{B/\rho_0}$ as the velocity of propagation of sound waves.

### 6.3.1 Newton's Formula

Newton gave the first theoretical expression of the velocity of sound wave in a gas. He assumed that when sound wave travels through a gaseous medium, the temperature variations in the regions of compression and rarefaction are negligibe. The conditions are, therefore, isothermal and Boyle's law can be applied.

Consider a volume $V$ of a gas at pressure $P$. As a result of sound having travelled in it, let the change in pressure be $\Delta P$ and the corresponding change in volume be $\Delta V$. From Boyle' law we have

$$PV = \text{constant} \qquad \text{(isothermal change)}$$

Differentiating partially we have

$$P\Delta V + \Delta P V = 0$$

or, $$P = -\frac{\Delta P}{\Delta V/V} = \frac{\text{excess pressure}}{\text{volume strain}}$$

$$= Ei \qquad \text{(by definition)}$$

where E is the volume elasticity of the gas when isothermal conditions hold. Therefore, under isothermal conditions,

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

This is Newton's formula for the velocity of sound in a gas. For air at standard temperature and pressure (STP), density $\rho = 1.29$ kg m$^{-3}$ and pressure $P = 0.76$ m of $Hg = 0.76 \times (13.6 \times 10^3) \times 9.8$ Nm$^{-2}$, Substituting these values in Newton's formula gives

$$v = 280 \text{ ms}^{-1}$$

The experimental value, from various experiments, for the velocity of sound at STP

is $\simeq$ 332 ms$^{-1}$ which is considerably higher (by about 16%) than the above theoretical value obtained from Newton's formula. Newton was unable to give a satisfactory explanation for this discrepancy. The first satisfactory explanation was given by a French scientish Laplace, in 1816.

### 6.3.2. Laplace's correction

According to Laplace when a sound wave passes through a gas the compression and rarefaction occurs so rapidly that the heat produced in the regions of compression finds no time to dessipate in the body of the gas. This makes the process adiabatic and we can use the adiabatic relation

$$PV^l = \text{constant}$$

whosh $\gamma$ is the ratio of specific heat of the gas at constant pressure to that at constant volume. Differentiating the above equation we get

$$V^\gamma.dP - \gamma PV^{\gamma-1}dV = 0$$

$$\text{or,} \quad \gamma P = -\frac{dP}{dV/V} = B$$

Therefore, the velocity of sound in gas is

$$v = \sqrt{B/\rho_0} = \sqrt{\gamma P/\rho_0} \qquad \qquad \text{... (6.3.5)}$$

For air at NTP, $\rho_0$ = 1.293 kg.m$^{-3}$, $P$ = 1.013 × 10$^5$ N.m$^{-2}$ (or $Pa$) and $\gamma$ = 1.4. Substituting all these values in Eqn. (6.3.5) we get $v$ = 331.6 m.s$^{-1}$. The result agrees well with experliment.

The Eqn. (6.3.5) can be used to discuss the dependence of the velocity of sound in gases on various factors such as pressure, temperature, humidity, frequency, amplitude and nature of the gases. Using the perfect gas equation $PV = RT$ we can write from Eqn. (6.3.5)

$$\text{Therefore the velocity} \quad v = \sqrt{\frac{\gamma RT}{M}} \qquad \qquad \text{... (5.3.6)}$$

where $M = \rho_0 V$ is the molecular weight. Eqn. (6.3.6) shows that velocity $v$ is independent of pressure, frequency and amplitude but it is proportional to the square root of the absolute temperature $T$. The presence of moisture lowers the density and hence increases the velocity of sound. $\gamma$ depends on the molecular structure and $M$ is different for different gases. Hence the velocity of sound through a gas depends on its molecular structure and molecular weight.

## 6.4 Relationship between the speed of sound and rms speed of gas molecules

According to the kinetic theory of gases the pressure exerted by a gas is given by

$$P = \frac{1}{3}\rho_0 v_r^2$$

where $v_r$, is the root mean square (rms) speed of the gas molecules and $\rho_0$ is the density of the gas. Therefore

$$v_r = \sqrt{\frac{3P}{\rho_0}} = \sqrt{\frac{3}{\gamma}}\sqrt{\frac{\gamma P}{\rho_0}} = \sqrt{\frac{3}{\gamma}}.v$$

where $v$ is the speed of sound For air or any diatomic gases $\gamma = 1.4$ and hence

$v_r = \sqrt{\frac{3}{1.4}}$ $v = 1.46v$. Thus the rms speed of molecules in air is slightly greater than the speed of sound.

The particle speed $(\partial \xi / \partial t)$ in sound waves is small and has a value of the order of $0.1$ m.s$^{-1}$, but the rms molecular speed due to thermal agitation of the order of $10^2$ to $10^3$ m.s.$^{-1}$. When the molecules of a gas come in contact with a vibrating source they receive small amount of additional momentum in a definite direction. The rate of transfer of this additional momentum in a definite direction from molecule to molecule is the sound speed. Obviously this sound speed cannot exceed the thermal speed. The speed of sound is roughly 1/2 of the rms speed.

## 6.5 Summary

1. Velocity of transverse vibration of a stretched string comes out to be $v = \sqrt{\frac{T}{\rho}}$ where $T$ = tension of the string.

2. Velocity of longitudinal waves is found to be $v = \sqrt{\frac{B}{\rho_0}}$ where B = bulk modulus of the medium and $\rho_0$ = density of the medium.

3. Newton's formula for the velocity of sound in a gas $v = \sqrt{\frac{P}{\rho}}$ where $P$ = pressure of the gas and $\rho$ = density of the gas at a particular temperature.

4. According to Laplace correction, $v = \sqrt{\dfrac{\gamma P}{\rho}}$ where $\gamma$ = ratio of specific heat of the gas at constant pressure to that at constant volume.

5. Relation between the speed of sound ($v$) and rms speed of gass molecules ($v_r$) is given by $v_r = \sqrt{\dfrac{3}{\gamma}} \cdot v$. The rms speed of gas molecules is greater than the speed of sound.

---

## 6.6 Review Questions and Answers

1. A heavy chain of length $l$ and mass per unit length $m$ is suspended vertically from one end. A transverse wave is initiated at the upper most end. Show that the time taken by the wave to travel down to the lower end is $2\sqrt{l/g}$.

2. Transverse wave are generated in two uniform steel wires $A$ and $B$ of diameters $10^{-3}$ m and $0.5 \times 10^{-3}$ m respectively, by attaching their free end to a vibrating source of frequency 500 Hz. Find the ratio of the wavelengths if they are stretched with the same tension.

3. Calculate the velocity of sound in (a) water and (b) steel. Given density of steel = 7800 kgm$^{-3}$, Young's modulus of steel = $20 \times 10^{10}$ Nm$^{-2}$ and bulk modulus of water = $0.20 \times 10^{10}$ Nm$^{-2}$,

4. Compare the velocities of sound in hydrogen ($H_2$) and carbon dioxide ($CO_2$). the ratio ($\gamma$) of specific heats of $H_2$ and $CO_2$ are respectively 1.4 and 1.3.

### ANSWER

1. Consider an element $dx$ of the chain at a distance $x$ from the upper end. The tension $T_x$ at this point is equal to the weight of the lower part of the chain of length $(l - x)$. Thus

$$T_x = (l - x). \text{ m.g.}$$

The velocity of transverse wave at this point is

$$v_x = \sqrt{\frac{T_x}{m}} = \sqrt{(l - x)g}$$

The time taken by the wave to travel the distance $dx$ is

$$dt = \frac{dx}{v_x} = \frac{dx}{\sqrt{(l - x)g}}$$

Required time $t = \int_0^t \dfrac{dx}{\sqrt{(l-x)g}} = \dfrac{1}{\sqrt{g}} \int_l^0 \dfrac{(-dz)}{\sqrt{z}}$

where we have put $t - x = z$.

$$\therefore \quad t = \dfrac{1}{\sqrt{g}} \left[ 2.z^{1/2} \right]_0^t = 2\sqrt{\dfrac{l}{g}}$$

2.  The density $\rho$ of a wire of mass $M$, length $L$ and diameter $d$ is given by

$$\rho = \dfrac{4M}{\pi d^2 L} = \dfrac{4\mu}{\pi d^2} \quad \text{where} \quad \mu = \dfrac{M}{L}$$

Now $\quad \vartheta_A = \sqrt{\dfrac{T}{\mu_A}} \quad$ and $\quad \vartheta_B = \sqrt{\dfrac{T}{\mu_B}}$

$$\therefore \quad \dfrac{\vartheta_A}{\vartheta_B} = \sqrt{\dfrac{\mu_B}{\mu_A}} = \dfrac{d_B}{d_A}$$

$\vartheta$

but $\vartheta_A = v\lambda_A$ and $\vartheta_B = v\lambda_B$ , $v$ being the frequency of the source.

Hence, $\quad \dfrac{\lambda_A}{\lambda_B} = \dfrac{\vartheta_A}{\vartheta_B} = \dfrac{d_B}{d_A} = \dfrac{0.5 \times 10^{-3}}{10^{-3}} = 0.5 = \dfrac{1}{2}$

3.  (a) $\quad E_w = 0.20 \times 10^{10}$ Nm$^{-2}$

$\rho_w = 1000$ kgm$^{-3}$

$$\therefore \quad \vartheta_w = \sqrt{\dfrac{E_w}{\rho_w}} = 1414 \ \text{ms}^{-1}$$

(b) $\quad Y_s = 20 \times 10^{10}$ Nm$^{-2}$

$\rho_s = 7800$ kgm$^{-3}$

$$\therefore \quad \vartheta_s = \sqrt{\dfrac{Y_s}{\rho_s}} = 5060 \ \text{ms}^{-1}$$

4.
$$\vartheta H_2 = \sqrt{\frac{\gamma H_2 P}{\rho H_2}} \; ; \; \vartheta CO_2 = \sqrt{\frac{\gamma CO_2 P}{\rho CO_2}}$$

$$\therefore \quad \frac{\vartheta H_2}{\vartheta CO_2} = \sqrt{\frac{\gamma H_2}{\gamma CO_2} \cdot \frac{\rho CO_2}{\rho H_2}}$$

Since density of a gas is proportional to its molecular weight

$$\frac{\rho CO_2}{\rho H_2} = \frac{44.01}{2.016}$$

$$\therefore \quad \frac{\vartheta H_2}{\vartheta CO_2} \sqrt{\frac{1.4}{1.3} \times \frac{44.01}{2.016}} \simeq 4.85$$

Velocity of sound in hydrogen is 4.85 times that in carbon dioxide.

$_B$

# Unit-7 : Superposition of Waves

**Structure**

## 7.0  Objectives

In this unit, you will learn

- Standing (stationary) waves in a stretched string and analytical treatment to find eigen values, eigen functions and eigen frequency
- energy of a vibrating string
- plucked and struck strings
- some important points about stringed musical instruments

- Meldes experiment to find frequency of a tuning fork
- Relation between wave velocity and group velocity
- Longitudinal standing waves and normal modes in open and closed pipes

## 7.1  Introduction

In the Unit-5 and Unit-6, we have discussed waves which travel uninterrupted in open or unbounded systems. We assumed that the medium was homogeneous and infinite in extent. This unit is devoted to closed or bounded systems. The study of the vibrations of strings and air columns is very important because the sound emitted from many sound sources is due to vibrations of one of these kinds.

The study of vibration of strings is important both from theoretical and practical points of view. In a simple vibrating system, the mass is assumed to be concentrated at one point and the motion of the system can be completely specified by giving displacement of one or more mass points as a function of time. But most practical vibrators are not of this type. For example, in the strings of a muscial instrument and air column of organ pipes the masses are distributed. In case of wave motion the inertia and elasticity are distributed throught the medium. The transverse vibration of a string and longitudinal vibration of air column are simple examples of a vibrator with distributed characteristics and are also simple examples of a medium of wave propagation. There is, therefore, much theoretical interest in the study of the vibration of strings and air columns. From practical point of view, we note that the vibrations of string and air column play important roles in the world of music. String instruments like guitar, sitar, piano are all based on the transverse vibration of stretched strings.

Only longitudinal waves are possible in air columns. the vibrations of air columsn in open and closed pipes serve as the source of sound in many musical instruments like organ, flute, clarinet etc.

In all the cases of above vibrations generated waves are reflected from the boundary and stationary waves or standing waves are produced due to superpostion of the reflected and incident waves.

## 7.2  Wave equation : stretched string

We have obtained one dimensional differential wave equation in Unit-6 (see Art. 6.2) for a stretched string given by

$$\frac{\partial^2 y}{dt^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

... (7.2.1)

where c is the velocity of transverse waves along the string.

Eq. (7.2.1) is the well-known differential wave equation and hence $c$ can be identified as the velocity of transverse waves along the string.

## 7.3 Harmonic Solution of the wave equation

A general solution of the wave equation (7.2.1) may be obtained by using the method of separation of variables. Let

$$y(x,t) = F_1(x).F_2(t) \qquad \qquad ...(7.3.1)$$

where $F_1$ is a function of $x$ only and $F_2$ is a function of $t$ only. Substituting (7.3.1) in Eqn. (7.2.1) and dividing throught by $F_1F_2$ we get

$$\frac{1}{F_2} \cdot \frac{d^2F_2}{dt^2} = \frac{c^2}{F_1} \cdot \frac{d^2F_1}{dx^2} \qquad \qquad ...(7.3.2)$$

Left hand side of this equation is a function of $t$ only and the right hand side is a function of $x$ only. So for equality each side must be equal to some constant, say, $-\omega^2$. the negative sign is chosen in order to get periodic solution. Thus we get

$$\frac{d^2F_1}{dx^2} + k^2 F_1 = 0 \qquad \qquad ...(7.3.3)$$

and $$\frac{d^2F_2}{dt^2} + \omega^2 F_2 = 0 \qquad \qquad ...(7.3.4)$$

where $k = \omega/c$.

The general solutions to these equations are

$$F_1 = A\cos kx + B\sin kx \qquad \qquad ...(7.3.5)$$

and $$F_2 = C\cos \omega t + D\sin \omega t \qquad \qquad ...(7.3.6)$$

where $A$, $B$, $C$ and $D$ are arbitrary constants which can be determined from the given initial and boundary conditions. Thus the solution to the wave Eqn. (7.2.1) may be written as

$$y(x,t) = (A\cos kx + B\sin kx)(C\cos \omega t + D\sin \omega t) \qquad ...(7.3.7)$$

In this solution there is no restriction on the values of $\omega$. This happens so because in the above discussion we have tacitly assumed stretched string of infinite length. Thus *a string of infinite length can sustain any arbitrary frequncy of oscillation.*

With a veiw to practical system we consider a streched string of finte length $l$ reigidly fixed at both ends. Suppose the string is fixed at $x = 0$ and $x = 1$. Then $y = 0$ at $x = 0$ and $l$ for all values of $t$. Substituting the condition $y = 0$ at $x = 0$ in the solution (7.3.7) we get

$$A \ (C \cos \omega t + D \sin \omega t) = 0$$

This indicates that $A = 0$. the condition $y = 0$ at $x = l$ requires that

$$B \sin kl \left( C \cos \omega t + D \sin \omega t \right) = 0$$

This is to be valid for all $t$. So we must have

sin $kl = 0$

or,     $kl = n\pi,$              where   $n = 1, 2, 3$

or,     $k = n\pi/l = k_n$ (say)                        ... (7.3.8)

and    $\omega = kc = n\pi c/l = \omega_n$ (say)

Note that the possibilities $B = 0$ and $n = 0$ are not physically intersting because they make $y = 0$ for all $t$. This corresponds to the situation when the string is at rest. The solution (7.3.7) now can be written as

$$y = \sin k_n x \ (a_n \cos \omega_n t + b_n \sin \omega_n t)$$

where we have put $BC = a_n$ and $BD = b_n$. For each value of $n$ we get a solution. Since the wave equation is linear and homogeneous the general solution is obtained by the principle of superposition as

$$y = \sum_{n=1}^{\infty} \sin k_n x \left( a_n \cos \omega_n t + b_n \sin \omega_n t \right) \qquad \text{... (7.3.9)}$$

It can also be written as

$$y = \sum_{n=1}^{\infty} c_n \sin k_n x \cos \left( \omega_n t - \phi_n \right) \qquad \text{... (7.3.10)}$$

where $c_n = \sqrt{a_n^2 + b_n^2}$ and $\phi_n = \tan^{-1} b_n/a_n$.

The solution (7.3.10) indicates that a *stretched string of finite length and rigidly fixed at both ends can sustain only certain discrete frequencies*. In the $n$th mode an element of the string vibrates simple harmonically with amplitude $c_n \sin k_n x$ and andgular frequency $\omega_n$. The frequency of vibration in the $n$th mode is

$$v_n = \frac{\omega_n}{2\pi} = \frac{1}{2\pi} \cdot \frac{n\pi c}{l} = \frac{nc}{2l} = \frac{n}{2l} \sqrt{\frac{T}{m}} \qquad \text{... (7.3.11)}$$

The mode corresponding to $n = 1$ is called *fundamental mode* of vibration and its frequency $v_1 = \dfrac{1}{2l}\sqrt{T/m}$ is called the *fundamental* or first harmonic frequncy. The higher frequency modes have frequencies which are initegral multiple of the fundamental frequency and are called harmonics. All the normal modes may be present at the same time and the general displacement is the superpostion of the displacements due to all the modes.

## 7.4 Eigen values, eigen functions and eigen frequencies

Suppose we have an operator A and functions $\psi_i$ such that

$$\hat{A}\psi_i = a_i\psi_i$$

Then $\psi_i$ are called eigen functions and $a_i$ eigen values of the operator $\hat{A}$. For example,

$$\frac{d}{dx}\left(e^{mx}\right) = m.e^{mx}$$

In this case $e^{mx}$ is an eigen function and $m$ is an eigen value of the operator $d/dx$. the space part of the solution (7.3.10) is $X_n(x) = \sin k_n x$.

Now $$\left(-\frac{d^2}{dx^2}\right)X_n = k_n^2 X_n$$

Thus $X_n$ are eigen functions and $k_n^2$ are eigen values corresponding to the operator $-d^2/dx^2$.

Similarly taking the time dependent part of the solution (7.3.10) we find that

$$\left(-\frac{d^2}{dt^2}\right)\cos\left(\omega_n t - \phi_n\right) = \omega_n^2.\cos\left(\omega_n t - \phi_n\right)$$

Thus $\omega_n^2$ are the eigen values and $\cos\left(\omega_n t - \phi_n\right)$ are the eigne functions of the operator $-d^2/dt^2$. These frequencies $\omega_n$ are called eigen frequencies.

The eigen functions $X_n\left(x\right) = \sin k_n x$ form a *complete set of orthogonal functions.* Two eigen functions $X_n(x)$ and $X_m(x)$ are said to be orthogonal if

$$\int\limits_{all\,x} X_n\left(x\right)X_m\left(x\right)dx = 0 \text{ for } n \neq m$$

Here $\qquad \int\limits_0^l \sin k_n x \sin k_m x dx = \int\limits_0^l \sin \frac{n\pi x}{l} \sin \frac{m\pi x}{l} dx$

Thus $X_n(x)$ form a set of orthogonal functions. These functions $X_n(x)$ form a *complete set* in the sense that any arbitrary function $f(x)$ satisfying the same boundary conditions can be expressed as a linear combinations of these functions i.e.,

$$f(x) = \sum_{n=1}^{\infty} C_n X_n(x)$$

## 7.5  Stationary waves in a vibrating string

The solution (7.3.10) can be rewritten as $y = \sum\limits_{n=1}^{\infty} y_n$ where

$$y_n = c_n \sin k_n x . \cos (\omega_n t - \phi_n)$$

$$= \frac{1}{2} c_n \sin (\omega_n t + k_n x - \phi_n) - \frac{1}{2} c_n \sin (\omega_n t - k_n x - \phi_n)$$

This gives the displacement of the string at any point $x$ at any time instant $t$ when it is vibrating in the $n$th mode. Thus the vibration of the $n$th mode may be considered as the superpostion of two waves propagating in two opposite directions. The resultant amplitude at any point is $c_n \sin k_n x$. It varies with $x$ and has maximum values (*antinodes*) when

$$k_n x = \frac{n\pi x}{l} = (2p+1)\pi/2$$

where p = 0, 1, 2, 3, ... $(n-1)$

or, $\qquad x = \frac{l}{2n}, 3\frac{1}{2n}, \cdots$

and has minimum (*nodes*) when

$$k_n x = \frac{n\pi x}{l} = s\pi \qquad \text{where s = 0, 1, 2, 3, ... } n$$

or, $\qquad x = 0, l/n, 2l/n, ... l$
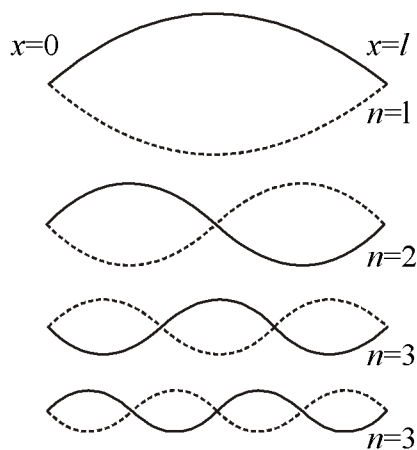
**Fig. 7.1 The first four modes of vibration in the string**

Note that the cases $s = 0$ and $s = n$ corresponds to the original boundary conditions that $y = 0$ at the ends of the string. Thus in the $n$th mode the string vibrates in n segments. The distance between two consecutive nodal points is $l/n$. The neighbouring segments will vibrate in opposite phase. The stationary wave patterns for the first few modes of vibration are shown in Fig. 7.1.

## 7.6 Energy of a vibrating string

For a vibrating stretched string rigidly fixed at both ends. the displacement y at any point $x$ at any instant of time $t$ is given by

$$y(x,t) = \sum_{n=1}^{\infty} c_n \sin k_n x \cos(\omega_n t - \phi_n)$$

... (7.6.1)

where the sum is over all possible modes of vibration.
For brevity let us put

$$c_n \cos(\omega_n t - \phi_n) = \xi_n$$

Then

$$\xi_n = \frac{d\xi_n}{dt} = -\omega_n c_n \sin(\omega_n t - \phi_n)$$

... (7.6.2)

and

$$y = \sum_n \sin k_n x . \xi_n$$

... (7.6.3)

Now kinetic energi of the string is

$$E_k \int_0^l \frac{1}{2}(mdx) y^2$$

$$E_k = \frac{m}{2} \int_0^l \frac{1}{2} \left( \sum_{n=1}^{\infty} \sin k_n x . \xi_n \right)^2 dx$$

$$= \frac{n}{2} \int_0^l \frac{1}{2} \sum_{n=1}^{\infty} \sum_{s=1}^{\infty} \xi_n \xi_s \sin k_n x \sin k_s x \, dx$$

Since $\int_0^l \sin k_n x \sin k_s x dx = 0$      for $n \neq s$

$= l/2$      for $n = s$

We can write

$$E_k = \frac{m}{2} \cdot \frac{l}{2} \sum_{n=1}^{\infty} \xi_n^2 = \frac{M}{4} \sum_{n=1}^{\infty} \omega_n^2 c_n^2 \sin^2 \left(\omega_n t - \phi_n\right) \qquad \dots (7.6.4)$$

where $M = m.l$ is the total mass of the string.

The potential energy $\delta E_p$ associated with an element $\delta l$ of the string (Fig. 6.1) can be calculated by considering the work done against the tension $T$ in stretching the element $\delta x$ to the length

$$\delta l = \sqrt{(\delta x)^2 + (\delta y)^2} = \sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2} \cdot \delta x = \left\{1 + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right\} \delta x$$

Since the tension remains practically constant,

$$\delta E_p = T\left(\delta l - \delta x\right) = T \cdot \frac{1}{2} \delta x \left(\frac{\partial y}{\partial x}\right)^2$$

Using Eqn. (7.6-3) we get,

$$\delta F_p = \frac{T}{2} \cdot \delta x \left(\sum_{n=1}^{\infty} k_n \cos k_n x . \xi_n\right)^2$$

$$= \frac{T}{2} \cdot \delta x \sum_{n=1}^{\infty} \sum_{s=1}^{\infty} k_n k_s \xi_n \xi_s \cos k_n x \cos k_s x$$

Integrating over 0 to $l$ we get the total potential energy as

$$E_p = \frac{T}{2} \int_0^l \sum_{n=1}^{\infty} \sum_{s=1}^{\infty} k_n k_s \xi_n \xi_s \cos k_n x \cos k_s x . \delta x$$

Since $\qquad \int_0^l \cos k_n x \cos k_s x dx = 0 \qquad\qquad$ for $n \neq s$

$$= l/2 \qquad\qquad \text{for } n = s$$

We get $\quad E_p = \frac{T}{2} \cdot \frac{l}{2} \sum_{n=1}^{\infty} k_n^2 \xi_n^2$

$$= \frac{Tl}{4c^2} \sum_{n=1}^{\infty} \omega_n^2 \xi_n^2 \qquad\qquad \left[ \because k_n^2 = \frac{\omega_n^2}{c^2} \right]$$

$$= \frac{M}{4} \sum_{n=1}^{\infty} \omega_n^2 \xi_n^2 \qquad\qquad [ \because s^2 = T/m \text{ and } M = ml]$$

$$= \frac{M}{4} \sum_{n=1}^{\infty} \omega_n^2 c_n^2 \cos^2 \left( \omega_n t - \phi_n \right) \qquad\qquad ... (7.6.5)$$

$\therefore$ The total energy of the vibrating string is

$$E = E_k + E_p$$

$$= \frac{M}{4} \sum_n \left( \xi_n^2 + \omega_n^2 \xi_n^2 \right) = \frac{M}{4} \sum_n \omega_n^2 c_n^2 \qquad\qquad (7.6.6)$$

This expression shows that the total energy is the sum of the energies associated with the individual modes excited separately. The energy associated with different modes are entirely independent of each other. The energy is never exchanged among the modes. If the string is set into oscillation in a particular mode other modes will not be excited by acquiring energy from the vibrating mode.

A string may be considered as an assembley of infinite number of strongly coupled particle oscillators. So it will have infinite nimber of normal modes of vibration. For the $n$th mode the quantity $\xi_n$ can be identified as a *normal coordinate* and $\omega_n$ as the *normal frequency*.

## 7.7 Plucked string

A stretched string may be set into transverse vibration in a variety of ways. Most common methods are the excitations by plucking, striking and bowing. The number of modes excited and their relative magnitudes are determined by the method of excitation. Thus the quality of the emitted note depends on the way in which the string is set into vibration.
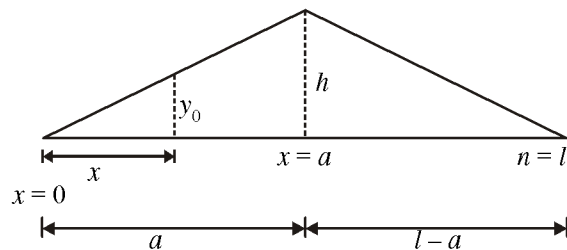


**Fig. 7.2 Plucked string**

If a point of a stretched string rigidly fixed at both ends is pulled in a transverse

direction and then released the string is said to be plucked. It can be found in musical instruments like guitar, sitar, tanpura etc. Suppose a stretched string of legnth $l$ rigidly fixed at points $x = 0$ and $x = l$ is plucked at a point $x = a$ so that the initial displacement at $x = a$ is $h$. From Fig. 7.2 we find that the initial configuration of the string may be represented analytically as follows.

$$\frac{y_0}{x} = \frac{h}{a}$$

or, $\qquad y_0 = \frac{hx}{a}$ $\qquad\qquad\qquad\qquad$ for $0 < x < a$

Similarly

$$y_0 = \frac{h(l-x)}{l-a} \qquad\qquad\qquad\qquad \text{for } a < x < 1 \quad ... (7.7.1)$$

Here $y_0$ is the initial displacement at any point $x$. The initial velocity at every point is zero i.e.,

$$\dot{y}_0 = 0 \qquad\qquad\qquad\qquad\qquad ... (7.7.2)$$

Now if the string is released the transverse disturbance travels alone the string satisfying the differential wave equation

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

which has a general solution in this case as

$$y(x,t) = \sum_{n=1}^{\infty} \sin k_n x \left( a_n \cos \omega_n t + b_n \sin \omega_n t \right) \qquad ... (7.7.3)$$

where $k_n = n\pi/l$, $\omega_n = n\pi c/l$ and $c = \sqrt{T/m}$
The initial condition (7.7.2) indicates that

$$\dot{y}(x,0) = \dot{y}_0 = \sum_{n=1}^{\infty} \omega_n b_n \sin k_n x = 0$$

Therefore, $\quad b_n = 0$ $\qquad\qquad\qquad\qquad\qquad\qquad ... (7.7.4)$
Now from Eqn. (7.7.3) we can write

$$y(x,0) = y_0 = \sum_{n=1}^{\infty} a_n \sin k_n x \qquad\qquad ... (7.7.5)$$

The constant $a_n$ can be determined by the Fourier series method. Multiplying both sides of (7.7.5) by $\sin k_s x$ and and integrating from $x = 0$ to $x = l$ we get

$$\int_0^l y_0 \sin k_s x dx = \sum_{n=1}^{\infty} a_n \int_0^l \sin k_n x \sin k_s x \, dx$$

Since $\qquad \int_0^l \sin k_s x \sin k_s x \, dx = 0 \qquad\qquad$ for $n \neq s$

$$= l/2 \qquad\qquad \text{for } n = s$$

We can write

$$\int_0^l y_0 \sin k_s x \, dx = a_s \cdot \frac{l}{2}$$

This is valid for any integer value of $s$.

$$a_n = \frac{2}{l} \int_0^l y_0 \sin k_n x \, dx$$

Using the value of $y_0$ from Eqn. (7.7.1) we get

$$a_n = \frac{2}{l} \int_0^a \frac{hx}{a} \sin k_n x dx + \frac{2}{l} \int_0^l \frac{h(l-x)x}{l-a} \cdot \sin k_n x \, dx$$

Integrating by parts,

$$a_n = \frac{2}{l} \cdot \frac{h}{a} \left[ -\frac{x}{k_n} \cdot \cos k_n x + \frac{\sin k_n x}{k_n^2} \right]_0^a$$

$$+ \frac{2}{l} \cdot \frac{h}{(l-a)} \left[ -\frac{l \cos k_n x}{k_n} + \frac{x}{k_n} \cos k_n x - \frac{\sin k_n x}{k_n^2} \right]_0^l$$

$$= \frac{2h \sin k_n a}{lk_n^2} \left( \frac{1}{a} + \frac{1}{l-a} \right) = \frac{2h \sin k_n a}{a(l-a).k_n^2} \qquad\qquad \dots (7.7.6)$$

Now putting the values of $a_n$, $b_n$, $k_n$ and $\omega_n$ in Eqn. (7.7.3) we get the general expression for displacement in a plucked string as,

$$y(x,t) = \frac{2hl^2}{a(l-a)\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\frac{n\pi a}{l} . \sin\frac{n\pi x}{l} \cos\frac{n\pi ct}{l} \qquad ... (7.7.7)$$

It shows that the maximum amplitude of the $n$th harmonic is inversely proportional to $n^2$. So the amplitudes of higher harmonics fall off rapidly and only first few harmonics become audible. The term $\sin(n\pi a/l)$ indicates that the $n$th harmonic will be absent if the plucking point is such that

$$\sin\left(\frac{n\pi a}{l}\right) = 0 \quad \text{or,} \quad \frac{n\pi a}{l} = s\pi$$

where s = 1, 2, 3, ...

$$\text{or,} \qquad n = \frac{sl}{a}$$

For example, if the string is plucked at the mid-point i.e., if $a = l/2$ then $n = 2s$ = 2, 4, 6, ... etc. harmonics will be absent from the resultant vibration. Note that all these harmonics require the plucking point as a node. Similarly if a = $l/3$ then n = 3, 6. 9,... etc., harmonics requiring a node at the point $x = l/3$ will be absent. Thus we can make the following general statement :

*When any point of a string is plucked, struck or bowed all the harmonics which require that point as a node will be absent from the resultant vibration.* This is known as *Young's law or Young-Helmoholtz law.*

If the plucked point is touched immediately after plucking all the vibrations will be stopped. This is due to the fact that all the harmonics having node at the point of plucking are absent from Young's law and other harmonics not having a node at that point will be damped at due to touching.

## 7.8 Struck string

In the musical instruments like a piano a stretched string is set into vibration by striking it at a point with a hammer. A rigorous analysis of the motion of a struck string requires to consider a number of factors such as the duration of contact, the velocity of the hammer, the relative masseses of the hammer and the string. However, for the sake of simplicity we assume that time of contact is so small that the impact ceases before the disturbance finds time to spread over any appreciable length of the

string. Thereafter the motion can be considered to be almost free with the following initial conditions :

(i) At $t = 0$, the displacement $y(x, 0) = y_0 = 0$ for all $x$.

(ii) At $t = 0$, the velocity $\dot{y}(x, 0) = \dot{y}_0 = 0$ for all $x$ excepting the small region of striking.

Let us consider a stretched string of length $l$ rigidly fixed at two ends. It is struck by a hammer over an infinitesimally small region from $x = a$ to $x = a + da$. The disturbance created travels along the string satisfying the wave equation

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

which has a general solution in this case as

$$y(x,t) = \sum_{n=1}^{\infty} \sin k_n x \left( a_n \cos \omega_n t + b_n \sin \omega_n t \right) \qquad \dots (7.8.1)$$

where $k_n = \dfrac{n\pi}{l}$, $\omega_n = \dfrac{n\pi c}{l}$ and c is the velocity of transverse wave. $\quad \dots (7.8.2)$

The condition $y(x, 0) = 0$ indicates that $a_n = 0$.

Now $\qquad \dot{y}(x,0) = \dot{y}_0 = \sum_{n=1}^{\infty} \omega_n b_n \sin k_n x \qquad \dots (7.8.3)$

the constant $b_n$ can be determined by the Fourier series method. Multiplying both sides of Eqn. (7.8.3) byt $\sin k_s x$ and integrating from $x = 0$ to $x = l$ we get

$$\int_0^l \dot{y}_0 \sin k_s x \, dx = \sum_{n=1}^{\infty} \int_0^l \omega_n b_n \sin k_n x \sin k_s x \, dx$$

Since $\qquad \displaystyle\int_0^l \sin k_s x \sin k_s x \, dx = 0 \qquad\qquad$ for $n \neq s$

$$= l/2 \qquad\qquad \text{for } n = s$$

We can write

$$\int_0^l \dot{y}_0 \sin k_s x \, dx = \omega_s b_s \cdot \frac{l}{2}$$

This is valid for any integer value of $s$.

$$b_n = \frac{2}{l\omega_n} \int_0^l \dot{y}_0 \sin k_n x \, dx$$

Since $\dot{y}_0 = 0$ for all points excepting the region of contact from $a$ to $a + da$, the above relation reduces to

$$b_n = \frac{2}{l\omega_n} \int_a^{a+da} \dot{y}_0 \sin k_n x \, dx$$

Since $da$ is infinitesimally small, the sine function may be assumed remain consant over this region and hence we may write

$$b_n = \frac{2}{l\omega} \sin k_n a \int_a^{a+da} \dot{y}_0 dx$$

$$= \frac{2A}{\pi c n} \sin \frac{n\pi a}{l}$$

where $A = \int_a^{a+da} \dot{y}_0 \, dx$. If the law of variation of $\dot{y}_0$ over da is known then A can be evaluated.

Now putting the values of $a_n$, $b_n$, $k_n$ and $\omega_n$ in Eqn. (7.8.1) we get the general expression for diaplacement in a struck string as

$$y(x,t) = \frac{2A}{\pi c} \sum_{n=1}^{\infty} \frac{1}{n} \sin \frac{n\pi a}{l} . \sin \frac{n\pi x}{l} . \sin \frac{n\pi c t}{l} \qquad \text{... (7.8.4)}$$

It shows that the maximum amplitude of the $n$th harmonic is inversely proportional to $n$. In case of plucked string this variation is as $1/n^2$. So in case of struck string greater number of harmonics will have appreciable amplitudes. Hence *sound emitted by a struck string will be richer in harmonics as compard to that emitted by a plucked string.*

The term sin ($n\pi a/l$) in Eqn. (7.8.4) indicates that the nth harmonic will be absent if the striking point is such that

$$\sin \frac{n\pi a}{l} = 0 \quad \text{or,} \quad \frac{n\pi a}{l} = s\pi$$

where $s = 1, 2, 3, ...$

or,    $n = \dfrac{sl}{a}$

For example, if the string is strucked plucked at the mid-point i.e., if $a = l/2$ then $n = 2s = 2, 4, 6, ...$ etc. harmonics will be absent from the resultant vibration. Note that all these harmonics require the striking point as a node. Similarly if $a = l/3$ then $n = 3, 6. 9,...$ etc., harmonics requiring a node at the point $x = l/3$ will be absent. So in general, all the harmonics requiring the striking point as a node will be absent from the resultant vibration. Thus Young's law is also obeyed by the struck string.

**Discussions :**

the above is a simplitied analysis. The actual motion of a struck string is complicated and we need to consider a number of factors such as the duration of contact, the velocity of the hammer, the relative mass of the hammer and the string. A considerable amount of research has been done to account for these factors. It has been found that for more abrupt and localised blow the relative intensities of higher harmonics become greater. In a piano the *harmonics above the sixth are considered undesirable as they tend to produce dissonance.* Consequently the hammer is covered with felt to redue the abruptness and localisation of the impact and *the string is struck at a point nearly one-seventh the length to eliminate the 7th harmonic. The amplitude of the fundamental incresses with the increase in the mass of the hammer and for a striking point nearer the bridge.*

## 7.9   A few remarks about stringed musical instruments

The stringed musical instruments can be classified into three distinct types— plucked, struck and bowed. Guitar, mandiline, sitar, tanpura, sarode, vina etc. are the plucked type instruments. Piano and santur are struck type instruments whereas violine, esraj, sarengi etc. are bowed type musical instruments. In plucked and bowed type of instruments the amplitude of $n$th harmonic is proportional to $1/n^2$ whereas in struck type instuments it is proportional to $1/n$. So more number of harmonics will have appreciable amplitude in the case of struck string. Hence sound emitted from struck string will be richer in harmonics.

In most stringed musical instruments the ends of a stretched string are fixed to screws or pegs in a board. As the string vibrates the vibration is communicated to the board via these serews or pegs. The vibrations of the board and the adjoining air responsible for the radiation of sound energy. The string itself is a poor radiator of sound energy. The quality of the sound is modified to a large extent by the body of

the instument. The resonant vibration of the body may increase the intensity of some of the frequency components relative to the other.

We have made all the analysis by considering an ideal string. A real string is never an ideal one. A practical string may be non-unform . In that case tension will be a function of $x$. The two ends of the string are not perfectly fixed due to yielding of the support. This has the same effect as that of an increase of length of the string. A practical string always possess some amount of rigidity and elastic forces come into play. Due to this stiffness there is an extra restoring force. Its effect is to increase the frequencey of each component by different proportions and hence they now do not form a harmonic series.

## 7.10 Melde's Experiement

This is a scientific experiment carried out in 1859 by the German physicist Franz Melde on the standing waves produced in a tense cable originally set oscillating by a turning fork, later improved with connection to an electric vibrator. This experiment attempted to demonstrate that mechanical waves undergo interference phenomena. Mechanical waves traveled in opposite directions form immobile points, called nodes. These waves were called standing waves by Melde since the position of the nodes and loops (points where the cord vibrated) stayed static. In this experiment the change in frequency produced when the tension is increased in the string similar to the change in pitch when a guitar string is tuned will be measured. From this the mass per unit length of the string / wire can be derived.
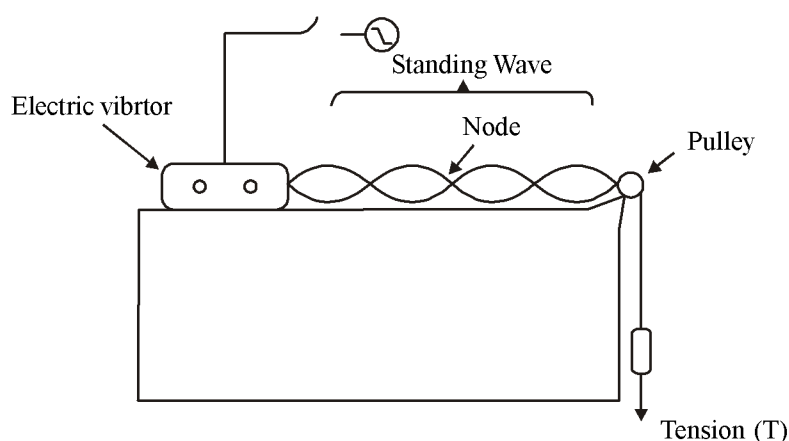


**Fig. 7.3**

This is a scientific experiment carried out in 1859 by the German physicist Franz Melde on the standing waves produced in a tense cable originally set oscillating by a tuning fork, later improved with connection to an electric vibrator. This experiment attempted to demonstrate that mechanical waves undergo interference phenomena. Mechanical waves traveled in opposite directions form immobile points, called nodes. These waves were called standing waves by Melde since the position of the nodes and loops (points where the cord vibrated) stayed static. in this experiment the change in frequncy produced when the tension is increased in the string similar to the change in pitch when a guitar string is tuned will be meaured. From this the mass per unit length of the string / wire can be derived.

**Melde's experiment in longitudinal/parallel position :**



(a) Longitudinal wave



(b) Transverse wave
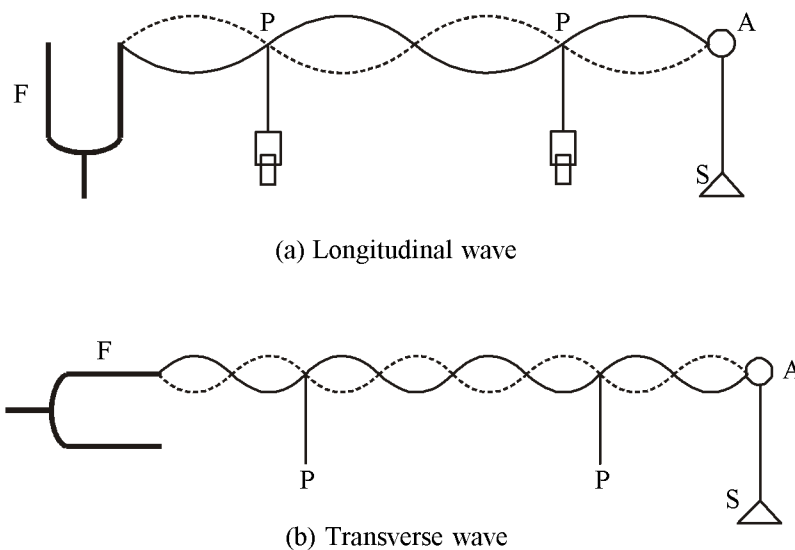
**Fig. 7.4**

Melde's experimental set up consists of a light string tied to one of the prongs of a tuning fork which is mounting on a sounding board. The other end of the string is passed over a horizontal pulley and a light pan is suspended from the free end. The tension of the string can be adjusted by changing the weights placed in the pan while the length of the string can be adjusted by changing the position of the pulley towards or away from the fork. If they fork is adjusted so that its arms vibrate parallel to the length of the string. Through adjusting the length or the tension number of clear loops can be visible as the standing wave is formed. When the prongs vibrate parallel to

the length of the string, the frequency of the fork is twice the frequency of the string. Since the string is vibrating in the fundamental mode, its frequency $n$ is given by,

$$n = \frac{1}{2l}\sqrt{\frac{T}{m}}$$

Therefore, the frequency of the tuning fork will be

$$N = 2n = \frac{1}{l}\sqrt{\frac{T}{m}} \qquad \qquad ... (7.10.1)$$

If number of loops formed is p, then

$$N = \frac{p}{l}\sqrt{\frac{T}{m}} \qquad \qquad ... (7.10.2)$$

**Melde's experiment in transverse/perpendicular position :**

When the fork is adjusted so that its arms vibrate perpendicular to the length of the string, the frequency of the fork is same as the frequency of the string. Since the string is vibrating in the fundamental mode, its frequency $n$ is given by

$$n = \frac{1}{2l}\sqrt{\frac{T}{m}}$$

Therefore, the frequency of the tuning fork will be

$$N = n = \frac{1}{2l}\sqrt{\frac{T}{m}} \qquad \qquad ... (7.10.3)$$

If number of loops formed is $p$, then

$$N = \frac{p}{2l}\sqrt{\frac{T}{m}} \qquad \qquad ... (7.10.4)$$

## 7.11 Vibration of air columns

The vibration of air columns enclosed in a pipe serves as the source of sound in a large class of wind instruments like organ, flute, clarinet etc. The phenomena are also of great theoretical interest. There are two types of pipe which have practical importance. The pipe open at both ends is known as *open pipe* and the pipe closed at one end is known as *closed pipe*. Any distrubance created at the mouth of a pipe, travels down the pipe and is reflected back from the other end. Stationary waves are

formed by the superposition of reflected and incident waves. At the open end there must be formed an antinode and at the closed end there must be a node. Only certain descrete frequencies depending on the size of the pipe and the velocity of sound can be excited in the air column. At the closed end a forward displacement of air particles is reflected by the rigid boundary as a backward displacement of equal magnitude and makes the displacement zero at the closed end. Here  a compression is reflected as compression i.e., with no phase reversal. When a compression reaches at the open end, the front part of the compression spreads out rapidly causing a fall of pressure. The rear part of the copression having higher pressure moves rapidly in the outward direction causing a fall of pressure behind it. In this way a pulse of compression is reflected back as rarefaction save i.e., with a reversal of phase.

In order to develop the thneory of vibration of air columns in a pipe we start with the following simplifying assumptions :

(i) the diameter of the pipe is small compared with the length of the pipe and the wavelength of sound.

(ii) walls of the pipe are perfectly rigid.

(iii) The effects of viscosity and thermal conduction are negligible.

(iv) There is only small amplitude adiabatic change of pressure without rotatory or eddy motion.

(v) Plane waves are produced so that the motion is uniform over any particular cross-acction.

Under these conditions the differential equation representing the vibration of air column will be given by

$$\frac{\partial^2 \xi}{\partial t^2} = c^2 \frac{\partial^2 \xi}{\partial x^2}$$

where $c = \sqrt{B/\rho}$ is the velocity of the wave, B = bulk modulus and $\rho$ = density of the air.

The equation can be solved by the method of separation of variables as done in Art. 7.3. Thus

$$\xi(x,t) = (A\cos kx + B\sin kx)(C\cos \omega t + D\sin \omega t) \qquad \text{... (7.11.1)}$$

where $k = \omega/c$. The constants A, B, C and D are to be evaluated from the given boundary and initial conditions.

### 7.11.1 Open pipe :

For a pipe open at both ends ($x = 0$ and $x = l$) there must have displacement antinodes (maximum vibrations) at the two ends i.e., (i) $\dfrac{\partial \xi}{\partial x} = 0$ at $x = 0$ and (ii) $\dfrac{\partial \xi}{\partial x} = 0$ at $x = l$.

Using the condition (i) in Eqn. 7.11.1 we get $B = 0$

Therefore, $\xi(x, t) = \cos kx (a \cos \omega t + b \sin \omega t)$

where $a = AC$ and $b = AD$

Now applying the condition (ii) we get

$\sin kl = 0$ or, $kl = n\pi$ where $n = 1, 2, 3$ etc.

or, $\quad k = n\pi / l = k_n$ (say)

Also $\quad \omega = kc = n\pi c / l = \omega_n$ (say)

For each value of n we get a solution. Hence, by the principle of superposition the general solution is given by

$$\xi(x, t) = \sum_{n=1}^{\infty} \cos k_n x (a_n \cos \omega_n t + b_n \sin \omega_n t) \qquad \dots (7.11.2)$$

The solution indicates that the air column can vibrate with certain characteristic frequencies. The frequency of the $n$th mode is

$$v_n = \frac{\omega_n}{2\pi} = \frac{nc}{2l} = \frac{n}{2l} \sqrt{\frac{B}{\rho}}$$

So the fundamental mode will have a frequency $v_1 = \dfrac{c}{2l}$ and a wavelength $\lambda_1 = c/v_1 = 2l$. As $n$ can take up all integer values *an open pipe gives rise to all harmonics,* both odd and even Eqn. (7.11.2) represents a stationary wave pattern formed inside the pipe. The stationary wave patterns of first few harmonics are shown in Fig. 7.5.
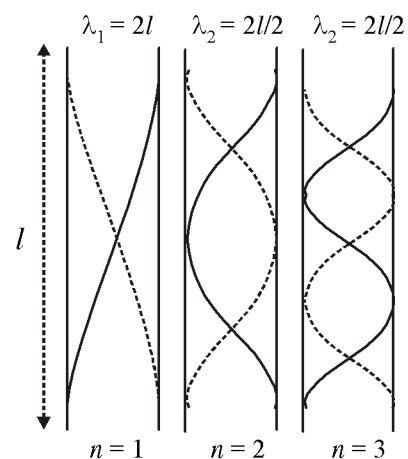


**Fig. 7.5 Modes of vibration in open pipe**

### 7.11.2. Pipe closed at one end

For a pipe closed at one end there must be formed a node at the closed end $x = 0$ and an antinode at the open end $x = l$. Hence the boundary conditions to be satisfied are (i) $\xi = 0$ at $x = 0$ and (ii) $\dfrac{\partial E}{\partial x} = 0$ at $x = l$.

The condition (i) when applied to the Eqn. 7.11.1 gives A = 0. So

$$\xi(x,t) = \sin kx (a \cos \varepsilon t + b \sin \omega t)$$

where a = BC and b = BD

Now condition (ii) requires that $\cos kl = 0$

$$\text{or,} \quad kl = \frac{n\pi}{2} \quad \text{where n = 1, 3, 5, ...}$$

$$\text{or,} \quad k = \frac{n\pi}{2l} = k_n \quad \text{(say)}$$

$$\text{Also} \quad \omega = kc = \frac{n\pi c}{2l} = \omega_n \quad \text{(say)}$$

For each value of $n$ we get a solution and hence the general solution is given by the principle of superposition as

$$\xi(x,t) = \sum_{n=1,3}^{\infty} \sin k_n x (a_n \cos \omega_n t + b_n \sin \omega_n t) \qquad \text{... (7.11.3)}$$

The solution indicates that the air column in a pipe closed at one end can vibrate only with certain characteristic frequencies. The frequency of the $n$th mode is

$$v_n = \frac{\omega_n}{2\pi} = \frac{nc}{4l} = \frac{n}{4l}\sqrt{\frac{B}{\rho}}$$

The frequency of the fundamental is $v_1 = c/4l$ and its wavelength is $\lambda_1 = c/v_1 = 4l$. As here $n$ can take up only odd integer values, *a pipe closed at one end gives rise to only odd harmonics.* Eqn. (7.11.3) represents a stationary wave pattern formed inside the pipe. The stationary wave patterns of first few modes of vibration are shown in Fig. 7.6.

$\lambda_1 = 4l \quad \lambda_3 = 4l/3 \quad \lambda_5 = 4l/5$

$l$

$n = 1 \qquad n = 3 \qquad n = 5$
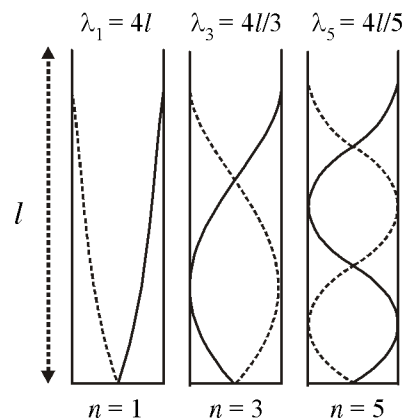
**Fig. 7.6 Modes of vibration in an a pipe closed at one end**

### 7.11.3 End correction

In the above discussion we have assumed that antinodes occur at the open ends of a pipe and that the sound energy in the pipe is perfectly reflected from the open ends. This is not strictly correct. Some amount of energy escapes at each reflection from the open end. The plane waves inside the pipe changes into spherical waves at the open end. The air beyond the open end is, therefore, in vibration and the effective length of the pipe becomes greater than the geohetrical length of the pipe. Under certain assumptions Helmholtz deduced a value $\pi r/4$ as the open as the open end correction where $r$ is the radius of the pipe. Rayleigh obtained a value 0.82 $r$ for this. Experimental values are 0.8 r for pipe with large flange and 0.6 $r$ for an unflanged pipe.

The end effect lowers the frequencies of all tones in a pipe.

## 7.12 Wave train and the concept of wave group

A monochromatic wave may be represented as $\psi(x,t) = \text{Re}\, A e^{i(\omega t - kx)}$. It extends from $-\infty$ to $+\infty$ both in space and time. It is an idealized concept because no practical source vibrates indefinitely and its amplitude does not remain constant. In practice we get a train of waves of finite duration and of damped amplitude. It may be called a *wave train*. A wave train of finite length may be considered to be formed by the superposition of a theoretically infinite number of plane harmonic waves having continuously differing frequencies. For practical purposes the frequencies are limited with a finite range depending on the length of the wave train. The shorter is the length of the wave train the wider is the effective frequency range. In this sense a short enough wave train may be considered as a group of harmonic waves. It may be called a *wave group* or a *wave packet*. If such a wave group is examined by spectrum analyser it would be found to contain a range of frequencies. Maximum energy is associated with some central frequency $\omega_0$ in the range and the energy content of other frequencies falls off as the frequency differs from $\omega_0$.

## 7.13 Phase velocity and group velocity

When a monochromatic harmonic wave moves through a medium the velocity with which a point of constant phase moves is called the *phase velocity* or *wave velocity*.

A plane harmonic wave may be represented by $\psi(x,t) = \text{Re}\, A e^{i(\omega t - kx)}$

For points of constant phase,

$$\omega t - kx = \text{constnat}$$

Differentiating,

$$\frac{d}{dt}(\omega t - kx) = 0$$

or, $$\frac{dx}{dt} = \frac{\omega}{k} = c$$

where $c$ is called the phase velocity.

**Group velocity :**

A wave group may be considered as formed by the superpositin of an infinite number of plane harmonic waves. Practical waves are of finite duration and are not truly monochromatic. A short enough wave train may be considered as a *wave group* formed by the superpositon of an infinite number of plane harmonic waves having slighly differing frequencies and phases. The wave group has a maximum of amplitude which falls off to zero not far from the maximum (Fig. 7.7). If the medium is dispersive i.e., the phase velocity depends on frequency then the shape of the wave group changes as it Fig.7.7(a) Wave group formed by infinite number of harmonic waves having slightly differing frequencies (b) its frequency spectrum.
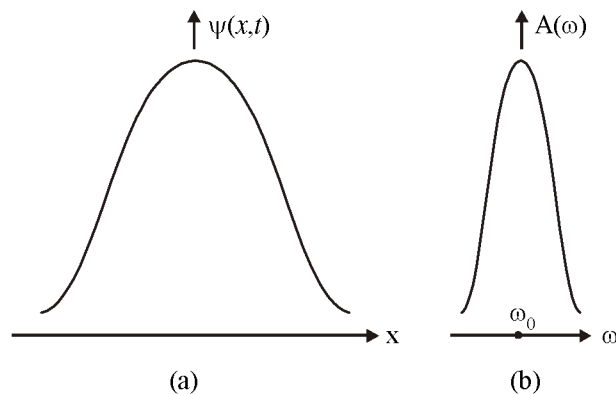


Fig. 7.7 (a) Wave group formed by infinite number of harmonic waves having slightly differing frequencies, (b) its frequency spectrum

travels through the medium. The maximum of the wave group travels with a velocity that is different from the velocities of the component waves. This velocity is known as group velocity ($c_g$).

## 7.14 Relation between phase velocity and group velocity

For a simpler treatment we consider a wave group formed by the superpostion of two waves of equal amplitudes but of slightly different frequency and wavelength. Let us represent the waves by

$$\psi_1 = a\cos(\omega_1 t - k_1 x) \text{ and } \psi_2 = a\cos(\omega_2 t - k_2 x)$$

The superposition of these two waves gives the resultant wave as

$$\psi = \psi_1 + \psi_2$$

$$= 2a\cos\left\{\frac{(\omega_1 - \omega_2)t}{2} - \frac{(k_1 - k_2)x}{2}\right\}\cos\left\{\frac{(\omega_1 + \omega_2)t}{2} - \frac{(k_1 + k_2)x}{2}\right\}$$

$$= A_m \cos(\omega_a t - k_a x) \qquad\qquad\qquad ...(7.14.1)$$

where $\omega_a = \frac{1}{2}(\omega_1 + \omega_2)$, $k_a = \frac{1}{2}(k_1 + k_2)$ and $A_m = 2a\cos(\omega_m t - k_m x)$ with

$\omega_m = \frac{1}{2}(\omega_1 - \omega_2)$ and $k_m = \frac{1}{2}(k_1 - k_2)$. If $\omega_1$ and $\omega_2$ differs lightly then Eqn. (7.14.1) may be interpreted as a high frequency ($\omega_a$) wave whose amplitude varies slowly with a frequency $\omega_m$ (Fig. 7.8). The high frequency part moves with a velocity $\omega_a/k_a$ which
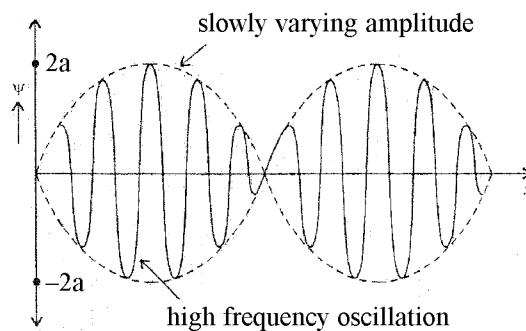


**Fig. 7.8 Wave group formed by two waves of slightly different frequencies**

is practically the phase velocity of either wave. The velocity with which the modulation represneted by the term $A_m = 2a \cos(\omega_m t - k_m x)$ propagates is the group velocity ($c_g$).

$$\therefore \qquad c_g = \frac{\omega_m}{k_m} = \frac{\omega_1 - \omega_2}{k_1 - k_2}$$

Now $\omega$ and $k$ are related by the dispersion relation, $\omega = \omega(k)$

Therefore, $\qquad c_g = \frac{\omega(k_1) - \omega(k_2)}{k_1 - k_2} \qquad\qquad ...(7.14.2)$

Let $k_1 - k_2 = \Delta k$ and $\frac{k_1 + k_2}{2} = k_a$

Therefore, $k_1 = k_a + \dfrac{\Delta k}{2}$ and $k_2 = k_a - \dfrac{\Delta k}{2}$.

Now by making Taylor series expansion we can write

$$\omega(k_1) = \omega\left(k_a + \frac{\Delta k}{2}\right) \approx \omega(k_a) + \frac{d\omega}{dk}\bigg|_{k_a} \cdot \frac{\Delta k}{2}$$

$$\omega(k_2) = \omega\left(k_a - \frac{\Delta k}{2}\right) \approx \omega(k_a) - \frac{d\omega}{dk}\bigg|_{k_a} \cdot \frac{\Delta k}{2}$$

Putting all these is Eqn. (7.14.2) we get

$$c_g = \frac{d\omega}{dk}\bigg|_{k_a}$$

Since phase velocity $c = \omega/k$ or $\omega = ck$ we can write

$$c_g = \frac{d}{dk}(ck) = c + k\frac{dc}{dk} = c + k\frac{dc}{d\lambda} \cdot \frac{d\lambda}{dk}$$

Now $k = \dfrac{2\pi}{\lambda}$ or, $\dfrac{d\lambda}{dk} = -\dfrac{2\pi}{k^2}$.

$$\therefore \qquad c_g = c - \frac{2\pi}{k} \cdot \frac{dc}{d\lambda} = c - \lambda\frac{dc}{d\lambda}$$

This is the required relation between phase velocity and group velocity for a dispersive medium. For a non-dispersive medium $dc/d\lambda = 0$. Then $c_g = c$ i.e., group velocity equals phase velocity. Note that sound waves in the audible range do not show dispersion in any medium but ultrasonic waves show dispersion.

---

## 7.15 Importance of group velocity

An important use of waves is that they can be used to carry energy or information. A single monochromatic harmonic wave cannot be used to send a signal because such a wave travels unchanged. In order to send a message we must modulate some properties of the wave at the sending end in such a way that it can be decoded at the receiving end. Now a modulated wave may be considered as a mixture or a group of harmonic waves. Such a wave group travels with the group velocity. The energy or the signal is thus propagated with the group and has the velocity of the group. In some cases the phase velocity may become greater than the speed of light (for example, the

phase velocity of electromagnetic waves in the ionosphere is greater than the speed of light). It does not violate the principle of relativity as the energy is propagated not with phase velocity but with group velocity which is always less than the speed of light.

## 7.16 Summary

1. According to the principle of superpostion of waves, the resultant displacement (at a particular point (produced by a number of waves is the vector sum of the displacements produced by each one of the disturbances.

2. The stationary waves on a string and in an air column are formed by the superpostion of waves travelling in opposite direction.

3. Harmonic solution of the wave equation has been found and eigen value, eigen functions and eigen frequencies have been evaluated.

4. We have made all the analysis by considering an ideal string.

5. Constants $a_n$ and $b_n$ for plucked string and struck string, respectively have been determined using the Fourier series method. It is to be noted that $b_n = 0$ for plucked string and $a_n = 0$ struck string.

6. When any point of a string is plucked, struck, or bowed all the harmonics which require that point as a node will be absent from the resultant vibration. This is known as *Young's law or Young Helmholtz law.*

7. In plucked type instruments the amplitude of the $n$th harmonic is proportional to $1/n^2$ whereas in struck type instruments it is proportional to $1/n$. Hence sound emitted from struck string will be richer in harmonics.

8. Frequency of a tuning fork can be determined by Melde's experiment both in longitudinal and transverse positions.

9. The theory of vibration of air columns for both open and closed pipes with end correction have been discussed.

10. It is noticed that as n can take up all integer values as open pipe gives rise to all harmonics, both odd and even. Sound emitted from open pipe is richer in harmonics as compared to closed pipe.

11. A wave train of finite length may be considered to be formed by the superposition of a theoretically infinite number of plane harmonic waves having continuously differing frequencies. A wave packet may be considered as a wave train of short length.

12. For a wave packet, phase velocity $c = \dfrac{\omega}{k}$ and group velocity $c_g = \dfrac{d\omega}{dk}$ and their relationship is given by

$$c_g = c - \frac{2\pi}{k} \cdot \frac{dc}{d\lambda} = c - \lambda \frac{dc}{d\lambda}$$

## 7.17  Review Questions and Answers

1. For gravity waves in a liquid the phase velocity c depends on the wavelength $\lambda$ according to the formula $c = A\sqrt{\lambda}$, A being a constant.

   Show that the group velocity is half the phase velocity.

2. The phase velocity $c$ of deep water waves is given by

$$c^2 = \frac{g\lambda}{2\pi} + \frac{2\pi S}{\rho\lambda}$$

   where $g$ is the acceleration due to gravity; $\rho$ is the density of water, $S$ is the surface tension of water. Find the wavelength $\lambda_0$ at which the waves do not disperse in water. Show that for $\lambda \ll \lambda_0$, $c_g = 3c/2$ and for $\lambda \gg \lambda_0$, $c_g = c/2$.

3. The dispersion relation for microwaves in ionoshere is given by

$$\omega^2 = \omega_p^2 + c^2 k^2$$

   where c is the velocity of light in free space and $\omega_p$ is a constant depending on the electron density of the ionosphere. Show that the phase velocity $c_p$ is greater than $c$. Does it violate the principle of relativity?

4. The phase velocity c of transverse waves of wavelength $\lambda$ in a crystal is given by

$$c = c_0 \frac{\sin(\pi a/\lambda)}{(\pi a/\lambda)}$$

   where a is the average interatomic separation and $c_0$ is a constant. Show that the ratio of phase velocity and group velocity is given by

$$\frac{c}{c_g} = \frac{\tan(\pi a/\lambda)}{(\pi a/\lambda)}$$

5. Two harmonic waves represented by

$$\xi_1 = 3 \cos(7t - 12x)m \quad \text{and} \quad \xi_2 = 3 \cos(5t - 8x)m$$

are superposed to form a wave group. Find the group velocity.

6. A bridge is placed near the middle of a sonomter wire. When two parts are plucked simultaneously 3 beats per second are produced with a stretching load of 4 kg-wt. Find the number of beats produced when the load is increased to 16 kg-wt.

7. A string of length $l$ fixed at both ends has a fundamental frequency $v_0$. The string is plucked at a point $l/4$ from one end and at the same time the string is lightly touched at the mid-point. What is the lowest frequency excited in such a case? What will be the frequency of the next harmonic in this situation?

8. The stationary waves on a string is represented by

$$y(x, t) = 0.15 \sin 5x \cos 300t \ m.$$

Find the (i) amplitude of vibration at the antinode, (ii) distance between consecutive nodes, (iii) wavelength, (iv) frequency and (v) the speed of the wave.

9. Suppose one end ($x = 0$) of a long stretched string is given a sinusoidal motion with frequency 500Hz and amplitude 0.001 m. At time t = 0 the end has zero displacement and is moving towards + $y$ direction. At any given instant of time the displacement at $x = 0.1$ m is –0.005 m. and that at $x = 0.2$ m is + 0.005 m. Calculate the wavelength and velocity of the wave. Also, write down an equation for the wave.

10. Consider two strings of same length and smae material. Tensions in the two strings are in the ratio 4 : 1 and diameters of the wires are in the ratio 1 : 2. Compare the frequencies of the fundamental modes of vibration.

11. Two open organ pipes, one 1 m long and the other 1.05 m long, are sounded together. If the velocity of sound in air is 340 m/s how many beats per second will be produced between the fundamental tones?

12. An air column in a closed pipe of length 80cm and radius 5 cm produces resonarce with a tuning fork. If the speed of sound in iar is 340 m/s, what would be the frequency of the tuning form?

13. Which of the following represents a progressive wave and which one represents stationary waves :

(a) $y = 5 \sin 200\pi t \, e^{2\pi x}$

(b) $y = 5 \cos 2\pi x \tan 2\pi t$

(c) $y = 3\tan(\omega t - kx)$

(d) $y = 3\cos\left(2000\pi t - 0.02\pi x + 3\dfrac{\pi}{5}\right)$

(e) $y = 2e^{i\omega t}.\cos(0.2\pi x)$

14. The equation of a wave is $y = 20\cos\left(2000\pi t - 0.02\pi x + \dfrac{\pi}{3}\right)m$. Find the frequency, phase velocity and wavelength of the wave.

15. If the phase velocity of a transverse wave in crystal medium is $v_p = c\,\dfrac{\sin\dfrac{kd}{2}}{\left(\dfrac{kd}{2}\right)}$.

    where c = constant then find group velocity.

16. Why we do not hear beats when two waves of frequency $\gamma_1$ = 200 Hz and $\gamma_2$ = 250 Hz are superposed on each other.

17. Two waves are propagating along x-axis such that, they are $y_1 = 5\cos(2000\pi t - 0.04\pi x_1)$ and $y_2 = 5\cos(2000\pi t - 0.04\pi x_2)$. For what values of $x_2 - x_1$, we will get maximum intensity and minimum intensities.

## ANSWER

1.      $c = A\sqrt{\lambda}$

        $\dfrac{dc}{d\lambda} = \dfrac{1}{2}A\lambda^{-1/2}$

    Therefore     $c_g = c - \lambda\dfrac{dc}{d\lambda} = c - \dfrac{1}{2}A\sqrt{\lambda} = c - \dfrac{c}{2} = \dfrac{1}{2}c$

2. For no dispersion $\dfrac{dc}{d\lambda} = 0$

    Diferentiating the given expression with respect to $\lambda$ and putting $\dfrac{dc}{d\lambda} = 0$ and

    $\lambda = \lambda_0$ we get $\lambda_0 = 2\pi\sqrt{S/\rho g}$

For $\lambda << \lambda_0$, $c^2 = \dfrac{2\pi S}{\rho \lambda}$

or, $c = \dfrac{A}{\sqrt{\lambda}}$ where $A = \sqrt{2\pi S / \rho}$

$$\dfrac{dc}{d\lambda} = -\dfrac{1}{2}A\lambda^{-3/2}$$

Hence $c_g = c - \lambda \dfrac{dc}{d\lambda} = c + \dfrac{1}{2}A\lambda^{-1} = c + \dfrac{1}{2}c = \dfrac{3}{2}c$

For $\lambda >> \lambda_0$, $c^2 = \dfrac{g\lambda}{2\pi}$ or, $c = B\sqrt{\lambda}$ where $B = \sqrt{g / 2\pi}$

$$\dfrac{dc}{d\lambda} = \dfrac{1}{2}B\lambda^{-1/2}$$

Hence, $c_g = c - \lambda \dfrac{dc}{d\lambda} = c - \dfrac{1}{2}B\lambda^{1/2} = c - \dfrac{1}{2}c = \dfrac{1}{2}c$

$$\omega = \sqrt{\omega_p^2 + c^2 k^2}$$

∴ Phase velocity $c_p = \dfrac{\omega}{k} = \sqrt{c^2 + \dfrac{\omega_p^2}{k^2}}$ which is greater than $c$.

Differentiating the dispersion relation with respect to $k$ we get

$$2\omega \dfrac{d\omega}{dk} = c^2.2k$$

or, $\dfrac{\omega}{k} \cdot \dfrac{d\omega}{dk} = c^2$

or, $c_p c_g = c^2$ [since $c_p > c$, $c_g < c$]

It does not voilate the principle of relativity because energy or signal is transmitted not at the phase velocity but at group velocity which is less than $c$.

4.    $\dfrac{dc}{d\lambda} = c_0 \dfrac{\sin \pi a/\lambda}{\pi a} + c_0 \dfrac{\cos(\pi a/\lambda)}{(\pi a/\lambda)} \cdot \left(-\dfrac{\pi a}{\lambda^2}\right)$

Now $\quad c_g = c - \lambda \dfrac{dc}{d\lambda} = c - c_0 \dfrac{\sin(\pi a/\lambda)}{(\pi a/\lambda)} + c_0 \dfrac{\cos(\pi a/\lambda)}{(\pi a/\lambda)} \cdot (\pi a/\lambda)$

$$= c - c + c \frac{(\pi a/\lambda)}{\sin(\pi a/\lambda)} \cdot \frac{\cos(\pi a/\lambda)}{(\pi a/\lambda)} \cdot (\pi a/\lambda)$$

$$= c \cdot \frac{(\pi a/\lambda)}{\tan(\pi a/\lambda)}$$

$$\frac{c}{c_g} = \frac{\tan(\pi a/\lambda)}{(\pi a/\lambda)}$$

5. Angular frequencies are $\omega_1 = 7$ rad/s and $\omega_2 = 5$ rad/s. Wave numbers are $k_1 = 10$ m$^{-1}$ and $k_2 = 8$ m$^{-1}$.

$$\therefore \quad c_g = \frac{\Delta\omega}{\Delta k} = \frac{7-5}{10-8} = 1 \quad \text{m/s.}$$

6. Let the length of the sonometer wire be $2l$ and a bridge is placed near the middle such that the length of one part is $l + x$ and that of the other part is $l - x$ where $x \ll l$. Frequencies of the fundamental tones under a tension $T$ are

$$v_1 = \frac{1}{2(l+x)}\sqrt{\frac{T}{m}} \quad \text{and} \quad v_2 = \frac{1}{2(l-x)}\sqrt{\frac{T}{m}}$$

According to the problem $v_2 - v_1 = 3$ with $T = 4$ kg-wt.

Now, if $T$ is increased to $T' = 16$ kg-wt $= 4T$ then new fundamental frequencies would be

$$v_1' = \frac{1}{2(l+x)}\sqrt{\frac{T'}{m}} = \frac{1}{2(l+x)}\sqrt{\frac{4T}{m}} = 2v_1$$

Similarly $\quad v_2' = 2v_2$

$\therefore$ The number of beats per second would now be

$$v_2' - v_1' = 2(v_2 - v_1) = 2 \times 3 = 6$$

7. From Young's law all the harmonic having a node at the plucked point $l/4$ will be absent i.e. 4th. 8th, 12th, ...etc. harmonics are absent. Since the string is touched at the mid-point only those harmonics having a node at the mid-point can exist. Thus only the 2nd, 4th, 6th, ...etc. harmonics may be present. Out of

these 4th, 8th etc. modes are absent due to Young's law. So the lowest two frequencies excited would be $2v_0$ and $6v_0$.

8. (i) Comparing with the form

$$y = A \sin kx \cos (\omega t - \phi)$$

We get amplitude A = 0.15m

(ii) For modes $\sin 5x = 0$ or, $5x = 0$, $\pi$, $2\pi$ ... etc.

∴ The nodes are at $x = 0$, $\pi/5$, $2\pi/5$.

Therefore, the distance betwen two consecutive nodes is $\pi/5 = 0.628$ $m$.

(iii) Wevelength $\lambda = 2 \times$ distances between two consecutive nodes

$$= 2 \times \pi/5 = 1.26 \text{ m}$$

(iv) $2\pi v = \omega = 300$ or, $v = 300/2\pi = 47.7$ Hz.

9. Let $y(x,t) = A\sin(\omega t - kx)$        ... (1)

where A = 0.01 m, $\omega = 2\pi v = 2\pi \times 500 = 1000\pi$ rad/s and k = $2\pi/\lambda$.

It represents a wave travelling along $+ x$ direction and it satisfies the condition that y (0, 0) = 0.

Now        $y(0.1, t) = 0.01\sin(1000\pi t - 0.1k) = -0.005$

$$y(0.2, t) = 0.01\sin(1000\pi t - 0.2k) = +0.005$$

$$1000\pi t - 0.1k = \sin^{-1}(-1/2) = 7\pi/6 \qquad \text{... (2)}$$

$$1000\pi t - 0.2k = \sin^{-1}(1/2) = \pi/6 \qquad \text{... (3)}$$

Subtracting (3) from (2) we get

$$0.1\,k = \pi \text{ or, } 2\pi/\lambda = 10\pi \text{ or, } \lambda = 1/5 \text{ m} = 0.2 \text{ m}$$

Velocity $c = v.\lambda = 500 \times 0.2 = 100$ m/s

Putting A = 0.01 m, $\omega = 1000\ \pi$ and $k = 10\pi$ we get from (1) the equation for the wave as

$$y(x, t) = 0.01 \sin(1000\ \pi t - 10\ \pi x) \text{ m}$$

10. Fundamental frequency

$$v = \frac{1}{2l}\sqrt{\frac{T}{m}} = \frac{1}{2l}\sqrt{\frac{T}{\pi(D/2)^2\rho}}$$

where $T$ = tension, $D$ = diameter and $\rho$ = density

Now for the two strings,

$$\frac{v_1}{v_2} = \sqrt{\frac{T_1}{T_2} \cdot \frac{D_2^2}{D_1^2}} \quad [\because l \text{ and } \rho \text{ are the same}]$$

$$= \sqrt{\frac{4}{1} \cdot \frac{4}{1}} = \frac{4}{1}$$

11. In the fundamental mode in an open pipe the length $l = \lambda/2$ or $\lambda = 2l$ and frequency $v_1 = c/\lambda = c/2l$.

$$\therefore \text{ No of beats per second } = \frac{c}{2l_1} - \frac{c}{2l_2} = \frac{340}{2}\left[\frac{1}{1} - \frac{1}{1.05}\right] = 8.09$$

12. For resonance with the fundamental mode there is a node at the closed end and an antinode at the open end. Therefore, length of the pipe $l = \lambda/4$ or $\lambda = 4l$. If the end correction 0.6 r is taken into accounts, then $\lambda = 4(l + 0.6r)$

$$\therefore \text{ Frequency } v_1 = \frac{c}{4(l+0.6r)} = \frac{340}{4(0.8+0.6\times 0.05)} \text{ Hz} = 102.4 \text{ Hz}.$$

13. The general expression of a progressive wave is $y = a \sin(\omega t + kx)$ or $y = a \cos(\omega t - kx)$ i.e. both the variable must remain within a periodic function which must have finite values all the time. But for stationary wave the variable $x$ and $t$ are in separate finite periodic function. i.e. $y = A \sin \omega t \cos kx$ or, $A \cos \omega t \sin kx$ etc.

Thus (d) $\Rightarrow y = 3\cos\left\{2000\pi t - 0.02\pi x + \frac{3\pi}{5}\right\}$ represents progressive wave.

and (e) $y = 2\cos(0.2\pi x)e^{i\omega t}$ represents stationary wave.

14. Here $\omega = 2000\pi = 2\pi\gamma$

$$\therefore \quad \gamma = 1000 \text{ Hz. phase velocities } v_p = \frac{\omega}{R} = \frac{2000\pi}{0.02\pi} = 100,000 \text{ m/sec}$$

and $k = \frac{2\pi}{\lambda} = 0.02\pi$

$$\therefore \quad \lambda = \frac{2}{0.02} = 100 \text{ m}$$

15. $v_g = \frac{d\omega}{dk}$, $\frac{\omega}{k} = v_p$ $\therefore v_g = \frac{d}{dk}(kv_p) = v_p + k\frac{dv_p}{dk}$

16. We know that due to persistance of hearing we will not hear beats greater than 10 Where.

17. For maximum frequency the path difference of two waves must be $x_2 - x_1 = 2n.\dfrac{\lambda}{2}$.

    Where $n = 0, 1, 2, 3 \dots$

    Here $\dfrac{2\pi}{\lambda} = k = 0.04\pi$ $\therefore$ $\lambda = \dfrac{2}{0.04} = 50\,m$

    Hence for $x_2 - x_1 = 2n\dfrac{50}{2} = 50n$ metre where $n = 0, 1, 2, 3\dots$

    We get maximum

    For minimum intensity $x_2 - x_1 = (2n+1)\dfrac{\lambda}{2}$

# Unit-8 : Wave Optics

**Structure**

## 8.1   Objectives

In going through this unit, you will learn :

1. Electromagnetic nature of light
2. Definition and properties of wavefront

3.  Huygens' Principle and propagation of wavefronts

4.  Wave theory of reflection and refraction both at a plane and spherical surface

5.  Concept of coherence - temporal and spatial

6.  Coherence length and coherence time

7.  Line width : Purity of spectral line

10. Visibility of fringes and degree of coherence

## 8.2  Introduction

Optics is a branch of Physics in which the nature and properties of light are studied. The term light is commonly used to mean radiant energy that causes the sensation of vision. However, at present the term light is used to mean all kinds of radiations, both visible and invisible. The subject of optics is conveniently divided into three distinct branches :

1.  Geometrical or ray optics

2.  Physical or wave optics

3.  Quantum optics

In **geometrical optics**, many basic principles concerning light are studied by geometrical methods, without assuming anything regarding the nature of light. It assumes rectilinear propagation of light which is considered as the ray of light. Geometrical optics is therefore also known as **ray optics**.

In **Physical optics** many experimental results are explained by considering primarily the wave nature of light and is thus called the **wave optics**.

**Quantum optics** deals with the interaction of light with atomic entities of matter. Here one considers the particle nature, the so-called **photons**, and exact analysis, one must take recourse to the method of quantum mechanics.

To understand different optical phenomena, we are to discuss the **nature of light**. Historically different theories have been proposed regarding the nature of light. In 1675 Newton first proposed the **corpuscular theory** of light. The theory suggests that light consists of a stream of minute particles that move in straight lines with extremely high but finite speed. Different colours are due to different size of the corpuscles. While this model could explain many of the facts of geometrical optics, it failed to explain phenomena like interference, diffraction, and polarization etc.

The **mechanical wave model** of light, in contradistinction to the corpuscular theory of Newton, was first proposed by the Dutch physicist C. Huygens in 1690.

Huygens assumed that, like sound waves, light waves were also longitudinal in character. But this supposition was subsequently proved to be wrong. Though wave theory was extremely successful in explaining the phenomenon of interference, and diffraction, there was no satisfactory answer to the objection raised by Newton as to the nature of light wave and of the medium through which it travels. It also failed to explain the phenomena of polarization. To overcome these difficulties Fresnel and Young introduced the concept of **transverse** waves and Fresnel's idea of **half period zones**.

All the above difficulties were finally resolved in 1864 by the English physicist Clark **Maxwell** when he put forward his **electromagnetic theory of light** considering **transverseelectromagnetic wave**.

We shall keep in mind that the modern concept (which includes quantum theory) of light, pioneered by M. Planck, A. Einstein, N. Bohr, and P.A.M. Dirac, says that electromagnetic energy is *quantized* which means that light can only be imparted to or taken from the electromagnetic field in discreate amounts called *photons*. Light is said to have a *dual* nature. When the wavelength of electromagnetic wave becomes very small compared to the dimension of the material with which it interacts, as in the case of x-rays, and $\gamma$-rays, it exhibits the particle aspect of light. The electromagnetic theory of Maxwell and the quantum theory of Heisenberg-Schrödinger has been combined by Dirac and is known as *quantum electrodynamics.*

With this modern concept we can fairly assume that the nature of light is well understood, at least within the context of a mathematical framework that accurately accounts for present experimental observations. The question as to the **true** or **ultimate** nature of light, although yet **unanswered**, is quite irrelevant to the study of *optical phenomena.*

There are many physical processes that can be described in terms of wave motion. Throughout the study of optics, one uses the idea of plane, spherical and cylindrical waves. We learned in **Part-I** some simple ideas regarding the propagation of **disturbance** through any medium. Here the term disturbance is used in a general sense; it may be the electric field of an electromagnetic wave or a particle displacement in any medium.

Finally, we consider **Huygens' principle** and its applications. It is important to point out that Huygens' principle is an approximate descriptions of wave propagation in three dimension which contains some physical truth and at same time avoids mathematical complexities.

**Coherence** is the key concept in the study of **interference** phenomena. It should be kept in mind that monochromatic light wave, unlike radio or micro - waves, is a mathematical artifice and not a physical reality. Actual light disturbances are not as

simple as the sinusoidal variations with constant amplitude and phase. **Laser** source may approximate such disturbances, but not the light from candles, incandescent bulbs, or stars. These lights are the resultant of a large numbers of amplitudes from individual atomic emissions, which occur independently. Light from a real physical source is never strictly monochromatic. Even the sharpest spectral line (viz., green of a Hg-vapour, yellow of Na-vapour, etc.) has a finite **line width**. Moreover, a physical source is never a point source, but has a finite dimension, consisting of billions of elementary **radiators** (atoms). According to **Fourier's theorem**, the disturbance produced by such a source may be expressed as the sum of strictly monochromatic and, therefore, of infinitely long wave trains. The elementary **monochromatic theory** is, then, essentially concerned with a single component of this **Fourier representation**, expecting for the best that the amplitudes of the higher harmonics vanish quickly.

The fluctuations in the two beams originating in the same source are, in general correlated, and the beams are said to be **completely or partially coherent** depending on whether the correlation is complete or partial. The fluctuations are completely uncorrelated in beams from different sources, and the beams are said to be **mutually incoherent**. No interference is seen under ordinary experimental conditions if mutually incoherent beams are superposed, and the total intensity will be everywhere the sum of the intensities of the individual beams.

Hence, **coherence** of a wave describes the accuracy with which it can be represented by a pure sine wave for which the wave vector **k** and the frequency $\omega$ can be exactly defined.

## 8.3 Electromagnetic nature of light

According to **Maxwell's electromagnetic (EM) theory of light** waves are transverse electromagnetic waves consisted of varying and coupled electric and magnetic fields, perpendicular to each other and to the direction of propagation of the wave. A changing magnetic field produces a changing electric field and a changing electric field in turn produces a changing magnetic field. This makes the propagation of EM wave possible even through vacuum. Hertz first experimentally verified the existence of such EM waves. The **velocity of propagation** of an EM wave is given by $\sqrt{1/\mu\varepsilon}$, where $\mu$ is the permeability and $\varepsilon$ the permittivity of the medium. Since $\varepsilon$ and $\mu$ have definite values for vacuum, it was unnecessary to assume the existence of the hypothetical ether medium.

Here we will not discuss Maxwell's EM theory in detail. We will just mention that these equations can be written in *differential and integral forms.*

## 8.4  Definition and properties of wavefront

**Definition of Wavefront**–*a wavefront is defined as a surface on which the phase of disturbance is the same at any given instant of time.* All points on it are in the same state of vibration (Fig. 8.1).
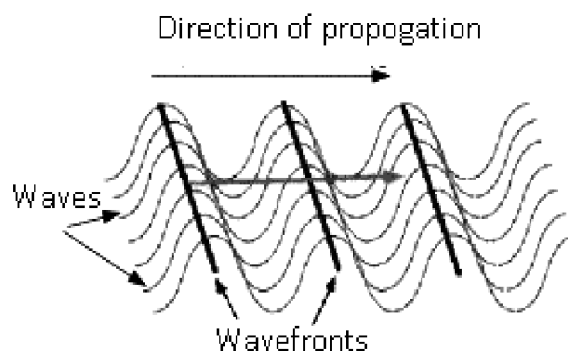


**Fig. 8.1**

*The direction along which the disturbance propagates in a homogeneous medium is called the ray.* So, *a ray is always normal to the wavefront.*

There are three types of wavefronts :

1.  Spherical wavefront (spherical in shape)
2.  Plane Wavefront (linear in shape)
3.  Cylindrical wavefront (cylinder in shape)

**Spherical Wavefront** (See Fig. 8.2a)

➢   When the source of light is a point source the wavefront formed will be spherical wavefront.

➢   Point source means the source of light is so small that it is considered as point. It can be considered as dimensionless.

➢   For example : Ripples in water are in the form of concentric circles which are spherical wavefronts.

**Plane Wavefront** (See Fig. 8.2b)

➢   When the small part of a spherical or cylindrical wavefront originates from a distant source like infinity then the wavefront which is obtained is known as plane wavefront.

➢   For example: Rays coming from infinity like Sun.

### Cylindrical Wavefront

➢ When the source of disturbance is a slit (i.e., line source) then the wavefront is cylindrical because all the points are equidistant from the source, and they lie on the surface of the cylinder.

➢ For example: When rays of light fall on a lens after coming out of lens, they will converge at a point.

➢ The waves are bending and converging at a point, so the shape of the wavefront is in the form of cylinder.

➢ Many concentric circles are formed, and the wavefront is in the form of cylinder.

### Properties of Wavefront :

1. It is the locus of all points having same phase at a given instant of time.

2. The shape of wavefront depends on the shape of the source of disturbance.

3. A wavefront is always normal to the light rays.

4. A wavefront does not propagate in the backward direction.

## 8.5 Huygens' principle of wavefront propagation

According to Huygens, if we have a point light source at S in a homogeneous isotropic medium sends out light waves in every direction and theses waves travel with equal velocity to carry energy with them to be transmitted in all directions and will, therefore, reach simultaneously to all points lying on the surface of a sphere drawn with the S as centre (Fig. 8.2a). Thus, the wavefront is spherical in this case. After a short interval of time $\Delta t$, all the particles in the medium will be on the spherical surface AB and lie in the same state of vibration, i.e., the particles will be in the same phase. Plainly, AB is drawn with S as centre and radius $c\Delta t$, where c is the velocity of propagation of the wave. The surface AB is called the primary wavefront. At long distances it may be considered as a plane wavefront (Fig. 8.2b). As the wave advances, the wavefront moves parallel to itself in a direction normal to the wavefront.

Now, an important question arises. Suppose we know the wavefront AB at an instant t. Then how to construct the wavefront at a later instant $t + \delta t$ ? The answer to this question is the **Huygens' principle**. According to this principle, *in an isotropic homogeneous medium, each point on the primary wavefront serves as the centre of a new disturbance that sends out secondary wavefronts. So, a series of secondary wavelets propagate through space in all directions with the same velocity and frequency*

*as* those of the primary wave. After an interval of time, *the envelope of the secondary wavelets gives the new wavefront.*
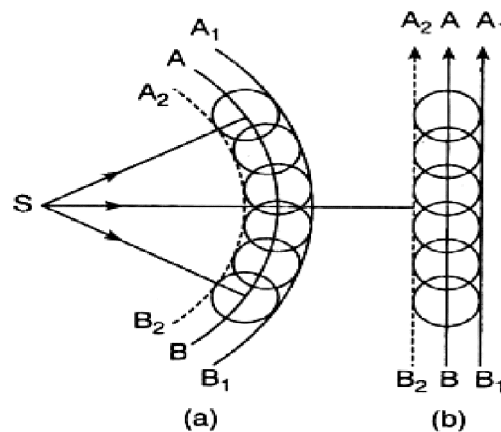


**Fig. 8.2**

So, to locate the position of the new wavefront after an interval $\Delta t$, several points are taken on *AB* and with each point as centre and radius $c\Delta t$, spheres are drawn in turn. These are the secondary wavelets, starting from those points respectively. The sections of these secondary waves by a plane will be circles as shown in figure (Fig. 8.2). The surface $A_1B_1$ touching all such spheres in the **forward direction**, is the *new wavefront.* This is how one wavefront can be constructed from another.

■  In case the medium is not homogeneous and isotropic, appropriate velocity is to be used in different points and directions to construct the new wavefront.

■  Huygens' principle is open to **two objections** : (i) Why does not a wavefront propagate backward giving rise to $A_2B_2$? (ii) What is the justification of ignoring arbitrarily the large part of the secondary waves not tangent to the enveloping surface?

Subsequently, Fresnel considered the superposition and interference of the wavelets and showed that they are effective at the points of tangency only. And Kirchhoff showed mathematically that the amplitude due to a wavelet in a direction making an angle $\theta$ with the ray is proportional to $(1 + \cos\theta)$. For backwave, $\theta = \pi$ which gives zero amplitude. Hence the wavefront does not propagate backward. The quantity $(1 + \cos\theta)$ is often referred to as *obliquity factor.*

## 8.6 Laws of reflection from the wave theory

Let *PQ* be the reflecting surface and *AB* the plane wavefront incident on it (Fig. 8.3). When the wavefront strikes the reflector at *A* at an angle *i* with it (wavefront and reflector being perpendicular to the plane of paper), *i* is the *angle of incidence*, being equal to the angle between the normal and the incident ray (direction of arrows) at *A*. Now A becomes the source of secondary wavelets tending to spread out in surrounding space. By the time the disturbance from *B* reaches the reflector at *C*, the secondary waves from *A* will have a radius *BC*. Draw a sphere with *A* as centre and radius = *BC*. A plane through *C*, touching the sphere at *D*, then gives *reflected wavefront CD*.
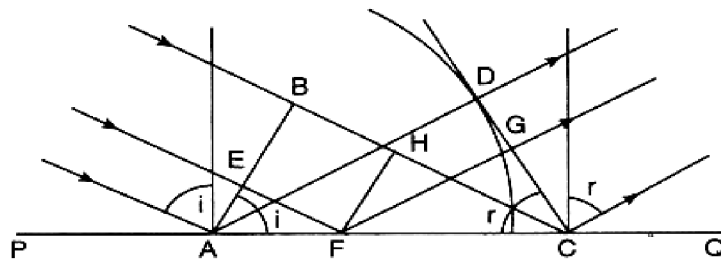


**Fig. 8.3**

The disturbance from any point *E* on the incident wavefront, after reflection from the reflector at *F*, must then reach the reflected wavefront in the time the disturbance from *A* reaches *D*. If *FG* is drawn perpendicular to *CD*, then *EF* + *FG* must equal *AD* (or *BC*). We shall now prove this to ensure that *CD* is the true reflected wavefront.

**Proof.** Draw *FH* ⊥ *BC* and consider right-angled $\triangle ABC$ and $\triangle ADC$. Since *AD* = *BC* and *AC* is common,

$$\therefore \quad \triangle ABC = \triangle ADC \Rightarrow \angle BAC = \angle DCA = r \qquad \qquad ... (8.1)$$

Again, $FH \parallel AB \Rightarrow \angle HFC = \angle BAC = \angle DCA = \angle GCF$.

Now in right angled $\triangle FHC$ and $\triangle FGC$, $\angle GCF = \angle HFC$ and *FC* is common.

$$\therefore \quad \triangle HFC \equiv \triangle FGC \Rightarrow FG = HC$$

$$\therefore \quad EF + FG = BH + HC = BC. \text{ So, } CD \text{ is the } true \text{ reflected wavefront.}$$

Now, i = $\angle BAC$ = $\angle DCA$ = r $\Rightarrow$ *i* = *r*, i.e., **angle of incidence = angle of reflection**.

Also, plainly, from the very construction the *incident ray*, the *reflected ray and the normal at the point of incidence, all lie in the same plane.*

Hence the **laws of reflection** are *proved* from wave theory.

## 8.7 Laws of refraction from the wave theory

Let a plane wavefront *AB* be incident on the plane refracting surface *PQ* that separates two media *a* and *b* (Fig. 8.4) in which the velocities of the waves are $c_1$ and $c_2$ ($c_1 > c_2$).
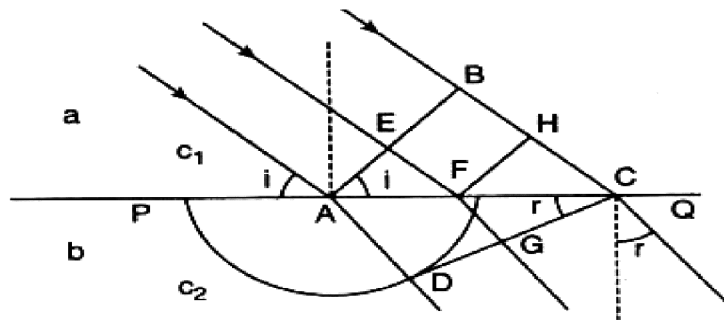


**Fig. 8.4**

When the point *A* of the wavefront touches the refracting surface, the point *B* is away from it and by the time the disturbance from *B* reaches *C* at the surface of separation, secondary waves from *A* have acquired a radius *AD* in the medium b.

$$\therefore \quad \frac{AD}{c_2} = \frac{BC}{c_1} \qquad\qquad ...(8.2)$$

Draw a sphere with *A* as centre and radius *AD*. The tangent plane *CD* through *C* and touching the sphere gives the *refracted wavefront*. To confirm that *CD* is the *true refracted* wavefront, the disturbance from any point *E* on the incident wavefront should, after refraction at *F* on the surface of separation, reach *G* on *CD* at the time the disturbance from *A* reaches *D*, i.e., $EF / c_1 + FG / c_2$ must equal $AD/c_2$ (or $BC/c_1$). We shall now prove this.

Proof. Draw *FH* $\perp$ *BC*. But *BC* = *BH* + *HC*

$$\therefore \quad \frac{BC}{c_1} = \frac{BH}{c_1} + \frac{HC}{c_1} = \frac{EF}{c_1} + \frac{HC}{c_1} \qquad (\because BH = EF)$$

Suffice then if we show : $HC/c_1 = FG/c_2$. Now $\Delta ABC$ and $\Delta FHC$ are similar. Also, $\Delta ADC$ and $\Delta FGC$ are similar.

$$\therefore \quad \frac{FC}{AC} = \frac{HC}{BC} \; ; \; \frac{FC}{AC} = \frac{FG}{AD} \qquad \qquad \dots (8.3)$$

$$\Rightarrow \frac{HC}{BC} = \frac{FG}{AD} \Rightarrow \frac{AD}{BC} = \frac{FG}{HC}$$

$\therefore$ Using (8.2), we obtain

$$\frac{c_2}{c_1} = \frac{FG}{GC} \Rightarrow \frac{HC}{c_1} = \frac{FG}{c_2}, \text{ as required}$$

It is thus confirmed that $CD$ is the *true refracted wavefront.*

Also, $\angle BAC = i$ = angle of incidence; $\angle ACD = r$ = angle of refraction.

$$\therefore \quad \frac{\sin i}{\sin r} = \frac{\sin BAC}{\sin ACD} = \frac{BC}{AC} \cdot \frac{AC}{AD} = \frac{BC}{AD} = \frac{c_1}{c_2} = \text{constant} \qquad \dots (8.4)$$

Using the relation (8.2).

This is **Snell's law**. The constant is called the *refractive index,* $a^n b$, of the medium $b$ with respect to medium $a$.

Plainly, from the manner of construction, the incident ray, the refracted ray and the normal to the surface of separation of the media at the point of incidence, all lie in the same plane. Thus, the *laws of refraction* are established from the wave theory.

From (1.4), we may write that

$$a^n b = \frac{\text{velocity of light in medium a}}{\text{velocity of light in medium b}}$$

If the medium $a$ is vacuum, $a^n b$ is written as $n_b$ or simply $n$, the refractive index of medium $b$.

$$\therefore \quad n \text{ (of a medium)} = \frac{\text{velocity of light in vacuum}}{\text{velocity of light in medium}}$$

The refractive index $n$ of a medium does not depend on the *frequency* of the wave, for when light travels through a medium, it is the wavelength that changes but the frequency remains unaltered.

## 8.8 Reflection of spherical wave at a plane surface

Let *LOS* be a *spherical* wavefront that diverges from the point source *P* to be incident at *O* on a plane *reflector AOB* (Fig. 8.5). In the absence of the reflector, the wavefront after a time $\Delta t$ would have been in the position *AMB*. Thus, $OM = v\Delta t$ where *v* is the velocity of light in the incident medium.
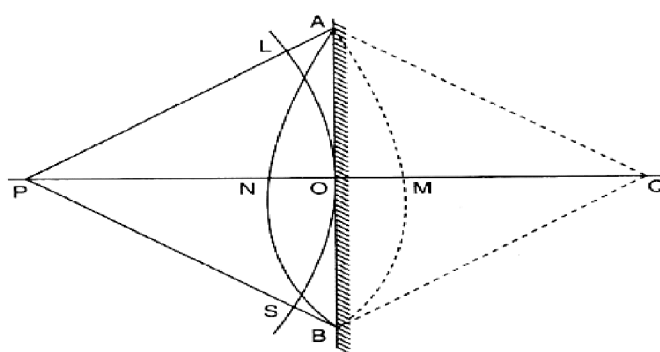


**Fig. 8.5**

The presence of the reflector however causes the secondary wavelets generated at *L,O* and *S* on the wavefront *LOS* to be at *A, N* and *B* respectively after time $\Delta t$. Thus, the common envelope of these secondary waves, i.e., *ANB* will be the *reflected wave front* after time $\Delta t$. We thus have

$$M = ON = AL$$
$$= BS = v\Delta t$$

Since the medium is homogeneous, the *reflected wavefront* ANB will also be spherical and appear to diverge from Q. Q is thus the *image* of P.

From geometry of Fig. 8.5, we obtain,

$$ON = AO^2/2QN$$

And $\quad OM = AO^2/2PM.$

Since, $ON = OM$, we have, from above,

$$MP = QN$$
$$\Rightarrow OM + OP = ON + OQ$$
$$\Rightarrow OP = OQ \ (\because OM = ON)$$

∴ **Object distance = Image distance** (For a plane mirror)

## 8.9 Reflection of spherical wave at a spherical surface

Let a spherical waveform LBS diverge from a point object P and be incident on a concave spherical reflector LAS. The aperture of the reflector is small, its pole is at A and the centre of curvature is at C (Fig. 8.6).
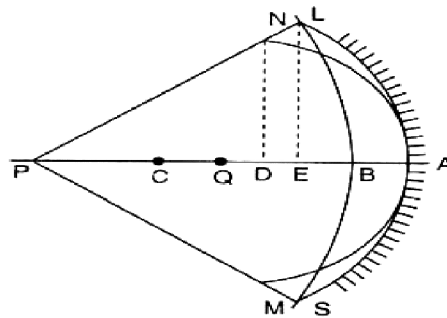


**Fig. 8.6**

In time-interval $\Delta t$, say, the secondary wave from $B$ reaches $A$, when those from $L$ *and* $S$ arrive, after reflection from the reflector, at $N$ and $M$ respectively.

$$\therefore \quad AB = LN = SM = v\Delta t$$

where $\vartheta$ is the velocity of light in the incident medium.

*NAM*, the common envelope of the reflected secondary waves represents the *reflected wavefront* after time $\Delta t$. Like *LBS*, the reflected wavefront *NAM* will also be *spherical* since the object and image spaces are homogeneous and converge at $Q$ to form a real image. Draw $LE \perp AP$ and $ND \perp AP$. $LE$ and $ND$ are essentially equal.

Now, from geometry, we obtain

$$BE = LE^2 / 2BP = LE^2 / (-2u) \qquad \qquad \ldots (8.5)$$

$$\left( \because BP \simeq AP = -u \right)$$

$$AE = LE^2 / 2AC = LE^2 / (-2r) \qquad \qquad \ldots (8.6)$$

$$AD = DN^2 / 2AQ = DN^2 / (-2\vartheta) \qquad \qquad \ldots (8.7)$$

Now, from figure, we obtain,

$$AD = AE + ED = AE + AB = AE + (AE - BE)$$

$$\text{or,} \quad AD + BE = 2AE \quad (\because \quad ED = LN = AB, \text{ practically})$$

or, $\dfrac{DN^2}{-2v} + \dfrac{LE^2}{-2u} = 2.\dfrac{LE^2}{-2r}$

or, $\boxed{\dfrac{1}{v} + \dfrac{1}{u} = \dfrac{2}{r}}$

which is the relation for reflection of light in a *concave* spherical reflector.

■ The students are advised to work out the corresponding relation for reflection in a *convex* spherical reflector.

■ The students should study the **"Total internal reflection from wave theory"**. In this context they may consult any textbook on light.

■ Note that the following *two conditions* must be satisfied for total reflection.

1. *The ray must travel from a denser to a rarer medium.*

2. *The angle of incidence* in the denser medium *must be greater than the critical angle* for the pair of media.

## 8.10 Refraction of spherical wave at a spherical surface

In Fig. 8.7 let $CAB$ be a *concave* spherical refracting surface separating the media of refractive indices $n_1$ and $n_2$ $(n_2 > n_1)$. $P$ is the centre of curvature of surface and $r$ ( $= PA$) the radius of curvature.
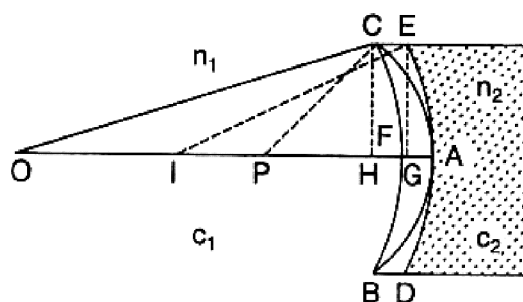


**Fig. 8.7**

Let $O$ be a point object on the principal axis $OA$ and $BFC$ the portion of spherical wave touching the surface at $C$ and $B$.

By Huygens' construction, each point on the wavefront $BFC$ is a source of secondary waves. While $B$ and $C$ send disturbances into medium $n_2$, the other points send them

partly in medium $n_1$ and then in medium $n_2$. If $t$ be the time taken by the disturbance from $F$ to reach $A$ in medium $n_1$, then at the same time, the disturbance from $C$ reaches, say $E$, in medium $n_2$.

$$\therefore \quad \frac{AF}{c_1} = \frac{CE}{c_2} \Rightarrow CE = AF \times \frac{c_2}{c_2} = AF \times \frac{n_1}{n_2} \qquad \qquad ... (8.8)$$

where $c_1$, $c_2$ are wave velocities in medium $n_1$ and medium $n_2$ and $c_1/c_2 = n_1/n_2$.

In the same time $t$, the disturbance from $B$ reaches $D$ in medium $n_2$. So, after time $t$, the disturbance from all points on *incident wavefront* will be on $DAE$, the *refracted wavefront*, originating from $I$. $I$ is thus the virtual image $O$.

Draw $CH \perp OA$, $EG \perp OA$. Clearly, $CH = EG = y$, say.

Now, $AO = u$, $AI = v$ and $AP = r$. Also, from geometry, we have

$$CH^2 = AH \times 2r \Rightarrow AH = CH^2/2r = y^2/2r) \qquad ... (8.9)$$

Similarly, $FH = y^2 / 2u$ and $AG = y^2/2v$ $\qquad ... (8.10)$

From Fig. 8.7, $CE = HG$ and so from (8.8)

$$n_1.AF = n_2.CE = n_2.HG$$

$$\therefore \quad n_1 \left( AH - HF \right) = n_2 \left( AH - AG \right)$$

$$\Rightarrow \quad n_1 \frac{y^2}{2r} - n_1 \frac{y^2}{2u} = n_2 \frac{y^2}{2r} - n_2 \frac{y^2}{2v}, \quad \text{using (8.9) and (8.10)}$$

$$\Rightarrow \quad \boxed{\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{r}} \qquad \qquad ... (1.11)$$

which is the required relation; here $u$, $v$, $r$ are all negative by sign convention.

■ When $n_1 = 1$ (i.e., air), the boxed equation (8.11) takes the usual form

$$\frac{n}{v} - \frac{1}{u} = \frac{n-1}{r}, \quad (n_2 = n)$$

■ The students are advised to work out the corresponding relation for refraction in a convex spherical refracting surface.

## 8.11 Refraction of spherical wave through thin lens

In Fig. 8.8 let $A$ and $D$, be the poles of two spherical surfaces of radii of curvature $r_1$ and $r_2$ respectively of the thin *convex* lens $L$ of index $n_2$. Let $O$ be a point object in the surrounding medium of index $n_1$, on the principal axis, giving out spherical

waves. Also, let $c_1$, $c_2$ be the wave-velocities in medium $n_1$ and $n_2$ $(n_2 > n_1 \Rightarrow c_1 > c_2)$ and $BAC$ a portion of a spherical wavefront touching the pole $A$ at an instant.

By Huygens' principle, $A$ becomes the origin of secondary waves that move over the thickness $AD$ in lens-medium in time t. So, $t = AD/c_2$. The disturbance from $C$ and $B$ travels in medium $n_1$ and reach, say $F$ and $E$ respectively following paths $CF$ and $BE$ in time $t$. So, $t = CF/c_1$ and $FDE$ is the spherical wavefront *after refraction* through the lens and converging to $I$, which is the *real image* of the object $O$.
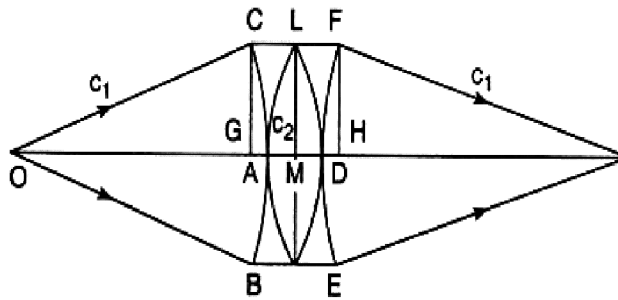


**Fig. 8.8**

Now, $\quad \dfrac{AD}{c_2} = \dfrac{CF}{c_1} = t \;\Rightarrow CF = \left(c_1/c_2\right)AD = \left(n_2/n_1\right)AD$ $\qquad\qquad$ ... (8.12)

Draw, $\quad CG \perp OA, \quad LM \perp OI$.

But, simply,

$\qquad GH = CF = (n_2/n_1)AD$, using (8.12)

$\therefore\quad n_1(GA + AD + DH) = n_2 AD$

$\Rightarrow\quad n_1\left(GA + DH\right) = \left(n_2 - n_1\right)AD = \left(n_2 - n_1\right)\left(AM + MD\right)$ $\qquad$ ... (8.13)

Now, $\qquad AO = u$, $ID = v$ and $CG = LM = FH = y$, say. Also, from geometry

$\qquad GA = y^2/2u; \quad AM = y^2/2r_1; \quad MD = y^2/2r_2$ and $DH = y^2/2v$

Substituting the above values in (8.13), we obtain

$\qquad n_1\ [y^2/2u + y^2/2v] = (n_2 - n_1)[y^2/2r_1 + y^2/2r_2]$

$\Rightarrow\quad \dfrac{1}{u} + \dfrac{1}{v} = \left(\dfrac{n_2 - n_1}{n_1}\right)\left(\dfrac{1}{r_1} + \dfrac{1}{r_2}\right)$

By sign convention, $v$ and $r_1$ are positive and $u$ and $r_2$ are negative

$$\therefore \quad \frac{1}{v} - \frac{1}{u} = \left(\frac{n_2 - n_1}{n_1}\right)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad \text{... (8.14)}$$

When u = ∞ (i.e., object is at infinity), $v = f$ (i.e., image is at the focus : second focal length).

$$\therefore \quad \boxed{\frac{1}{f} = \left(\frac{n_2}{n_1} - 1\right)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)} \qquad \text{... (8.15)}$$

From (8.14) and (8.15), we obtain

$$\boxed{\frac{1}{v} - \frac{1}{u} = \frac{1}{f}} \qquad \text{... (8.16)}$$

the usual lens formula.

If the surrounding medium be air, $n_1 = 1$ and from (8.15) using $n_2 = n$,

$$\boxed{\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)} \qquad \text{... (8.17)}$$

the **lens-maker's formula**.

- ■ The students are advised to work out the formula corresponding to thin *concave lens.*

- ■ It is to be noted that Huygens extended his principle to explain the phenomenon of **double refraction** in the case of *uniaxial crystal* (e.g., *calcite*). He used his new ideas regarding the ordinary and extraordinary wavelets to find the direction of the refracted rays inside the crystal.

## 8.12  Limitations of Huygens' principle

We have seen that Huygens' principle, in its original form, was able to give a satisfactory account of the laws of reflection, refraction and double refraction. It enabled a series of wave surfaces to be constructed when one was given. However, Huygens' principle by itself, was insufficient to enable the distribution of illumination in diffraction patterns (Unit-11) to be calculated.

We have also seen that Huygens' principle, by itself, cannot explain the phenomena of non-occurrence of backwave. This was avoided by an arbitrary, at the time of Huygens, assumption that the amplitude of the secondary wavelets was not uniform in all directions; it was maximum in the forward direction and zero in the backward direction. Almost two hundred years later, this assumption was found to be correct.

At the time of Huygens, light was known to travel in straight lines and Huygens explained this by assuming that the secondary wavelets do not have any amplitude at any point not enveloped by the wavefront which was again open to criticism.

The occurrence of illumination in the geometrical shadow was later (after 150 years) explained satisfactorily be Fresnel on the postulate that the secondary wavelets mutually interfere. At the time of Newton and Huygens, the phenomenon of interference of light was not known. In fact, the formation of Newton's rings, a phenomenon observed by Newton,was not explained either by Newton or by Huygens.

Huygens' principle gained perfection in three stages (distributed over two hundred years).

- Stage I : Solely contributed by Huygens.

- Stage II : Fresnel and his followers extended Huygens' principle. This is what is known as *Huygens-Fresnel principle.*

- Stage III : *Fresnel-Kirchhoff integral formula. This formula gives us spherical monochromatic waves travelling outward from the source. Finally, this equation has a significant importance in the theory of diffraction.*

## 8.13 Geometrical optics as a limit of wave optics

Geometrical optics constitutes our day-to-day guide to the outside world. It is the basis for construction of the image-forming devices like mirrors, telescopes, photographic and microscopic lenses, etc. However, geometrical optics does not embrace such studies as the nature of light, the mode of its propagation or such phenomena as interference, diffraction, and polarization, which form the subject of study in *physical optics* or more precisely wave optics.

It can be shown that

- *geometrical optics is the limiting case of wave optics as the wavelength of light* $\lambda \to 0$.

- *In the limiting case as* $\lambda \to 0$ *there is no diffraction i.e., light travels in straight line.*

## 8.14 Concept of coherence

If a wave appears to be a pure sine wave for an infinitely extended time or an infinitely extended space, it is called a *perfectly coherent wave*. The *characteristic* of such a wave is that a definite phase relationship at different points of time and space. *A perfectly coherent wave* however is an *idealization*, for a real light source always emits light in short pulses. Light waves that are pure sine waves only in a limited space or for a limited time are only *partially coherent waves*. For a source to be coherent it must emit radiation of single frequency, or the frequency spread must be very small, and the wavefront must have shape which remains constant in time.

From above, it is apparent that there are two different criteria as regards coherence - the criterion of *time* or **temporal coherence** and the criterion of space or **spatial coherence**.

## 8.15 Temporal coherence

The oscillating electric field $E$ of a perfectly coherent light wave has a *constant* amplitude at any point in space, but its phase will vary with time linearly. As a function of $t$, the field $E$ will appear as illustrated in Fig. 8.9. Obviously, it is an ideal sinusoidal function of time and does not agree with real cases.
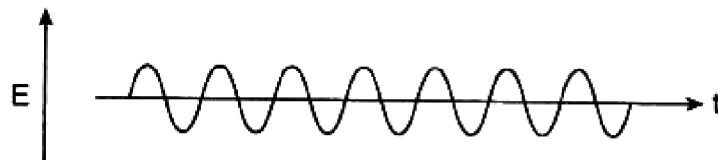


**Fig. 8.9**

No real source can ever produce an ideal sinusoidal field for all time, since when an excited atom makes a transition back to its initial state, it emits a light pulse of extremely short duration $\Delta t \sim 10^{-8}$ s. The field therefore is sinusoidal for an interval of $\Delta t \sim 10^{-8}$ s only. It undergoes abrupt phase changes thereafter for, different atoms of the source emit radiation in a random fashion. Fig. 8.10 illustrates the field due to a real source of light.
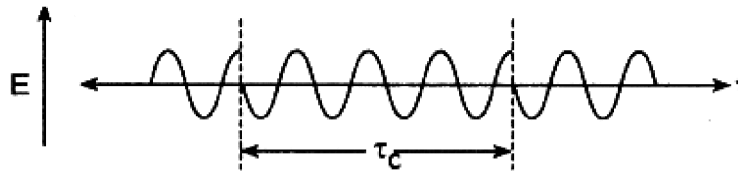
**Fig. 8.10**

**Coherence time :** *It is the average time-interval for which the wavefield of light vector remains sinusoidal, and therefore a definite phase relationship exists. It is also termed as temporal* **coherence** of the light beam and symbolized by $\tau_c$ (or $\Delta t$).

**Coherence length ($L_c$) :** *The distance over which the wave field is sinusoidal* and is given by

$$L_c = \tau_c \times c \qquad \qquad ... (8.18)$$

where $c$ is the speed of light in vacuum.

In an interference experiment we consider two wave fields arriving at a particular space point from a point source via two different optical paths. If in any interference experiment the path difference between the two interfering beams becomes greater than $L_c$ then the wavetrains are derived from two different wavetrains having no definite phase relationship. So, no definite interference pattern will be observed in this case.

Temporal coherence can be related to the line width. A Fourier analysis of the wavetrains of finite duration shows that it is equivalent to many harmonic waves having frequencies within a certain interval $\Delta v$ about a central frequency ($v$). The frequency spread or line bandwidth $\Delta v$ is related to the coherence time $\tau_c$ by the relation

$$\Delta v = \frac{1}{\tau_c}$$

Thus, narrow bandwidth implies long coherence time. For a perfectly monochromatic wave $\Delta v = 0$ and $\tau_c = \infty$. Thus, a monochromatic wavefield is truly temporally coherent.

**Illustration :** In order to clarify the concept of temporal coherence let us consider the *Michelson interferometer experiment* (Fig. 8.11) about which we shall read in Unit-10 *Interferometers*.

The two beams 1 and 2 can produce stationary interference pattern only when there exists a definite phase relationship between them. With $M_2{}'$, the image of the mirror

$M_2$, the Michelson's arrangement is equivalent to an air-film sandwiched between two parallel reflecting surfaces–$M_2'$ and the mirror surface $M_1$. If $d$ be the separation between $M_1$ and $M_2$, $2d$ will be the path difference between the two interfering beams. Now, if

$$2d \ll L_c$$

then a definite phase relationship exists between the two beams and well-defined interference fringes are observed. On the other hand, if

$$2d \gg L_c$$

then, in general, there is no definite phase relationship between the two beams and no interference pattern is observed.

So, starting with equal path lengths, if by moving one of the mirrors, the separation d is gradually increased, the fringes become poorer and poorer in contrast and eventually at a certain stage completely disappear. The path difference corresponding to the disappearance gives an estimate of coherent length.
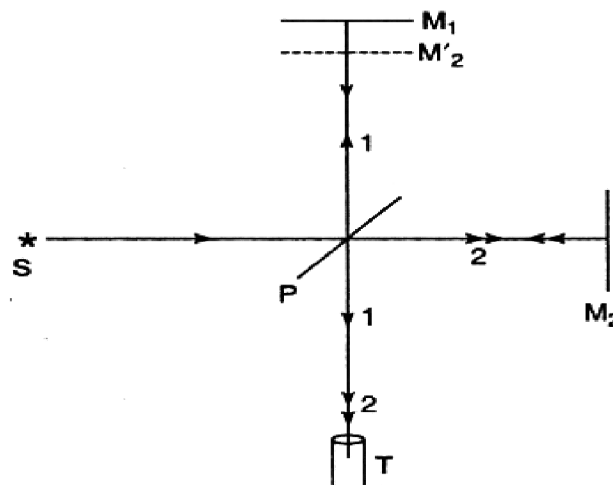


**Fig. 8.11**

- For example, for the neon line ($\lambda = 6328\text{Å}$), $\tau_c \sim 10^{-10}$ sec and for the red cadmium line ($\lambda = 6328\text{Å}$), $\tau_c \sim 10^{-10}$ sec; the corresponding coherent lengths are 3 cm and 30 cm respectively and few kilometers for laser light. Thus, using a laser beam, say commercially available helium-neon lasers, sharp contrast *interference fringes can be obtained even for a path difference of a few meters.*

■ The finite value of the coherence time $\tau_c$ could be due to many factors; for example, if a radiating atom undergoes collision with another atom, then the wavetrain undergoes an abrupt phase shift. The finite coherence time could also be on account of the random motion of atoms or since an atom has a finite lifetime in the energy level from which it drops to the lower energy level while radiating.

## 8.16 Spatial coherence

The **spatial coherence** *refers to the phase relationship between the radiation fields at different space points.*

Let the light waves be emitted from the source $S$ (Fig. 8.12) and $A$, $B$ be two space points on a line joining them with $S$. The phase relationship between the points $A$, $B$ depends on (i) the distance $AB$ and (ii) the temporal coherence of the beam.

If $AB \ll L_c$, the coherent length, there will be a definite phase relationship between $A$ and $B$. Consequently, there will be a high degree of coherence between the points $A$ and $B$.
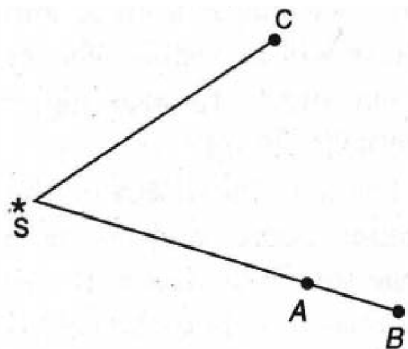


**Fig. 8.12**

Conversely, if the distance $AB \gg L_c$, the coherent length, there will be no coherence between the points $A$ and $B$.

We now consider the points $A$ and $C$ which are *equidistant* from the source $S$, but unlike the previous case, not on the same line joining with $S$. If $S$ is a true *point source*, light waves will reach $A$ and $C$ exactly in the same phase. In other words, the two space points will be in *perfect spatial coherence*. In case the source is an extended one, the two points $A$ and $C$ will no longer be in coherence as illustrated by Young's double slit experiment.

**Illustration :** In order to understand the concept of spatial coherence we consider *Young's double hole experiment* as shown in Fig. 8.13; the interference pattern produced by this experimental arrangement will be discussed in considerable detail in Unit-9 *Interference of light.*
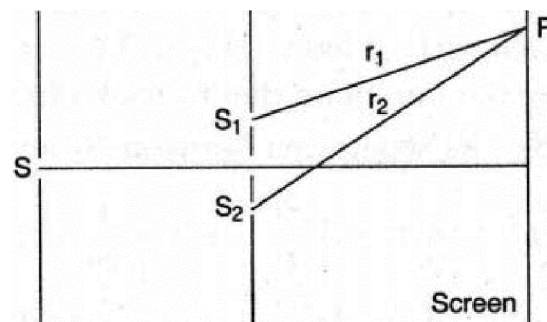


**Fig. 8.13**

Light from a *narrow* slit $S$ falls on two slits $S_1$ and $S_2$, situated *symmetrically* with respect to $S$ and  equidistant from it. Since the light beams from $S_1$ and $S_2$ are from the *same* original beam, a *constant* phase difference will be maintained at all points on the screen and a stationary interference pattern will be observed.

Now, the interference pattern observed around the point $P$ at time $t$ is due to the superposition of waves emanating from $S_1$ and $S_2$ at times $t - r_1/c$ and $t - r_2/c$ respectively, where $S_1P = r_1$ and $S_2P = r_2$. Obviously, if

$$\frac{r_2 - r_1}{c} \ll \tau_c$$

then the waves arriving at $P$ from $S_1$ and $S_2$ will have a definite phase relationship and an interference pattern of good contrast will be obtained. On the other hand, if the path difference $(r_2 - r_1)$ is large enough such that

$$\frac{r_2 - r_1}{c} \ll \tau_c$$

then the waves arriving at $P$ from $S_1$ and $S_2$ will have no fixed phase relationship and *no* interference pattern will be observed. Thus, the central fringe for which $r_1 = r_2$ will, in general, have a good contrast. As we move towards higher order fringes the contrast of the fringes will gradually become poorer and poorer and eventually disappear.

Let us gradually increase the slit width. Then also the pattern becomes poorer and poorer in contrast and eventually disappear. As the size of the source increases, the spatial coherence on the screen changes into incoherence. The reason is that when $S$ becomes wide, the slits $S_1$ and $S_2$ receive light waves from *different*, and hence *independent*, parts of the source and do not remain coherent with each other.

**Calculation of lateral coherence width :**

Let the double-slit be illuminated by two independent point sources $S$ and $S'$, separated by $l$ (Fig. 8.14). To find the least value of $l$ at which the pattern on the screen would disappear.
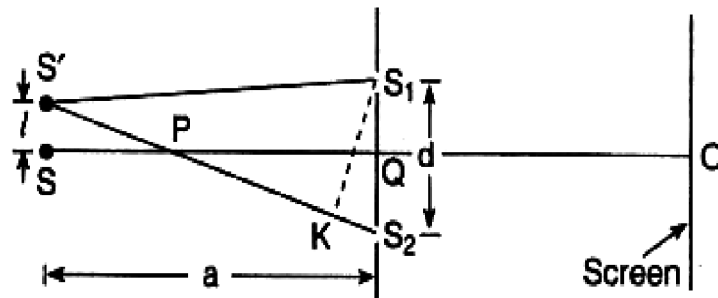


**Fig. 8.14**

There would be a bright fringe at $O$ on the screen due to interference of waves from $S$ reaching $O$ via $S_1$ and $S_2$, having zero path difference. Waves from $S'$ reaching $O$ via $S_1$ and $S_2$, have path difference $KS_2$. So, at $O$ a dark fringe is obtained $KS_2 = \dfrac{\lambda}{2}$, $\lambda$ being the wavelength of light. When this occurs, the maximum due to waves from $S$ is destroyed by the minimum due to waves from $S'$.

Let a $= SQ$ and $d = S_1 S_2$, the separation between $S_1$ and $S_2$. Then,

$$S'S_2 = \left[a^2 + (d/2+l)^2\right]^{1/2} \simeq a + \frac{1}{2a}(d/2+l)^2$$

$$S'S_2 = \left[a^2 + (d/2-l)^2\right]^{1/2} \simeq a + \frac{1}{2a}(d/2+l)^2$$

assuming a $>> d$ and also $l$,

$$\therefore \quad KS_2 = S'S_2 - S'S_1 = ld/a.$$

So, for the fringes to disappear, we must have

$$KS_2 = \lambda/2 \simeq ld/a$$

$$\Rightarrow \boxed{l = \frac{\lambda a}{2d}}$$

Now, if we have an extended incoherent source whose linear dimension is ~ $\lambda a/d$ then for every point on the source, there is a point at a distance of ~ $\lambda a/2d$ which produces fringes which are shifted by half a fringe width. Therefore, the interference pattern will not be observed. Thus, for an extended incoherent source, interference fringes of good contrast will be observed only when

$$l \ll \frac{\lambda a}{2d} \qquad \qquad ... (8.19)$$

If θ be the angle subtended by the extended source $SS'$ at $Q$, then θ = $l$/a and (8.19) becomes,

$$d \ll \lambda/\theta \qquad \qquad ... (8.20)$$

On the other hand, if

$$d \sim \lambda/\theta \qquad \qquad ... (8.21)$$

$= \dfrac{\lambda}{2}$

the fringes will be of very poor contrast. Indeed, a more rigorous diffraction theory shows that the interference pattern disappears when

$$d = 1.22\frac{\lambda}{\theta}, \ 2.25\frac{\lambda}{\theta}, \ 3.24\frac{\lambda}{\theta} \qquad \qquad ... (8.22)$$

Thus, as the separation of the pinholes is increased from zero, the interference fringes disappear when $d = 1.22\lambda/\theta$; if d is further increased the fringes reappear with relatively poor contrast and they vanish again when d = $2.25\lambda/\theta$, and so on. The distance

$$l_w = \lambda / \theta \text{ (Lateral coherence width)} \qquad \qquad (8.23)$$

gives the lateral distance over which the beam may be assumed to be spatially coherent and is known as the **lateral coherence width**.

For a circular source $l_w$ is given by

$$\boxed{l_w = 1.22\frac{\lambda}{\theta}} \qquad \qquad ... (8.24)$$

From the above discussion it is evident that by using ordinary non-coherent sources such as electric bulb, the sun etc., it is possible to make spatially coherent beam by passing the light through a small hole of size << $l_w$.

## 8.17 Linewidth: Purity of spectral line

A spectral line is said to be *perfectly pure* if it consists of a *single frequency*. This is only an ideal condition and is true only theoretically. Every spectral line has a finite width implying that it corresponds to a continuous distribution of wavelength in an interval, however narrow, between $\lambda$ and $\lambda + d\lambda$. The width of a spectral line is a measure of the *purity* of line; smaller the width, pure is the line.

In Michelson interferometer, the fringe pattern disappears, as already mentioned, when the path difference between the interfering beams exceeds the coherence length. Thus, the concept of temporal coherence is directly related to the width or purity of the spectral line.

**Relation with temporal coherence :**

Let us consider the Michelson's interferometer experiment using two closely spaced wavelengths $\lambda_1$ and $\lambda_2$ (like the $D_1$ and $D_2$ lines of sodium). It is known that the criterion that when the path difference $2d$ between the interfering beams equals the coherence length $L_c$, at the centre a bright ring of $\lambda_1$ coincides with a *darkring* of $\lambda_2$ i.e., the interference pattern will disappear if

$$\frac{L_c}{\lambda_2} - \frac{L_c}{\lambda_1} = \frac{1}{2} \qquad\qquad \text{... (8.25)}$$

$$L_c = \frac{\lambda_1 \lambda_2}{2(\lambda_1 - \lambda_2)} \qquad\qquad \text{... (8.26)}$$

Instead of two discrete wavelengths, if we assume that the beam consists of all wavelengths lying between $\lambda$ and $\lambda + d\lambda$, then the interference pattern produced by the wavelengths $\lambda + \frac{1}{2}\Delta\lambda$ will disappear if

$$L_c = \frac{\lambda^2}{2\left(\frac{1}{2}\Delta\lambda\right)} = \frac{\lambda^2}{\Delta\lambda} \qquad\qquad \text{... (8.27)}$$

Thus, we may conclude that the *spectral width or spread* $\Delta\lambda$ of the spectral line of mean wavelength $\lambda$ is

$$\Delta\lambda \sim \frac{\lambda^2}{L_2} = \frac{\lambda^2}{c\tau_c} \qquad\qquad \text{... (8.28)}$$

Thus, the temporal coherence $\tau_c$ of the beam is directly related to the spectral width $\Delta\lambda$.

Further, since $v = c/\lambda$, the frequency spread $\Delta v$ of a line would be

$$\Delta v \sim \frac{c}{\lambda^2}\Delta\lambda \sim \frac{c}{L_c} \qquad \text{... (8.29)}$$

where we have disregarded the sign. Since $\tau_c = L_c/c$, we obtain

$$\Delta v = \frac{1}{\tau_c} \qquad \text{... (8.30)}$$

Thus, the frequency spread $\Delta v$ of a spectral line is of the order of the inverse of the coherence time $\tau_c$.

■ **Note : Monochromaticity or spectral purity** of the source $= \dfrac{\Delta v}{v}$

■ For a commercially available laser beam, $\tau_c \sim 50$ $n$sec implying
$\quad\quad \Delta v/v \sim 4\times10^{-8}$.

## 8.18 Visibility of fringes and degree of coherence

The *visibility* $V$ of interference fringes is defined as follows :

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

where $I_{\max}$ and $I_{\min}$ are respectively the maximum and minimum intensity.

The visibility $V$ is a measure of the degree of coherence of the interfering light waves.

When two wavetrains of equal intensities *overlap* in in the entire finite lengths, the minimum intensity $I_{\min} = 0$. So, the visibility $V = 1$ or 100% and the waves are said to be in *perfect coherence* with highest contrast between the maxima and the minima.

If, however, the two wavetrains do not *overlap* at all $I_{\max} = I_{\min}$ and so V = 0. The waves have no coherence and fringes are not visible.

The intermediate situation is the *partial overlapping* of the two wavetrains as illustrated in Fig. 8.15.
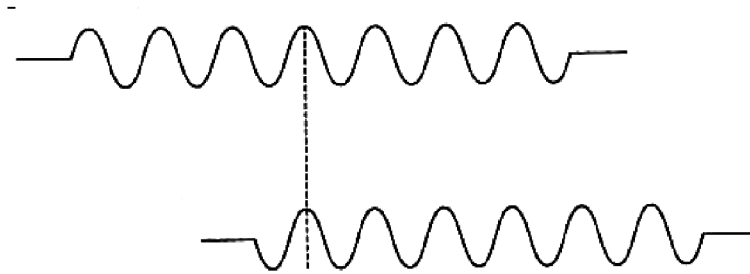
**Fig. 8.15**

In this case, interference is possible but the contrast between the fringes is far less. The degree of contrast depends on the degree of overlapping.

**Relation between visibility and coherence :**

Let two points on a screen be illuminated by two bundles of light to produce equal intensities $I_0$ and assume that each bundle has two parts $a$ and $b$. Of these, the parts $a$ are completely *coherent* and parts $b$ completely *incoherent*. Let the parts a produce an intensity

$$I_a = \beta I_0$$

where $\beta$ is known as the *degree of coherence*.

$\therefore$ Parts $b$ will produce an intensity, $I_b = (1 - \beta)\, I_0$.

Obviously, interference occurs due to parts $a$, only and give fringes whose maxima is 4 times the individual contributions (see Note below).

$\therefore$ Maximum intensity $= 4I_a$; Minimum intensity $= 0$

Due to the $b$ parts, a uniform intensity, a uniform intensity $2I_b$ will be superposed on the above interference pattern. In the pattern, therefore,

$$I_{max} = 4I_a + 2I_b = 4\beta I_0 + 2(1 - b)I_0$$
$$= 2(1 + \beta)I_0$$
$$I_{min} = 0 + 2I_0 = 2(1 - \beta)I_0$$

$$\therefore \quad V = \frac{2(1+\beta)I_0 - 2(1-\beta)I_0}{2(1+\beta)I_0 + 2(1-\beta)I_0} = \frac{4\beta I_0}{4I_0} = \beta$$

*The visibility or the degree of contrast of the fringes produced by two waves of equal intensities is equal to the degree of coherence between the waves.*

**Note :** $I_{max} = (a_1 + a_2)^2$, where $a$ = amplitude. If $a_1 = a_2 = $ a, say then we have $I_{max} = 4a^2 = 4 \times I_{indiv}$.

## 8.19 Summary

- Using corpuscular model, Descartes derived Snell's law of refraction.

- Using the wave model, Huygens could explain the laws of reflection and refraction and he could also interpret the phenomenon of double refraction.

- Maxwell propounded his famous electromagnetic theory of light according to which, **light waves are electromagnetic waves**.

- In 1905, Einstein interpreted the photoelectric effect by putting forward his famous photon theory according to which the energy in a light beam of frequency $v$ was concentrated in corpuscles of energy h$v$, where h represents Planck's constant.

- In the Young's double hole interference pattern, the corpuscular nature of the radiation is evident from its detection in the form of single photons and never a fraction of a photon. The appearance of the interference pattern is because a photon interferes with itself. The quantum theory tells us that a photon partially passes through the two holes. This is not the splitting of the photon into two halves but only implies that the photon is in a state which is a superposition of two states, one corresponding to the wave emanating from the first hole and the other to the one emanating from the second hole. The superposed state will give rise to an intensity distribution like that obtained by considering the superposition of two waves.

- **Huygens' principle** along with the fact that the secondary wavelets mutually interfere, is known as the **Huygens - Fresnel principle**.

- Laws of reflection and Snell's law of refraction can be derived using Huygens' principle.

- Using Huygens' principle one can obtain the mirror equation, relation for spherical refracting surface, lens formula and lens-maker formula.

  The **sign convention** is as follows :

  1. The rays are always incident from the left on the refracting surface.

  2. All distances to the right of the refracting surface are positive and distances to the left of the refracting surface are negative.

- Limitations of Huygens' principle discussed.

- The coherence time $\tau_c$ represents the average duration of the wavetrains, i.e., the electric field remains sinusoidal for times of the order of $\tau_c$.

- The length of the wavetrain, given by

$$L_c = \tau_c \times c$$

(where $c$ is the speed of the light in free space) is referred to as coherence length. For example, for the red cadmium line ($\lambda = 6438$ Å), $\tau_c \sim 10^{-9}$ sec; the corresponding coherent lengths are 3 cm and 30 cm.

- The lateral coherence width ($l_w$) of an extended incoherent source represents the distance over which the beam may be assumed to be spatially coherent; it is given by

$$l_w \approx \frac{\lambda}{\theta}$$

where $\theta$ is the angle subtended by the source at the point of observation.

- Using the concept of spatial coherence, Michelson developed an ingenious method for determining the angular diameter of stars. The method is based on the result that for a distant circular source, the interference fringes (formed by two pinholes) will disappear if the distance between the two pinholes is given by

$$d = 1.22 \frac{\lambda}{\theta}$$

where $\theta$ is the angle subtended by the circular source.

- In the two beams interference pattern, the contrast of the interference fringes varies as the optical path difference $\Delta$ is varied, beginning from an extremely good contrast for $\Delta \ll L_c$ to a very poor contrast for $\Delta \gg L_c$.

- Monochromaticity or spectral purity of the source $= \dfrac{\Delta v}{v}$.

## 8.20 Review Questions and Answers

1. **Deduce from Huygens' principle of wave propagation, the laws of reflection and refraction of plane waves at plane surface.**

**Ans.** See Articles 8.6 and 8.7.

2. **Apply Huygens' principle to study the reflection of the spherical waves from a concave mirror and hence prove the mirror equation**

$$\frac{1}{v} + \frac{1}{u} = \frac{2}{r}$$

**Ans.** See Article 8.9.

3. **Establish, from wave theory, the relation**

$$\frac{n}{v} - \frac{1}{u} = \frac{n-1}{r}$$

for refraction of spherical waves through spherical refracting surface.

**Ans.** See Article 8.10.

4. **Account for the formation of images by refraction through lenses on wave theory and prove the relation**

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

where the symbols have their usual significance.

**Ans.** See Article 8.11.

5. **Explain the concept of coherence. Discuss temporal coherence and spatial coherence, illustrating them with the help of suitable experiments.**

**Ans.** See Articles 8.14, 8.15 and 8.16.

6. **Explain how the width and purity of a spectral line is related with the concept of temporal coherence. Show that the frequency-spread of a spectral line is of the order of inverse of coherence time.**

**Ans.** See Article 8.17.

7. **How does the visibility, of fringes depend on the degree of coherence?**

**Ans.** See Article 8.18.

8. **Distinguish between the spatial coherence and temporal coherence. What are coherence length and coherence time?**

**Ans.** See Articles 8.15 and 8.16.

9. **Explain what you understand by line width and the frequency spread.**

**Ans.** See Article 8.17.

10. **Explain spatial coherence due to a point source and an extended source of light.**

**Ans.** See Article 8.16.

## 8.21  Illustrated Examples

**Example 1 :** The Doppler width of a source at $\lambda = 6058$ Å is 0.0055 Å.  Find out the coherent length.

**Solution** : Here frequency $v = c/\lambda \Rightarrow |\Delta v| = \dfrac{c}{\lambda^2}\Delta\lambda$

$\therefore$ Coherence time, $\tau_c = \dfrac{1}{\Delta v} \Rightarrow$ coherent length,

$$L_c = c \times \tau_c = c\,\dfrac{1}{\Delta v} = \dfrac{c.\lambda^2}{c.\Delta\lambda} = \dfrac{\lambda^2}{\Delta\lambda}$$

$$L_c = \dfrac{\left(6058\times10^{-10}\right)^2}{0.0055\times10^{-10}} = \mathbf{0.67\,m}$$

**Example 2** : For sodium light of wavelength $\lambda$ = 5890 Å and, coherence time $\tau_c$ = $10^{-10}$ s; find the monochromaticity of the source.

**Solution :** $\tau_c = 10^{-10}\,s \Rightarrow \Delta v = \dfrac{1}{\tau_c} = 10^{10}\,Hz$

$$v = \dfrac{c}{\lambda} = \dfrac{3\times10^{10}}{5.89\times10^{-5}} = 5\times10^{14}\,Hz$$

Now, monochromaticity,

$$\dfrac{\Delta v}{v} = \dfrac{10^{10}}{5\times10^{14}} = \mathbf{2 \times 10^{-5}}$$

**Example 3** : A laser beam of wavelength $\lambda$ = 6000 Å, power 10 mW and angular spread $1.5 \times 10^{-4}$ rad is focused by a lens of focal length 10 cm. Find (i) radius, (ii) power density of the image and (iii) the coherence width.

**Solution :** The radius of the image is given by

$r = f.\,\theta$ where $f$ = 10 $cm$ and $\theta$ = 1.5 × $10^{-4}$ $rad$

$\therefore$ $r$ = 10 × 1.5 × $10^{-4}$ $cm$ = 1.5 × $10^{-3}$ $cm$

Area of the image = $\pi r^2$ and the power density of the image

$$= \dfrac{10\times10^{-3}W}{\pi\left(1.5\times10^{-3}\right)^2 cm^2} = 1.4\ kW/cm^2$$

Coherence width $l_w = \dfrac{\lambda}{\theta} = \dfrac{6000\times10^{-10}m}{1.5\times10^{-4}} = \mathbf{4\,mm}$

**Example 4 :** Consider an optical resonator of length 10 cm. Determine the number of modes present in the wavelength range 4 nm about a central wavelength $\lambda$ = 500 nm.

**Solution :** Frequency separation between the consecutive modes is,

$$\Delta v = \frac{c}{2L_c} \text{ (assuming refractive index n = 1)}$$

Now $v = \frac{c}{\lambda}$; or, $\Delta v = \frac{c}{\lambda^2}\Delta\lambda$

$\therefore$ Wavelength separation between the consecutive modes is,

$$|\Delta\lambda| = \frac{\lambda^2}{c}\Delta v = \frac{\lambda^2}{2L_c}$$

Thus, the required number of modes is,

$$\frac{4\,nm}{|\Delta\lambda|} = \frac{4\,nm \times 2L_c}{\lambda^2} = \frac{4\times10^{-9}\times2\times0.1}{\left(500\times10^{-9}\right)^2} = \textbf{3200}$$

**Example 5 :** Find the separation between holes illuminated by the sun of angular diameter 32′ through a filter which passes light of wavelength $\lambda$ = 5000 Å, such that interference fringes of good contrast can be obtained.

**Solution :** The separation must be smaller than the lateral coherence width

$$l_w \approx \frac{1.22\lambda}{\theta} = \frac{1.22\times5000\times10^{-10}}{\dfrac{32}{60}\times\dfrac{\pi}{180}}\,m = \textbf{0.0065 cm}$$

**Example 6 :** Two light beams having intensities in the ratio 1:9 produce interference fringes of intensity 0.3. What information do you get about the degree of coherence?

**Solution :** If the intensity ratio of the two coherent beams be $\alpha$, then we have

$$V = \frac{2\sqrt{\alpha}}{1+\alpha}$$

**Proof :** With a's as amplitudes,

$$I_{\max} = \left(a_1+a_2\right)^2;\ I_{\min} = \left(a_1-a_2\right)^2$$

$$\Rightarrow I_{\max} - I_{\min} = 4a_1a_2 \text{ and } I_{\max} + I_{\min} = 2(a_1^2 + a_2^2).$$

Also, $\alpha = a_1^{\,2} / a_2^{\,2}$

$$\therefore \quad V = \frac{4a_1 a_2}{2\left(a_1^{\,2} + a_2^{\,2}\right)} = \frac{2a_1 a_2}{a_2^{\,2}\left(1 + a_1^{\,2} / a_2^{\,2}\right)} = \frac{2\left(a_1 / a_2\right)}{\left(1 + a_1^{\,2} / a_2^{\,2}\right)} = \frac{2\sqrt{\alpha}}{1+\alpha}$$

Here, $\alpha = 1/9 \Rightarrow V = \dfrac{2\sqrt{(1/9)}}{1+(1/9)} = 0.6$

But the given value of visibility is **0.3** i.e., half the value of two coherent beams. So, the degree of coherence is only **50%**.

**Example 7 :** A magnetic-field technique for stabilizing a He-Ne laser to 2 parts in $10^{10}$ has been patented. At 632.8 nm, what would be the coherence length of a laser with such a frequency stability?

**Solution :** Frequency stability is given by

$$\frac{\Delta v}{v} = \frac{2}{10^{10}} \; ; \quad c = v\lambda$$

So, $v = \dfrac{c}{\lambda} = \dfrac{3 \times 10^8 \, m/s}{632.8 \times 10^{-9} \, m} = 4.74 \times 10^{14} \, Hz$

Coherence length $L_c = c \times \tau_c$

Frequency range $\Delta v = \dfrac{\Delta v}{v} \times v = \dfrac{2}{10^{10}} \times 4.74 \times 10^{14} \, Hz$

$$= \pm 2(4.74 \times 10^4 \, \text{Hz}) \;\; \text{or} \;\; 9.48 \times 10^4 \, \text{Hz}$$

$$\tau_c = \frac{1}{\Delta v} = \frac{1}{9.48 \times 10^4 \, Hz} \approx 1.05 \times 10^{-5} \, s$$

$L_c = (3 \times 10^8 \, m/s) \times (1.05 \times 10^{-5} s) = \textbf{3.15} \times \textbf{10}^3 \, \boldsymbol{m}$

# Unit-9 : Interference

**Structure**

## 9.1  Objectives

In going through this unit, you will learn :
1. Interference of light : Young's Double slit experiment (YDSE)
2. Interference Pattern : Fringe width and shape of fringes
3. Intensity Distribution of Interference Pattern
4. Conditions for Interference
5. Types of Interference
6. Division of wavefront : Fresnel's Biprism
7. Determination of unknown wavelength
8. Measurement of the acute angle of biprism
9. Measurement of the thickness of a thin film
10. Interference with White light : Colour effect
11. Lloyd's Mirror
12. Comparison of biprism and Lloyd's mirror fringes
13. Phase change on Reflection : Stoke's treatment
14. Division of amplitude : Interference in Thin Film (wedge-shaped)
15. Fringe width in wedge-shaped film
16. Fringes of equal thickness and equal inclination
17. Newton's rings
18. Applications of Newton's rings

## 9.2  Introduction

In general, whenever two waves superpose, one obtains an intensity distribution which is known as the **interference pattern**. In this unit, we will consider the interference pattern produced by waves emanating from two-point sources. It may be mentioned that with sound waves the interference pattern can be observed without much difficulty because the two interfering waves maintain a constant phase relationship; this is, also the case for microwaves. However, for light waves, due to the very process of emission, one cannot observe interference between the waves from two independent sources.

Thus, one tries to derive interfering waves from a single wave so that the phase relationship is maintained. The methods to achieve this can be classified under **two**

**broad categories**. Under the first category, in a typical arrangement, a beam is allowed to fall on two closely spaced holes and the two beams emanating from the holes interfere. This method is known as **division of wavefront** and will be discussed in detail in this unit. In the other method, known as **division of amplitude**, a beam is divided at two or more reflecting surfaces and the reflected beams interfere. For example, if a plane wave falls on a thin film, then the wave reflected from the lower surface. Such studies have many practical applications and explain phenomena like the formation of beautiful colours produced by a soap film illuminated by white light.

We must, however, emphasize that the present and the following units are based on one underlying principle, namely the **superposition principle**.

It may be mentioned that it is also possible to observe interference using multiple beams; this is known as **multiple beam interferometry** and will be discussed in Unit 10. It will be shown that multiple beam interferometry offers some unique advantages over two beam interferometry.

**Family-tree of Interference** : See Fig. 9.1
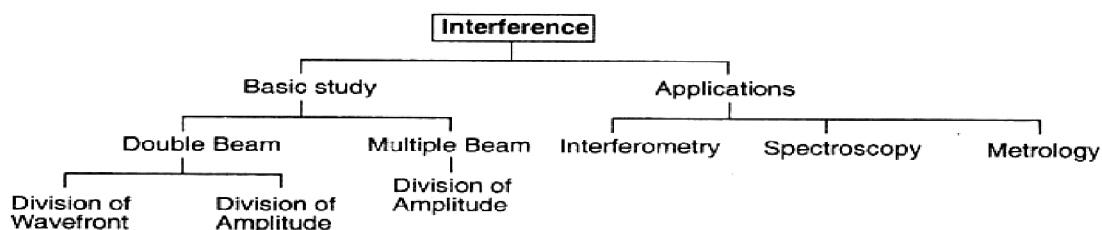


**Fig. 9.1**

**Principle of superposition :** According to the principle of superposition of waves, the resultant displacement (at a particular point) produced by several waves is the vector sum of the displacements produced by each one of the disturbances.

Because of this principle, when monochromatic light waves, coming from two sources, proceed almost in the same direction and superpose at a point, the intensity of light at that point will be maximum or minimum according as the waves at the point are in the same or opposite phase. This phenomenon is called interference of light and is rather important since the phenomenon requires for its explanation the wave nature of light.

## 9.3 Interference of light : Young's double slit experiment (YDSE)

*The modification in the distribution of light intensity in the region of superposition of two or more waves of light is called the* **interference of light**.

When the resultant intensity is the sum of the intensities, the interference is said to be **constructive**, and the resultant intensity is equal to the difference of the intensities of the interfering waves, the interference is said to be **destructive**.

**Young Double slit experiment :** Thomas Young in 1801 devised an ingenious but simple method to lock the phase relationship between the two sources. The trick lies, in the division of a single wavefront into two; these two split wavefronts act as if they emanated from two sources having a fixed phase relationship and, therefore, when these two waves were allowed to interfere, a stationary interference pattern was obtained. In the actual experiment a light source illuminates the pinhole $S$ (see Fig. 9.2). Light diverging from this pinhole fall on a barrier which contained two pinholes $S_1$ and $S_2$ which were very close to one another and were located equidistant from $S$. **Spherical waves** emanating from $S_1$ and $S_2$ were coherent and on the screen beautiful interference fringes were obtained. To show that this was indeed an interference effect, Young showed that the fringes on the screen disappear when $S_1$ (or $S_2$) is covered up. Young explained the interference pattern by considering the **principle of superposition**, and by measuring the distance between the fringes he calculated the wavelength.
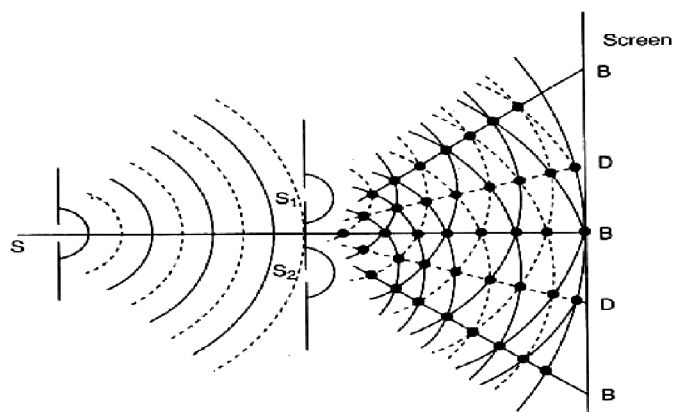


**Fig. 9.2**

## 9.4  Mathematics of interference (YDSE)

Let us consider two plane harmonic waves of the same frequency ω. They meet at a point in time t. Three different methods are available for carrying out the calculations for the resultant. These are

1.   the algebraic method

2.   the method using complex quantities, and

3.   the vector method or phasor diagram.

The choice of the method to be used in solving a particular problem is a matter of personal choice or mathematical convenience since each of the three methods must give the same result if we used correctly. In the following we shall use complex method for two sources called the two-beam interference.

## 9.5  Interference Pattern : Fringe width and shape of fringes

**Fringe width :** Let $S_1$ and $S_2$ represent the two pinholes of the Young's interference experiment. We would determine the positions of maxima and of minima on the line $LL'$ which is parallel to the y-axis and lies in the plane containing the points $S_1$ and $S_2$ (see Fig. 9.3). We will show that the interference pattern (around the point $O$) consists of a series of dark and bright lines perpendicular to the plane of Fig. 9.3; $O$ being the foot of the perpendicular from the point $S$ on the screen.
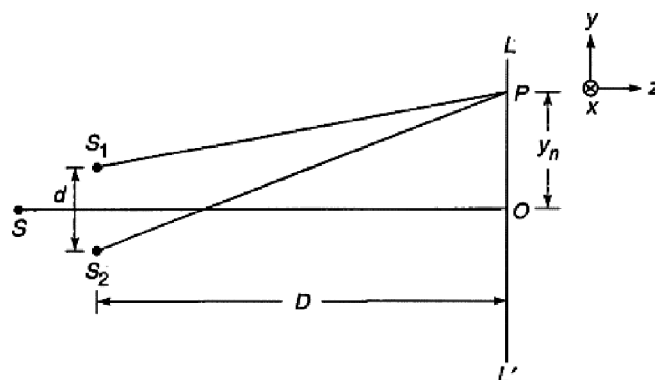


**Fig. 9.3**

For an arbitrary point $P$ (on the line $LL'$) to correspond to a maximum we must have

$$S_2P - S_1P = n\lambda; \ n = 0,1,2,\ldots \qquad \ldots (9.1)$$

Now, $(S_2P)^2 - (S_1P)^2 = \left[D^2 + \left(y_n + \frac{d}{2}\right)^2\right] - \left[D^2 + \left(y_n - \frac{d}{2}\right)^2\right] = 2y_n d$

where $S_1S_2 = d$ and $OP = y_n$

Thus $S_2P - S_1P = \dfrac{2y_n d}{S_2P + S_1P}$ $\qquad \ldots (9.2)$

If $y_n$, $d \ll D$ then negligible error will be introduced if $S_2P + S_1P$ is replaced by $2D$. For example, for $d = 0.02$ cm, $D = 50$ cm, $OP = 0.5$ cm (which corresponds to typical values for a light interference experiment)

$$S_2P + S_1\ P = [(50)^2 + (0.51)^2\ ]^{1/2} + [(50)^2 + (0.49)^2]1/2$$

$$\approx 100.005 \text{ cm}$$

Thus, if we replace $S_2P + S_1P$ by $2D$, the error involved is about 0.000%. in this approximation, (9.2) becomes

$$S_2P - S_1P \approx \frac{2y_n d}{S_2P + S_1P} \approx \frac{y_n d}{D} \qquad \ldots (9.3)$$

Using (9.1) we obtain

$$y_n = \frac{n\lambda D}{d} \qquad \ldots (9.4)$$

Thus, the dark and bright fringes are equally spaced and the distance between two consecutive dark (or bright) fringes is given by

$$\beta = y_{n+1} - y_n = \frac{(n+1)\lambda D}{d} - \frac{n\lambda D}{d}$$

or, $\beta = \dfrac{n\lambda D}{d} = $ **Fringe width** $\qquad \ldots (9.5)$

**Shape of fringes :** To determine the shape of the interference pattern we first note that the locus of the point $P$ such that

$$S_2P - S_1P = D \qquad \ldots (9.6)$$

is a hyperbola in any plane containing the points $S_1$ and $S_2$. Consequently, the locus is a hyperbola of revolution obtained by rotating the hyperbola about the axis $S_1S_2$.
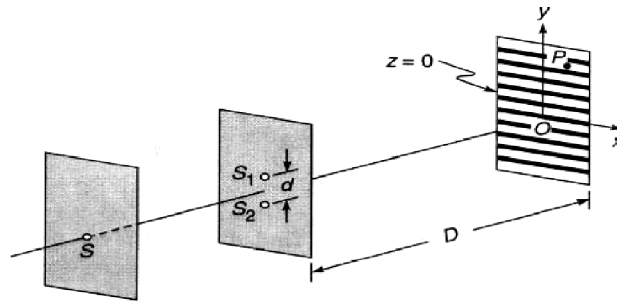


**Fig. 9.4**

To find the shape of the fringe on the screen we assume the origin to be at the point $O$ and the $z$-axis to be perpendicular to the plane of the screen as shown in Fig. 9.4. The $y$-axis is assumed to be parallel to $S_1S_2$. We consider an arbitrary point $P$ on the plane of the screen (i.e., z = 0) (see Fig. 9.4) Let its coordinates be $(x, y, 0)$. The coordinates of the points $S_1$ and $S_2$ are $\left(0, \dfrac{d}{2}, D\right)$ and $\left(0, -\dfrac{d}{2}, D\right)$ respectively. Thus

$$S_2P - S_1P = \left[x^2 + \left(y + \frac{d}{2}\right)^2 + D^2\right]^{1/2} - \left[x^2 + \left(y - \frac{d}{2}\right)^2 + D^2\right]^{1/2} = \Delta \text{ (say)}$$

or, $\left[x^2 + \left(y + \dfrac{d}{2}\right)^2 + D^2\right] = \left\{\Delta + \left[x^2 + \left(y - \dfrac{d}{2}\right)^2 + D^2\right]^{1/2}\right\}^{1/2}$

or, $\left[2yd - \Delta^2\right] = (2\Delta)^2\left[x^2 + \left(y - \dfrac{d}{2}\right)^2 + D^2\right]$

Hence, $\left(d^2 - \Delta^2\right)y^2 - \Delta^2 x^2 = \Delta^2\left[D^2 + \dfrac{1}{4}\left(d^2 - \Delta^2\right)\right]$

which is the *equation of a hyperbola*. Thus, the **shape of the fringes is hyperbolic**. On rearranging we get

$$y = \pm \left( \frac{D^2}{d^2 - \Delta^2} \right)^{1/2} \left[ x^2 + D^2 + \frac{1}{4} \left( d^2 - \Delta^2 \right) \right]^{1/4} \qquad \ldots (9.7)$$

For values of $x$ such that

$$x^2 \ll D^2 \qquad \ldots (9.8)$$

the *loci are straight lines parallel to the x-axis*. Thus, we obtain approximately **straight-line fringes** on the screen. It should be emphasized that **the fringes are straight lines although the sources $S_1$ and $S_2$** are point sources. It is easy to see that if we had slits instead of the point sources, we would have obtained again straight-line fringes with increased intensities.

The fringes so produced are said to be **non-localized**; they can be photographed by just placing a film on the screen; they can also be seen through an eyepiece.

If instead of slits, two coherent point sources $S_1$ and $S_2$ are used, then in three-dimensional space, the loci of maxima of different orders will be a system of confocal hyperboloids with $S_1$, $S_2$ as foci (Fig. 9.5). On a screen placed parallel to the line $S_1$ $S_2$, short straight-line fringes are obtained. If, however, the screen be at right angles to the line $S_1S_2$, several alternately bright and dark concentric circles with their common centre on the intersection of $S_1S_2$ line with the screen.
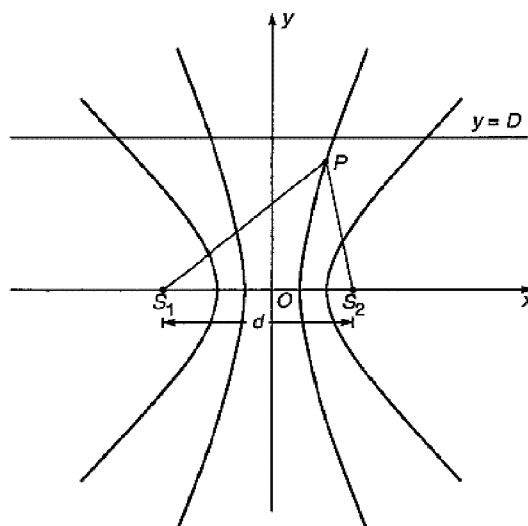


**Fig. 9.5**

## 9.6 Intensity distribution of interference pattern

Let $y_1$ and $y_2$ be the disturbances produced at the point $P$ by $S_1$ and $S_2$ respectively (see Fig. 9.3). The disturbances $y_1$ and $y_2$ will, in general, have different directions and different magnitudes. However, if the distance $S_1P$ and $S_2P$ are very large in comparison to the distance $S_1S_2$, the two disturbances will almost be in the same direction. Thus, we may write

$$\left. \begin{array}{l} y_1 = a_1 e^{i(\omega t - kx_1)} \\ y_2 = a_2 e^{i(\omega t - kx_2)} \end{array} \right\} \qquad \qquad \dots (9.9)$$

The resultant disturbance at $P$, according to the principle of superposition, is given by

$$y = y_1 + y_2 = e^{i\omega t}\left[ a_1 e^{-ikx_1} + a_2 e^{-ikx_2} \right]$$

$$= e^{i\omega t}\left[ a_1\left(\cos kx_1 - i\sin kx_1\right) + a_2\left(\cos kx_2 - i\sin kx_2\right)\right]$$

$$= e^{i\omega t}\left[a - jb\right] = e^{i\omega t} \cdot \sqrt{a^2 + b^2}.e^{i\phi}; \quad \phi = \tan^{-1}\frac{b}{a} \qquad \dots (9.10)$$

where $\quad a = a_1\cos kx_1 + a_2\cos kx_2; \ b = a_1\sin kx_1 + a_2\sin kx_2$

Therefore, $\quad y = Ae^{i(\omega t - \phi)}$

where $\quad A^2 = a^2 + b^2 = a_1^2 + a_2^2 + 2a_1a_2\cos k(x_2 - x_1) \qquad \dots (9.11)$

Here $A$ is the resultant amplitude of the resultant disturbance. Hence the **intensity** of light at $P$ is given as

$$I \propto A^2 = a_1^2 + a_2^2 + 2a_1a_2\cos\delta \qquad \dots (9.12)$$

$$I = I_1 + I_2 + 2\sqrt{I_1I_2}\cos\delta \qquad \dots (9.13)$$

where **phase difference**

$$\delta = k.\left(x_2 - x_1\right) = \frac{2\pi}{\lambda}\left(x_2 - x_1\right) = \frac{2\pi}{\lambda}\times \text{ path difference} \qquad \dots (9.14)$$

Obviously, this phase difference depends on the **path difference** $\Delta = (x_2 - x_1)$ and hence on the position of the point $P$.

From (9.13) we may deduce the following :

1. **Maxima or minima :** The maximum and minimum values of cos δ are +1 and −1 respectively, as such the maximum and minimum values of $I$ are given by

$$\left.\begin{array}{l} I_{\max} = \left(\sqrt{I_1} + \sqrt{I_2}\right)^2 \\[2mm] I_{\min} = \left(\sqrt{I_1} - \sqrt{I_2}\right)^2 \end{array}\right\} \qquad \text{... (9.15)}$$

The **maximum intensity** occurs when

$$\delta = 2n\pi, \qquad n = 0,1,2,...$$

$$\Delta = n\lambda$$

and the **minimum intensity** occurs when

$$\delta = (2n + 1)\pi, \qquad n = 0,1,2,...$$

$$\Delta = \left(n + \frac{1}{2}\right)\lambda$$

Notice that when $I_1 = I_2$, the intensity minimum is zero. In general, $I_1 \neq I_2$ and the minimum intensity is not zero.

2. If the holes $S_1$ and $S_2$ are illuminated by different sources (see Fig. 9.2), then the phase difference $\delta$ will remain constant for about $10^{-10}$ sec (see discussion in Art. 8.16, Unit 8) and thus $\delta$ would also vary with time in random way. If we carry out the averaging over time scales which are of the order of $10^{-8}$ sec, then

$$\langle \cos \delta \rangle = 0 \quad \text{and we obtain } I = I_1 + I_2$$

Thus, for two **incoherent sources**, the resultant intensity is the sum of intensities produced by each one of the sources independently and **no interference pattern** is observed. Hence, we get general illumination.

3. In the arrangement shown in Fig. 9.4, if the distances $S_2P$ and $S_1P$ are extremely large in comparison to $d$, then

$$I_1 = I_2 = I_0 \text{ (say)}$$

and

$$I = 2I_0 + 2I_0 \cos \delta = 4I_0 \cos^2 \frac{\delta}{2} \qquad \text{... (9.16)}$$

The intensity distribution (which is often termed as the $\cos^2$ pattern) is shown in Fig. 9.6. The intensity distribution curve is known as the **interference pattern** and the alternate bright and dark regions the **interference fringes**.
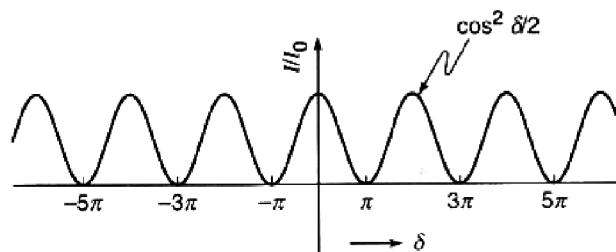
**Fig. 9.6**

## 9.7 Interference and conservation of energy

At first sight, it appears that the law of conservation of energy is violated in the phenomena of interference, for the energy at the minima points is lost. But this is only *apparent* as would be clear from the following discussion.

In absence of interference, the intensity due to two waves should be $I_0 + I_0 = 2I_0$, but the intensity at maxima points is $4I_0$, while that at minima points is zero. Fig. 9.6 shows the variation of intensity $I$ as a function of phase difference $\delta$.

As discussed in Art. 9.6, the average value of intensity in the range $\delta = 0$ to $\delta = 2\pi$ is $2I_0$. So, the energy which apparently disappears at minima has been transferred to the maxima where the intensity is $4I_0 > 2I_0$, the sum of the intensity of the two beams acting separately. There is thus *no destruction of energy because of interference.* The energyis simply **redistributed**. *The formation of interference fringes is thus in accord with the law of conservation of energy.*

## 9.8  Conditions for interference

In order to have well defined observable interference pattern the following conditions must be satisfied.

1. *The two beams of light which interfere must be coherent* (see Arts. 8.16 and 8.17, Unit 8).

Two sources are said to be coherent if the phase difference $\phi$ between the sources remains constant in time. If the sources are incoherent $\phi$ changes continually and we get uniform general illumination.

2. *The interfering waves must have the same frequency. Also, their amplitudes must be equal or very nearly equal.*

If the intensities $I_1$ and $I_2$ differ widely then the intensity, $I_{max} = \left(\sqrt{I_1} + \sqrt{I_2}\right)^2$, in

the bright region and that, $I_{min} = \left(\sqrt{I_1} - \sqrt{I_2}\right)^2$ in the dark region will not differ significantly and hence intensity variation cannot be recognized.

3. *The original source must be monochromatic or very nearly monochromatic. If the light source is heterogeneous the optical path difference between the interfering beams must be kept very small.*

The spacing $\lambda D/d$ between consecutive bright or dark fringes is a function of $\lambda$. So, fringes for different colours will be in step only at the central fringe and soon get out of step on either side of central fringe. If the path difference is large, then dark fringes for some wavelengths may be masked by the bright fringes of some other wavelengths.

4. *The two interfering beams must propagate along the same direction or must intersect at a very small angle.*

If the angle between the two interfering wavefronts is large or the distance between the coherent sources is large the spacing between the interference fringes becomes small and may become indistinguishable even under high magnification.

5. *For interference with polarized light the waves must be in the same state of polarization.*

## 9.9  Types of interference

Interference may be divided into two main classes based on how the coherent sources are produced. These are

1. By the **division of wavefront** 2. By **the division of amplitude**

**Division of wavefront :** Here, the incident *wavefront is divided into two parts* by reflection, refraction etc. to produce two coherent interfering beams. The two beams are made to travel unequal distances and reunite at some angles to produce interference fringes. To maintain spatial coherence, it is essential to use narrow line or point sources in this type of interference. Examples of this type are the fringes formed by biprism, Lloyd's mirror etc.

**Division of amplitude :** In this class, the *amplitude* of the incident beam is *divided into two or more parts* by partial reflection or refraction to produce two or more coherent interfering beams. The beams travel over different paths and reunite to produce interference fringes. A broad light source is to be employed here (Why? To be explained in a subsequent article) to produce brighter fringes. Examples of this class are the fringes formed by thin films, Newton's rings, Michelson's interferometer etc.

## 9.10 Division of wavefront : Fresnel's biprism

Fresnel produced two coherent sources by division of wavefront using this device.

A *biprism essentially* consists of two prisms, each of very *small refracting angle*, placed base to base. In practice, however, it is made from a single optically plane glass plate by proper grinding and polishing. The obtuse angle of the prism is about 179°, and the other angles are about 30' each.

**Production of fringes : Set up** - The arrangement for obtaining interference fringes using biprism is illustrated in Fig. 9.7.
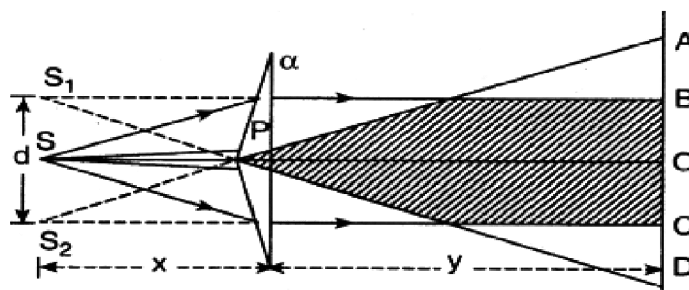


**Fig. 9.7**

$S$ is a narrow vertical slit and is illuminated by monochromatic light. The light emerging from $S$ is allowed to fall *symmetrically* on the biprism $P$, placed at a small distance from the slit $S$, with its refracting edge parallel to it. The incident wavefront is divided into two parts and suffer separate refractions from the upper and lower halves of the biprism to diverge from two virtual images $S_1$ and $S_2$, respectively, of the slit $S$. $S_1$ and $S_2$ thus act as *two coherent (virtual) sources*. The light cones $BS_1D$ and $AS_2C$, diverging respectively from $S_1$ and $S_2$ are superposed. $BC$ being the overlapping region where interference occurs. Interference fringes can be easily seen through an eyepiece. The fringes are not localized (non-localized fringes) in any plane and that is why the screen can be placed anywhere within a suitable distance. Typical fringe pattern is shown in Fig. 9.8.
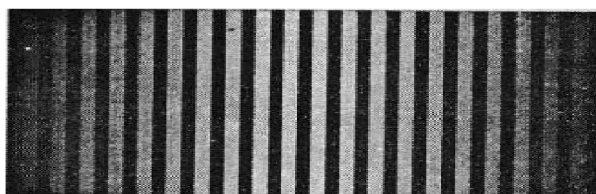


**Fig. 9.8**

**Necessity of narrow source :**

A broad source of light is equivalent to many narrow sources placed side by side. Now if the slit is broad the two virtual coherent sources will also be broad. Now each pair of conjugate points on the virtual sources will give rise to an interference pattern.These interference patterns are slightly displaced from one another. An overlapping of such patterns results in general illumination.

Fresnel's biprism can be used for various optical measurements  as discussed below.

### 9.10.1 Determination of unknown wavelength

**Theory :**

The theory of biprism is essentially the same as that of the double slit; it is already discussed and would not be repeated.

The unknown wavelength $\lambda$ of monochromatic light can be determined using biprismfringes from the formula

$$\beta = \frac{\lambda D}{d} \Rightarrow \lambda = \beta \frac{b}{D} \qquad \qquad \text{... (9.17)}$$

where $\beta$ = fringe-width, $d$ = distance between the virtual sources and $D$, the separation between the slit $S$ and the eye-piece at $O$ fitted with a micrometer. Thus, measuring $\beta$, $d$ and $D$ we can determine $\lambda$.

**Experiment :**

The experiment can be conducted by using a suitable form of optical bench along the bed of which can slide several uprights carrying the linear slit $S$, the biprism $P$ and Ramsden's eye-piece fitted with a micrometer screw.

The slit is illuminated by the monochromatic light and adjustments are made to make the slit $S$, the edge of the biprism and one of the cross-wires of the eye-piece perfectly vertical and all in the same vertical plane and at the same height from the bench. At this time the fringes will be very distinct. The biprism and eye-piece stands are given proper lateral movement so that fringes do not suffer lateral shift relative to the cross-wire as the eye-piece is moved.

**Measurement of $\beta$ :** The fringe width $\beta$ is now measured by setting the cross-wire at successive fringes with the help of micrometer screw fitted with the eye-piece.

**Measurement of D :** Distance $D$ can be measured directly from the bench scale as the distance between the slit and the eye-piece.

**Measurement of d :** To measure $d$ a convex lens of suitable focal length is placed on another upright inserted between the biprism and the eye-piece. The focal length

of the convex lens is such that the distance between the slit and the eye-piece is greater than four times the focal length of the lens. Under this condition, there are two conjugate positions of the convex lens for which real images of $S_1$ and $S_2$ will be seen by the eye-piece kept at the same place. The distances $d_1$ and $d_2$ between the real images of $S_1$ and $S_2$, for the first and second positions of the convex lens respectively, are measured by moving the eye-piece perpendicular to the bench. Now magnification at one position will be inverse of magnification at the second position i.e.,

$$\frac{d_1}{d} = \frac{d}{d_2} \quad \text{or,} \quad d = \sqrt{d_1 d_2}$$

There may be index error between slit stand and eye-piece stand. It can be corrected for or avoided by measuring fringe widths $\beta_1$ and $\beta_2$ at two different distances $D_1$ and $D_2$ respectively. Then, instead of (9.17), $\lambda$ is given by

$$\lambda = d\frac{\beta_2 - \beta_1}{D_2 - D_1} \qquad \qquad \text{... (9.18)}$$

As all the quantities on the right hand are known, $\lambda$ can be easily determined.

### 9.10.2 Measurement of the acute angle of biprism

The separation $d$ between the virtual sources $S_1$, $S_2$ can be expressed, as follows, in terms of refracting angle $\alpha$ of prism, the refractive index $n$ of its material and the slit-prism separation $x$. Since $\alpha$ is very small, the deviation of a ray produced by each of the prism is $(n - 1)\alpha$. If, $\theta$ be the total deviation of the ray striking the centre of the biprism (Fig. 9.9)

$$\frac{1}{2}\theta = (n-1)\alpha\,; \text{ also, } d = x\theta$$

$$\therefore \quad d = 2x(n - 1)\alpha \qquad \qquad \text{... (9.19)}$$

Measuring $d$ and $x$ and knowing the refractive index $n$ for the monochromatic light used, $\alpha$ can be evaluated.
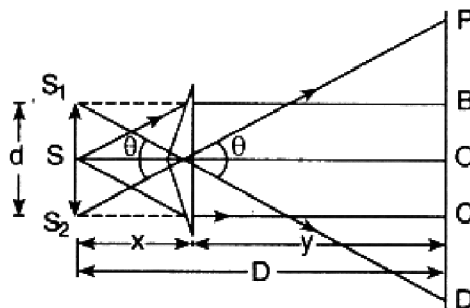


**Fig. 9.9**

### 9.10.3 Measurement of the thickness of a thin film

Fresnel's biprism can be used to measure the thickness of a given thin sheet of transparent material.

Let $S_1$ and $S_2$, be the two virtual coherent sources which are producing interference fringes on the screen so that $C$ is the position of the central bright band of zero optical path difference i.e., $S_1C = S_2C$ (Fig. 9.10). If a thin film of thickness t be introduced into one of the paths (say $S_1P$) of the interfering rays, then the positions of central fringe will be shifted from $C$ to $P$ (say), so that the optical path $S_1P$ is again equal to the optical path $S_2P$.
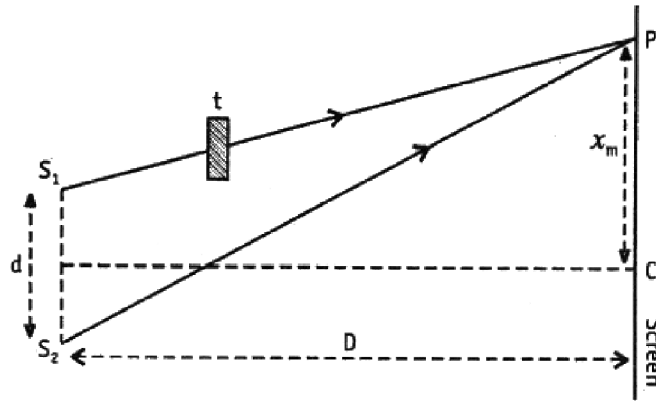


**Fig. 9.10**

The time taken by light in going from $S_1$ to $P$ and from $S_2$ to $P$ will be equal. Thus,

$$\frac{S_2P}{c} = \frac{S_1P - t}{c} + \frac{t}{v}$$

or, $S_2P - S_1P = (n - 1)t$       ... (9.20)

where $c$ and $v$ are the velocity of light in air and in the film; $n = c/v$ is the refractive index of the material of the film.

If $P$ is the position originally occupied by the mth order fringe, then

$$S_2P - S_1P = m\lambda$$       ... (9.21)

Therefore, from (9.20) and (9.21) we get,

$$(n - 1)\, t = m\lambda$$       ... (9.22)

The lateral shift of the central fringe of zero optical path difference is given as

$$x_m = CP = m\beta \qquad \qquad ...(9.23)$$

where $\beta = \dfrac{\lambda D}{d}$ is the fringe width $\qquad \qquad ...(9.24)$

From (9.22) and (9.23) we get

$$(n-1)t = \frac{x_m}{\beta}.\lambda$$

$$t = \frac{x_m.\lambda}{\beta(n-1)} \qquad \qquad ...(9.25)$$

Finding $x_m$, the **displacement** *of the central fringe due to introduction of the thin film* and β, the distance between two consecutive bright bands, we can find $t$, the thickness of the film by (9.25), when the wavelength λ of light and the refractive index $n$ of the film are known.

From (9.24) and (9.25) we may write,

$$t = \frac{x_m}{(n-1)}.\frac{d}{D} \qquad \qquad ...(9.26)$$

With a monochromatic source, fringes appear to be similar, and it is difficult to locate the position where the central fringe is shifted. For this purpose, a white light source is used. With a white source the central fringe is white and all other order fringes are coloured.

**Velocity test :** The relation (9.26) may be written as

$$x_m = \frac{D}{d}\left(\frac{c}{v}-1\right)t \qquad \qquad ...(9.27)$$

If the light travels with smaller velocity in the denser medium (film) then $c/v$ is greater than 1 and $x_m$ is positive. Hence the central fringe will be *shifted towards the side of the film.* Experimental result agrees with this inference arrived at. Hence the conclusion is that light travels with smaller velocity in denser medium.

## 9.11  Interference with white light : Colour effect

The distance of the $m$th bright fringe from the central one is given by

$$x_m = m.\frac{\lambda D}{d}$$

where $d$ is the separation between the two coherent sources and $D$ is the distance of the screen from the sources, $x_m$ is a function of wavelength $\lambda$. So, the fringes of different colours will be in step only for the central band ($m = 0$). In this case $x_m = 0$ and bright fringe of all wavelengths will coincide, and the **central band will be white** with white light. For higher order bright fringes ($m > 0$) $x_m$ will be greater for a light of longer wavelength and less for a light of shorter wavelength. As the wavelength $\lambda_r$ for red light is longer than the wavelength $\lambda_v$ for violet we infer that *all bright bands, excepting the central bright band, will be* **coloured** in which *red will be in the outermost position while violet will be in the innermost position.*

When the path difference between interfering waves is large the condition for constructive interference for one wavelength and the condition for destructive interference for another wavelength may be satisfied at the same point. In that case the resultant illumination cannot be distinguished from white light. So, *for observable fringes with white light the path difference should be kept very small. In this case we get a few coloured fringes on either side of the central white fringe.*

## 9.12 Lloyd's mirror

In Lloyd's single mirror arrangement (Fig. 9.11) light from a narrow slit $S_1$ illuminated with a monochromatic light is partly incident at a grazing angle on a metallic mirror $M_1M_2$ while the rest reaches the screen $AB$ directly. The reflected light appears to diverge from a virtual source $S_2$. Hence $S_1$ and act as coherent sources and interference fringes are formed in the region of overlapping $EF$.
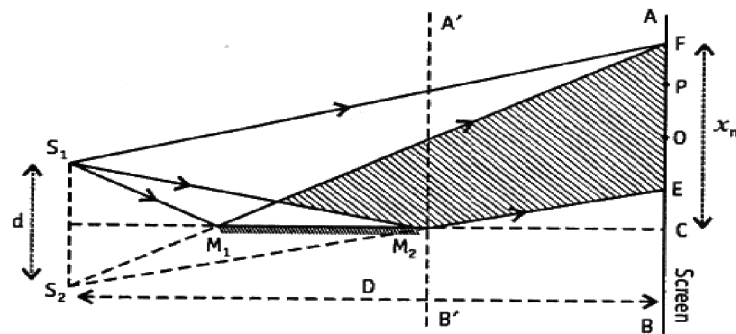


**Fig. 9.11**

The central point $C$ on the screen for which $CS_1 = CS_2$ receives only the direct light and for this the central fringe of zero path difference is not visible here. However, if

the screen is displaced to the position A'B' the central fringe can be brought into view. The central fringe can also be brought into view by introducing a thin film of mica or glass in the path of direct light. In this case the fringe system shifts in direction in which the film is introduced. If t is the thickness of the film and n is the refractive index of the film then central fringe will be formed at a point such as O for which the condition $S_2O = S_1O + (n - 1)t$ is satisfied. The *central fringe is found to be dark*. This indicates that *light reflected from the mirror suffers a sudden phase change of* $\pi$.

### 9.12.1 Comparison of biprism and Lloyd's mirror fringes

(i) In biprism experiment fringes are formed on both sides of the central fringe, whereas in Lloyd's mirror arrangement less than half of the fringes are obtained on one side of the central line.

(ii) In biprism the central fringe is bright whereas in Lloyd's mirror it is dark.

(iii) In Fresnel's biprism the separation (*d*) between every pair of corresponding points of the coherent sources is same and hence fringe width is same for all parts of the source. But in Lloyd's mirror due to lateral inversion in the mirror *d* is different for different pair of coherent point sources. Consequently, in Lloyd's mirror fringe width is not same for all parts of the source slit.

### 9.12.2 Measurement of wavelength

Lloyd's mirror can be used to measure the wavelength ($\lambda$) of a monochromatic light.

Fringe width $\beta$ is given by (for deduction see Art. 9.6)

$$\beta = \frac{\lambda D}{d} \Rightarrow \lambda = \beta \frac{d}{D} \qquad \qquad ...\,(9.28)$$

where $d$ = distance between the coherent sources and $D$ = distance between the slit and the screen. Thus, measuring $\beta$, $d$ and $D$ we can determine $\lambda$.

## 9.13 Phase change on reflection : Stoke's treatment

We will now investigate the reflection of light at an interface between two media using the **principle of optical reversibility**. According to this principle, *in the absence of any absorption, a light ray that is reflected or refracted will retrace its original path if its direction is reversed.*

Consider a light ray incident on an interface of two media of refractive indices $n_1$ and $n_2$ as shown in Fig. 9.12(a). Let the amplitude reflection and transmission

coefficients be $r_1$ and $t_1$ respectively. Thus, if the amplitude of the incident ray is $a$, then the amplitudes of the reflected and refracted rays would be $ar_1$ and $at_1$ respectively.
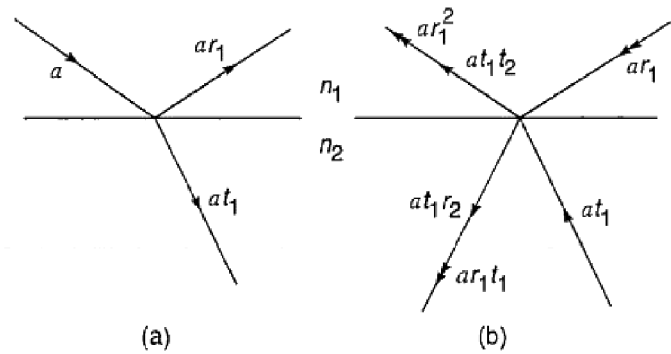


**Fig. 9.12**

We now reverse the rays, and we consider a ray of amplitude $at_1$ incident on medium 1 and a ray of amplitude $ar_1$ incident on medium 2 as shown in Fig. 9.12(b). The ray of amplitude $at_1$ will give rise to a reflected ray of amplitude $at_1 r_2$ and a transmitted ray of amplitude $at_1 t_2$ where $r_2$ and $t_2$ are the amplitude reflection and transmission coefficients when a ray is incident from medium 2 on medium 1. Similarly, the ray of amplitude $ar_1$ will give rise to a ray of amplitude $ar_1^2$ and a refracted ray of amplitude $ar_1 t_1$. According to the principle of optical reversibility the two rays of amplitudes $ar_1^2$ and $at_1 t_2$ must combine to give the incident ray of Fig. 9.12(a).

Thus

$$ar_1^2 + at_1 t_2 = a$$

or, $\quad t_1 t_2 = 1 - r_1^2$ $\qquad\qquad$ ... (9.29)

Further, the two rays of amplitudes $at_1 r_2$ and $ar_1 t_1$ must cancel each other, i.e.,

$$at_1 r_2 + ar_1 t_1 = 0$$

or, $\quad rr_2 = -r_1$ $\qquad\qquad$ ... (9.30)

Since we know from the Lloyd's mirror experiment that an abrupt phase change of $\pi$ occurs when light gets reflected by a denser medium, we may infer from (9.30) that no such abrupt phase change occurs when light gets reflected by a rarer medium. This is indeed borne out by experiments. Equations (9.29) and (9.30) are known as Stokes' relations.

## 9.14 Division of amplitude : Interference in thin film (wedge-shaped)

Let a ray $PQ$ of monochromatic light of wavelength $\lambda$, be incident at an angle $i$ on a transparent **wedge-shaped film**, bounded by the surfaces $AB$ and $CD$ and enclosing an angle $\alpha$, as shown in Fig. 9.13. At $Q$, the ray is partly reflected along $QO$ and partly refracted along $QR$ at an angle $r$. At $R$, it is again partly reflected along $RS$ and partly refracted along $RT_1$. The above process of *multiple reflections and refractions are* repeated at $S$, $V$ etc. As a result, we have a set *reflected rays $QO$, $SO$*, ... etc. and another set of transmitted rays $RT_1$, $VT_2$, ... etc. At each point $Q$, $R$, $S$, ... etc., only a fraction of light is reflected; the rest is refracted so that rays $QO$ and $SO$, suffering one reflection each, have almost equal intensities. The rest are rapidly attenuated and are ignorable. $QO$ and $SO$, both derived from the same incident ray, are *coherent* and hence can *interfere*.
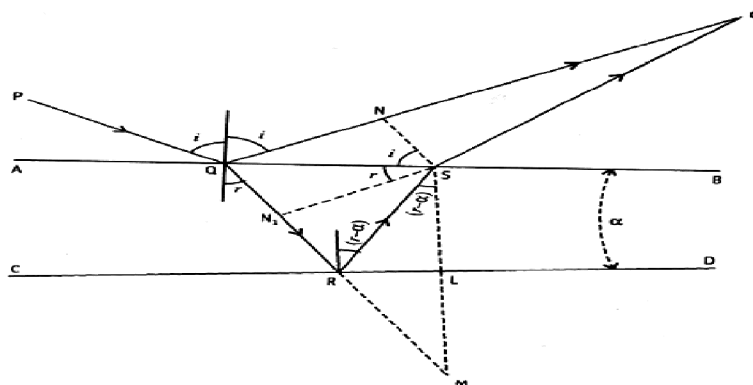


**Fig. 9.13**

**Computation of path difference : Reflected system**

Draw $SN \perp QO$, $SN_1 \perp QR$ and $SLM \perp CD$

Produce $QR$ and $SL$ to meet at $M$. The paths of the two reflected beams ($QO$ and $SO$) which are going to meet at $O$, will be equal from the dotted line $SN$ upto $O$ (for they are very close to each other). Hence the path difference of the two reflected interfering beams would be,

$$\Delta = n(QN_1 + N_1R + RS) - QN \qquad (9.31)$$

where $n$ = refractive index of the film.

From the geometry of Fig. 9.13, $RS = RM$ and $SL = LM = t =$ thickness of the film at $S$. By Snell's law we get,

$$n = \frac{\sin i}{\sin r} = \frac{QN/QS}{QN_1/QS_1}$$

or,  $QN = n.QN_1$

Thus (9.31) now reduces to,

$$\Delta = n(N_1R + RS) = n(N_1R + RM) = n.N_1M$$

$$= n.SM \cos(r - \alpha)$$

$$\therefore \quad \Delta = 2nt \cos(r - \alpha) \qquad\qquad \dots (9.32)$$

In addition to this path difference, there is an extra phase difference of $\pi$ equivalent to a path difference $\pm \lambda/2$, caused by reflection at $Q$, from the surface backed by denser medium. The reflection at $R$, from the surface backed by rarer medium, will not cause any change of phase.

So, **in reflected system**, the net path difference between the above two rays ($QO$ and $SO$) is

$$\Delta = 2nt \cos(r - \alpha) \pm \frac{\lambda}{2} \qquad\qquad \dots (9.33)$$

### 9.14.1 Interference in reflected light

**Conditions for maxima and minima :**

For **maxima** or **brightness** at $O$,

$$2nt \cos(r - \alpha) \pm \frac{\lambda}{2} = \text{even multiples of } \frac{\lambda}{2}$$

$$\Rightarrow \quad 2nt \cos(r - \alpha) = \text{odd multiples of } \frac{\lambda}{2}$$

$$= (2m+1)\frac{\lambda}{2} \quad m = 0,1,2,\dots \qquad\qquad \dots (9.34)$$

Similarly, for **minima** or **darkness** at $O$, we must have

$$2nt \cos(r - \alpha) = 2m\frac{\lambda}{2} \quad m = 0,1,2,\dots \qquad\qquad \dots (9.35)$$

**Note that for a parallel film $\alpha = 0$.**

### 1. Fringes with monochromatic light :

(i) When a parallel beam of monochromatic light falls on a film of *varying thickness*, $n$, $\lambda$ and $r$ all are constants; so, we shall get different orders numbers ($m$) of fringes for different thickness $t$ of the film.

With increasing $t$, the order number of the fringe increases, and a given order corresponds to the locus of points of equal $t$. The fringes are therefore called **fringes of equal thickness** or **FET** fringes.

At the thin edge of the film, $t$ is essentially zero so that $\Delta = \lambda/2$. The thin edge would thus be dark, for all $\lambda$-values (lights of all colours).

(ii) For extremely thin film, $t \rightarrow 0$ for all parts of the film and the path difference between the reflected rays, $\Delta = \lambda/2$. The film surface then appears completely dark even with white light.

(iii) If the **film** surfaces are **plane-parallel** (i.e., $\alpha = 0$) and a parallel beam of monochromatic light is incident obliquely, $n$, $\lambda$ and $t$ are all constants but $r$ will vary with the inclination of the beam. The surface would thus be bright or dark according as $r$ satisfies the conditions of brightness or darkness.

The fringe of a given order here corresponds to the locus of points of equal $r$. Such fringes are thus known as **fringes of equal inclination** or **FEI fringes**.

For *normally incident* parallel beam, $\cos r = 1$, so that the film surface will be uniformly bright or dark according as $t$ satisfies the brightness or darkness condition.

### 2. Fringes with white light : Colour of thin films

When a parallel beam of white light falls on a thin wedge-shaped film, $n$, $\lambda$ and $r$ will be different for light of different colours in the white light. Since $\lambda_v < \lambda_r$, the first order bright fringe of violet light occurs at a smaller $t$ of the film than the $t$-value for the corresponding bright fringe of red. We shall thus get *differently coloured fringes at different thickness*. This explains the *colour of thin films*. Such **fringes are called fringes of equal chromatic order** or **FECO** fringes.

At the *thin edge* of the film, $t$ is essentially zero and the edge will be practically dark since $\Delta = \lambda/2$ for all $\lambda$'s, as explained earlier.

It may so happen that we get another thickness where the brightness condition is simultaneously satisfied by two or more colours leading to a coloured band by overlapping. When the film thickness is sufficient, the overlapping is to such an extent that *uniform illumination results*.

### 9.14.2 Interference in transmitted light

A similar interference phenomenon occurs with rays transmitted through the film and emerging on the side CD of it.

Calculations, following a procedure like that adopted in reflected beam, yield for the path difference $\Delta$ between two successive rays and is given by

$$\Delta = 2nt \cos(r - 2\alpha) \qquad \qquad ... (9.36)$$

Proof. In reference to Fig. 9.14 one part of the ray $QR$ incident on the boundary $CD$ emerges out along $RO_1$, the other part undergoes *two successive internal reflections* at $R$ and $S$ to emerge finally out of $CD$ along $TO_1$.

Draw $TN_1 \perp RS$, $TN \perp RO_1$ and $TLM \perp AB$ $RS$ and $TL$ meet at $M$ when produced. From geometry, $ST = SM$; $LT = LM = t =$ thickness of film at $L$.
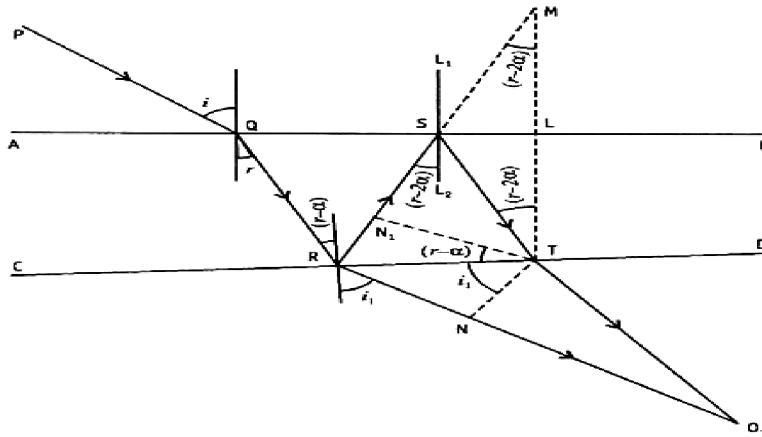


**Fig. 9.14**

Also, since $NO_1 - TO_1$, the path difference of the two transmitted interfering rays is

$$\Delta = n(RN_1 + N_1S + ST) - RN \qquad \qquad ... (9.37)$$

Now, $$n = \frac{\sin i_1}{\sin(r - \alpha)} = \frac{RN/RT}{RN_1/RT} = \frac{RN}{RN_1}$$

$$\Rightarrow RN = n.RN_1$$

$\therefore$ From (9.37) $\Delta = n(N_1S + SM) = n.N_1M = n.TM \cos(r - 2\alpha)$

$$= 2nt \cos(r - 2\alpha) \qquad \qquad ... (9.38)$$

### Conditions for maxima and minima :

In the transmitted beam, the reflection at different points *occurs inside the film at the surface backed by rarer medium* (usually air) and so there is *no phase change by* $\pi$, unlike in the case of reflected system.

For **Maxima** or **brightness**, the condition is

$$2nt\cos\left(r-2\alpha\right)=2m\frac{\lambda}{2}=m\lambda \quad m=0,1,2,...$$

And for **Minima** or **darkness**, the condition is

$$2nt\cos\left(r-2\alpha\right)=\left(2m+1\right)\frac{\lambda}{2} \quad m=0,1,2,...$$

*Thus, the conditions of maxima and minima in the transmitted pattern are just the reverse of the conditions for reflected pattern. So, the pattern due to the transmitted beam is* **complementary** *to that due to the reflected beam.*

### 9.14.3 Fringe width in wedge-shaped film

Let *BAC* be the thin wedge-shaped film of a very small angle $\theta$ and index *n* (Fig. 9.15). A parallel beam of light of wavelength $\lambda$ falls normally on the upper surface and viewed by reflected light. As already discussed, alternate bright and dark bands will be observed.

**Fringe width :** Let t be thickness at any point *P*, distant $x_m$ from the end *A* of the wedge. The condition of **brightness** at *P* gives

$$2nt=\left(2m+1\right)\frac{\lambda}{2}$$

But, $\qquad t=x_m\theta$ ($\because$ $\theta$ small)

$$\therefore \quad 2n\theta x_m=\left(2m+1\right)\frac{\lambda}{2}$$

$$\Rightarrow \quad x_m=\frac{\left(2m+1\right)\lambda}{4n\theta}$$

Similarly, for (*m* + 1)th bright band, we obtain

$$x_{m+1}=\frac{\left[2\left(m+1\right)+1\right]\lambda}{4n\theta}$$

∴ **Fringe width,** β = distance between two consecutive bright bands is

$$\beta = x_{m+1} - x_m = \frac{\lambda}{2n\theta} \qquad \qquad ... (9.39)$$

Therefore, for a given film ($n$, θ const.) and light of a given wavelength λ, β = **constant**. The fringes are thus *equi-spaced*. When θ increases, β *decreases* implying fringes become *narrower* and conversely.

The same expression is obtained with dark fringes as well.
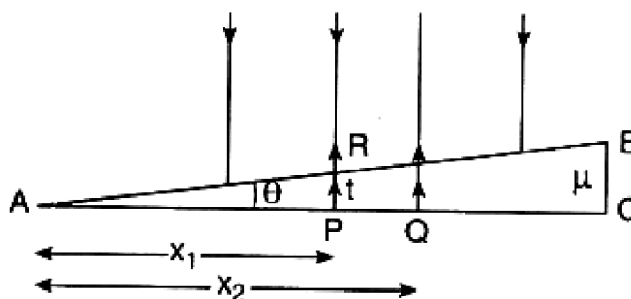


**Fig. 9.15**

### 9.14.4 Difference in film thickness

Let $t_1$, $t_2$ be the thicknesses at $P$ and $Q$respectively and $m$ = number of fringes between these points. So, $t_2 = x_2\theta$; $t_1 = x_1\theta$, $x$'s being distances of $P$ and $Q$ from the end $A$ of the wedge (see Fig. 9.15).

$$\therefore \quad \frac{t_2 - t_1}{\theta} = x_2 - x_1$$

But $(x_2 - x_1)$ / m = λ / $2n\theta$, from (9.39) and therefore,

$$\therefore \quad \frac{t_2 - t_1}{m\theta} = \frac{x_2 - x_1}{m} = \frac{\lambda}{2n\theta}$$

$$\therefore \quad (t_2 - t_1) = \frac{\lambda m}{2n} \qquad \qquad ... (9.40)$$

### 9.14.5 Need of an extended source

An extended source is necessary to observe colours in thin films, for the narrow source limits the visibility of the film.

The situation is explained in reference to Fig. 9.16(a) and let *SA* be a ray that starts from a *narrow source* and be incident on the film. On reflection from the upper surface of the film and suffering one internal reflection, it enters the eye *E*. *SB* is another ray from *S* but is incident at a different angle. On reflection and having an internal reflection, it *cannot* finally reach the eye (Fig. 9.16 (a)). So, only a limited portion of the film becomes visible. To observe the whole of the film, the eye is to be continuously shifted from one position to another.



(a)                    (b)
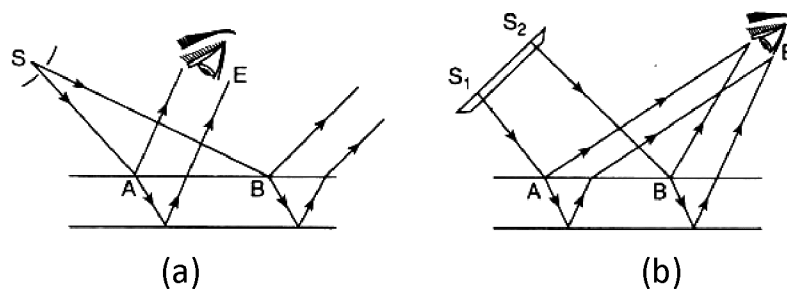
**Fig. 9.16**

If, however, an extended source $S_1 S_2$ is used (Fig. 9.16(b)), the light reflected from a large part of the film reaches the eye, placed in a suitable position. One can thus see the *entire film* simultaneously. Therefore, an extended source is used to view the film.

### 9.14.6 Fringes of equal thickness and equal inclination

The conditions for maxima and minima of brightness of the fringes formed by the light reflected from a thin wedge-shaped film are respectively given by equations (9.34) and (9.35). These conditions suggest the existence of two distinct types of fringes as mentioned in Art. 9.15.1.

1. **Fringes of equal width or thickness(FET) called Fizeau fringes**

2. **Fringes of equal inclination (FEI) called Haidinger fringes**

**Fizeau fringes** of equal thickness are employed to test the optical planeness of a surface. For this purpose, an air film is formed between the working surface and a standard optically flat surface. The fringes of equal thickness formed by the air film are repeatedly observed with a monochromatic light and the polishing of the working surface is continued until the fringes are perfectly straight and parallel to the line of intersection of the surfaces of the air film.

**Haidinger fringes** of equal inclination can be produced by transmitted light from a thick transparent plate. As the plate is thick, the pair of adjacent transmitted rays will be wide apart and hence they cannot enter the eye through its small pupil to produce any interference phenomena there. But if a telescope with a bigger diameter objective be employed, then the objective will be able to collect those transmitted rays which are very close to the normal ray and consequently those collected rays will produce interference phenomena at the focal plane of the objective. The fringe pattern obtained in the focal plane of the telescope consists of concentric bright and dark rings. They are fringes of equal inclination type. These fringes are called Haidinger fringes.

Haidinger fringes are employed **to test the flatness of a plate** to a high degree of accuracy. For accurately plane-parallel Haidinger fringes will be perfectly circular but any deviation from the parallelism of the surfaces will be indicated by the distortion in the rings.

## 9.15  Newton's rings

Newton's rings constitute a special instance of interference in a film of slowly varying thickness. It is the example of **fringes of equal width or thickness**. It was Hooke, and not Newton, who had first observed them.

If a *plano-convex* lens of *large radius of curvature be* placed with the *convex surface* in contact with a *plane glass plate*, an air-film of small thickness is formed between the two. The thickness of the film at the point of contact is zero and gradually increases outwards. If monochromatic light be allowed to fall normally on the film and viewed normally with a low-power microscope in reflected light, **a system of alternate bright and dark concentric rings** *around the point of contact*, with their centre dark, is observed to form in the air-film. These are called **Newton's rings**.



(a)                              (b)
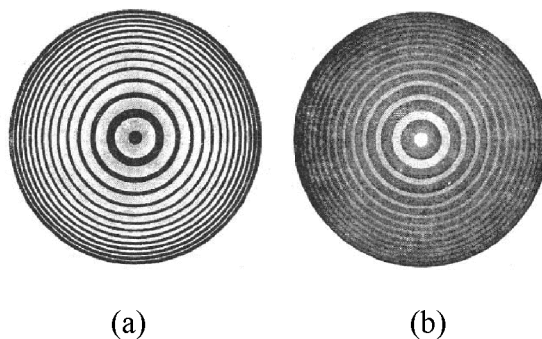
**Fig. 9.17**

These fringes are the loci of points of *equal film thickness* and *the rings are localized* in the film. Typical Newton's rings (both in reflected Fig. 9.17(a) and transmitted light Fig. 9.17(b)) are shown in Fig. 9.17.

**Experimental arrangement :** The experimental arrangement for observing Newton's rings is illustrated in Fig. 9.18. L is the plano-convex lens, PC the glass plate on which L is placed with convex surface in contact. Light from an extended monochromatic source S fall on a glass plate G held at 45° with the vertical. The light reflected from G falls normally on air-film enclosed between L and PC. Interference occurs between the rays reflected from the upper and the lower surfaces of the film giving rise to interference pattern in the form of circular rings localized in the film. The rings are viewed by a low-power microscope $M$ focused on the film.
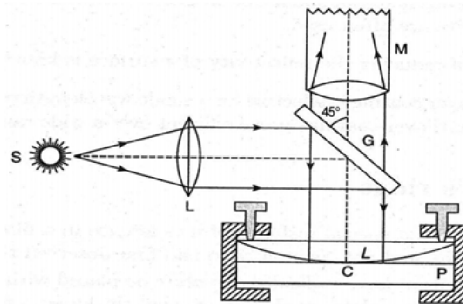


**Fig. 9.18**

**Theory :** In Fig. 9.19 corresponding to the incident ray, the interfering reflected rays are 1 and 2. The effective path difference between the rays is

$$\Delta = 2nt \cos r \pm \frac{\lambda}{2}$$

where $n$ = index of film (air), $t$ = thickness of the film at $B$ and $r$ = inclination of the ray. The factor $\lambda/2$ is due to phase change of $\pi$ on reflection at lower surface of the film - glass plate.

For normal incidence, r = 0

$$\Delta = 2nt \pm \frac{\lambda}{2}$$

At the point of contact, $t = 0 \Rightarrow \Delta = \lambda/2$ which corresponds to the condition of minimum intensity. Hence the central point is dark.
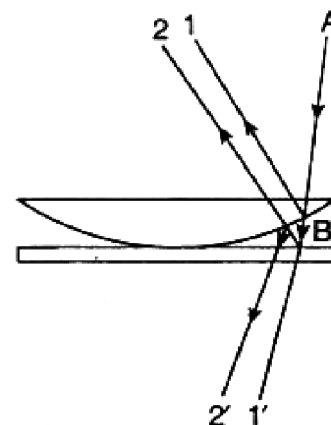


**Fig. 9.19**

The condition for $m$th order maximum intensity, i.e., **bright fringe** is

$$\Delta = m\lambda \Rightarrow 2nt \pm \lambda / 2 = m\lambda \quad m = 0, 1, 2, \dots$$

$$\Rightarrow 2nt = 2(m \pm 1)\lambda / 2$$

The condition for mth order minimum intensity, i.e., **dark fringe** is

$$\Delta = (2m \pm 1)\lambda / 2 \Rightarrow 2nt \pm \lambda / 2 = (2m \pm 1)\lambda / 2$$

$$\Rightarrow 2nt = 2m\frac{\lambda}{2} = m$$

The above expressions show that, bright or dark, a fringe of any order m occurs for a constant $t$-value. In a film, $t$ = constant *along a circle with centre at the point of contact.* So, the *fringes appear* in the form of *concentric circles*—each circle corresponds to a definite t-value, and different circles to different t's. Since each fringe is the locus of constant thickness ($t$), Newton's rings are **fringes of equal thickness (FET)**.

**Diameter of rings :** We shall now find an expression for the diameters of the Newton's rings.
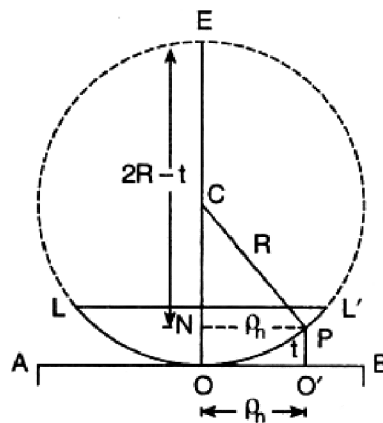


**Fig. 9.20**

Let $AB$ be the glass plate, $LOL'$ the lens placed on it with $O$ as contact point (Fig. 9.20) and $R$ the radius of curvature of the curved surface of the lens. Also let $\rho_m$ be the radius of the $m$th Newton's ring corresponding to the point $P$ where thickness is $t$.

Now, from geometry,

$$PN^2 = ON \times NE$$

$$\Rightarrow \rho_m^2 = t \times (2R - t)$$

$$= 2Rt - t^2 = 2Rt \quad (\because t \text{ very small})$$

$$\therefore \quad 2t = \rho_m^2 / R \qquad \qquad \dots (9.41)$$

Since, for a **bright ring**,

$$2nt = (2m + 1) \lambda/2 \qquad \qquad \dots (9.42)$$

Using (9.41) and (9.42) we get

$$\rho_m^2 / R = (2m+1)\lambda / 2$$

$$\Rightarrow \rho_m^2 = (2m+1)\frac{\lambda R}{2n} \qquad \qquad \dots (9.43)$$

Similarly, for the $m$th **dark ring**,

$$\rho_m^2 = 2m.\frac{\lambda R}{2n} \qquad \qquad \dots (9.44)$$

$\therefore$ **Diameter of $m$th bright ring**, $D_m$ is given by

$$D_m = 2\rho_m = 2\sqrt{2m+1} \times \sqrt{\lambda R / 2n}$$

$$D_m = \sqrt{2\lambda R}\sqrt{(2m+1)/n} \qquad \qquad \dots (9.45)$$

$$\alpha \sqrt{2m+1}$$

Similarly, **Diameter of $m$th dark ring**, $D_m$ is given by

$$D_m = \sqrt{2\lambda R}\sqrt{2m/n} \qquad \qquad \dots (9.46)$$

$$\propto \sqrt{m}$$

Thus, we note that *the diameter of bright rings are proportional to the square root of odd natural numbers and that of the dark rings proportional to the square root of natural numbers.*

The difference in diameters of the $m$th and $(m + 1)$th order **dark rings** is

$$D_{m+1} - D_m = \sqrt{\frac{4\lambda R}{n}} \left[\sqrt{m+1} - \sqrt{m}\right]$$

Thus, *as the order number m increases this difference decreases meaning that the rings gradually become narrower as their diameter increase. There will be more crowding of the rings as we move outward from the centre.*

**Fringe width :** If $D_m$ and $D_{m+1}$ are the diameters of two successive bright rings then we can write from (2.45),

$$D_{m+1}^2 - D_m^2 = \frac{4\lambda R}{n}$$

$$\text{or} \quad D_{m+1} - D_m = \frac{4\lambda R}{n(D_{m+1} + D_m)}$$

Writing $D_{m+1} + D_m \approx 2D_{m'}$ we have fringe width $\beta$ as,

$$\beta = \frac{D_{m+1} - D_m}{2} = \frac{\lambda R}{n.D_m} \qquad \ldots (9.47)$$

Thus, fringe width $\beta$ decreases as the diameter $D_m$ of the ring increases.

**Central fringe :** At the point of contact $O$, we have $t = 0$. Consequently, the condition for destructive interference will be satisfied. This indicates that the **central fringe is dark**. It appears as a dark spot. Normally, with the presence of minute dust particles the point of contact is not perfect, and the central spot may not be perfectly dark.

**Newton's rings with white light :** With a white light the central spot will be black, and it will be surrounded by a few coloured rings and beyond this there will be general illumination due to overlapping of different coloured rings.

## 9.16 Applications of Newton's rings

**1. Determination of wavelength :** Newton's rings can be used for the measurement of the wavelength of almost monochromatic light (emitting from a sodium lamp). The diameters of mth Newton's rings for an air film ($n = 1$) are given from (9.45) and (9.46) as

$$D_m^2 = 2(2m+1)\lambda R \quad \text{[Bright ring]}$$

$$D_m^2 = 4m\lambda R \qquad \text{[Dark ring]} \qquad \ldots (9.48)$$

Thus, if $D_m$ and $D_{m+p}$ are respectively the diameters of the mth and $(m+p)$th rings (bright or dark) then the wavelength $\lambda$ is given by

$$\lambda = \frac{D_{m+p}^2 - D_m^2}{4pR} \qquad \qquad \dots (9.49)$$

**2. Determination of refractive index of a liquid :** Newton's rings can be used for the measurement of the refractive index $(n)$ of a liquid. At first the diameters of the mth and $(m+p)$th rings (bright or dark) are measured with air film. Then the diameters of these rings are measured again by forming a liquid film between the lens and the glass plate. Introduction of the liquid decreases the diameters of the rings.

For air film,

$$\left( D_{m+p}^2 - D_m^2 \right)_{air} = 4p\lambda R$$

For liquid film,

$$\left( D_{m+p}^2 - D_m^2 \right)_{liq} = \frac{4p\lambda R}{n}$$

Hence,

$$n = \frac{\left( D_{m+p}^2 - D_m^2 \right)_{air}}{\left( D_{m+p}^2 - D_m^2 \right)_{liq}} \qquad \qquad \dots (9.50)$$

**3. Determination of optical flatness of a glass plate :** Consider a glass surface placed on another surface whose flatness is known. If a monochromatic light beam is allowed to fall on this combination and the reflected light is viewed by a microscope, then, in general, dark, and bright patches will be seen. The space between the two glass surfaces forms an air film of varying thickness and whenever this thickness becomes $m \, \lambda/2$, we see a dark spot and when this thickness becomes $\left( m + \frac{1}{2} \right) \lambda / 2$ we see a bright spot. Two consecutive dark fringes will be separated by the air film whose thickness will differ by $\lambda/2$. Consequently, by measuring the distance between consecutive dark and bright fringes one can calculate the optical flatness of a glass plate.

**4. Newton's rings under different configurations :**

    1. Lens in contact with a concave surface

    2. Lens in contact with a convex surface

    3. Lens separated from the plate by some distance.

The first two cases have been discussed in 'Illustrated Examples', Example 7.

## 9.17 Summary

- In 1801, Thomas Young devised an ingenious but simple method to lock the phase relationship between two sources of light. The trick lies in the division of a single wavefront into two; these two split wavefronts act as if they emanated from two sources having a fixed phase relationship and, therefore, when these two waves were allowed to interfere, a stationary interference pattern is obtained.

- For two coherent point sources, almost straight-line interference fringes are formed on some planes and by measuring the fringe width (which represents the distance between 2 consecutive fringes) one can calculate the wavelength.

- On a plane which is normal to the line joining the two coherent point sources, the fringe pattern is circular.

- In the Young's double slit interference pattern, if we use a white light source, one gets a white central fringe at the point of zero path difference along with a few coloured fringes on both the sides, the colour soon fading off to white. If we now introduce a very thin slice of transparent material (like mica) in the path of one of the interfering beams, the fringes get displaced and by measuring the displacement of fringes, one can calculate the thickness of the mica sheet.

- If a plane wave is incident normally on a thin film of uniform thickness t then the waves reflected from the upper surface interfere with the waves reflected from the lower surface.

- If we place a plano-convex lens on a plane glass surface, a thin film of air is formed between the curved surface of the lens and the plane glass plate. If we allow monochromatic light (such as from a sodium lamp) to fall (almost normally) on the surface of the lens, then the light reflected from the curved surface interferes with the light reflected from the plane surface. Since the convex side of the lens is a spherical surface, the thickness of the air film will be constant over a circle, and we will see concentric dark and bright rings. These rings are known as **Newton's rings**. The radii of the concentric rings are such that the difference between the square of the radii of successive fringes is very nearly a constant.

## 9.18 Review Questions and Answers

1. **What are coherent sources? How are they produced? Explain.**

**Ans.** See Article 9.8.

2. **Find an expression for the fringe-width in case of Young's double slit. Show that the dark and bright fringes are equally spaced.**

**Ans.** See Article 9.5.

3. **Explain what happens in biprism experiment when (a) white light illuminates the slit and (b) the edge of the prism not parallel to the slit.**

**Ans.** (a) See Article 9.11.

(b) If the edge of the prism is not parallel to the slit, then the edge will not divide the incident wavefront into two equal parts and the fringes observed will be indistinct.

4. **Calculate the displacement of fringes when a thin transparent sheet is introduced in the path of one of the interfering beams in biprism. Show how the method is used for finding the thickness of mica plate.**

**Ans.** See Article 9.10.3

5. **Give Stokes' treatment to explain the change of phase when reflection takes place at a denser medium.**

**Ans.** See Article 9.13.

6. **In what respect do Lloyd's mirror fringes differ from biprism fringes?**

**Ans.** See Article 9.12.1.

7. **While it is necessary to use narrow source for Fresnel's biprism, extended source is needed for Newton's rings. Explain.**

**Ans.** See Articles 9.10 and 9.14.5.

8. **How will you test the flatness of a surface by interference?**

**Ans.** The interference fringes obtained from a film of varying thickness are used to test the flatness of a surface. If a wedge-shaped air film is formed between an optically plane glass plate and the surface under examination, the fringes will be straight if the surface is perfectly plane, otherwise the fringes will be irregular in shape.

9. **Why are the fringes in Newton's rings arrangement circular and in an air-wedge straight and parallel?**

**Ans.** In both, each fringe is the locus of points of equal thickness of t]re film. In Newton's rings arrangement, the points of equal thickness of film lie on

circles with point of contact of the lens and plate as centre. So, the fringes are concentric circles. In wedge-shaped air-film, the loci of equal thickness are straight lines parallel to the edge of the wedge.

**10. What will happen in Newton's ring experiment if the glass plate is replaced by a plane mirror?**

**Ans.** If the glass plate is replaced by a plane mirror, the fringes will disappear, and we shall get a uniform illumination. It is because the intensity of light reflected from the surface of the mirror is very large as compared to the intensity of light reflected from the curved surface of the air film. Hence the intensity at a dark fringe (difference of two intensities) and at the bright fringe (sum of the two intensities) will be practically the same.

**11. Why do we prefer a convex lens of large radius of curvature for producing Newton's rings?**

**Ans.** In Newton's ring arrangement, we prefer a lens of large radius of curvature because?

(i) It results in the increase of the diameter of Newton's rings which increases the accuracy in the measurement of their diameters.

(ii) Large value of R results in the decrease of the thickness of the air film at any point and hence we are justified in neglecting $t^2$ as compared to $2Rt$ in the theory.

(iii) The angle of wedge-shaped film enclosed between the glass plate and the lower surface of lens is very small and hence can be neglected.

**12. Explain what will happen in Newton's ring experiment when air in the inter-space is replaced by a transparent liquid (water)?**

**Ans.** If $d_a$ and $d_l$ are the diameters of the $m$th dark rings in the two cases, then

$$d_a^2 = 4mR\lambda \text{ and } d_l^2 = \frac{4mR\lambda}{\mu}$$

$$\therefore \quad \left(\frac{d_1}{d_a}\right)^2 = \frac{1}{\mu}$$

$$\text{or, } \frac{d_1}{d_a} = \sqrt{\frac{1}{\mu}} = 0.867$$

As $\mu = 1.33$ for water, hence $d_1 < d_a$

Thus, the rings will shrink.

## 9.19  Illustrated Examples

**Example 1 :** Interference fringes are produced by prism in the focal plane of a reading microscope which is 1 m away from the slit. A lens interposed between biprism, and microscope gives two images of the slit in two positions: 4.05 mm in one position and 2.90 mm in the other. If the wavelength of light used is $\lambda$ =5893 Å, find the distance between consecutive  bands.

**Solution :**  Here, $\lambda$ =5893Å = 5893 $\times$ $10^{-10}$ m; D = 1 $m$;

$$d_1 = 4.05 \ mm = 4.05 \times 10^{-3} \ m; \ d_2 = 2.90 \ mm = 2.90 \times 10^{-3} \ m$$

$$\therefore \quad d = \sqrt{d_1 d_2} = \sqrt{4.05 \times 2.90} \times 10^{-3} m = 3.427 \times 10^{-3} m$$

$\therefore$ Distance between consecutive bands i.e., fringe width is

$$\beta = \frac{D}{d}\lambda = \frac{1.0 \times 5893 \times 10^{-10}}{3.427 \times 10^{-3}} = \mathbf{1.72 \times 10^{-4}} \, \boldsymbol{m}$$

**Example 2 :** Fringes are produced with monochromatic light of wavelength 689 nm. A thin film of glass of $n$ = 1.52 is placed normally in the path of one of the interfering beams. The central bright fringe is found to be shifted to a position occupied by the fifth bright band from the centre. Calculate the thickness of the glass film.

**Solution :**  Here $\lambda$ = 689 $\times$ $10^{-9}$ $m$, n = 1.52, $t$ = ?

By the problem, shift in central fringe = 5$\beta$

From the relation,

$$x_0 = \beta(n-1)t/\lambda$$

$$\therefore \quad 5\beta = \frac{\beta}{\lambda}(n-1)t \Rightarrow 5\lambda = (n-1)t$$

$$\therefore \quad t = \frac{5\lambda}{n-1} = \frac{5 \times 689 \times 10^{-9}}{1.52-1} = \mathbf{6.625 \times 10^{-3}} \textbf{ mm}$$

**Example 3 :** In Newton's ring arrangement, a source emitting two wavelengths $\lambda_1$ = 6 $\times$ $10^{-7}$ m and $\lambda_2$ = 5.3 $\times$ $10^{-7}$ m it is found that $m$th dark ring due to one wavelength coincides with ($m$ + 1)th dark ring due to the other. Find the diameter of $m$th dark ring if the radius of curvature of the lens is 0.9 m.

**Solution :** By the problem, $\left(D_m^2\right)_{\lambda_1} = \left(D_{m+1}^2\right)_{\lambda_2}$, We have therefore

$$4mR\lambda_1 = 4(m + 1)R\lambda_2 \Rightarrow m\lambda_1 = (m + 1)\lambda_2$$

$$\therefore \quad m \times 6 \times 10^{-7} = (m + 1) \times 5.3 \times 10^{-7}$$

Solving, we get, $m = 59$

$$\therefore \quad D_m = D_{59} = \sqrt{4 \times 59 \times 0.9 \times 6 \times 10^{-7}} = 1.13 \times 10^{-2}\ m$$

**Example 4 :** The fringes of equal thickness are formed when two glass plates are kept over each other with a small gap in between. If a parallel beam of light of wavelength 6000 Å is used and fringe separation is 3 mm then what is the angle between the plates in seconds?

**Solution :** Here $\quad \lambda = 6000 \times 10^{-10}\ m$, $\beta = 3\ mm = 3 \times 10^{-3}\ m$

$$n = 1, \theta = \text{angle of plate} = ?$$

From the relation,

$$\beta = \frac{\lambda}{2n\theta}, \quad \Rightarrow \theta = \frac{\lambda}{2n\beta} \text{ so that}$$

$$\theta = \frac{6000 \times 10^{-10}}{2 \times 1 \times 3 \times 10^{-3}} = 10^{-4}\ rad = \frac{10^{-4} \times 180}{\pi} \text{ degree}$$

$$= \frac{10^{-4} \times 180}{\pi} \times 3600\ seconds = \mathbf{20.62''}$$

**Example 5 :** Two plane surfaces of glass are in contact at one edge while they are separated by a thin piece of mica at the other edge. If 20 interference bright fringes are observed between these edges when illuminated normally by sodium light ($\lambda = 589.3$ nm), find the thickness of mica.

**Solution :** We know $2t = (2m + 1)\ \lambda/2 \Rightarrow t = (2m + 1)\ \lambda/4$

Here $\quad\quad m = 20, \lambda = 589.3 \times 10^{-9}\ m$

$$t = \frac{41 \times 589.3 \times 10^{-9}}{4} = \mathbf{6.04 \times 10^{-3}\ mm}$$

**Example 6 :** Newton's rings are formed with a source of a light containing two wavelengths $\lambda_1$ and $\lambda_2$. If $m$th order dark ring due to $\lambda_1$ coincides with the ($m$ + 1)th order dark ring due to $\lambda_2$ prove that the radius of the mth dark ring of $\lambda_1$ is equal to

$$\sqrt{\frac{\lambda_1\lambda_2 R}{\lambda_1-\lambda_2}}$$

where $R$ is the radius of curvature of the lower curved surface.

**Solution :** By the problem, mth order dark ring due to $\lambda_1$ coincides with the $(m + 1)$th order dark ring due to $\lambda_2$. So

$$m\lambda_1 = (m + 1)\ \lambda_2$$

$$\therefore\quad m = \frac{\lambda_2}{\lambda_1-\lambda_2}$$

radius of the mth dark ring of $\lambda_1$ is

$$\rho_m = \sqrt{m\lambda_1 R} = \sqrt{\frac{\lambda_1\lambda_2 R}{\lambda_1-\lambda_2}}$$

**Example 7 :** Show that the diameter $D_m$ of the $m$th Newton's ring, when two surfaces of radii $R_1$ and $R_2$ are placed in contact, is given by

$$\frac{1}{R_1}\pm\frac{1}{R_2}=\frac{4m\lambda}{D_m^2}$$

**Solution :** Newton's rings result from interference between waves reflected from upper and lower surfaces of air-film of varying thickness enclosed between the two surfaces (Fig. 9.21).



(a)                                      (b)
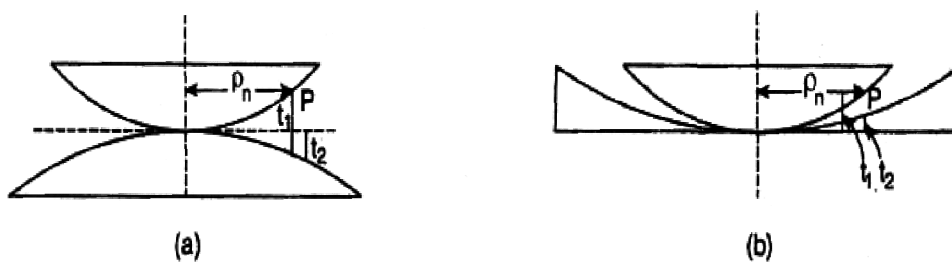
**Fig. 9.21**

If the radii of curvature of the surfaces be large and incidence of light almost normal, then the effective path difference between interfering waves is

$$\Delta = 2t + \lambda/2$$

where $t$ is the film-thickness at the point under reference.

∴ For a dark ring say,

$$2t + \frac{\lambda}{2} = (2m+1)\frac{\lambda}{2} \Rightarrow 2t = m\lambda \qquad \qquad \dots (9.51)$$

Let $\rho_m$ be the radius of a ring corresponding to a point $P$, where thickness is $t$ and $R_1$, $R_2$ be the (large) radii of curvature of the two surfaces.

∴ $t = t_1 + t_2$ (Fig. (a) for + sign; Fig. (b) for - sign)

$$= \frac{\rho_m^2}{2R_1} \pm \frac{\rho_m^2}{2R_2} \Rightarrow 2t = \rho_m^2 \left( \frac{1}{R_1} \pm \frac{1}{R_2} \right)$$

Substituting this value of $2t$ in (9.51), we get

$$\rho_m^2 \left( \frac{1}{R_1} \pm \frac{1}{R_2} \right) = m\lambda$$

$$\frac{D_m^2}{4} \left( \frac{1}{R_1} \pm \frac{1}{R_2} \right) = m\lambda \quad (\because \text{ diameter } D_m = 2\rho_m)$$

∴ $$\frac{1}{R_1} \pm \frac{1}{R_2} = \frac{4m\lambda}{D_m^2} \quad \text{(Proved)}$$

**Example 8 :** An oil film (n = 1.2) on water (n = 1.33) is viewed from directly above with light of wavelength 600 mm in air. The film appears circular and has a centre thickness 1 μm decreasing to zero thickness at the edge. Explain if the edge will appear bright or dark. How many dark rings appear in the fringe?

**Solution :** Since the incident rays suffer identical change in phase, the condition for mth bright fringe is

$$2nt = m\lambda \Rightarrow m = \frac{2nt}{\lambda} = \frac{2 \times 1.2 \times 10^{-6}}{600 \times 10^{-9}} = 4$$

Also, at the edge, the optical path difference is zero implying that the edge will appear **bright** ($m = 0$) so, 4 dark rings will be seen.

# Unit-10 : Interferometers

**Structure**

## 10.1 Objectives

In going through this unit, you will learn :

1. Michelson Interferometer
2. Idea of formation of fringes
3. Determination of Wavelength
4. Wavelength Difference
5. Refractive Index
6. Visibility of Fringes
7. Fabry-Perot interferometer

## 10.2 Introduction

An instrument which depends on the interference of EM waves in general, and light, is called an **interferometer**. Thus, an interferometer is an arrangement for obtaining interference effects with a large path difference, compared to the wavelength of the interfering light, between the two interfering rays. The simplest and the oldest arrangement is due to Young which we already have studied (Art.9.3, Unit 9).

209

All interferometers divide a beam of light into two or more parts, which travel by different paths and recombine, forming interference fringes. It is thus a difference in optical paths that an interferometer measures, in units of wavelength of light.

In general, the purpose of any interferometer is to isolate a specific portion of the EM spectrum and to make spectroscopic study such as (i) analysis of hyperfine structure of spectral lines, (ii) the variation of refractive index (r.i.) of gases or salt solutions with density, temperature, (iii) the linear diameters of stars etc.

Unlike grating or prism spectroscopes, *interferometers are not dispersive instruments but use interference to selectively transmit a certain wavelength.*

The connection between *interferometers and stellar astrophysics* is amazing. These are principally used for the measurement of the angular diameter of stars and for the measurement of binary star orbits. Experimental astrophysicists generally choose a small portion of the EM spectrum to study, preferably one that has the least noise from other sources in the sky. The types of waves they want to study are EM waves (now-a-days mostly microwaves and is called CMB meaning *cosmic microwave background*), because that is the characteristic wavelength of the photons *left over from Big Bang* In this context, remember that photons are timeless and don't die until they impart the information they were carrying by interaction with matter (mostly *electrons and phonons, the quanta of crystal vibration*).

The number of interfering beams in an interferometer is 2 or more. Those with two beams are called **double-beam** and those with many beams are called **multiple-beam** interferometers. In the former, the distribution of intensity in an interference band *obey cosine* (*square*) law, as in equation (9.16). In a multiple-beam instrument the bands have *very sharp maxima* which makes it very useful for the study of spectral lines.

There are two basic classes of interferometers depending on interference by (see Art. 9.10) :

1. **Division of wavefront :** such as, Rayleigh's interferometer, Michelson stellar interferometer, Young's interferometer, Fresnel bi-prism, etc.

2. **Division of amplitude :** it is further subdivided into two categories:

   (a) **Double-beam :** such as Michelson interferometer, Twyman-Green interferometer, Mach-Zehnder, Jamin, etc.

    (b) **Multiple-beam :** such as Fabry-Perot interferometer, Lummer-Gehrcke plate,etc.

When the interferometer is employed to study specifically the refractive index of a substance, it is called a refractometer.

*Interferometers are usually instruments of very high resolving power.*

In the following articles we shall briefly describe only Michelson interferometer and Fabry-Perot interferometer, state their principles of action, and mention their uses based on their resolving powers.

## 10.3 Michelson interferometer

Michelson interferometer is an ingenious arrangement which was designed to test the ether-hypothesis in the famous Michelson-Morley experiment. However, it is now primarily used to find wavelength of light, fine structure of spectral lines and some other measurements based on the phenomenon of interference of light.

**Construction :** It essentially consists of two highly polished and optically plane mirrors $M_1$, and $M_2$ (Fig. 10.1) placed *at right angles* to each other. There are *two* other optically flat glass plates $G_1$ and $G_2$ of the *same material and thickness*, placed *parallel* to each other and *inclined at* 45° to the plane mirrors $M_1$, and $M_2$. The plate $G_1$ is *semi-silvered* (half-silvered) at the back, i.e., the face towards $G_2$. The mirror $M_1$ is mounted on a carriage and can be moved back and forth, *exactly* parallel to itself, with the help of a fine micrometer screw that can read displacements up to $10^{-5}$ cm. By the levelling screws at their backs, the mirrors $M_1$, and $M_2$ can be tilted about both vertical and horizontal axes to make them exactly at right angles to each other. The interference bands produced by the interferometer are observed in the field of view of the telescope T.
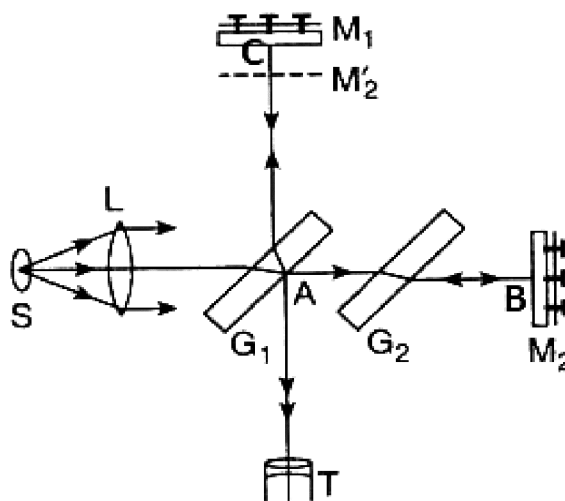


**Fig. 10.1**

**Principle :** Light from an extended monochromatic source $S$ is rendered parallel by a collimating lens L and is allowed to fall on the plate $G_1$ (Fig. 10.1). It gets divided

into two parts of equal intensities : one being reflected at the semi-silvered back surface of $G_1$ to move along $AC$ towards $M_1$, and the other being transmitted along $AB$ towards $M_2$. Both the rays $AC$ and $AB$ fall normally on $M_1$ and $M_2$ respectively and retrace their paths after reflection to enter finally into the telescope $T$ along $AT$. Since the two rays are derived originally from the same single beam by division of amplitude, they are able to produce interference fringes in the field of view of telescope $T$ under suitable conditions.

What is the **role of glass plate $G_2$**? For without it, the ray $AC$ passes through $G_1$ twice, while the ray $AB$ none at all. So, the paths of the two rays $AC$ and $AB$ in the glass are not equal. To compensate for this, an exactly similar glass plate $G_2$ and made of the same material is introduced in the path of $AB$. Hence $G_2$ is called a compensator. While $G_2$ is not essential in working with monochromatic light, it is a must for white light.

**Production and form of fringes :** If one looks towards $M_1$ from T, one sees $M_1$ and also a virtual image $M_2$ of $M_2$ formed by reflection at $G_1$ (Fig. 10.1). Although the two interfering beams, received by telescope $T$, come by way of reflections from $M_1$ and $M_2$, behave functionally as if reflected from $M_1$ and $M_2$. So, the interferometer is optically equivalent to an air-film enclosed between $M_1$ and $M_2$. Depending on the path-difference and the angles between the mirrors $M_1$ and $M_2$ (virtual mirror), the form of the fringes may be different - circular, straight etc.
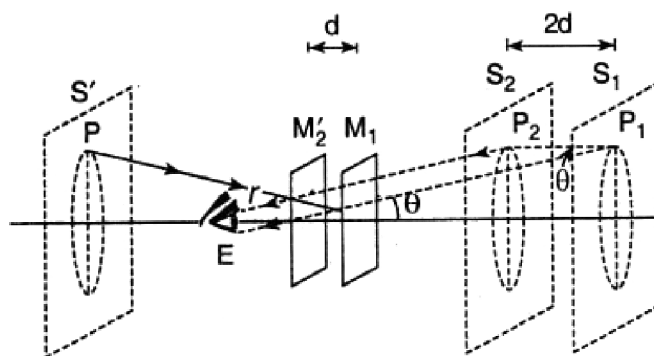


**Fig. 10.2**

**Circular fringes :** Circular fringes are produced in Michelson interferometer with monochromatic light when $M_1$ and $M_2$ are perfectly vertical and exactly at right angles to each other. The production of such fringes is illustrated in Fig. 10.2.

$M_1$ and $M_2$ (image of $M_2$) are parallel reflecting surfaces, the real extended source being replaced by its virtual image $S'$, formed by reflection at $G_1$, behind the eye E.

This virtual source $S'$ again forms two virtual images $S_1$ and $S_2$ in $M_1$ and $M_2'$. Light from a point $P$, say on the extended source, appears to the eye $E$ to come from the corresponding coherent points $P_1$ and $P_2$ and on $S_1$ and $S_2$ respectively. If $d$ be the separation between $M_1$ and $M_2'$ then that between $S_1$ and $S_2$ will be $2d$. The path difference between the rays entering the eye is obviously $2d \cos \theta$ where $\theta$ is the angle the reflected beam makes with the normal.

If $2d \cos \theta = m\lambda$ ($m$ = 0,1,2,...),          P **appears bright** (maximum)

If $2d \cos \theta = (2m + 1) \lambda/2$     ($m$ = 0,1,2,...),     P **dark** (minimum)

For a given $d$, $m$ and $\lambda$, *the angle $\theta$ will be constant so that the locus of P is a circle with centre on the axis.* So, a series of alternate bright or dark **circular fringes** are observed. The fringe order, however, decreases as $\theta$ increases, i.e., as we move away from the centre. Since the interfering rays are parallel, these fringes are formed at infinity and hence observed through a telescope focused for infinity.

**Localized fringes :** When the mirror $M_1$ and the image $M_2'$ of $M_2$ are not exactly parallel, they enclose a wedge-shaped film of air between them (Fig. 10.3). Hence the ray $PQ$ starting from the extended source $S'$, after reflection from $M_2'$ and $M_1$ will give rise to two reflected rays $QE_1$ and $RE_2$ which when produced backward will meet at $O$ where interference fringes will be produced. To see these fringes the eye must be *focused near to the mirror $M_1$* and *not for infinity* as in the case of circular fringes. Hence these fringes are called *localized fringes.*
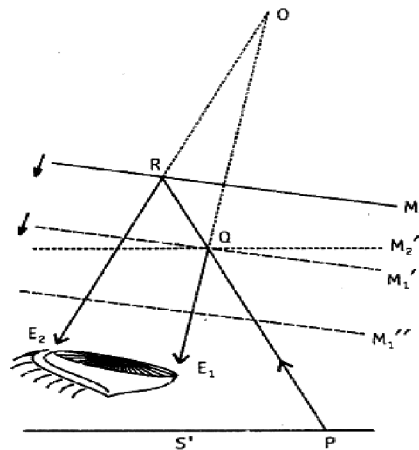


**Fig. 10.3**

**Shapes of the localized fringes :** Since the light is incident on the film at various angles, *curved fringes* with convexity towards the thin edge of the wedge are observed. The shape of the fringes for different path differences are shown in Fig. 10.4. Curved fringes are seen in Fig. 10.4(a) and Fig. 10.4(c).

A fringe of given order is a line of constant path difference. When the mirrors $M_1$ and $M_2'$, intersect at the middle and film-thickness is small, and hence the path difference $2d \cos \theta$ is practically zero. Here we shall get **straight fringes** parallel to the line of intersection of $M_1$ and $M_2'$ [Fig. 10.4 (b)]. These fringes are called *fringes of equal width*. The fringes that corresponds to thickness $d = 0$ is the one which is *perfectly straight*. These fringes are visible up to path difference relatively much smaller than those in circular fringes.



(a)        (b)        (c)
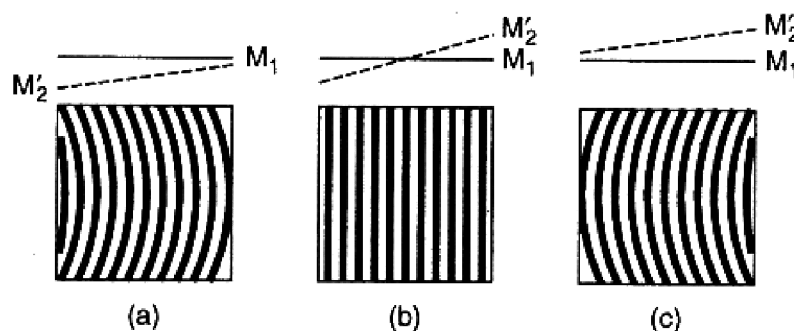
**Fig. 10.4**

**White light fringes :** When monochromatic light source is replaced by white light and the thickness of film is small, a few curved and coloured localized fringes are seen. The fringe of zero thickness ($d = 0$) is again perfectly *achromatic and straight*. If, however, the film is thick, uniform illumination is observed. White light fringes are useful in finding the zero path difference, particularly in standardizing a metre.

● In Michelson interferometer, the intensityis given by

$$I = 4I_o \cos^2 \frac{\delta}{2}, \quad \text{where} \quad \delta = \frac{2\pi}{\lambda}(2d \cos \theta)$$

$d$ being the distance between $M_1$ and $M_2'$.

$$I = I_{max} \quad \text{when} \quad \delta = 2m \, \lambda/2$$

and $I = I_{min}$ when $\delta = (2m + 1) \, \lambda/2, \quad m = 0,1,2,...$

## 10.4 Applications of Michelson interferometer

1. Determination of unknown wavelength of light.
2. Determination of wavelength difference of two very close spectral lines.
3. Determination of thickness or refractive index of a thin transparent film.
4. Standardization of 'metre' in terms of wavelength.

### 1. Determination of wavelength of a monochromatic light :

The interferometer is first adjusted for circular fringes in the field of view of the telescope. If the mirrors $M_1$ and $M_2$ are equidistant from the glass plate $G_1$, the field of view will be perfectly dark. $M_1$ is next adjusted in positive to obtain a particular bright fringe in the field view of telescope with its centre coinciding with the cross-wire. From this position when $M_1$ is moved away through a distance $\lambda/2$, the path-difference changes by $\lambda$ and the position of a particular bright fringe is occupied by the next bright fringe. If m be the number of fringes that cross the centre of field when $M_1$ is moved from initial position $x_1$ to a final position $x_2$, then

$$x_2 - x_1 = m\frac{\lambda}{2} \Rightarrow \lambda = \frac{2(x_2 - x_1)}{m} \qquad \dots (10.1)$$

The difference $x_2 - x_1$ can be measured correct up to $10^{-4}$ mm with the help of the micrometer screw and $m$ is actually counted. Then the wavelength $\lambda$ is obtained by using the above formula.

### 2. Determination of difference of two very closed wavelengths:

Let the source emit two very close wavelengths $\lambda_1 - \lambda_2$ ($\lambda_1 > \lambda_2$), e.g., sodium $D$-lines, and the interferometer be adjusted to obtain circular fringes. There will be two sets of fringes - one corresponding to $\lambda_1$ and the other to $\lambda_2$. Since $\lambda_1$ is very close to $\lambda_2$, the two sets will practically coincide. As $M_1$ is moved slowly the two sets will separate out slowly and when the dark fringe of $\lambda_1$ coincides with the bright fringe of $\lambda_2$, the result is maximum indistinctness (poor visibility). On further movement of $M_1$, the *next indistinct* position is reached after a distance $x$, say. In this position, if $m$ fringes of $\lambda_1$ appear at the centre, then ($m + 1$) fringes of $\lambda_2$ will appear at the centre of the field of view.

$$\therefore \quad x = m\frac{\lambda_1}{2} \Rightarrow m = \frac{2x}{\lambda_1}$$

$$\text{and} \quad x = (m+1)\frac{\lambda_2}{2} \Rightarrow m+1 = \frac{2x}{\lambda_2}$$

$$\therefore \quad (m+1)-m = 2x\left(\frac{1}{\lambda_2}-\frac{1}{\lambda_1}\right) = \frac{2x(\lambda_1-\lambda_2)}{\lambda_1\lambda_2}$$

$$\lambda_1 - \lambda_1 = \frac{\lambda_1\lambda_2}{2x} = \frac{\lambda_m^2}{2x} \qquad\qquad \text{... (10.2)}$$

where $\lambda_m = \frac{1}{2}(\lambda_1 + \lambda_2)$ = mean of two wavelengths $\lambda_1$, $\lambda_2$.

Thus, measuring the distance $x$ moved over by $M_1$ between two consecutive positions of maximum indistinctness, the difference of two close wavelengths (Resolution of spectral lines) can be determined.

**3. Determination of thickness or refractive index of a thin transparent film :**

The interferometer is set for localized white light fringes and the cross-wires made to coincide with the central dark fringe. The path AB and AC are exactly equal.

The thin film is next interposed between the glass plate $G_1$ and the mirror $M_2$ such that the light passes through it normally. The introduction of the sheet of thickness $t$ and refractive index $n$ increases the optical path by $2(n-1)t$. The fringes will thus get shifted from their positions. The mirror $M_1$ is next moved till again a dark fringe of zero path coincides with the cross-wire. The distance $x$ through which $M_1$ is moved is noted.

The white light is next replaced by monochromatic light of wavelength $\lambda$ and the mirror $M_1$ is moved slowly to count the number of fringes contained in $x$. Then,

$$2(n-1)t = m\lambda \Rightarrow t = \frac{m\lambda}{2(n-1)} \qquad\qquad \text{... (10.3)}$$

Thus, the thickness $t$ can be readily evaluated from a knowledge of $n$, $m$ and $\lambda$.

**Note 1 :** It may appear that when the formula

$$2(n-1)t = m\lambda$$

is being used, localized white fringes are unnecessary. In fact, however, it is not so, for the fringes with monochromatic light are exactly alike and it is difficult to count the number of fringes shifted, the shift being sudden on insertion of the sheet.

**Note 2 :** The same method or working formula may be used to determinethe refractive indexn of the material of the sheet, provided $m$, $\lambda$ and $t$ are known.

## 10.5 Fabry-Perot interferometer

It is an interferometer of high resolving power in which fringes of constant inclination, called **Haidinger fringes**, are produced by transmitted light after multiple reflections in an air-film enclosed between two *parallel* and highly reflecting glass plates. It is based on the *principle of multiple beam interference.*

**Construction :** It consists essentially of two *optically plane* glass plates *A* and *B* (Fig. 10.5) placed *accurately parallel* to each other and separated by a distance *d*. The inner surfaces of *A* and *B* are thinly silvered so as to reflect 80-90% of the incident light. The *outer surfaces* of the plates are also *parallel but inclined to their respective inner surfaces,* (i.e., the plates are slightly prismatic) to avoid interference of the rays reflected at the outer unsilvered surface.
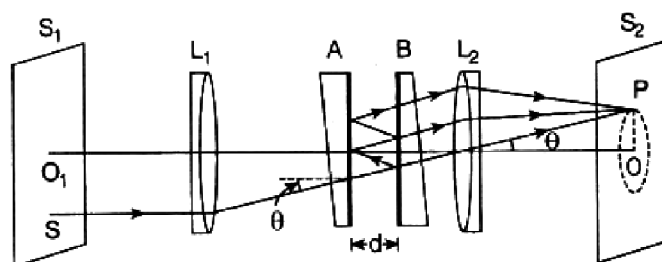


**Fig. 10.5**

**Set-up :** Monochromatic light from an extended source $S_1$ is made parallel by the collimating lens $L_1$. Each parallel ray suffers multiple reflections successively at the two silvered surfaces as shown. At each reflection, a small fraction of light is also transmitted so that each incident ray produces a group of coherent, parallel transmitted rays. There is a constant path difference between any two successive transmitted rays. A second converging lens $L_2$ brings these rays to focus on *P* in its focal plane where they interfere. Thus, the rays from all points of the source produce an interference pattern on a screen $S_2$ at the focal plane of $L_2$. This is known as **multiple beam interference**.

**Formation of fringes :** If $\theta$ be the inclination of a particular ray with the normal to the silvered surface of *A* and *d* the separation between the plates, then the path difference between two successive transmitted rays corresponding to the incident one is $2d \cos \theta$.

For maxima

$$2d \cos\theta = m\lambda \quad (m = 0,1,2, \ldots)$$

where $m$ is the order of interference and $\lambda$ the wavelength of light.

The locus of points in the source giving rays of constant inclination $\theta$ is a *circle*. Thus, with an extended source the interference pattern will be a series of bright concentric rings (Fig. 10.6) on a dark background. Each of the rings will correspond to a particular $\theta$-value.
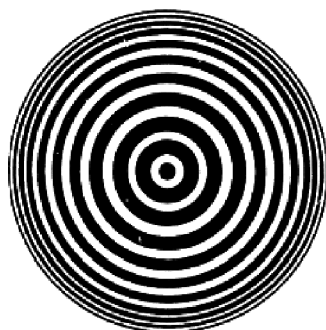


**Fig. 10.6**

**Visibility of fringes :** The *visibility* $V$ of the fringes measures the distinctness of fringes to the eye and is defined as

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{R}{1 + R^2} \qquad \ldots (10.4)$$

where R being the fraction of the intensity of incident light which is reflected at each silvered surface. R is also called *reflecting power.*

## 10.6 Applications of Fabry-Perot interferometer

Fabry-Perot interferometer is mainly used for the (i) comparison or measurement of wavelengths of spectral lines and for (ii) the investigation of hyperfine structure of spectral lines.

**1. Measurement of wavelength :** From the condition of maxima, $2d \cos\theta = m\lambda$, we get at the centre of the fringe pattern

$$2d = m_0 \lambda$$

where $m_0$ is the order of interference at the centre.

If at the centre, we get a bright fringe of order $m_1$ for a separation $d_1$, and again a bright fringe of order $m_2$ for a separation $d_2$, then

$$2d_1 = m_1\lambda; \qquad 2d_2 = m_2\lambda$$

$$\therefore \quad 2(d_2 - d_1) = (m_2 - m_1)\lambda = N\lambda$$

where $N = m_2\ m_1$, i.e., the number of fringes crossing the centre of the fringe system, when the separation changes from $d_1$ to $d_2$

$$\therefore \quad \lambda = \frac{2(d_2 - d_1)}{N} \qquad\qquad \text{... (10.5)}$$

**2. Determination of $\Delta\lambda$ (fine structure) :** With two very close wavelengths $\lambda_1$ and $\lambda_2$ emitted from a source we get two sets of fringes, one for each wavelength. Let mth order maximum for $\lambda_1$ coincide with pth order maximum for $\lambda_2$, with plate separation $d_1$ in a direction $\theta$.

$$\therefore \quad 2d_1\ \cos\theta = m\lambda_1 = p\lambda_2$$

$$\Rightarrow\ \lambda_1/\lambda_2 = p/m \qquad\qquad \text{... (10.6)}$$

If now we move from centre outward, the condition for next order to coincide is

$$\lambda_1/\lambda_2 = (p - 1)\ /\ (m - 1)\ \text{(order decreases outward)}$$

Since $m$ and $p$ are *rather large,* $p/m \approx (p - 1)/(m - 1)$ which implies that the maxima of $\lambda_1$ and $\lambda_2$ coincide over a *wide range of angle* covering the whole *field of view* appearing *distinct maximum.*

On increasing the plate separation to $d_2$ till the two sets of fringes again coincide showing maximum *distinct* field, then

$$2d_2\ \cos\theta = (m + s)\lambda_1 = (p + s + 1)\lambda_2 \qquad\qquad \text{... (10.7)}$$

$s$ being the increase in order of $\lambda_1$ and $s + 1$ that for $\lambda_2$

$$\therefore \quad 2(d_2 - d_1)\cos\theta = s\lambda_1 = (s + 1)\lambda_2$$

using (10.6) and (10.7).

At the centre $\theta = 0$, $\cos\theta = 1$; so, from above, we have

$$2(d_2 - d_1) = s\lambda_1 = (s + 1)\lambda_2 \Rightarrow s(\lambda_1 = \lambda_2) = \lambda_2$$

$$\therefore \quad s = \frac{\lambda_2}{\lambda_1 - \lambda_2}$$

$$\therefore \quad 2(d_2 - d_1) = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} = \frac{\lambda_m^2}{\lambda_1 - \lambda_2} = \frac{\lambda_m^2}{\Delta \lambda}$$

$$\Rightarrow \quad \Delta \lambda = \frac{\lambda_m^2}{2(d_2 - d_1)} \qquad \qquad ... (10.8)$$

where $\lambda_m$ = mean of $\lambda_1$ and $\lambda_2$ and $(d_2 - d_1)$ = distance moved over by the movable plate for two consecutive positions of maximum distinctness of the field of view.

**Note :** The method described is known as the method of coincidence and is not very accurate. The other method - the method of exact function - is due to Benoit.

## 10.7 Summary

■ The Michelson interferometer was used by Michelson for the standardization of the meter. He had found that the red cadmium line ($\lambda$ = 6438.4696 Å) is one of the ideal monochromatic sources and as such this wavelength was used as a reference for the standardization of the meter. In fact, he defined the meter by the following relation :

   1 meter = 1553164.13 red cadmium wavelengths, the accuracy is almost one part in $10^9$.

■ Michelson interferometer can also be used in the measurement of two closely spaced wavelengths.

■ If a plane wave falls on a plane parallel film, then the beam would undergo multiple reflections at the two surfaces and a large number of beams of successively diminishing amplitude will emerge on both sides of the plate. These beams (on either side) interfere to produce an interference pattern at infinity. If the reflectivity $R$ at each surface is close to unity, then the fringes so formed are much sharper than those by two beam interference and, therefore, the interferometers involving multiple beam interference have a high resolving power and hence find applications in high resolution spectroscopy. The Fabry-Perot interferometer is characterized by a high resolving power.

## 10.8  Review Questions and Answers

**1.**  **Explain how circular fringes are produced in Michelson interferometer? Show that the radii of circular fringes obtained by the interferometer are proportional to the square root of natural numbers.**

**Ans.** First part : see Article 10.3.

Second Part : Radii of circular fringes are proportional to the square root of natural numbers.

The path difference between the two rays reflected normally is 2d. when this path difference is an integral multiple of $\lambda$, all the rays reflected normally will be in phase. Hence for maximum intensity

$$2d = m'\lambda \qquad\qquad\qquad \text{(i)}$$

The path difference between the two rays coming to the eye from corresponding points $P_1$ and $P_2$ will be $2d \cos\theta$ where $\theta$ is the angle which the reflected beam make with the normal. Hence for maximum intensity

$$2d \cos\theta = m\lambda \qquad\qquad\qquad \text{(ii)}$$

From (i) and (ii), we have

$$2d - 2d\cos\theta = \left(n' - n\right)\lambda = n''\lambda \quad \left(n < n'\right)$$

or,    $2d\left(1 - \cos\theta\right) = n''\lambda$

where $n''$ is also an integer.

Now     $1 - \cos\theta = 1 - \left(1 - \dfrac{\theta^2}{2!} + \dfrac{\theta^4}{4!} - \ldots\right)$

Neglecting $\dfrac{\theta^4}{4!}$ and higher powers of $\theta$, we have

$$\frac{\theta^2}{2} = \frac{n''\lambda}{2d}$$

$$\therefore \quad \theta \propto \sqrt{n''}$$

If $\theta_1$, $\theta_2$, $\theta_3$ are the angles and $r_1$, $r_2$, $r_3$ ... the radii of the first, second, third ... rings respectively, then

$$\theta_1 = \frac{r_1}{x}, \ \ \theta_2 = \frac{r_2}{x}, \ \ \theta_3 = \frac{r_3}{x} \ldots$$

where $x$ is the distance of $M_2'$ from $S'$. As x *is constant* $r \propto \theta$, *but as* $\theta \propto \sqrt{n''}$

$$\therefore \quad r \propto \sqrt{n''}$$

It may be noted that as $M_1$ is moved forward or backward by half a wavelength, keeping $M_2'$ fixed a path change of $\lambda$ is introduced, hence a circular fringe will appear or disappear at the centre of the field. However, if $M_1$ and $M_2'$ are exactly coincident, the field of view will be dark. Under these conditions the path difference will be zero for all angles of incidence.

2. **What is the shape of interference fringes obtained with Michelson interferometer?**

**Ans.** When the mirrors $M_1$ and $M_2$ are perfectly perpendicular to the direction of two beams, circular fringes are obtained. If the two mirrors are made slightly inclined, straight localized fringes are obtained.

3. **How will you measure small difference in the wavelength of two waves with Michelson interferometer?**

**Ans.** See Article 10.4.

4. **How will you use Michelson interferometer to determine the thickness of a thin transparent sheet of refractive index µ?**

**Ans.** See Article 10.4.

5. **Explain the visibility of the fringes obtained with Fabry-Perot interferometer.**

**Ans.** See Article 10.5.

---

## 10.9 Illustrated Examples

**Example 1 :** The movable mirror of Michelson interferometer is moved through 0.02603 mm. If the wavelength used in 5206 Å, find the number of fringes shifted across the cross-wire of the eye-piece of the telescope.

**Solution :** Here $\lambda = 5206 \times 10^{-10} m$, $d = 0.02603 \times 10^{-3} m$ and $m = ?$

From the relation $2d = m\lambda$

$$m = \frac{2d}{\lambda} = \frac{2 \times 0.02603 \times 10^{-3}}{5206 \times 10^{-10}} = \mathbf{100}$$

**Example 2 :** A Michelson interferometer is set for white light straight fringes. When a mica sheet of thickness 0.005 cm is put in front of the fixed mirror, then to bring back the coloured fringes to their original position, the movable mirror is movedby 0.0025 cm. Calculate the refractive index of mica.

**Solution :** Let $n$ be the index of mica (sheet) of thickness $t$. So, the increase in optical path is $(n - 1)t$. If $d$ be the distance moved. by movable mirror, then to bring the coloured fringes in the original position,

$$(n - 1)t = d \Rightarrow n - 1 = d/t$$

$$\therefore \quad n = 1 + \frac{d}{t} = 1 + \frac{0.0025}{0.005} = 1 + 0.5 = \mathbf{1.5}$$

**Example 3 :** The wavelengths of two components of D-lines of sodium are 5890 Å and 5896 Å. By how much distance one of the mirrors of Michelson interferometer be moved so as to obtain consecutive positions of maximum distinctness?

**Solution :** Here $\lambda_1 = 5890 \times 10^{-10}\,m$, $\lambda_2 = 5896 \times 10^{-10}\,m$ and $d$ = distance to be moved by one of the mirrors in passing from one maximum intensity to the next.

$$\therefore \quad \lambda_1 \sim \lambda_2 = \frac{\lambda_1\lambda_2}{2d} \Rightarrow d = \frac{\lambda_1\lambda_2}{2(\lambda_1 \sim \lambda_2)}$$

$$\therefore \quad d = \frac{5890\times10^{-10}\times5896\times10^{-10}}{2\times(5896-5890)\times10^{-10}} - \frac{5890\times5896}{2\times6}\times10^{-10}$$

$$= 0.2894 \times 10^{-3}\ \text{m}$$

**Example 4 :** In an experiment for determining the index of a gas by Michelson interferometer, a shift of 140 fringes is observed when all the gas is removed from the tube. If the wavelength of light used is 5460 Å and the length of the tube is 20 cm, calculate the refractive index of the gas.

**Solution :**  Here $\lambda = 5460$ Å $= 5460 \times 10^{-10}m$, $m = 140$, $d = 20$ cm $= 0.2m$, $n$ = ?

$$(n - 1)\ d = m\lambda/2$$

$$n - 1 = \frac{m\lambda}{2d} = \frac{140\times5460\times10^{-10}}{2\times0.2} = 0.000191$$

$$\therefore \quad n = 1 + 0.000191 = \mathbf{1.000191}$$

**Example 5 :** Find the order number and angular radius of the 8th bright fringe in a Michelson interferometer. Assume that the two arms differ in length by 3 mm and the central fringe is bright. Wavelength of light is 589.6 nm.

**Solution :** For the central fringe, $2d = m\lambda$

$\therefore$ Order number, $m = \dfrac{2d}{\lambda} = \dfrac{2 \times 3 \times 10^{-3}}{589.6 \times 10^{-9}} \approx 10176$

$\therefore$ Order of 8th bright fringe $= m - 8 = 10176 - 8 = \mathbf{10168}$

If $\theta$ = angular radius of $8^{th}$ bright fringe, then

$$2d \cos\theta = 10168 \times \lambda \Rightarrow \cos\theta = \dfrac{10168 \times \lambda}{2d}$$

$\therefore \quad \cos\theta = \dfrac{10168}{10176}$ $(\because 26/\lambda = 10176)$

$\therefore \quad \theta = \cos^{-1}(10168 / 10176) = \mathbf{2.12^{o}}$

# Unit-11 : Diffraction

## Structure

## 11.1 Objectives

In going through this unit, you will learn :

   **(a) Fraunhofer diffraction :**

     1. Diffraction in a single slit

     2. Diffraction in a double slit

     3. Single slit vs. Double slit diffraction

     4. Circular aperture

     5. Multiple slits : Plane Diffraction Grating

     6. Absent spectra and Ghosts in a Diffraction Grating

     7. Overlapping of Spectral Lines

     8. Comparison of Grating and Prism Spectra

     9. Determination of wavelength of light by using Plane Diffraction Grating

     10. Resolving power and Rayleigh Criterion of Resolution

     11. Resolving power of a Grating

   **(b) Fresnel Diffraction :**

     12. Fresnel Half-period Zones of a Plane Wavefront

     13. Explanation of Rectilinear Propagation of Light

     14. Theory of Zone plate : Multiple Foci of a Zone Plate

     15. Zone plate vs. Convex lens

     16. Fresnel diffraction pattern of a straight edge, a slit and a wire

## 11.2 Introduction

Consider a plane light wave incident on a long narrow slit. According to geometrical optics one expects the region of the screen to be illuminated and the remaining portion (known as the geometrical shadow) to be absolutely dark. However, if the observations are made carefully then one finds that if the width of the slit is not very large compared to the wavelength, then the light intensity is not uniform and there is also some intensity inside the geometrical shadow. Further, if the width of the slit is made smaller, larger amounts of energy reach the geometrical shadow. This spreading-out of a wave when it passes through a narrow opening is usually referred to as diffraction

and the intensity distribution on the screen is known as the **diffraction pattern**. We will discuss the phenomenon of diffraction in this unit and will show that the spreading out decreases with decrease in wavelength. Indeed, since the light wavelengths are very small ($\lambda \sim 5 \times 10^{-5}$ cm), the effects due to diffraction are not readily observed.

We should point out that there is not much of a difference, between the phenomena of interference and diffraction, indeed, *interference corresponds to the situation when we consider the superposition of waves coming out from a number of point sources and diffraction corresponds to the situation when we consider waves coming out from an area source like a circular or rectangular aperture or even a large number of rectangular apertures* (like the diffraction grating).

The diffraction phenomena are usually divided into two categories : (i) **Fresnel diffraction** and (ii) **Fraunhofer diffraction**.

In the Fresnel class of diffraction, the source of light and the screen are, in general, at a *finite distance* from the diffracting aperture. In the Fraunhofer class of diffraction, the source and the screen are at *infinite distances* from the aperture. It turns out that it is much easier to calculate the intensity distribution of a Fraunhofer diffraction pattern which we plan to do in this unit. Further, the Fraunhofer diffraction pattern is not difficult to observe; all that one needs is an ordinary laboratory spectrometer; the collimator renders a parallel beam of light, and the telescope receives parallel beams of light on its focal plane. The diffracting aperture is placed on the prism table. In this unit we will also study the Fresnel class of diffraction. The underlying principle in the entire analysis is the **Huygens-Fresnel principle** according to which:

*Each point on a wavefront is a source of secondary disturbance and the secondary wavelets emanating from different points mutually interfere.*

We will first introduce the concept of **Fresnel half-period zones** to have a qualitative understanding of the Fresnel diffraction pattern.

## 11.3 Fraunhofer diffraction at a single slit

Let a plane wavefront of monochromatic light of wavelength $\lambda$ propagating normal to a slit $S_1 S_2$ be incident on it (Fig. 11.1). According to Huygens-Fresnel principle each point of the wavefront in the plane of the slit may be regarded as the source of secondary spherical wavelets. The secondary wavelets travelling normal to the slit are brought to focus by a convex lens $L$ on the screen at $C$. The wavelets travelling at an angle $\theta$ with the normal are brought to focus on Q. Let us calculate the intensity

of light at Q. Let the complex light disturbance at any instant due to secondary waves from the origin $O$ (the midpoint of the slit) be represented by $Ae^{i\omega t}$ where $A$ is the amplitude and $\omega$ is the circular frequency of the wave. The phase difference between the waves at $Q$ coming from $O$ and from apoint $Pat$ a distance $x$ from $O$ is given by
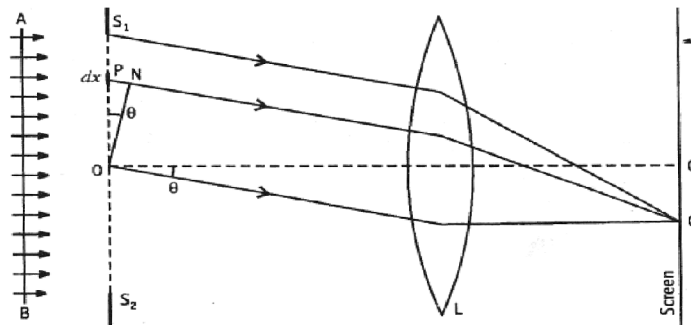


**Fig. 11.1**

$$\frac{2\pi}{\lambda} \times PN = \frac{2\pi}{\lambda} \cdot x \sin \theta = lx \qquad \qquad \dots (11.1)$$

where, $\qquad l = \frac{2\pi}{\lambda} \sin \theta$

Hence the disturbance at $Q$ due to secondary waves from will be proportional to $e^{i(\omega t - lx)}$. The disturbance at Qdue to the diffracting element can be written as

$$dy = C.A.dx.e^{i(\omega t - lx)}$$

where we assume the amplitude to be proportional to the width $dx$; $C$ is a proportionality constant.

If $a$ is the width of the slit, then the resulting complex disturbance at due to the waves coming from all the diffracting elements will be given by

$$y = \int_{-\frac{a}{2}}^{+\frac{a}{2}} CAe^{i(\omega t - lx)} dx$$

$$= CAe^{i\omega t} \left[ \frac{e^{-ilx}}{-il} \right]_{-\frac{a}{2}}^{+\frac{a}{2}}$$

$$= CAa \frac{\sin \frac{la}{2}}{\frac{la}{2}} . e^{i\omega t} \left[ \because \sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} \right]$$

The resultant intensity $l$ at $Q$ is obtained by multiplying $y$ by its complex conjugate $y^*$. Thus

$$I = y.y^* = I_0 \frac{\sin^2 \alpha}{\alpha^2} \qquad \dots (11.2)$$

where $I_0 = (CAa)^2$ and $\alpha = \frac{la}{2} = \frac{\pi}{\lambda} . a \sin \theta$ ... (11.3)

$\alpha$ depends on the angle of diffraction and thus (11.2) gives the intensity distribution i.e., it gives intensity for different values of $\theta$.

(a) **To find maxima and minima :**

The maxima or minima are given by

$$\frac{dI}{d\alpha} = 0$$

or, $\quad \dfrac{2 \sin \alpha \cos \alpha}{\alpha^2} - \dfrac{2 \sin^2 \alpha}{\alpha^3} = 0$

or, $\sin \alpha (\alpha \cos \alpha - \sin \alpha) = 0$

Hence either $\sin \alpha = 0$ ... (11.4)

or, $\alpha = \tan \alpha$ ... (11.5)

The value of $\dfrac{d^2 I}{d\alpha^2}$ is found to be positive for $\sin \alpha = 0$ and negative for $\alpha = \tan \alpha$. Hence the condition (11.4) will give **minima** and condition (11.5) will give **maxima**. Thus, for **minima**,

$\sin \alpha = 0$

or, $a = m\pi, \quad m = \pm 1, \pm 2, \pm 3 \dots$

or, $\alpha \sin \theta = m\lambda$ ... (11.6)

Here $m = 0$ is excluded because for this $\alpha = 0$ and since $\lim\limits_{\alpha \to 0} \dfrac{\sin \alpha}{\alpha} = 1$, it gives a maximum.
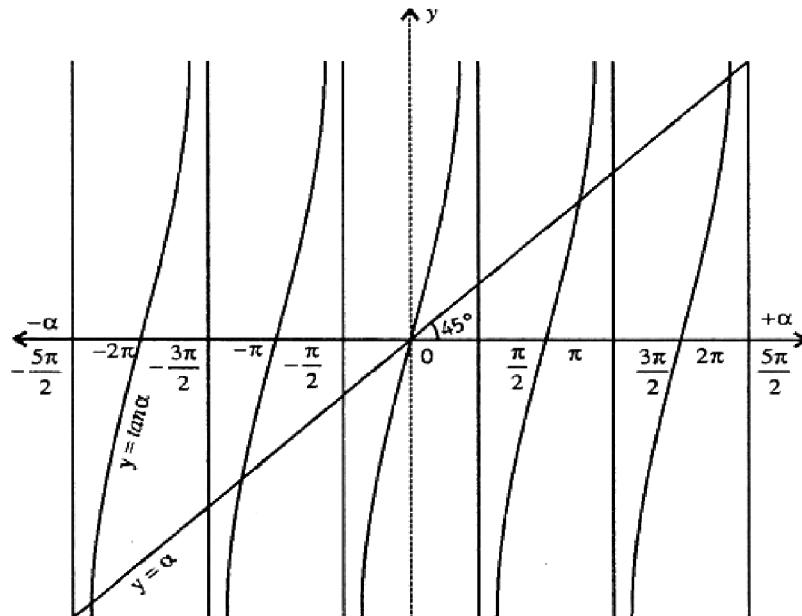
**Fig. 11.2**

The positions of maxima can be obtained by solving (11.5). It is transcendental equation and can be solved graphically by plotting the curves $y = \alpha$ and $y = \tan\alpha$ and finding the point of intersections as shown in Fig. 11.2. An inspection of the Fig. 11.2 shows that one value is $\alpha = 0$ and other values of $\alpha$ which will give maxima are less than but gradually approaching towards $\pm\dfrac{3\pi}{2}$, $\pm\dfrac{5\pi}{2}$, etc. $\alpha = 0$ gives the position of principal maximum and the other values gives secondary maxima.

The intensity of principal maximum is

$$I = \lim_{\alpha \to 0} I_0 \frac{\sin^2 \alpha}{\alpha^2} = I_0$$

The intensity of first secondary maximum is

$$I_1 \approx I_0 \frac{\sin^2 \dfrac{2\pi}{2}}{\left(\dfrac{3\pi}{2}\right)^2} = \frac{4}{9\pi^2} I_0 = \frac{I_0}{22}$$

The intensity of second secondary maximum is

$$I_2 \approx I_0 \frac{\sin^2 \frac{5\pi}{2}}{\left(\frac{5\pi}{2}\right)^2} = \frac{4}{25\pi^2} I_0 \approx \frac{I_0}{61}$$

Thus, the secondary maxima are very feeble compared with the principal maximum and their intensities diminish very rapidly with the increase in order number.

The intensity distribution in the **diffraction pattern** due to a single slit is shown in Fig. 11.3.
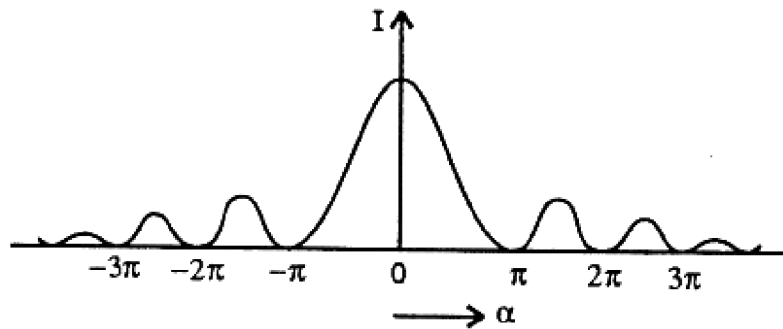


**Fig. 11.3**

The angle of diffraction $\theta_1$ for the first minima on either side of central maximum is given by (11.6) as

$$\alpha \sin \theta_1 = \lambda \ ; \quad \text{or,} \ \ \theta_1 \approx \sin \theta_1 \frac{\lambda}{a}$$

Therefore, **the angular width of central or principal maximum** $(2\theta_1 = 2\lambda/a)$ is *inversely proportional to the width (a) of the slit.* When $\alpha = \lambda$, we get from the relation $a \sin \theta = \lambda$, $\theta = 90^0$, i.e., the first minimum occurs at $90^0$, i.e., the *principal maximum fills the whole space.*

## 11.4 Diffraction at a circular aperture

Consider on *narrow circular* aperture $AB$ of diameter $d$ and centre at $C$, placed perpendicular to the plane of paper (Fig. 11.4) and a parallel beam of monochromatic light of wavelength $\lambda$ be incident *normally* on it. The light, in passing through the aperture, suffers diffractions. If a lens $L$ is placed in the path of the diffracted beam

very close to aperture a real image of the diffractive pattern is obtained on the screen *XY*, placed perpendicular to the plane of the paper, in the focal plane of the lens.
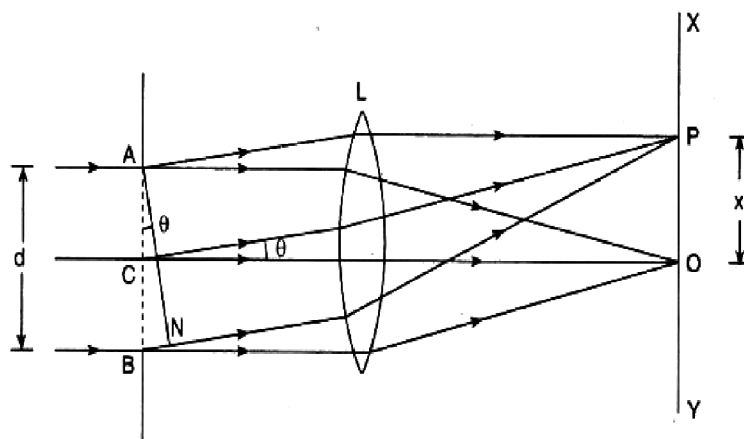


**Fig. 11.4**

The secondary waves travelling in the direction *CO* come to focus at *O* and bright image is formed as the secondary waves, equidistance from *C* in the upper and lower half travel the same distance before reaching *O* and reinforce each other.

The rays reaching a point *P* on the screen distant *x* from *O*, after diffraction at an angle θ from the extreme ends *A* and *B* of the aperture, have path difference

$$BN = AB \sin\theta = d \sin\theta$$

If this, path difference is an integral multiple of λ, then by reasoning similar to that as in the case of **single slit pattern**, the point *P* will have, *minimum intensity*, when

$$d \sin\theta = m\lambda = 2m\frac{\lambda}{2}, \text{ m = 0, 1, 2, ... etc. give } \textbf{minima}.$$

The point *P* will have maximum intensity if the path difference is equal to an odd multiple of λ/2 i.e.,

$$d \sin\theta = (2m+1)\frac{\lambda}{2}, \quad m = 0, 1, 2, ... \text{ etc. give } \textbf{maxima}.$$

As the circular aperture is **symmetrical** about the axis, the resultant pattern may be obtained by rotating Fig. 11.4 about the same axis. The point P thus will trace out a circular ring of uniform intensity, minimum or maximum depending upon, the conditions given above.

Thus, the diffraction pattern due to a circular aperture consists of a central bright disc called the **Airy's disc**, surrounded by alternate dark and bright concentric rings called the **Airy's rings** as shown in Fig. 11.5. The intensity of dark rings is zero and that of bright rings decreases gradually outward from $O$.
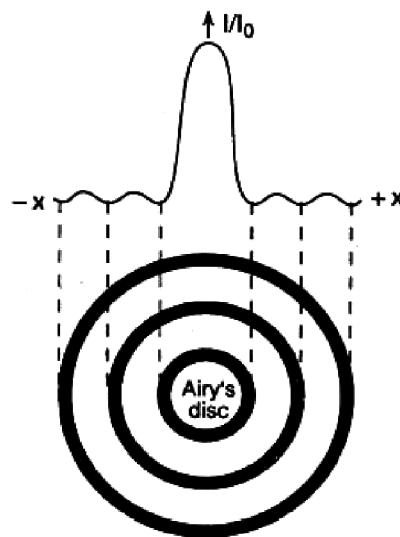


**Fig. 11.5**

**Radius of Airy's disc :** If the collecting lens $L$ lies very close to the aperture and the screen is at a very large distance, then

$$\sin\theta = \theta = \frac{x}{f} \qquad\qquad ... (11.7)$$

where $f$ is the focal length of the lens.
But for the first secondary minimum,

$$d\sin\theta = \lambda \Rightarrow \sin\theta = \theta = \frac{\lambda}{d} \qquad\qquad ... (11.8)$$

From (11.7) and (11.8), we get

$$\frac{x}{f} = \frac{\lambda}{d} \Rightarrow x = \frac{\lambda}{d}f \qquad\qquad ... (11.9)$$

where $x$ is the radius of the Airy's disc.

The radius of the first dark ring is slightly more than that given by the above relation (11.9). Airy has shown that

$$x = \frac{1.22\lambda}{d} f \qquad \qquad ...(11.10)$$

The intensity distribution of the bright and dark rings is similar to that given for a rectangular slit. Clearly, with increasing the diameter of the aperture, the radius of the central bright disc decreases.

**Importance :** The diffraction of a plane wave from anarrow circular aperture is of great importance, on account of its application to resolving power of telescopes and microscopes. A lens acts like a circular aperture and produces a diffraction pattern surrounded by alternate dark and bright rings. If there are two point objects lying close to each other, then two diffraction patterns are produced. The two objects are said to be resolved if the central maximum of one fall on the first secondary minimum of the other.

## 11.5  Diffraction at a double slit

Let a parallel beam of light of monochromatic light of wavelength $\lambda$ be incident normally on a surface containing two slits, each of width $a$ and separated by an opaque space of width $b$. The distance betweenany pair of corresponding points of the two slits is $d = a + b$. According to Huygens-Fresnel principle each point of the incident wavefront in the plane of the two slits may be regarded as the source of secondary spherical wavelets. The secondary wavelets travelling normal to the slits are brought to focus by a convex lens $L$ on the screen at $C$ (Fig. 11.6). The wavelets travelling at an angle $\theta$ with the normal are brought to focus at $Q$. Let us calculate the intensity of light at $Q$. Let the complex light disturbance at any instant due to secondary waves from the origin $O$ (mid-point of the first slit) be represented by $Ae^{i\omega t}$ where $A$ is the amplitude and $\omega$ is the circular frequency of the wave. The phase difference between the waves at $Q$ coming from $O$ and from a point $P$ at a distance $x$ from $O$ is given by

$$\frac{2\pi}{\lambda} \times PN = \frac{2\pi}{\lambda} \cdot x \sin \theta = lx \qquad \qquad ...(11.11)$$

where, $l = \frac{2\pi}{\lambda} \sin \theta$
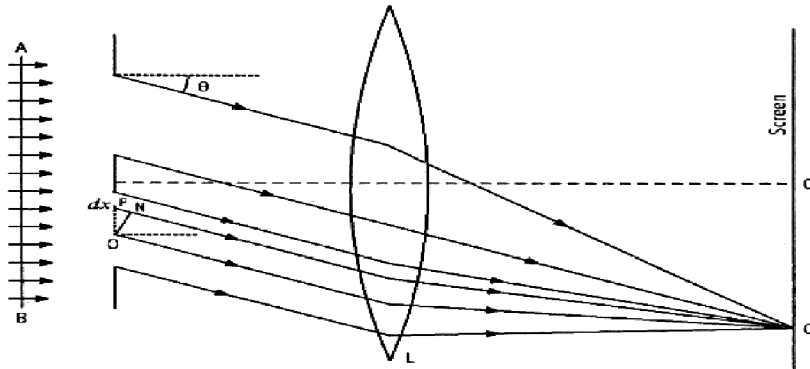
**Fig. 11.6**

Hence the disturbance at $Q$ due to secondary waves from will be proportional to $e^{i(\omega t - lx)}$ . The disturbance at $Q$ due to the diffracting element can be written as

$$dy = C.A.dx.e^{i(\omega t - lx)}$$

where we assume the amplitude to be proportional to the width $dx$; $C$ is a proportionality constant.

The resultant complex disturbance at $Q$ due to both the slits will be

$$y = \int\limits_{-\frac{a}{2}}^{+\frac{a}{2}} CAe^{i(\omega t - lx)}dx + \int\limits_{d-\frac{a}{2}}^{d+\frac{a}{2}} CAe^{i(\omega t - lx)}dx$$

$$= CAe^{i\omega t}\left[\frac{e^{-ilx}}{-il}\right]_{-\frac{a}{2}}^{+\frac{a}{2}} + CAe^{i\omega t}\left[\frac{e^{-ilx}}{-il}\right]_{d-\frac{a}{2}}^{d+\frac{a}{2}}$$

$$= CAa.\frac{\sin\frac{la}{2}}{\frac{la}{2}}.\left[1 + e^{-ild}\right].e^{i\omega t} \qquad\qquad ... (11.12)$$

The resultant intensity $I$ at $Q$ is obtained by multiplying $y$ by its complex conjugate $y^*$. Thus

$$y = yy^* = (CAa)^2 \frac{\sin^2 \frac{la}{2}}{\left(\frac{la}{2}\right)^2}\left(1+e^{-ild}\right)\left(1-e^{-ild}\right)$$

$$= (CAa)^2 \frac{\sin^2 \frac{la}{2}}{\left(\frac{la}{2}\right)^2}\left(2+e^{ild}+e^{-ild}\right)$$

$$= (CAa)^2 \frac{\sin^2 \frac{la}{2}}{\left(\frac{la}{2}\right)^2}\left(2+2\cos ld\right)$$

$$\therefore \quad e^{i\theta} = \cos\theta + i\sin\theta$$
$$\text{and} \quad e^{-i\theta} = \cos\theta - i\sin\theta$$

$$= 4I_0 \frac{\sin^2\alpha}{\alpha^2}.\cos^2\beta \qquad\qquad ... (11.13)$$

where $\quad I_0 = (CAa)^2, \quad \alpha = \dfrac{la}{2}$

$$= \frac{\pi}{\lambda}.a\sin\theta \quad \text{and} \quad \beta = \frac{ld}{2} - \frac{\pi}{\lambda}(a+b)\sin\theta \qquad\qquad ... (11.14)$$

Thus, we see that the resultant intensity in the diffraction pattern of two slits depends upon two factors :

(i) $I_0 \dfrac{\sin^2\alpha}{\alpha^2}$ which gives the diffraction pattern due to a single slit.

(ii) $\cos^2\beta$ which gives a system of interference fringes due to wavelets from the corresponding points of the two slits.

The set of diffraction fringes have greater dispersion while the set of interference fringes have lesser dispersion. These are superimposed on the diffraction set of fringes and are visible only in the region of central maxima.

**Condition for minima :** The resultant intensity would be zero when either factor

of (11.13) becomes zero. The factor $\dfrac{\sin^2\alpha}{\alpha^2}=0$ when $\sin\alpha = 0$ but $\alpha \neq 0$.

Therefore, $\alpha = m\pi$ ; $m = \pm 1, \pm 2, ...$

or, $a\sin\theta = m\pi$                                            ... (11.15)

These minima are known as **diffraction minima**.

The positions of minima due to the factor $\cos^2\beta$ are given by

$$\beta = \pm(2s+1)\pi/2 \,;\ s = 0,2,3,...$$

or,  $(a+b)\sin\theta = \pm(2s+1)\lambda/2$                    ... (11.16)

These minima are known as **interference minima**.

**Condition for maxima :** The positions of maxima due to the factor $\dfrac{\sin^2\alpha}{a^2}$ are

at $\alpha = 0$ and at values approaching $\pm\dfrac{3\pi}{2},\ \pm\dfrac{5\pi}{2}$, etc.

The positions of minima due to the factor $\cos^2\beta$ are given by

$\beta = p\pi;\ p = 0, 1, \pm 2, \pm 3...$

or, $(a + b)\sin\theta = p\lambda$                              ... (11.17)

**Missing order :** If the conditions for maxima of interference pattern (11.17) and minima of diffraction pattern (11.15) are simultaneously satisfied for a given value of θ then the corresponding interference maxima will be missing. If the slit width ais kept constant and separation $b$ is varied, then it is obvious from (11.17) the distance between consecutive interference maxima changes but the diffraction pattern due to single slit remains constant. Let for some value of θ conditions (11.15) and (11.17) are satisfied simultaneously. Then,

$$\frac{a+b}{a} = \frac{p}{m}$$                                  ... (11.18)

If $b = a$, $p = 2m = 2, 4, 6, ...,$ ($\because m = 1, 2, ...$) which implies that the 2nd, 4th, 6th, ...etc. order of interference maxima will be missing, i.e., coincide with the 1st, 2nd, 3rd, ..., etc. orders of diffraction minima.

If $b = 2a$, $p = 3m = 3, 6, 9, ...,$ $\Rightarrow$ 3rd, 6th, 9th, ... order of interference maxima will be missing, coinciding with the 1st, 2nd, 3rd, ..., etc. order of diffraction

minima.The central diffraction maximum will then have *five* interference maxima. And so on.
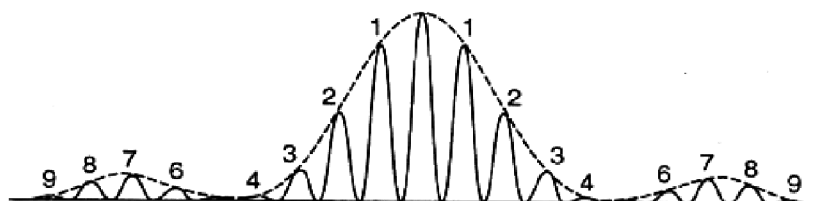


**Fig. 11.7**

Thus, as *b* increases to *a*, 2*a*, 3*a*, ..., the number of interference maxima within the central diffraction maximum increases due to missing orders.

For *b* = 4*a*, the central diffraction maxima will contain *nine* interference maxima, as shown in Fig. 11.7.

**What happens if a + b = a ?** This means *b* = 0 i.e., the two slits get joined and all the orders of interference maxima will be absent. The diffraction pattern observed will be that due to a *single slit* of width 2a.

### 11.5.1 Effect of various factors on double slit diffraction pattern

A number of factors such as slit width, slit separation, wavelength etc. may affect the double-slit diffraction pattern.

1. **Effect of increasing slit width :** If the slit width *a* is *increased*, the envelope of the fringe pattern so changes that the *central peak* becomes *sharper*. The spacing between the fringes however does not change as it depends only on the slit separation *b*. So, now within the central diffraction maximum, *a smaller number of interference maxima* will fall. The converse occurs on *decreasing* the slit-width.

2. **Effect of increasing slit separation :** If the slit separation *b* is *increased*, keeping the slit width *a* constant, the *fringes become closer together*, but the envelope of the pattern remains unaltered. So, *a greater number of interference maxima* will now fall within the central diffraction maximum.

3. **Effect of increasing wavelength :** If the wavelength $\lambda$ of the incident light is *increased*, the envelope of the fringe pattern becomes *broader*, and the fringes move *further apart*. The *converse effect* occurs on *decreasing* the wavelength.

### 11.5.2  Single slit vs. double slit diffraction

A comparison of the diffraction patterns due to a single slit and a double slit is worth noting. The main points of difference between the two are :

1. The single slit diffraction pattern consists of a central brightest maximum along with minima and secondary or subsidiary maxima of rapidly diminishing intensities on either side.

   The double slit pattern, on the other hand, consists of equally spaced interference maxima and minima within the central maximum. The intensity of the central maximum in double-slit pattern is four times that of the central maximum $(4I_0)$ due to a single slit diffraction. If, however, one of the slits is covered by an opaque screen, the pattern becomes similar to that due to a single slit.

2. The spacing of diffraction maxima and minima depends on the slit width $a$, but the spacing of the interference maxima and minima depends on both $a$ and $b$, where $b$ is the slit separation. The intensities of interference maxima are not constant but decreases to zero on either side of central maximum. The maxima reappear two or three times before the intensity becomes too low to be observable.

3. While in a double slit diffraction pattern there arises, under certain condition, missing or absent orders - missing of interference maxima, there is no such phenomenon in the single slit diffraction pattern.

## 11.6 Multiple slits: Plane diffraction grating

(a) **Construction :** Plane diffraction grating consists of a number of parallel and equidistant lines ruled on an optically plane and parallel glass plate by a fine diamond point. The number of such ruled lines per mm is of the order of 100. Each ruled line behaves as an opaque line while the transparent portion between two consecutive ruled lines behaves as a slit. If $a$ be the width of a clear space and $b$ be the width of a ruled line, then the distance $(a + b)$ is called **grating element or grating constant**. The two points in the consecutive clear spaces whose distance of separation is $(a + b)$, are called *corresponding points*.

Drawing of exactly parallel and equidistant lines on a glass plate by a diamond point is an extremely difficult task and hence the grating becomes too costly. To make it comparatively cheaper, a cast of this ruled surface is made with some transparent material and such casts are called *replica gratings*. Cellulose acetate, properly diluted, is poured on the surface of the master grating and then dried to a thin, tough film which can be easily separated from the master grating under water. This film can then

be mounted on a plane glass plate forming what is known as *replica grating*. Due to distortion and shrinkages of the film, replica seldom functions as well as its master.



(a)                                                                              (b)
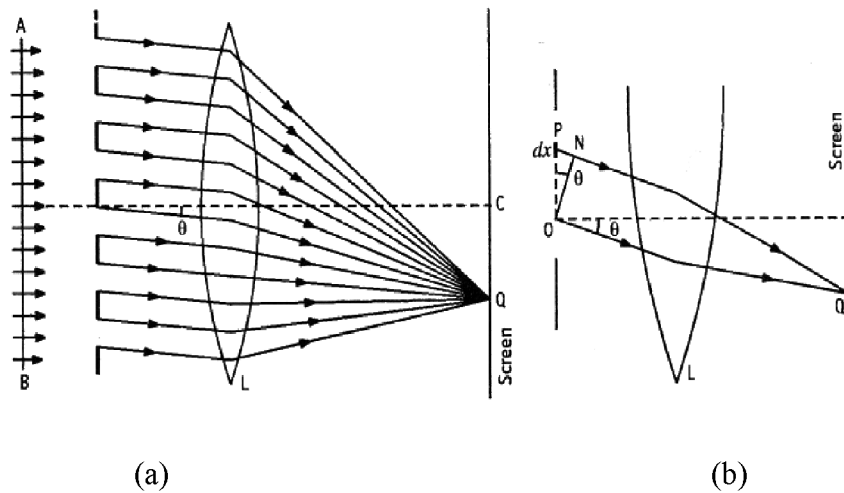
**Fig. 11.8**

**(b) Theory :** Let a parallel beam of monochromatic light of wavelength $\lambda$ be incident normally on a plane diffraction grating consisting of $N$ slits each of width $a$ and with equal opaque space $b$ between two successive slits. According to Huygens-Fresnel principle every point of the incident wavefront in the plane of the slits may be regarded as the origin of secondary spherical wavelets. The secondary wavelets travelling normal to the slits are brought to focus by a convex lens $L$ on the screen at $C$. The wavelets travelling at an angle $\theta$ with the normal are brought to focus at $Q$ [Fig. 11.8 (a)]. Let us calculate the intensity of light at $Q$. Let the complex light disturbance at $Q$ due to secondary wavelets from $O$ (the centre of the 1st slit) be represented by $Ae^{i\omega t}$ where $A$ is the amplitude and $\omega$ is the circular frequency. The phase difference between the waves at $Q$ [Fig. 11.8 (b)] coming from $O$ and from a point $P$ at a distance $x$ from $O$ is given by

$$\frac{2\pi}{\lambda} \times PN = \frac{2\pi}{\lambda}. x \sin\theta = lx \qquad \qquad ... (11.19)$$

where,          $l = \dfrac{2\pi}{\lambda}\sin\theta$

Hence the disturbance at $Q$ due to secondary waves from will be proportional to $e^{i(\omega t - lx)}$. The disturbance at $Q$ due to the diffracting element can be written as

$$dy = C.A.dx.e^{i(\omega t - lx)}$$

where we assume the amplitude to be proportional to the width $dx$; $C$ is a proportionality constant.

The resultant complex disturbance at $Q$ due to both the slits will be

$$y = z\int_{-\frac{a}{2}}^{+\frac{a}{2}} e^{-ilx}dx + z\int_{d-\frac{a}{2}}^{d+\frac{a}{2}} e^{-ilx}dx + \ldots + z\int_{(N-1)d-\frac{a}{2}}^{(N-1)d+\frac{a}{2}} e^{-ilx}dx$$

where $z = CAe^{i\omega t}$

$$= z\left[\frac{e^{-ilx}}{-il}\right]_{-\frac{a}{2}}^{+\frac{a}{2}} + z\left[\frac{e^{-ilx}}{-il}\right]_{d-\frac{a}{2}}^{d+\frac{a}{2}} + \ldots + z\left[\frac{e^{-ilx}}{-il}\right]_{(N-1)d-\frac{a}{2}}^{(N-1)d+\frac{a}{2}}$$

$$= CAa.\frac{\sin\dfrac{la}{2}}{\dfrac{la}{2}}\left[1 + e^{-ild} + \ldots + e^{i(N-1)ld}\right].e^{i\omega t} \qquad (\because z = CAei\omega t)$$

$$= CAa.\frac{\sin\dfrac{la}{2}}{\dfrac{la}{2}}.\frac{e^{-iNld}-1}{e^{-ild}-1}.e^{i\omega t} \qquad\qquad \ldots (11.20)$$

The resultant intensity $I$ at $Q$ is obtained by multiplying $y$ by its complex conjugate $y^*$. Thus

$$y = yy* = (CAa)^2.\frac{\sin^2\dfrac{la}{2}}{\left(\dfrac{la}{2}\right)^2}.\frac{e^{-iNld}-1}{e^{-ild}-1}.\frac{e^{+iNld}-1}{e^{+ild}-1}$$

$$= (CAa)^2\frac{\sin^2\dfrac{la}{2}}{\left(\dfrac{la}{2}\right)^2}.\frac{2-e^{-iNld}-e^{-iNld}}{2-e^{-ild}-e^{-ild}}$$

$$= (CAa)^2\frac{\sin^2\dfrac{la}{2}}{\left(\dfrac{la}{2}\right)^2}.\frac{2-2\cos Nld}{2-2\cos ld}$$

$$\therefore \quad e^{i\theta} = \cos\theta + i\sin\theta$$

and $$e^{-i\theta} = \cos\theta - i\sin\theta$$

$$= I_0 \frac{\sin^2\alpha}{\alpha^2} \cdot \frac{\sin^2 N\beta}{\sin^2\beta} \qquad \qquad \dots (11.21)$$

where

$$\left\{ \quad I_0 = (CAa)^2, \quad \alpha = \frac{la}{2} = \frac{\pi}{\lambda}a\sin\theta \right.$$

and

$$\left. \beta = \frac{\pi}{\lambda}d\sin\theta = \frac{\pi}{\lambda}(a+b)\sin\theta \quad \right\} \qquad \qquad \dots (11.22)$$

Thus, the resultant intensity depends on two factors :

(i) $I_0 \dfrac{\sin^2\alpha}{\alpha^2} = I_1$ which gives the diffraction pattern of a single slit.

(ii) $\dfrac{\sin^2 N\beta}{\sin^2\beta} = I_2$ which gives the interference pattern of the diffracted light beams from $N$ slits.

**Note :** *by putting N = 2 we can get the double slit intensity pattern.*

**Principal maxima :** If the slit width $a$ is very small and observation is confined to the neighbourhood of the central pattern the variation of the factor $\dfrac{\sin^2\alpha}{\alpha^2}$ is small and under this condition the maxima will be solely controlled by the factor $I_2 = \dfrac{\sin^2 N\beta}{\sin^2\beta}$. This factor is maximum when

$$\beta = m\pi; \quad m = 0, 1, \pm2, \pm3\dots$$
$$\text{or,} \quad (a+b)\sin\theta = m\lambda \qquad \qquad \dots (11.23)$$

These are known as **Principal maxima**.

For $\beta = m\pi$, $I_2 = \dfrac{0}{0}$

And hence indeterminate. But in the limit $\beta \to m\pi$ we get the maximum value of $I_2$.

$$\lim_{\beta \to m\pi} \frac{\sin N\beta}{\sin\beta} = \lim_{\beta \to m\pi} \frac{N\cos N\beta}{\cos\beta} = \frac{N\cos N(m\pi)}{\cos(m\pi)} = N$$

using $L$ Hospital rule. This gives

$$I_2 = N^2 \quad \text{and} \quad I = I_{\text{pm}} \text{ (say) } I_0 \frac{\sin^2 \alpha}{\alpha^2} \times N^2 = I_1 \times N^2 \qquad \text{... (11.24)}$$

Thus, the intensity of principal maxima increases as the number of slits ($N$) increases, but due to the presence of the factor $\frac{\sin^2 \alpha}{\alpha^2}$, whose value decreases with the increase of the angle of diffraction ($\theta$), the intensity of principal maxima decreases with the increase in the order number of bands.

**Conditions for secondary minima and maxima :**

The factor $I_2 = \frac{\sin^2 N\beta}{\sin^2 \beta}$ depends on $\beta$ and for maxima or minima we must have $\frac{dI_2}{d\beta} = 0$.

Now

$$\frac{dI_2}{d\beta} = \frac{2N \sin N\beta . \cos N\beta}{\sin^2 \beta} - \frac{2 \sin^2 N\beta . \cos \beta}{\sin^3 \beta}$$

$$= 2\frac{\sin^2 N\beta}{\sin^2 \beta}\left(N \cot N\beta - \cot \beta\right) \qquad \text{... (11.25)}$$

Hence for maxima or minima, either $\frac{\sin N\beta}{\sin \beta} = 0$ or, N cot N$\beta$ = cot $\beta$.

**(i) Secondary minima :**

When $\sin N\beta = 0$ but $\sin \beta \neq 0$ then $\frac{\sin N\beta}{\sin \beta} = 0$ and hence intensity becomes zero (i.e., minimum). Thus, for minima

$$N\beta = \pm s\pi$$

$$\text{or,} \quad (a+b)\sin \theta = \pm \frac{s}{N}\lambda \qquad (11.26)$$

where $s$ has integral values excepting $O$, $N$, $2N$, $3N$, etc. as for these values of $s$, sin $\beta$ = 0 and we obtain principal maxima. Thus, it is evident from (11.23) and (11.26) that between two consecutive principal maxima there are ($N - 1$) minima. Hence there will be $N - 2$ other maxima known as *secondary maxima* between any two adjacent principal maxima.

**(ii) Secondary maxima :**

The condition $N \cot N\beta = \cot \beta$ makes $\dfrac{dI_2}{d\beta} = 0$. Also, it can be shown that this condition makes $\dfrac{dI_2}{d\beta}$ negative. Thus, the values of $\beta$ which satisfy the condition $N \cot N\beta = \cot \beta$ will give the positions of secondary maxima, excepting $\beta = m\pi$ which gives principal maxima.

Now to find the intensity of secondary maxima we note that

$$N \cot N\beta = \cot \beta$$

or, $\quad N^2 \dfrac{\cos^2 N\beta}{\cos^2 \beta} = \dfrac{\sin^2 N\beta}{\sin^2 \beta}$

or, $\quad \dfrac{N^2 \left(1 - \sin^2 N\beta\right)}{1 - \sin^2 \beta} = \dfrac{N^2 \sin^2 N\beta}{N^2 \sin^2 \beta} = \dfrac{N^2}{1 + \left(N^2 - 1\right)\sin^2 \beta}$

Hence the intensity of secondary maxima is given by,

$$I_{sm} = I_1 \times \dfrac{N^2}{1 + \left(N^2 - 1\right)\sin^2 \beta} = \dfrac{I_{pm}}{1 + \left(N^2 - 1\right)\sin^2 \beta}$$

or, $\quad \dfrac{I_{sm}}{I_{pm}} = \dfrac{1}{1 + \left(N^2 - 1\right)\sin^2 \beta} \qquad\qquad \dots (11.27)$

This equation shows that as $N$ increases the intensity of secondary maxima relative to the principal maxima decreases. When $N$ is very large the secondary maxima become very weak. Therefore, secondary maxima are not generally observed with a grating having large $N$.

In Fig. 11.9 we have given a plot function $\dfrac{\sin^2 N\beta}{\sin^2 \beta}$ as a function of $\beta$ for $N = 5$ and $N = 11$. One can immediately see that as the value of $N$ becomes very large, the above function would become very sharply peaked at $\beta = 0, \pm\pi, \pm 2\pi. \dots$ The intensities of secondary maxima are also shown in Fig. 11.9(a). It is evident from this figure that

intensities of secondary fall off as we proceed towards the middle region between two consecutive principal maxima. These secondary maxima are in general unequally spaced and are not quite symmetrical [see Fig. 11.9(a)]. This lack of symmetry is greatest immediately adjacent to principal maxima and the secondary peaks are shifted a little towards the adjacent principal maxima.
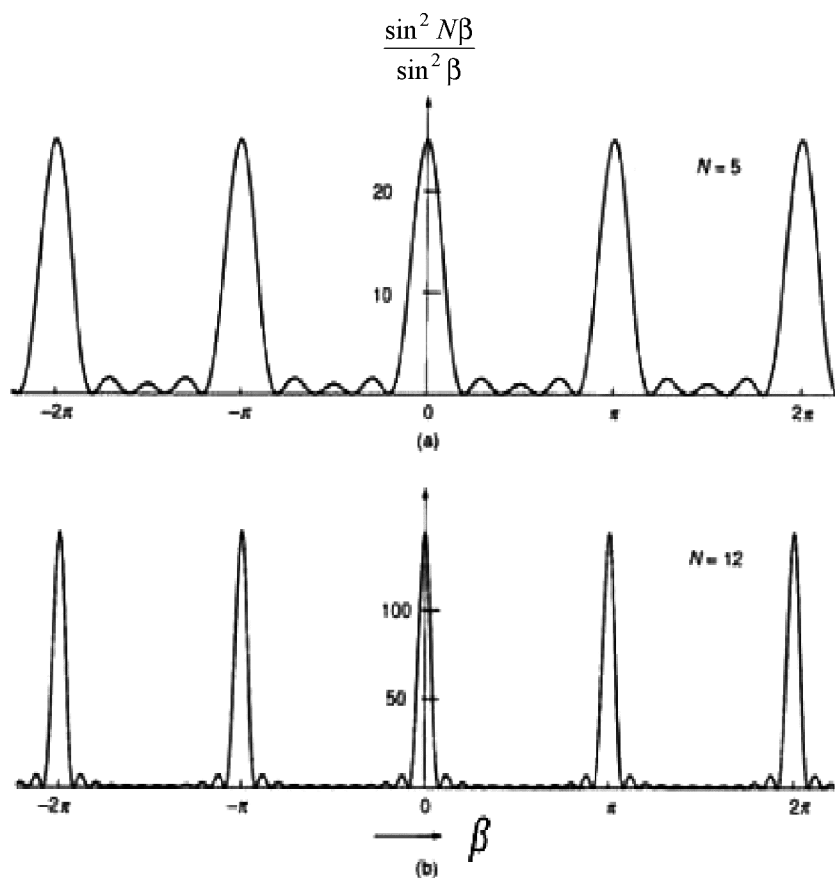


**Fig. 11.9**

### 4.6.1 Absent spectra and ghosts in a diffraction grating

For the $m$th order principal maximum in the direction $\theta$, we have the condition

$$(a + b) \sin \theta = m\lambda \qquad \qquad ...\,(11.28)$$

Suppose the value of $a$ is such that $s$th order diffraction minimum occurs in the same direction $\theta$ then

$$a \sin \theta = s\lambda \qquad \qquad ... (11.29)$$

If these two conditions are satisfied simultaneously then $m$th order principal maximum will be absent from the resulting spectra.

From (11.28) and (11.29) we get

$$\frac{a+b}{a} = \frac{m}{s}$$

For example, if $a = b$ then $m = 2s$. Then $m = 2, 4, 6$, etc. order of principal maxima will be absent corresponding to the diffraction minima $s = 1, 2, 3$, etc.

**Ghosts :** In an ideal grating the rulings should be equally spaced. But in practice there remains some errors in the rulings. If the error is *random* the grating gives a continuous background illumination. If the error is *progressive* in nature the spectral lines become sharper in planes which are different from the focal plane of the optical system.

The most common error is periodic in nature. It arises from defects in the driving mechanism of the ruling machine. It gives rise to false lines accompanying the principal maxima of ideal grating. These additional false lines are known as ghosts.

### 4.6.2 Overlapping of spectral lines

We know that for a grating having grating element $(a + b)$ the angle of diffraction $\theta$ in the $m$th order spectrum for a light of wavelength $\lambda$ is given by

$$(a + b) \sin \theta = m\lambda \qquad \qquad ... (11.30)$$

For a given grating $a + b$ is constant and hence $\theta$ will be the same when $m\lambda$ is constant. Thus, if the light incident on the grating surface consists of a large range of wavelengths, the spectral lines of longer wavelength and of lower order may overlap on the spectral lines of shorter wavelength and of higher order.

For example, the 3rd order maximum of light of wavelength $\lambda = 700$ nm and 4th order maximum of light of wavelength 525 nm will be formed in the same direction because

$$m\lambda = 3 \times 700 \; nm = 4 \times 525 \; nm = \text{constant}$$

## 11.7 Angular dispersive power of grating

The angular dispersive power of a grating is defined as the rate of change of the angle of diffraction ($\theta$) with the change in wavelength.

Thus, angular dispersive power $= \dfrac{d\theta}{d\lambda}$

If $\theta$ is the angle of diffraction for rays of wavelength $\lambda$ which form the $m$th order bright band then we have

$$(a + b) \sin \theta = m\lambda \qquad \qquad ... (11.31)$$

Differentiating with respect to $\lambda$ we get,

$$\frac{d\theta}{d\lambda} = \frac{m}{(a+b)\cos\theta} \qquad \qquad ... (11.32)$$

Thus, the dispersive power is

1. directly proportional to the order $m$ of the spectrum,

2. directly proportional to the number of lines per unit length; in other words, inversely proportional to the grating element $(a + b)$, and

3. large for large values of $\theta$.

■ Since in a given order $\theta$ increases with increase in $\lambda$ [see (11.31)], we conclude that the grating spectra are spread much more at red end than at the blue end of the spectrum.

■ When $\theta$ is small (not exceeding 6°) $\cos \theta$ remains almost constant. Under this situation $d\theta$ is proportional to $d\lambda$ i.e., angular separation between two spectral lines is proportional to the wavelength difference. Such a spectrum is known as **normal spectrum**.

## 11.8  Prism spectra vs. grating spectra

(i)    A prism produces only one spectrum, but a grating produces a number of spectra. Consequently, prism spectrum is brighter than the grating spectrum.

(ii)   In grating spectra there is chance of overlapping of different spectral lines in higher order. As there is only one prismatic spectrum there is no such overlapping in prismatic spectrum.

(iii)  In prism the deviation of violet colour is more than the deviation of red colour. But in grating spectrum the deviation of violet is smaller than that of red.

(iv)  Resolving power of grating is much greater than that of the prism.

(v)  Grating spectrum is nearly normal but the prism spectrum is never normal.

(vi)  Dispersion in the grating spectrum does not depend on the material of the grating, but dispersion in prism spectrum depends on the material of the prism.

(vii)  Some times, due to periodic error in ruling additional false spectra known as *ghosts* are produced by the grating. But prism spectrum is free from such ghosts.

## 11.9  Determination of wavelength of light bygrating

The wavelength of light can be measured by using a plane diffraction grating and ordinary spectrometer. Knowing $\theta$ and m the wavelength $\lambda$ can be calculated from the relation

$$(a + b) \sin \theta = m\lambda$$

$$\text{or,} \quad \lambda = \frac{(a+b)\sin\theta}{m} = \frac{\sin\theta}{pm}$$

The quantity $p = \frac{1}{a+b}$ represents the number of rulings per unit length, which is usually supplied by the manufacturer of the grating. If not known, it can be easily determined by measuring the angle of diffraction $\theta$ for a light of known wavelength and then using the

$$p = \frac{\sin\theta}{m\lambda}$$

## 11.10  Resolving power

By increasing the magnifying power of an optical instrument like a microscope or telescope we can increase the size of the image. But a larger image does not always reveal new details. The upper limit to the useful magnification is set by the fact that the image of a point object formed by an optical system with limited aperture consists of circular diffraction pattern. The **resolving power** *of an analyzing instrument is its ability to resolve two close point objects or to just separate two close spectral lines in their diffraction patterns.*

## 11.11  Rayleigh criterion of resolution

According to Lord Rayleigh two equally bright point sources could be just resolved by any optical system if the distance between themis such that the central maximum in the diffraction pattern due to one source coincides exactly with the first minimum in the diffraction pattern due to the other. This is known as **Rayleigh criterion of resolution**. In other words, the angular separation between the principal maxima of the two diffraction patterns should be equal to half the angular width of either principal maximum. Under this condition the resultant intensity distribution in the diffraction pattern shows a distinct dip (Fig. 11.10) at a point half- way between the two principal maxima and we are just able to identify them as separate. Accordingto Rayleigh the intensity at 'dip' is about $\dfrac{8}{\pi^2}$ times the intensity ofeither peak in the resultant intensity distribution.



Diffraction pattern of one object ⌐ ⌐ Diffraction pattern of other object
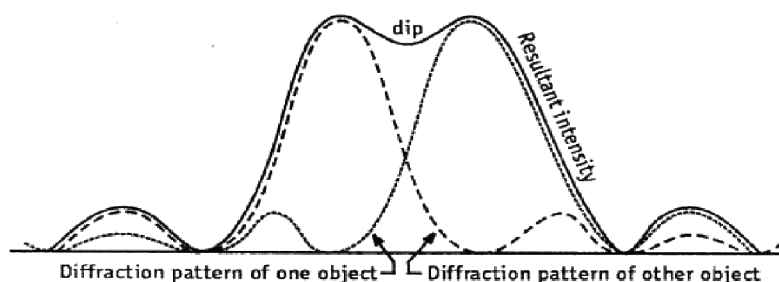
**Fig. 11.10**

The above criterion of resolution is also application to the resolution of two close spectral lines of equal intensity. Thus, two close spectral lines will be said to be just resolved when the angular separation between the principal maxima of two spectral lines in a given order is equal to half the angular width of either principal maximum.

## 11.12  Resolving power of a grating

The resolving power of a grating measures the ability to distinguish two very close spectral lines $\lambda$ and $\lambda + d\lambda$ and is defined by $\lambda/d\lambda$, $d\lambda$ being the smallest wavelength difference for which the spectral lines can be *just* resolved at wavelength $\lambda$.

**Expression for resolving power :** The $m$th order principal maximum of a spectral line of wavelength $\lambda$ of a grating for normal incidence is

$$(a + b) \sin \theta = m\lambda \qquad \ldots (11.33)$$

where $(a + b)$ is the grating element, $\theta$ the diffraction angle corresponding to $m$th order.
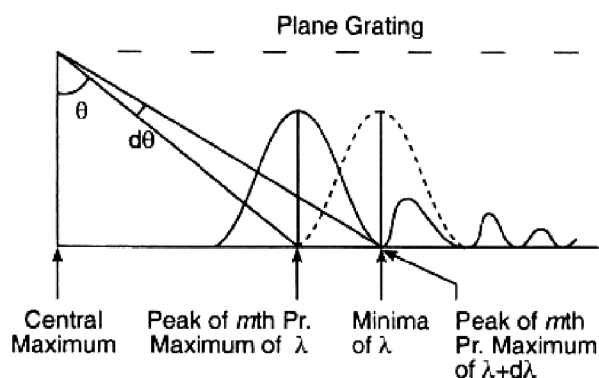


**Fig. 11.11**

Differentiating (11.33) we get

$$d\theta = \frac{md\lambda}{(a+b)\cos\theta} \qquad \ldots (11.34)$$

It gives the angular separation ($d\theta$) between the two mth principal maxima corresponding to $\lambda$ and $\lambda + d\lambda$ (Fig. 11.11). Eq. (11.33) can be written as,

$$N(a + b) \sin \theta = Nm\lambda \qquad \ldots (11.35)$$

where $N$ is the total number of slits in the grating.

Thus $N(a + b) \sin \theta = 0$, $N\lambda$, $2N\lambda$ ... $\qquad \ldots (11.36)$

gives the direction of 0, 1st, 2nd, ... order of principal maximum for the wavelength $\lambda$.

The positions of minimum intensity are given by

$$N(a + b) \sin \theta = \lambda, 2\lambda, \ldots (N - 1)\lambda, (N + 1)\lambda, (N + 2)\lambda,\ldots \quad (11.37)$$

An inspection of the (11.35) and (11.36) shows that the angle of diffraction $\theta + d\theta$ of the 1st minimum after the mth principal maximum will be given by

$$N(a + b) \sin (\theta + d\theta) = Nm\lambda + \lambda \qquad \ldots (11.38)$$

From (11.33) and (11.38) we can write

$$\frac{\sin(\theta + d\theta)}{\sin\theta} = \frac{Nm\lambda + \lambda}{Nm\lambda}$$

or, $\quad \dfrac{\sin\theta + d\theta\cos\theta}{\sin\theta} = 1 + \dfrac{1}{Mm}$

$$[\because d\theta \text{ is very small, } \sin d\theta \to d\theta \text{ and } \cos d\theta \to 1]$$

or, $\quad d\theta = \dfrac{1}{Nm\cot\theta}$ $\qquad$ ... (11.39)

It gives the angular half width of mth principal maximum.

Now equating (11.34) and (11.39) we get

$$\frac{1}{Nm} = \frac{md\lambda}{(a+b)\sin\theta}$$

Using (11.33) we get the resolving power of a grating as

$$\frac{\lambda}{d\lambda} = Nm \qquad\qquad .... (11.40)$$

Thus, the resolving power of a grating increases with the total number (N) of rulings effective in the formation of diffraction pattern and with the order number (m) of the spectrum.

Substituting the value of m from (11.33) we can also write

$$\frac{\lambda}{d\lambda} = \frac{N(a+b)\sin\theta}{\lambda}$$

$$= \frac{W.\sin\theta}{\lambda} \qquad\qquad ... (11.41)$$

where $W = N(a + b)$ is the total width of the ruled surface of the grating. Thus, a change in N (the number of rulings)for a given W will not change the resolving power. But note from (11.34) that the angular dispersive power of a grating increases with the increase in number of rulings per unit length (i.e., $\dfrac{1}{a+b}$). Thus, a grating with higher dispersive power does not necessarily possess a higher resolving power.

Putting $\theta = 90°$ in (11.41) we find that

$$\left(\frac{\lambda}{d\lambda}\right)_{max} = \frac{W}{\lambda}$$

For oblique incidence it can be easily shown that

$$\left(\frac{\lambda}{d\lambda}\right)_{max} = 2 \times \frac{W}{\lambda}$$

## 11.13 Resolving power of a telescope

The resolving power of a telescope is defined as the inverse of the least angle subtended at the objective by two distant close point objects which can be just distinguished as separate in its focal plane.

From the theory of Fraunhofer diffraction at circular aperture of diameter $D$ the angular separation $\theta$ of 1st minimum from the centre of diffraction pattern is given by

$$\sin\theta = \frac{1.22\lambda}{D}$$

As $\theta$ is very small practically,

$$\theta \approx \sin\theta = \frac{1.22\lambda}{D} \qquad\qquad ... (11.42)$$

This is the value of minimum angular separation between two point objects that can be just resolved by the telescope. So, the resolving power (R.P.) of the telescope would be

$$R.P. = \frac{1}{\theta} = \frac{D}{1.22\lambda} \qquad\qquad ... (11.43)$$

Therefore, telescope with larger objectives will have higher resolving power. The resolving power can also be increased by using light of smaller wavelength.

## 11.14 Fresnel's Assumptions

In Fresnel's diffraction, the source of light or screen or both are at finite distances from the obstacle or aperture. No lenses are used to make the rays parallel or convergent. The incident wavefront is not plane but is either spherical or cylindrical. As a result, the phase of secondary wavelets is not the same at all points in the plane of the

aperture or the obstacle causing diffraction. The resultant amplitude at any point of the screen is obtained by the mutual interference of secondary wavelets from different elements of unblocked portions of wavefront.

To explain this, Fresnel made the following **assumptions** :

(i) A wavefront could be divided into a large number of small elements or zones called Fresnel's zones, each of small area. The resultant effect at any point $P$ (Fig. 11.12) on the screen will depend on the combined effect of all the secondary waves originating from the various zones.

(ii) The effect at any point due to a particular zone will depend on the distance of the point from the zone.
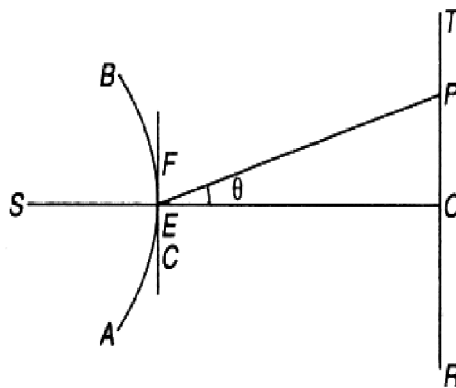


**Fig. 11.12**

(iii) The effect at $P$ will also depend upon the obliquity of the point with respect to the zone. The obliquity factor is proportional to $(1 + \cos \theta)$, where $\theta$ is the angle which $PE$ makes with $EO$. Considering an elementary wavefront at $E$, the effect is maximum at $O$ as $\theta = 0°$ and $\cos \theta = 1$. As we move from $O$ towards $P$ the value of $\theta$ increases and hence resultant effect decreases. Similarly in a direction $EF$ tangential to the wavefront, the resultant, effect is one half of that at $O$ as $\theta = 90°$ and $\cos \theta = $ . In the direction $ES$, the resultant effect is zero as $\theta = 180°$ and $\cos \theta = -1$. This explains the non-existence of the wave in the backward direction.

Fresnel type of diffraction is produced when light suffers diffraction at a straight edge, a narrow slit, a thin wire, a small circular hole, or an obstacle. An application of this type of diffraction is met with in the construction of a zone plate.

## 11.15    Fresnel's half-period zones : Plane wavefront

**Half-period zones :** In Fig. 11.13 consider a plane *ABCD* which represents a plane wavefront of monochromatic light of wavelength $\lambda$ travelling from left to right. According of *Huygen's principle* every point on the wavefront is regarded as the origin of secondary wavelets and at a given instant every one of these secondary wavelets pass through the point *O*. The resultant effect at *O* due to the whole wavefront will be equal to the disturbances reaching from different portions of the wavefront. To find the resultant effect we divide the entire wavefront into concentric zones as follow.
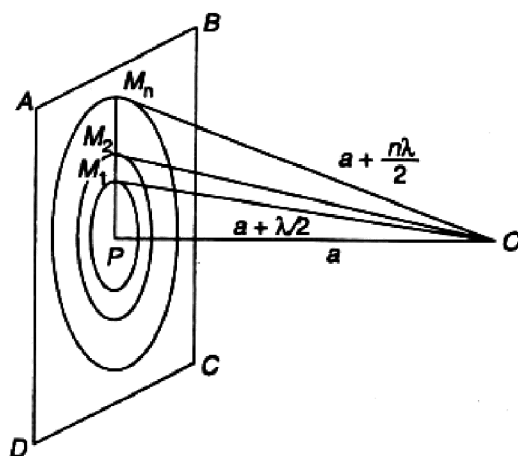


**Fig. 11.13**

From *O* drop a perpendicular on *ABCD* at the point *P*, the foot of the perpendicular is called the pole of the wave with respect to *O*. Let *OP* be equal to *a* and $\lambda$ be the wavelength of light waves. With *O* as centre and radius ($a+\lambda/2$) draw a sphere, cutting the wavefront in a circle at $M_1$ then

$$OM_1 = a + \lambda/2 \quad OM_1 - OP = \lambda/2$$

This simply means that the secondary wavelets originating from *P* and from the points on the circumference of the circle $M_1$, on reaching *O*, will differ in phase by

$$\frac{2\pi}{\lambda}\left(OM_1 - OP\right) = \pi$$

Now $\pi$ radian is equivalent to phase difference *T*/2. Hence the area enclosed by the circle $M_1$, is called **Fresnel's first half-period zone**.

Similarly, we can draw other spheres of radii $(a + 2\lambda/2)$, $(a + 3\lambda/2)$, $(a + 4\lambda/2)$... which will intersect the wavefront in circles $M_2$, $M_3$... This construction divides the wavefront into a number of Fresnel's half-period zones. It may be noted that the first half-period zone is a circle, while the second half-period zone is an annular ring between $M_1$ and $M_2$ and so on.

These are known as half-period zones because the waves reaching $O$ from $P$ and $M_1$ differ in phase by $\pi$ or $T/2$. Similarly, the waves reaching $O$ from $M_1$ and $M_2$ differ in phase by $\pi$ or $T/2$.

## 11.16    Explanation of rectilinear propagation of light

To explain rectilinear propagation of light we have to find the resultant effect of the whole wavefront at the point $O$ (Fig. 11.13). To do this divide the whole wavefront into half-period zones. The problem is then reduced to find the resultant of a large number of disturbances originating from the various annular rings into which the whole wavefront is divided. The amplitude due to wavelets produced by each zone reaching depends upon the following factors :

(i)   It is directly proportional to the area of the zone.

(ii)  It varies inversely as the distance of zone from the point $O$. Greater the distance smaller is the amplitude reaching $O$.

(iii) It varies with the obliquity factor $(1 + \cos \theta)$. The amplitude decreases with increasing obliquity.

The areas of each of the half period zones are approximately equal, as shown below.

The square of the radius of the *first half-period zone*

$$PM_1^2 = (a + \lambda/2)^2 - a^2 = a^2 + \lambda^2/4 + al - a^2 = a\lambda \qquad ... (11.44)$$

as $\lambda^2/4$ involves the square of a very small quantity and can be neglected as compared to $a\lambda$.

$\therefore$   Area of first half-period zone $= \pi a\lambda$

To find the area of the $n$th half-period zone.

Square of radius of the nth circle

$$PM_n^2 = (a + n\lambda/2)^2 - a^2 = an\lambda$$

Similarly Square of radius of the $(n-1)$th circle

$$PM_{n-1}^2 = a(n-1)\lambda$$

∴ Area of the nth half-period zone $= \pi\left(PM_n^2 - PM_{n-1}^2\right)$

$$= \pi\left[an\lambda - a(n-1)\lambda\right]$$

$$= \pi a\lambda \qquad\qquad\qquad ...\,(11.45)$$

This is the same as that for the first half period zone.

*Thus, we find that the areas of the various half period zones are independent of the order of the zone and are nearly equal. The radii of these zones are proportional to $\sqrt{n}$ where n = 1, 2, 3 etc.*

Suppose every point on the plane is in a state of vibration and sends out secondary wavetrains, then the light vibrations at $O$ are due to superposition of these waves (Fig. 11.14). Since the incident wave is plane one, all the points on the plane $ABCD$ are in the same phase. Their distances from $O$ are different and consequently the secondary waves which these points send out, reach $O$ in different phases.
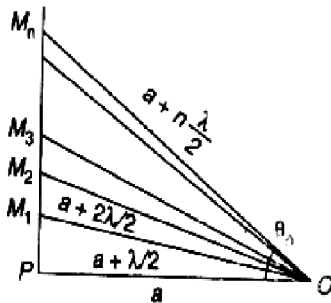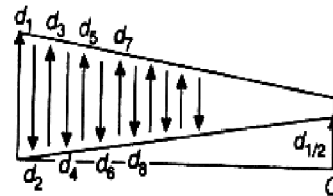


**Fig. 11.14**                    **Fig. 11.15**

The phase of the wavelets coming from $P$ is zero. The distance of $O$ from points in the first zone lies between a and $(a + \lambda/2)$. Hence the phase of wavelets from intermediate points between $P$ and $M_1$ will vary from 0 to $\pi$. Thus

Average phase of all wavelets from first zone $= \dfrac{0+\pi}{2} = \dfrac{\pi}{2}$ radian.

Similarly, the phase difference of wavelets from $M_1$ and $M_2$ will lie between $\pi$ and $2\pi$. Hence

Average phase/difference of wavelets from second zone $= \dfrac{\pi + 2\pi}{2} = \dfrac{3\pi}{2}$ radian. This is opposite to that due to the wavelets from the first half-period zone.

Similarly, the average phase difference of wavelets from the third zone will be $5\pi/2$ and fourth zone will be $7\pi/2$ and so on.

It is clear that the resultant phase difference between two consecutive zones is $\pi$ radian, while that between alternate zones is $2\pi$ radian. Thus, if the resultant amplitude of wavelets from first half-period zone is positive, that from the second half-period zone will be negative, that from third zone positive and fourth zone negative and so on (Fig. 11.15).

Let $d_1$, $d_2$, $d_3$, etc., be the resultant amplitudes at $O$ due to first, second, third etc., half-period zones respectively, then

Resultant amplitude $D = d_1 - d_2 + d_3 - d_4 \ldots \pm d_n$

The amplitudes of the successive zones depend upon the obliquity factor $(1 + \cos \theta_n)$ where $\theta_n$ is the angle which the direction of $O$ from the nth zone makes with $PO$. It is assumed that the obliquity factor varies slowly so that it may be regarded constant over a single half-period zone. As $\theta$ increases from zero, $\cos \theta$ decreases very slowly first, but more rapidly for larger values of $\theta$. Thus, successive amplitudes decrease at first, slowly but more rapidly for higher values of $n$. Hence the amplitude goes on decreasing with the order of zones. Thus, each term in the above series is slightly less than the one preceding it and is greater than the one succeeding it. To a first approximation, we can write

$$d_2 = \frac{d_1 + d_3}{2}, \quad d_4 = \frac{d_3 + d_5}{2} \text{ and so on.}$$

The above series can now be rewritten in the following form

$$D = \frac{d_1}{2} + \left( \frac{d_1 + d_3}{2} - d_2 \right) + \left( \frac{d_3 + d_5}{2} - d_4 \right) + \ldots \frac{d_n}{2}$$

$$= \frac{d_1}{2} + \frac{d_n}{2} \text{ (If } n \text{ is odd)}$$

$$D = \frac{d_1}{2} + \frac{d_{n-1}}{2} \text{ (If } n \text{ is even)}$$

If $n$ is sufficiently large, the effect due to the $n$th zone becomes negligible and resultant amplitude due to the whole wave

$$D = \frac{d_1}{2}$$

As intensity $I$ is proportional to the square of the amplitude

$$I = \frac{d_1^2}{4}$$

Hence the intensity at $O$ due to the wavelets from all the zones is equal to one fourth of the intensity due to the waves from the first half-period zone.

If an obstacle is placed at $P$, (Fig. 11.13) the resultant disturbance at $O$ is equal to half the disturbance due to the *first exposed zone*. As the displacement decreases rapidly with the order of the zone, if the obstacle at $P$ blocks a considerable number of half-period zones the effect will be negligible and practically no light will be received at $O$. In other words, **light travels approximately in a straight line**.

## 11.17 Theory of a zone plate : Multiple foci of a zone plate

**Zone plate :** It is a transparent plate on which circles whose radii, proportional to the square roots of natural numbers 1, 2, 3,... are drawn. The alternate annular zones thus formed are blocked. Such a plate behaves like a convex lens and produces an image of a source of light on the screen placed at a suitable distance.
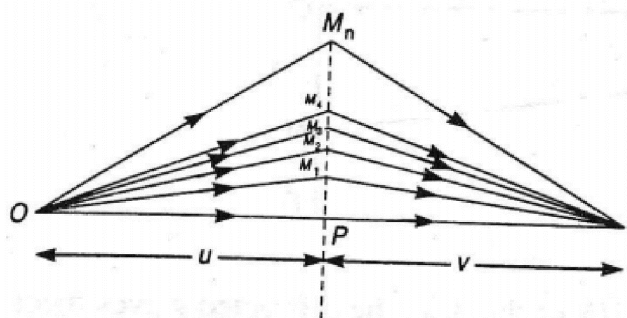


**Fig. 11.16**

**Theory :** Let $O$ be a luminous point object, emitting spherical waves of wavelength $\lambda$ whose effect at the point $I$ on the screen is required. Consider an imaginary plane through $P$ of a transparent medium lying perpendicular to the plane of the paper and

the line joining $OI$. Divide this plane into zones bounded by circles having centres at $P$ and radii $PM_1 = r_1$, $PM_2 = r_2$ ... $PM_n = r_n$ such that

$$OM_1 + IM_1 = OP + IP + \frac{\lambda}{2}$$

$$OM_2 + IM_2 = OP + IP + \frac{2\lambda}{2}$$

.............................................

$$OM_n + IM_n = OP + IP + \frac{n\lambda}{2}$$

The annular rings thus formed are half-period zones for the image $I$ because length of the path of light through the corresponding points of any two consecutive zones, differs by $\lambda/2$. To find the radius $r_n$ of the $n$th circle, we have

$$OM_n + IM_n = OP + IP + \frac{n\lambda}{2} \qquad\qquad ... (11.46)$$

Let $\qquad\qquad OP = u$ and $IP = v$

Now $\qquad\qquad OM_n = \left(OP^2 + PM_n^2\right)^{\frac{1}{2}} = \left(u^2 + r_n^2\right)^{\frac{1}{2}}$

$$= u\left(1 + \frac{r_n^2}{u^2}\right)^{\frac{1}{2}} = u + \frac{r_n^2}{2u}$$

Because $u$ is very large as compared to $r_n$ and higher powers of $\dfrac{r_n^2}{2u}$ can be neglected.

Similarly $\quad IM_n = \left(v^2 + r_n^2\right)^{\frac{1}{2}} = v + \frac{r_n^2}{2v}$

Substituting this value in (11.46), we have

$$u + \frac{r_n^2}{2u} + v + \frac{r_n^2}{2v} = u + v + \frac{n\lambda}{2}$$

or, $\quad r_n^2\left(\frac{1}{u} + \frac{1}{v}\right) = n\lambda \quad \Rightarrow \frac{1}{u} + \frac{1}{v} = \frac{n\lambda}{r_n^2}$

Applying the sign convention, we have

$$\therefore \quad \frac{1}{v} - \frac{1}{u} = \frac{n\lambda}{r_n^2} \qquad \qquad \dots (11.47)$$

$$\text{or} \quad r_n^2 = \frac{n\lambda uv}{u-v} \qquad \qquad \dots (11.48)$$

since $u$, $v$ and $\lambda$ are constants

$$r_n \propto \sqrt{n} \qquad \qquad \dots (11.49)$$

The area of the $n$th zone is given by

$$\pi\left(r_n^2 - r_{n-1}^2\right) = \left\{\frac{n\lambda uv}{u-v} - \frac{(n-1)\lambda uv}{u-v}\right\} = \frac{\pi\lambda uv}{u-v}$$

This is independent of $n$; hence for a given object and image, the areas of all the zones remain the same. Further, the area diminishes as $v$ and $u$ decreases i.e., as the plate is approached by the object or the image. The numerical magnitudes of the displacements $d_1$, $d_2$, $d_3$ ... at $I$ due to wavelets from the various zones, diminish only slightly with the order of the zone. Hence the resultant displacement at $I$

$$D = d_1 - d_2 + d_3 - d_4 \dots$$

**How zone plate acts like a convex lens :** If now the alternate zones say even are blocked, then the resultant displacement at $I$ will become

$$D = d_1 + d_3 + d_5 \dots$$

which is many times greater than that due to all the zones and hence the intensity at $I$ is very much increased. This explains the focusing action of the zone plate. Under these conditions I is said to be the image of the object $O$.

**Focal length :** The relation between $u$ and $v$, the respective distances of the object and the image is given by relation (11.47), i.e.,

$$\frac{1}{v} - \frac{1}{u} = \frac{n\lambda}{r_n^2}$$

This is a result similar to the one found for a convex lens, i.e.,

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

Thus, the focal length of the zone plate

$$f = \frac{r_n^2}{n\lambda} \qquad \qquad \dots (11.50)$$

If the radius of the half-period zone is 0.316 mm and the wavelength $\lambda = 5 \times 10^{-5}$ cm, then the zone plate will behave like a lens of focal length 20 cm.

**Construction of zone plate :** Zone plate is a system of areas corresponding to the half period zones. To construct a zone plate, concentric circles whose radii are proportional to the square roots of the natural numbers are drawn on a sheet of white paper. The alternate zones are painted black, and a very much reduced photograph of this drawing is obtained on a glass plate. The negative, when held in the light path from a distant point source, produces a large intensity at a point on its axis at a distance corresponding to the size of the zone and the wavelength of light used.



(a)          (b)
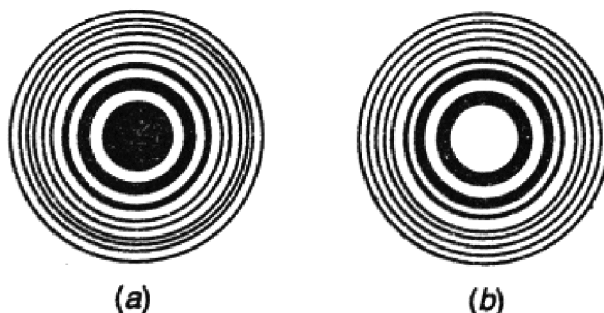
**Fig. 11.17**

If odd zones are transparent and even zones are opaque, as shown in Fig. 11.17(b) it is said to be a **positive zone plate**. If the even zones are transparent and odd zones are opaque, as shown in Fig. 11.17(a) it is said to be a **negative zone plate**.

### 4.17.1 Comparison between convex lens and zone plate

(i) Focal length of a lens is given by the relation $\frac{1}{f} = (\mu - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$ whereas that of the zone plate is given by $\frac{1}{f} = \frac{n\lambda}{r_n^2}$.

(ii) The focal length of a lens is directly proportional to the wavelength and hence is greater for red rays than for violet rays. The focal length of a zone plate is inversely proportional to the wavelength and hence is greater for violet rays than for red rays.

(iii) Light, in passing through the lens, takes the same time to go from $O$ to $I$ through any part of the lens, whereas in a zone plate disturbance from any transparent zone reach the point $I$ one period later than the disturbances from the next inner zone.

(iv) For a fixed distance of the object, a lens produces only one image, whereas a zone plate produces a number of images. In other words, it is **multiple foci**. It is because the number of half-period zones contained in an area depends upon the position of the screen. If therefore, a point $I_1$ lies near to $P$ than $I$ so that the transparent ring contains *three* half-period zones, then the point $I_1$ will have again maximum illumination and thus will be the image of $O$. Similar will be the case when 5, 7.... half-period zones are contained in a transparent ring. The position of the various foci is given by

$$f_1 = \frac{r_n^2}{n\lambda}, \quad f_2 = \frac{r_n^2}{3n\lambda}, \quad f_3 = \frac{r_n^2}{5n\lambda}, \cdots$$

The intensity of the image, however, decreases, as the focal length decreases.

## 11.18    Fresnel's Integrals : Cornu's spiral

The **Fresnel's integrals** are defined by the following equations :

$$X(v) = \int_0^v \cos\frac{\pi v^2}{2} dv \qquad \qquad \text{... (11.51)}$$

and
$$Y(v) = \int_0^v \sin\frac{\pi v^2}{2} dv \qquad \qquad \text{... (11.52)}$$

Since the integrands are even functions of $v$, the Fresnel's integrals $X(v)$ and $Y(v)$ are odd functions of $v$ :

$$X(-v) = -X(v) \text{ and } Y(-v) = Y(v) \qquad \qquad \text{... (11.53)}$$

Further, since

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \qquad \qquad \text{... (11.54)}$$

We have
$$\int_{-\infty}^{\infty} e^{i\pi v^2/2} dv = \sqrt{\frac{\pi}{-i\pi/2}} = \sqrt{2}e^{i\pi/4} = (1+i) \qquad \qquad \text{... (11.55)}$$

Now, 
$$\int_{-\infty}^{+\infty} \exp\left[i\frac{\pi v^2}{2}\right] dv = 2\left[\int_{0}^{\infty} \cos\frac{\pi v^2}{2} dv + i\int_{0}^{v} \sin\frac{\pi v^2}{2} dv\right]$$

$$= 2\left[X(\infty) + iY(\infty)\right]$$

Thus, using (11.55), we get $X(\infty) = 1/2 = Y(\infty)$.

To summarize, the Fresnel's integrals have the following important properties :

$$X(\infty) = Y(\infty) = 1/2 \; ; \; X(0) = Y(0) = 0 \qquad\qquad ... (11.56)$$

$$X(-v) = -X(v) \text{ and } Y(-v) = -Y(v)$$

Figure 11.18 gives a parametric representation of the Fresnel's integrals and is known as the Cornu's spiral (see Fig. 11.18) which is in the form of a double spiral. The horizontal and the vertical axes represent X(v) and Y(v) respectively. This spiral curve will be understood from the following discussion of the integrals.
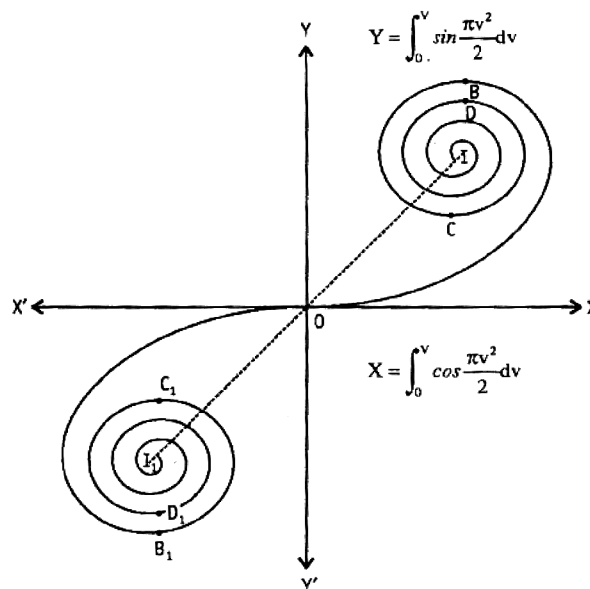


**Fig. 11.18**

**Discussions of Fresnel's integrals :**

(a) When $v = 0$, we get $X = Y = 0$. Thus, the curve passes through the origin.

(b) When the sign of v is changed, the signs of $X$ and $Y$ will change but their numerical values will remain unchanged. The curve connecting $X$ and $Y$ is, therefore, symmetrical about the origin.

(c) (i) If $\Psi$ be the angle which the tangent drawn to a point on the curve makes with $X$-axis then, $\tan\Psi = \dfrac{dY}{dX} = \tan\dfrac{\pi v^2}{2}$. Thus $\Psi = \dfrac{\pi v^2}{2}$; or $v^2 = \dfrac{2\Psi}{\pi}$. When $v = 0$, $\psi = 0$. The curve is, therefore, parallel to the $X$-axis at the origin.

(ii) For $v^2 = 1$, $\psi = \pi/2$. The curve is here parallel to $Y$-axis.

(iii) For $v^2 = 2$, $\psi = \pi$. The curve at this point is again parallel to $X$-axis.

(iv) For $v^2 = 3$, $\psi = 3\pi/2$. The curve is again parallel to $X$-axis.

(d) The radius of curvature of the curve at a point is given by, $\rho = \dfrac{dv}{d\psi} = \dfrac{1}{\pi v}$

$\left[\because v^2 = \dfrac{2\psi}{\pi}; \text{ or, } 2v\,dv = \dfrac{1}{\pi}d\psi\right]$. When $v = 0$, $\rho = \infty$. The curve has a point of inflection at this point. As the value of $v$ increases the value of $\rho$ decreases. Thus, the curve assumes the form of a *double spiral*. When $v = \infty$, $\rho = 0$; i.e., the curve reduces to points as are shown by $I$ and $I_1$ in Fig. 11.18.

## 11.19　Fresnel's diffraction at a straight edge

Consider a straight edge $C$ and an illuminated narrow slit $S$ parallel to each other and perpendicular to the plane of the paper (Fig. 11.19).
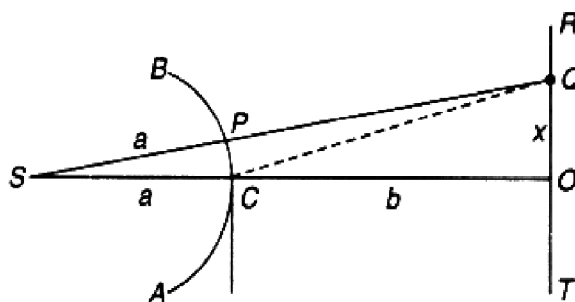


**Fig. 11.19**

It is required to determine the illumination of the screen $TR$ placed perpendicular to the plane of paper. Join $SC$ and produce it to meet the screen at $O$. A straight line

on the screen through $O$ perpendicular to the plane of the paper defines the limits of the geometrical shadow.

According to the laws of geometrical optics the illumination should start abruptly at $O$. In practice however dark and bright bands of unequal width and of varying intensity are observed in the region above $O$. The intensity goes on decreasing gradually, as we move into the geometrical shadow below $O$.

To study the intensity at any point $Q$ distant x above $O$ on the screen, let $ACB$ be the incident wavefront at $C$ originating from the source $S$. Join $QS$, cutting the wavefront at $P$, then $P$ is the pole of the wavefront with respect to the point $Q$. With $P$ as the pole construct Fresnel's half-period elements. The effect at $Q$ depends upon the number of half-period elements contained in $PC$ and the effect of the upper half of the wave $PB$.

The effect at $O$ is due to the upper half of the wavefront only. Hence the displacement at $O$ is half of the displacement which would have been there in the whole wavefront was effective. The intensity at $O$, therefore, is 1/4 of the intensity at a point far removed from $O$ where the whole of the wavefront is effective.

As we move into the geometrical shadow, the pole $P$ moves from $C$ towards A and first, second, third etc., half-period elements are intercepted respectively. The intensity, therefore, goes on decreasing gradually.

As we move from $O$ towards $R$, first, second, third, etc., half-period elements are exposed. The illumination at a point $Q$, distant x from $O$ on the screen, is due to the complete half of the wave surface $PB$ together with the resultant of the number of half-period elements contained in $CP$.
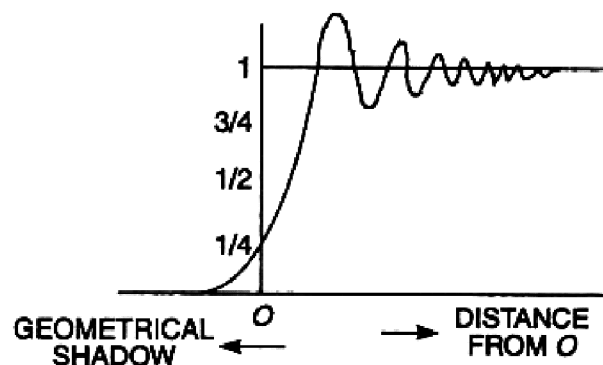


Fig. 11.20

The amplitude at $Q$ will be maximum or minimum, according as $CP$ contains odd or even number of half-period elements. The intensity at a maximum goes on decreasing and at a minimum goes on increasing and finally when we move a large distance from $O$, we get a uniform illumination. This is shown in Fig. 11.20.

The number of half-period elements contained in $CP$ depends upon the path difference $CQ - PQ$.

Suppose $SC = a$ and $CO = b$, then

$$CQ = \sqrt{b^2 + x^2}$$

$$= b\left(1 + \frac{x^2}{b^2}\right)^{\frac{1}{2}} = b + \frac{1}{2} \cdot \frac{x^2}{b}$$

Since $x$ is very small as compared to $b$, therefore, other terms are negligible.

Similarly $\qquad SQ = \sqrt{(a+b)^2 + x^2} = a + b + \dfrac{x^2}{2(a+b)}$

Hence $\qquad PQ = SQ - SP = a + b + \dfrac{x^2}{2(a+b)} - a$

$$= b + \frac{x^2}{2(a+b)}$$

$\therefore$ Path difference $\quad CQ - PQ = b + \dfrac{x^2}{2b} - b - \dfrac{x^2}{2(a+b)}$

$$= \frac{ax^2}{2b(a+b)}$$

For the displacement to be maximum

$$\frac{ax^2}{2b(a+b)} = (2n+1)\frac{\lambda}{2}$$

or, $\quad x = \sqrt{\dfrac{b(a+b)(2n+1)\lambda}{a}}\quad$ where $n = 0, 1, 2 \dots$

For the displacement to be minimum

$$\frac{ax^2}{2b(a+b)} = n\lambda$$

or, $x = \sqrt{\dfrac{b(a+b)2n\lambda}{a}}$ where $n$ = 0, 1, 2 ...

Since $x \propto \sqrt{n}$ for a dark band and $x \propto \sqrt{2n+1}$ for a bright band the fringes width goes on decreasing as we move away from O. This will be clear from the following example.

When      $n$ = 1      $x_1 = k\sqrt{1} = k$

For      $n$ = 2      $x_2 = k\sqrt{2} = 1.414k$

Width of the first bright fringe = 1.414k – k = 0.414 $k$

Similarly, width of the 2nd bright fringe $= k\sqrt{3} - k\sqrt{2}$

$$= 1.732k - 1.414k = 0.318k$$

and width of the 3rd bright fringe $= k\sqrt{4} - k\sqrt{3}$

$$= 2k - 1.732k = 0.268\,k$$

This shows that the fringe width goes on decreasing and hence the bands produced are not equally spaced.

**Determination of wavelength :** A straight edge, say a sharp razor blade is set with its edge parallel to the slit on one of the stands on the optical bench. The slit is illuminated with a monochromatic light. Diffraction bands of unequal width and decreasing intensity are observed in the field of view of the micrometer eye-piece. The position of the first maximum intensity band and the one (say nth) most distant and clearly visible band is noted, then

$$x_1 = \sqrt{\frac{b(a+b)\lambda}{a}} \qquad\qquad \because n = 0$$

For the nth maximum intensity band, we get

$$x_n = \sqrt{\frac{b(a+b)(2n+1)\lambda}{a}}$$

$$\therefore \quad x_n - x_1 = \sqrt{\frac{b(a+b)\lambda}{a}} \times \left(\sqrt{(2n+1)} - 1\right)$$

The distance $b$ from the straight edge to the eye-piece and $a$ between the slit and the straight edge are measured. Hence knowing $(x_n - x_1)$, we can find the wavelength of light.

## 11.20  Fresnel's diffraction by a slit

Let $C$ and $F$ be the edges of a rectangular aperture and $AB$ the section of a cylindrical wavefront of which the axis coincides with $S$. The edges $C$ and $F$ of the slit will cast a geometrical shadow above $M$ and below $N$ on the screen held normal to $SO$. The space between $M$ and $N$ will be illuminated.

(a) (i) **Width slit :** When the slit is wide enough so that it allows about a hundred  half-period elements from each half of the wavefront to pass through the aperture, the edges $C$ and $F$ will act like independent straight edges. This will result in producing  diffraction bands of unequal width at $M$ and $N$ inside the geometrical bright  portion, similar to  the  bands  produced  by  a  straight edge. As we  move  from  $M$  and $N$ towards $O$, the general illumination  increases  and the  visibility  of  the  bands  decreases, so that there is a uniform illumination about $O$. As we proceed beyond $M$ and $N$ the illumination rapidly falls to zero. This is shown in  the  side. (Fig.  11.21).
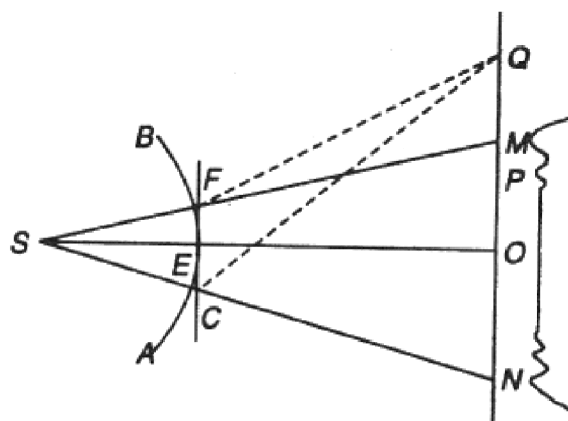


**Fig. 11.21**

(ii) **Narrow slit :** When  the  slit  is  narrow  so  that  it allows only a few half-period elements, say five (odd) to pass through each half of the wavefront with respect to $O$, the illumination at $O$ is again maximum. If the number of half-period elements contained in each half of the wavefront that passes through the slit is even, the point $O$ will have  a  minimum  intensity.

Keeping the screen fixed if we move to a point $P$ such that *four* half- period elements from the upper half of the wavefront and *six* half period elements from the lower half of the wavefront pass through the aperture, the resultant amplitude of the wavelets from the upper half of the wavefront will be $d_1 - d_2 + d_3 - d_4 = 0$ and that of the wavelets from the lower half will be $d_1 - d_2 + d_3 - d_4 + d_5 - d_6 = 0$. Hence the illumination at $P$ will be minimum.

If the point $P$ is moved a little further, so that three half-period elements from the upper half and seven from the lower half of the wavefront pass through the aperture, then the amplitude of the waves from three half-period elements will be $d_1 - d_2 + d_3$ $= d_3$ and that from the seven half-period elements will be $d_1 - d_2 + d_3 - d_4 + d_5 - d_6 + d_7 = d_7$

Hence the resultant amplitude is $d_3 + d_7$ which will produce a maximum intensity. In general—

*"As P moves up, the illumination is maximum or minimum according as odd or even number of half-period elements, from each half of the wavefront pass through the aperture and reach P".*

This shows that a set of alternately dark and bright bands parallel to the edges of the geometrical shadow will be seen inside *MN*.

For a point $Q$ within the geometrical shadow, one-half of the wavefront is entirely cut off. The intensity at $Q$ will be maximum or minimum according as $CF$ contains odd or even number of half-period elements with respect to $Q$. If the path difference $CQ - FQ = (2n + 1) \lambda/2$ i. e., *an odd multiple of* $\lambda/2$, there will be odd number of half period elements in $CF$ and hence the point $Q$ will be bright as one half-period element will be in excess. If the path difference, $CQ - FQ = 2n\lambda/2$ there will be even half-period elements in $CF$ and hence $Q$ will have minimum intensity. Thus, bands within the geometrical shadow beyond M and N are also diffraction bands of decreasing intensity and width.

(iii) **Aperture very narrow :** When the aperture comprises one or less than one-half period element with respect to $O$, it is always bright and there are no bands between $M$ and $N$. For a point $Q$ in the geometrical shadow the number of elements contained in the slit may be more than one. Hence the intensity at such a point is maximum or minimum according as the number of half-period elements contained in the aperture with respect to $O$ is odd or even.

(b) **Effect of movement of slit away from screen :** When the slit is moved gradually away from the screen, the number of half period elements in each half of the wavefront will change and hence the illumination will change. With increase in distance of screen from the slit, the number of half-period elements will decrease. As

an example, if the number decreases from four to three the displacement at $O$ will be maximum. If the slit is moved further away the number will become two with the result that the intensity will be again minimum. Thus, the point $O$ will undergo a change of maximum and minimum intensity. The same thing is observed when the distance is decreased gradually. If the distance is such that the slit comprises one or less than one half period element in each half of the wave surface with respect to $O$ on the screen, there will be a broad maximum at $O$.

## 11.21    Fresnel's diffraction by a thin wire

Let $S$ represent a narrow rectangular slit placed parallel to the wire of thickness $XY$ and perpendicular to the plane of the paper. This will give rise to a geometrical shadow MN on the screen $RT$ (see Fig. 11.22).

The effect at a point $Q$ outside the geometrical shadow is the same as that due to a straight edgeat $X$ and hence diffraction bands of unequal width and intensity will be seen above M and similarly below $N$. These bands are independent of the thickness of the wire, as on either side, the effect of the other half of the wave is negligible, because most important half-period elements are cut off due to the finite width of the wire.
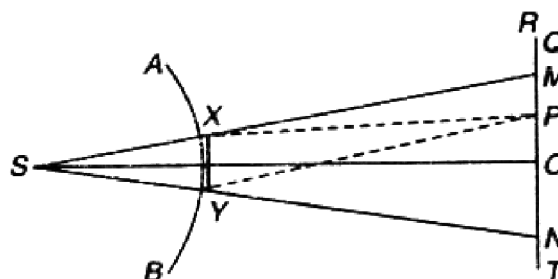


**Fig. 11.22**

Within the geometrical shadow, interference fringes are observed. The effect due to the portion $AX$ of the cylindrical wavefront, at any point $P$ within the geometrical shadow, is due entirely to a few half-period elements at $X$, so that it may be regarded as a small luminous source at $X$. Similarly, the effect due to the portion $BY$ of the wavefront is equal to the source at $Y$.

The effect at any point $P$ within the geometrical shadow $MN$ will depend upon the path difference $PY - PX$. The point $P$ will be bright if $PY - PX = n\lambda$ and dark if it

is equal to $(2n + 1)$ $\lambda/2$ . These fringes are of equal width and the fringe width

$$\beta = \frac{D}{d}\lambda$$

where $D$ is the distance of the screen from the obstacle $XY$ and $d$ the thickness of the obstacle, i.e., the diameter of the wire.
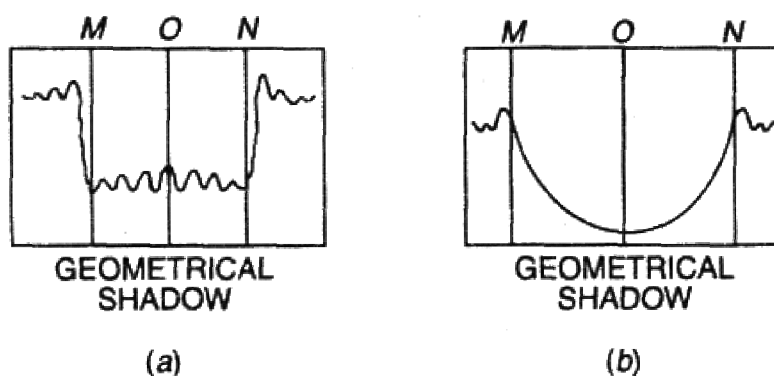


**Fig. 11.23**

The centre $O$ will be bright because at $O$ the waves from $X$ and $Y$ will always meet in phase.

The variation of intensity of the interference fringes produced within the geometrical shadow is shown in Fig. 11.23(a).

As the diameter of the wire is increased the fringe width decreases and when the thickness of the wire is sufficiently large, the interference fringes will disappear and only the diffraction bands outside the limits of the geometrical shadow will be visible as shown in Fig. 11.23(b).

**Diameter of the wire :** Focus the micrometer eye-piece on the interference fringes obtained within the geometrical shadow of the wire. The fringes are of equal width. Find the mean fringe width by measuring the shift of the cross-wire for a known number of fringes. The diameter of the wire can be found out from the expression—

$$\beta = \frac{D}{d}\lambda$$

Here $d$ is the diameter of the wire, $D$ the distance between the obstacle and cross-wires of micrometer eye-piece and $\lambda$ the wavelength of monochromatic light used.

## 11.22    Summary

■    Interference corresponds to the situation when we consider the superposition of waves coming out from a number of point sources and diffraction corresponds to the situation when we consider waves coming out from an area source like a circular or rectangular aperture or even a large number of rectangular apertures (like the diffraction grating).

■    When a plane wave is incident normally on $N$ parallel slits, the Fraunhofer diffraction pattern is given by

$$I = I_0 \frac{\sin^2 \alpha}{\alpha^2} \cdot \frac{\sin^2 N\beta}{\sin^2 \beta}$$

where $\alpha = \frac{\pi}{\lambda} a \sin\theta$, $\beta = \frac{\pi}{\lambda}(a+b)\sin\theta$

$\lambda$ is the wavelength of light, $\theta$ the angle of diffraction, a represents the width of each slit and $(a + b) = d$ the separation between two slits. When $N = 1$, we have the single slit diffraction pattern producing a central maximum at $\theta = 0$ and minima when $a \sin\theta = m\lambda$, $m = \pm 1, \pm 2, \ldots$ When $N \geq 2$, the intensity distribution is the product of the single slit diffraction pattern and the interference pattern produced by $N$ point sources separated by a distance d. For $N = 2$, we obtain the Young's double slit interference pattern. For large values of $N$, the principal maxima occur when $\beta = m\pi$ implying

$d \sin\theta = m\lambda$; $m = 0,1,2, \ldots$

which is usually referred to as the grating condition.

■    The resolving power of the grating is given by

$$R.P. = \frac{\lambda}{d\lambda} = Nm$$

where $N$ represents the total number of lines in the grating.

■    The underlying principle in the theory of diffraction is the Huygens-Fresnel principle.

■    The **Fresnel's integrals** are defined by the following equations:

$$X(v) = \int_0^v \cos\frac{\pi v^2}{2} dv \quad \text{and} \quad Y(v) = \int_0^v \sin\frac{\pi v^2}{2} dv$$

- ■ The intensity monotonically goes to zero as we go deep inside the geometrical shadow. As we move away from the edge of the geometrical shadow to the illuminated region, one obtains maxima at -1.22, -2.34, -3.08 … with gradual decrease in intensity and minima at -1.87, -2.74, -3.39 … with gradual increase in intensity.

## 11.23 Review Questions and Answers

1. **Discuss the distribution of intensity in the diffraction pattern due to a single slit.**

**Ans.** See Article 11.3.

2. **How does the diffraction pattern due to double slit change if the distance between the slit centres is varied, keeping slit width constant?**

**Ans.** In a double slit diffraction pattern, the slit width is taken as a and the separation between the slits is $b$. If the slit width is kept constant, the diffraction pattern remains the same. Keeping a constant, if the distance between the slit centres $b$ is varied, the spacing between the interference maxima changes, depending upon the relative values of $a$ and $b$. Certain orders of interference maxima will be missing in the resultant pattern.

The angular separation between two secondary maxima due to interference, is given by

$$\sin\theta_2 - \sin\theta_1 = \frac{\lambda}{a+b}$$

If $b$ increases, the angular separation decreases. Hence the spacing between the secondary interference maxima decreases.

The position of secondary minima is given by

$$a\sin\theta = \pm\, m\lambda$$

As a remains constant the diffraction pattern remains constant.

3. **What will be the effect on a diffraction pattern of a grating if the grating element is of the order of $\lambda$ or less?**

**Ans.** In a diffraction grating the principal maxima are given by the equation

$$(a+b)\,\sin\theta_m = m\lambda$$

For the first order maximum

$$\sin\theta_1 = \frac{\lambda}{a+b}$$

If $a + b = \lambda$ then $\sin\theta_1 = 1$ or $\theta = 90°$ ?

Thus, a theoretical limit is reached i.e., the first order maximum will not be visible.

If $(a + b) < \lambda$, $\sin\theta$ becomes greater than 1, which is not possible. Hence no spectrum is formed, and the grating is rendered useless.

**4.** **State and explain Rayleigh's criterion of resolution and discuss its significance in studying the spectral lines. Define resolving power and deduce an expression for it in the case of a telescope.**

**Ans.** See Articles 11.11, 11.10 and 11.13.

**5.** **Distinguish between the resolving power and the magnifying power of a telescope.**

**Ans.** The magnifying power of a telescope should be carefully distinguished from its resolving power. The following features are :

1. Magnifying power is its ability to show an enlarged view of the object to the eye: the R.P. on the other hand, is its ability to show the number of details or grains in the object.

2. Magnifying power increases with distance between the centres of diffraction pattern of two points but gives no idea of the size of central maximum; the R.P. provides ideas of both the distance between the centres of two maxima and their width.

3. The R.P. increases with the decrease of the wavelength of light. But the magnifying power is independent of the wavelength of light.

4. The magnifying power of a telescope is given by $F/f$, i.e., the ratio of the focal length of the objective to that of the eye-piece. Two telescopes of the same diameter for the objective will give different magnifying powers (as focal length can be different), but the same resolving power for a given wavelength since the resolving power is given by the ratio $D/\lambda$.

**6.** **What are the factors on which the amplitude of light waves from a half-period zone at observation point depend?**

**Ans.** (i) Resultant of secondary wavelets from the half-period zone.

(ii) On obliquity factor $(1 + \cos\theta)$, where $\theta$ is the angle which the line joining the centre of half-period zone makes with the horizontal.

**7.** **Why do we get bright light spot at the centre of the image of an opaque object when the size is extremely small?**

**Ans.** The resultant displacement dueto the whole wavefront is $d_1/2$. If the obstacle blocks first, second, third half-period zones the resultant displacement will be $-d_2/2, d_3/2, -d_4/2$. ... etc., and the intensity will be $d_2^2/4, d_3^2/4, d_4^2/4$ etc. Thus, the central point will always be bright if the obstacle is small.

## 11.24   Illustrated Examples

**Example 1 :** A diffraction pattern is observed using a double slit of wavelength 5000Å. If the slit-width is 0.02 mm and spacing between the slits is 0.10 mm, calculate (i) the distance between the central maximum and the first minimum of the fringe envelope and (ii) the distance between any two consecutive double slit dark fringes.Given : the slit/screen separation = 1 m.

**Solution :** Here $a$ = 0.02 mm = 0.02 × $10^{-3}$ m; b = 0.1 mm = 0.1 × $10^{-3}$ $m$ and $\lambda$ =5000Å = 5000 × $10^{-10}$ $m$; $D$ = 1$m$. So, a + b = 0.12 × $10^{-3}$ $m$, and let $x_1$ = distance between the central maximum and the first minimum of the fringe envelope.

(i)   Angular separation $\theta_1$ between central maximum and first minimumis given by

$$\sin\theta_1 = \theta_1 = \lambda_2 \, (a + b); \text{ also } \theta_1 = x_1/D$$

$$\therefore \quad \frac{x_1}{D} = \frac{\lambda}{2(a+b)} \Rightarrow x_1 = \frac{\lambda D}{2(a+b)} = \frac{5000 \times 10^{-10} \times 1}{2 \times 0.12 \times 10^{-3}}$$

$$= 2.08 \times 10^{-3} \, m = \textbf{2.08 mm}$$

(ii)   Angular separation between the consecutive  dark fringes is given by

$$\sin\theta_2 - \sin\theta_1 = \theta_2 - \theta_1 = \frac{3\lambda}{2(a+b)} - \frac{\lambda}{2(a+b)} = \frac{\lambda}{(a+b)}$$

If $x_2$ be the distance between the consecutive  dark fringes, then

$$\theta_2 - \theta_1 = \frac{x^2}{D} \Rightarrow x_2 = D(\theta_2 - \theta_1) = \frac{\lambda D}{a+b} = \frac{5000 \times 10^{-10} \times 1}{0.12 \times 10^{-3}}$$

$$= 4.16 \times 10^{-3} \, m = \textbf{4.16 } \textit{mm}$$

**Example 2 :** Fraunhofer double slit diffraction pattern is observed in the focal plane of a  lens of focal length 0.5 m. The wavelength of incident light is 500 nm. The distance between two maxima adjacent to zero order maximum is 5 mm and the

fourth order maximum is missing. Find the width of each slit and the distance between their centres.

**Solution :** Since the $4^{th}$ order maximum is missing, we must have

$$\frac{a+b}{a} = 4 \Rightarrow b = 3a$$

Now, for the first order maximum, $(a + b) \sin \theta_1 = \lambda$

$$\therefore \quad \theta_1 \approx \frac{\lambda}{a+b} = \frac{\lambda}{4a}$$

$\therefore$ Distance between two maxima adjacent to zero order maximum is

$$x = f \times 2\theta_1 = f \times \frac{2\lambda}{4a}$$

$$\therefore \quad a = \frac{f\lambda}{2x} = \frac{0.5 \times 500 \times 10^{-9}}{2 \times 5 \times 10^{-3}} m = \textbf{0.025 } \textbf{\textit{mm}}$$

$$\therefore \quad b = 3a = 3 \times 0.025 = \textbf{0.075 } \textbf{\textit{mm}}$$

**Example 3 :** How many orders would be visible if the wavelength of incident light is 589 nm and the number of lines in the grating is 100/mm?

**Solution :** Here $\lambda = 589 \times 10^{-9}$ $m$; $p = 1/a + b = 100 \times 10^3/m$;

$$\theta = 90° \text{ (maximum)} \Rightarrow \sin \theta = 1$$

From $\qquad (a + b) \sin \theta = m\lambda$

$$m = \frac{a+b}{\lambda} \sin \theta = \frac{1}{100 \times 10^3 \times 589 \times 10^{-9}} = 16.32$$

Thus, **16** orders would be visible.

**Example 4 :** Consider a grating of width 5 cm with slits of width $a = 0.001$ mm separated by a distance 0.002 mm? How many orders would be visible at $\lambda = 550$ nm? Find the width of the principal maximum. Is there any missing order?

**Solution :** Here $a = 0.001$ $mm$, $b = 0.002$ $mm$, $\lambda = 550 \times 10^{-9}$ $m = 550 \times 10^{-6}$ $mm$

From $(a + b) \sin \theta = m\lambda$, the maximum number of orders visible, is

$$m = \frac{(a+b)\sin\theta_{max}}{\lambda} = \frac{(0.001+0.002) \times \sin 90°}{550 \times 10^{-6}} = \textbf{5.45}$$

So, the **maximum** number of **orders** visible = **5**

Angular width of principal maximum is

$$2d\theta = \frac{2}{Nm\cot\theta} = \frac{2\lambda}{N(a+b)\cos\theta} = \frac{2\times550\times10^{-9}}{5\times10^{-2}}$$

$$= \textbf{2.2} \times \textbf{10}^{-5} \textbf{ rad}$$

Since $N(a + b) = $ width of grating $= 5 \times 10^{-2}$ $m$ and $\cos\theta = 1$ for principal maximum.

We have : $\dfrac{a+b}{a} = \dfrac{0.001+0.002}{0.001} = \dfrac{3}{1}$

Thus, **3rd**, **6th**, **9th** etc. order maxima will be missing.

**Example 5 :** Calculate the least width a grating must have to resolve the D-lines sodium (589.0 nm and 589.6 nm) in the second order. The number of lines per mm of the grating is 80.

**Solution :** Here $\lambda = \dfrac{589.0+589.6}{2} nm = 589.3 \ nm = 589.3 \times 10^{-9} \ m$

$m = 2$ and $d\lambda = 589.6 - 589.0 = 0.6$ nm $= 0.6 \times 10^{-9} m.$

R.P. of the grating is

$$\frac{\lambda}{d\lambda} = Nm \Rightarrow N = \frac{589.3\times10^{-9}}{2\times0.6\times10^{-9}} = 491.1 \text{ lines}$$

$\therefore$ Least number of lines required for resolution $= 492$

$\therefore$ Least width of the grating $= \dfrac{492}{80} = \textbf{6.15} \textbf{\textit{mm}}$

**Example 6 :** Calculate the inner and outer radii of the 10th half-period zone for a plane wavefront with respect to a point at a distance $0.5 \ m$ from it. Assume the wavelength of light $\lambda = 500$ nm.

**Solution :** For the $n$th half-period zone the outer radius is,

$$r_n = \sqrt{\left(b+\frac{n\lambda}{2}\right)^2 - b^2} \approx \sqrt{nb\lambda}$$

And the inner radius is,

$$r_{n-1} = \sqrt{\left(b+\frac{\overline{n-1}\lambda}{2}\right)^2 - b^2} \approx \sqrt{(n-1)b\lambda}$$

Here $n = 10$, $b = 0.5\ m$, $\lambda = 500 \times 10^{-9}\ m$

$$\therefore\quad r_{10} = 1.58\ mm,\ r_9 = 1.5\ mm$$

**Example 7 :** The primary focal length of a zone plate is 20 cm for light of wavelength 500 nm. Calculate the radius of the central zone on the zone plate.

**Solution :** We know that

$$f_n = \frac{r_n^2}{n\lambda} \Rightarrow f_1 = \frac{r_1^2}{\lambda}$$

or, $\quad r_1 = \sqrt{f_1\lambda} = \sqrt{0.2 \times 500 \times 10^{-9}}\, m = 0.316\ mm.$

**Example 8 :** Consider a plane wave of wavelength 500 nm incident on a circular aperture of fixed radius 0.5 mm. Calculate the position of the brightest and the darkest points on the axis.

**Solution :** If the distance $b$ of the axial point is such that the circular aperture exposes only the 1st half-period zone, then the point will be the brightest point. Thus

$$b^2 + r^2 = \left(b + \frac{\lambda}{2}\right)^2$$

or, $r^2 = b\lambda$;  or, $b = \dfrac{r^2}{\lambda} = \dfrac{\left(0.5 \times 10^{-3}\right)^2}{500 \times 10^{-9}}\, m = 0.5\ m$

For minimum intensity at an axial point the aperture must expose two half-period zones,

$$b^2 + r^2 = \left(b + \frac{2\lambda}{2}\right)^2 = b^2 + 2b\lambda$$

or, $\quad b = \dfrac{r^2}{\lambda} = \dfrac{0.5}{2}\, m = 0.25\ m$

# Unit-12 : Laser and Holography

**Structure**

## 12.1 Objectives

In going through this unit, you will learn:

1. Spontaneous and stimulated emission
2. Basic components of a laser
3. Population inversion
4. Meta-stable states
5. Optical resonator
6. Einstein A, B coefficients
7. Three level and four level LASERs
8. Ruby Laser
9. He-Neon Laser
10. Holography

## 12.2  Introduction

LASER is an acronym for Light Amplification by Stimulated Emission of Radiation. The light emitted from a laser often possesses some very special characteristics - some of these are :

(a) **Directionality :** The divergence of the laser beam is usually limited by diffraction and the actual divergence can be less than 10-5 radians; this leads to the application of the laser in surveying, remote sensing, lidar, etc.

(b) **High Power :** Continuous wave lasers having power levels ~ 105 W and pulsed lasers having a total energy ~ 50,000 J can have applications in welding, cutting, laser fusion, etc.

(c) **Tight Focusing :** Because of highly directional properties of the laser beams, they can be focused to areas ~ few $(\mu m)^2$ - this leads to applications in surgery, material processing, compact discs, etc.

(d) **Spectral Purity :** Laser beams can have an extremely small spectral width $\Delta\lambda \sim 10^{-6}$Å. Because of high spectral purity, lasers find applications in holography, optical communications, spectroscopy etc.

Because of such unique properties of the laser beam, it finds important applications in many diverse areas and indeed one can say that after the discovery of the laser, optics has become an extremely important field of study. Indeed, because a laser beam can be focused to very narrow areas, it has found applications in areas like eye surgery, laser cutting, etc.

The basic principle involved in the lasing action is the phenomenon of stimulated emission, which was predicted by Einstein in 1917. In this introductory section we will discuss this first, which will be followed by brief discussions of the main components of a laser and the underlying principle as to how the laser works. In Art. 12.7 we will have a slightly more detailed account of optical resonators. In Art. 12.8 we will discuss Einstein coefficients and optical amplification. In Art. 12.9 we will discuss the working of the ruby laser, which was the first laser to be fabricated. In Art 12.10 we will discuss the working of the helium-neon laser.

## 12.3 Spontaneous and Stimulated emissions

Atoms are characterized by discrete energy states. According to Einstein, there are three different ways in which an atom can interact with electromagnetic radiation :

(a) **Spontaneous emission :** Atoms in the energy state $E_2$ can make a (spontaneous) transition to the energy state $E_1$ with the emission of radiation of frequency

$$\omega = \frac{E_2 - E_1}{\hbar} \text{ where } \hbar = h/2\pi = 1.0546 \times 10^{-34} \text{ Js} \qquad \dots (12.1)$$

and $h$ ( $\approx$ $6.626 \times 10^{-34}$ $Js$) is known as the Planck's constant. Since this process can occur even in the absence of any radiation, this is called spontaneous emission [see Fig. 12.1 (a)]. The rate of spontaneous emission is proportional to the number of atoms in the excited state. In this case the photons are created at random. So, it creates waves of random phase, and the light is said to be incoherent.

(b) **Stimulated emission :** As put forward by Einstein, when an atom is in the excited state, it can also make a transition to a lower energy state through what is known as stimulated emission in which an incident signal of appropriate frequency triggers an atom in an excited state to emit radiation - this results in the amplification of the incident beam [see Fig. 12.1(b)]. The rate of stimulated emission depends both on the intensity of the external field and also on the number of atoms in the excited state. In this case, the light waves associated with the incident and emitted photons are in phase. Hence, they are said to be coherent.

(c) **Stimulated absorption :** Stimulated absorption (or simply absorption) is the process in which the electromagnetic radiation of an appropriate frequency (corresponding to the energy difference of the two atomic levels) can pump the atom to its excited state [see Fig. 12.1(c)]. The rate of stimulated absorption depends both on the intensity of the external field and also on the number of atoms in the lower energy state.
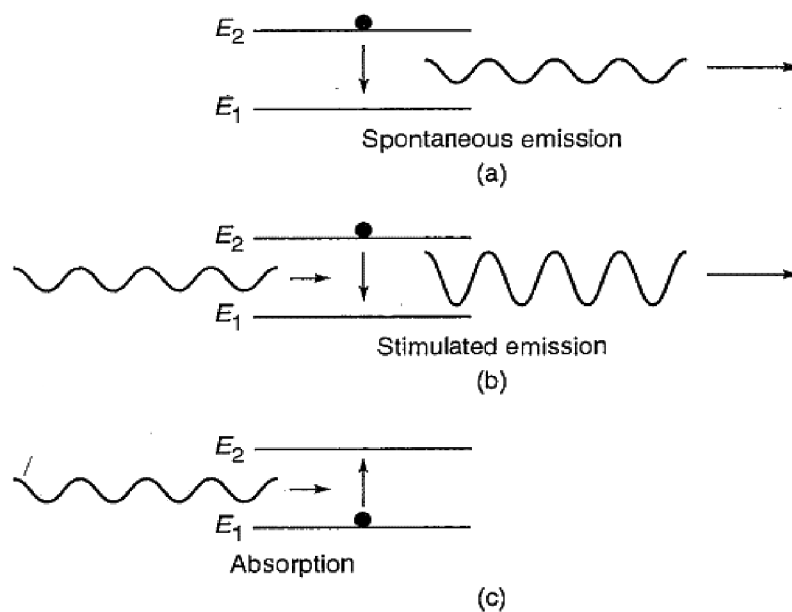


Fig. 12.1

When the atoms are in thermodynamic equilibrium, there are larger number of atoms in the lower state implying that the number of absorptions exceeds the number of stimulated emissions; this results in the attenuation of the beam [see Fig. 12.2(a)]. On the other hand, if we are able to create a state of population inversion in which there are larger number of atoms in the upper state then the number of stimulated emissions would exceed the number of absorptions resulting in the (optical) amplification of the beam [see Fig. 12.2(b)]. The amplification process due to stimulated transitions is phase coherent.
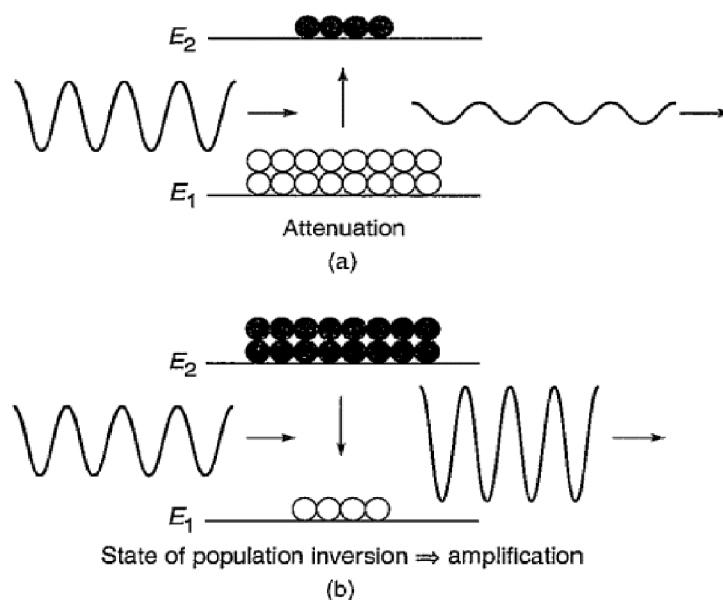


**Fig. 12.2**

## 12.4 Basic components of a laser

The three main components of any laser are (see Fig. 12.3) :

1. **Active medium :** The active medium consists of a collection of atoms, molecules, or ions (in solid, liquid, or gaseous form), which is capable of amplifying light waves. Under normal circumstances, there are always a larger number of atoms in the lower energy state than in the excited energy state. An electromagnetic wave passing through such a collection of atoms would get attenuated; this is discussed in detail in Article 12.7. In order to have optical amplification, the medium has to be kept in a state of population inversion, i.e., in a state in which the number of atoms in the upper energy

level is greater than that in the lower energy level - this is achieved by means of the pump.
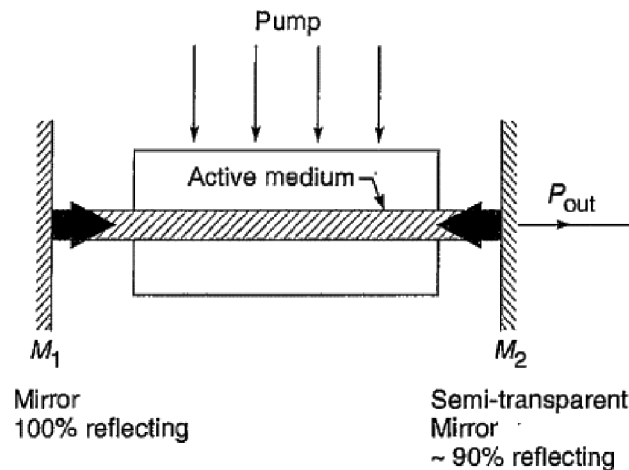


Fig. 12.3

2. **Pumping source** : The pumping mechanism provides for obtaining such a state of population inversion between a pair of energy levels of the atomic system and when we have a state of population inversion, the input light beam can get amplified by stimulated emission (see Fig. 12.4).

3. **Optical resonator** : A medium with population inversion is capable of amplification; however, in order that it acts as an oscillator, a part of the output energy must be fed back into the system. Such feedback isbrought about by placing the active medium in a resonator; the resonator could just be a pair of mirrors facing each other.
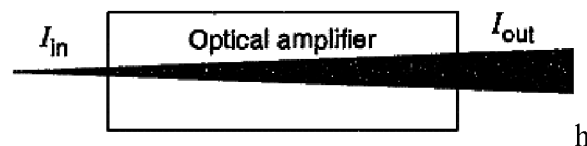


Fig. 12.4

### 12.4.1 Resonator

As mentioned earlier, a medium with population inversion is capable of amplification but in order that it act as an oscillator, a part of the output energy must be fed back

into the system. Such feedback is brought about by placing the active medium between a pair of mirrors facing each other (see Fig. 12.3). Such a system formed by a pair of mirrors is referred to as a resonator, a slightly more detailed account of which will be given in Art. 12.7. The sides of the cavity are usually open and hence such resonators are also referred to as open resonators. A resonator is characterized by various modes of oscillation with different field distributions and frequencies. One can visualize a mode as a wave having a well-defined transverse amplitude distribution which forms a standing wave pattern. The transverse intensity distribution of the fundamental mode is usually a Gaussian. Because of the open nature of the resonator, all modes have a finite loss due to the diffraction spill over of energy at the mirrors. In addition to this basic loss, scattering from the laser medium, absorption at the mirrors and output coupling at the mirrors also contribute to the cavity loss. In an actual laser, the modes that keep oscillating are those for which the gain provided by the laser medium compensates for the losses. When the laser oscillates in steady state, the losses are exactly compensated for by the gain. Since the gain provided by the medium depends on the extent of population inversion, for each mode there is a critical value of population inversion (known as the threshold population inversion) below which that particular mode would cease to oscillate in the laser (see Article 12.8 below).

### 12.4.2 Lasing action

The onset of oscillations in a laser cavity can be understood as follows: Through a pumping mechanism, one creates a state of population inversion in the laser medium placed inside the resonator system. Thus, the medium is prepared to be in a state in which it is capable of coherent amplification over a specified band of frequencies. The spontaneous emission occurring inside the resonator cavity excites the various modes of the cavity. For a given population inversion, each mode is characterized by a certain amplification coefficient due to the gain and a certain attenuation coefficient due to the losses in the cavity. The modes for which the losses in the cavity exceed the gain die out. On the other hand, the modes whose gain is higher than the losses get amplified by drawing energy from the laser medium. The amplitude of the mode increases rapidly until the upper level population reaches a value when the gain equals the losses, and the mode oscillates in steady state. When the laser oscillates in the steady state, the losses are exactly compensated for by the gain provided by the medium, and the wave coming out of the laser can be represented as a continuous wave.

## 12.5  Population inversion

Let us consider an assembly of independent atoms which can exist only in two levels 1 and 2 with energies $E_1$, and $E_2$. Let $N_1$, and $N_2$, be the number of atoms per unit volume in state 1 and 2, respectively (see Fig. 12.5). These numbers are called the population of the respective levels. According to Boltzmann distribution, under thermal equilibrium we can write

$$\frac{N_2}{N_1} = e^{-\left(E_2-E_1\right)/KT} \qquad \qquad ... (12.2)$$

where $K$ is the Boltzmann's constant, and $T$ is the absolute temperature and $E_2 > E_1$. Thus, the population of higher energy state is smaller than that of the lower energy state. For a transition to give visible light $E_2 - E_1 \sim 1.25$ $eV$. Now $KT\sim0.025$ $eV$ at room temperature. A substitution of these values in (12.2) shows that $N_2$ is almost negligible compared with $N_1$. Thus, under thermal equilibrium at room temperature if a photon of energy $hv = E_2 - E_1$ approaches the system it is very likely to be absorbed rather to cause a stimulated emission. For stimulated emission to dominate we must by some means make the population of the upper state greater than that of the lower state. The system is then said to possess population inversion. The method by which such a population inversion is achieved is called pumping.

Equation (12.2) indicates that $N_2$, becomes just equal $N_1$, when temperatureis made infinitely high $(T = \infty)$.To make $N_2 > N_1$ i.e., to get population inversion the temperature $T$ in (12.2) must be made negative. For this reason, a state of population inversion is sometimes misleadingly referred to as a negative temperature. In fact, population inversion can he achieved at ordinary temperature but only under nonequilibrium situation where Boltzmann's law is not applicable.

Energy  2 ———population $N_2$——— $E_2$

1 ———————$E_1$
population $N_1$
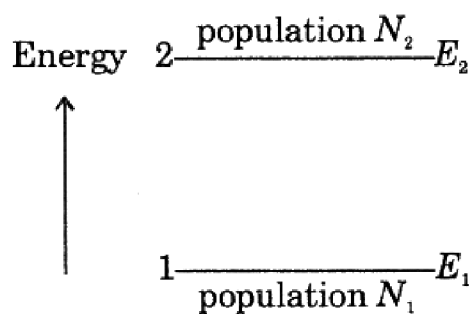
**Fig. 12.5**

**Mechanisms of population inversion :**

The mechanism of obtaining population inversion in a system is known as pumping. There are diverse techniques of achieving population inversion in a laser system. We discuss below some commonly used methods.

1. **Optical pumping :** Here the atoms are exposed to strong source of light such as flash lamp or arc lamp. By selective absorption of radiation, the atoms are excited from the lower state to the higher state. This method is known as optical pumping. It is used in ruby laser.

2. **Collision of first kind :** Here free electrons created by gas discharge are accelerated to high velocities by employing a strong external electric field. These electrons when make inelastic collision with gas atoms a part of the kinetic energy of the electrons is transferred to the atoms and the atoms get excited. The process may be represented as

$$A + e_{fast} \rightarrow A^* + e_{slow}$$
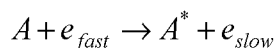
where $A$ and $A^*$ denote the ground state and excited state of the atom $A$ respectively. This method is employed in argon laser.

3. **Collision of second kind :** This method involves the use of a gas mixture in which there occurs in addition to excitation by electron impact, a resonant transfer of excitation between colliding atoms of two kinds. Consider a gas mixture of $A$ and $B$ atoms kept under electrical discharge. A may be excited to its **metastable state** $A^*$ (having a comparatively long lifetime than the normal mean lifetime $\sim 10^{-8}$ s of an excited atom) by collision with fast electrons and next it interacts with $B$, resulting $B$ in the excited state and $A$ itself goes to the ground state. Symbolically, the process may be represented as

$$A + e_{fast} \rightarrow A^* \left(metastable\right) + e_{slow}$$
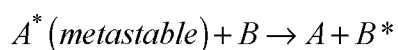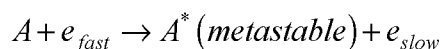
$$A^* \left(metastable\right) + B \rightarrow A + B^*$$

Here the A atom must have at least one metastable state that corresponds to the excited energy state of B. This method of pumping is used in the He-Ne laser.

4. **Chemical pumping :** It corresponds to an excitation mechanism through specially selected exothermic reaction. For example, the reaction $H_2 + F_2 \rightarrow 2HF$ generates enough energy to pump a $CO_2$ laser.

5. **Thermal pumping :** It corresponds to an excitation mechanism through heating of the active medium.

6. **Molecular dissociation :** When an excited atom collides with a molecule a dissociation of the molecule may take place and leave one of the dissociated fragments in an excited state which may then serve as a starting level of laser transition. For example, $Ne$ in an excited state dissociates $O_2$ molecule in this way :

$$Ne^* + O_2 \rightarrow Ne + O^* + O$$
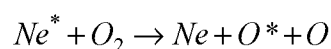
## 12.6 Optical resonator

In Article 12.5 we had briefly discussed that a light beam passing through a suitable medium with population inversion may be amplified. In order to construct an oscillator, which can supply light energy and act as a source of light, one must couple a part of the output back into the medium. This can be achieved by placing the active medium between two mirrors which reflect most of the output energy back to the system, see Fig. 12.3. Such a system of two mirrors represents a resonant cavity.

Now, in order to obtain an output beam, one of the mirrors is made partially reflecting. Thus, imagine a wave that starts from one of the mirrors and travels towards the other. In .passing through the active medium, it gets amplified. If the second mirror is partially reflecting, then the wave is partially transmitted, and the rest reflected back towards the first mirror. In traveling to the first mirror, it again gets amplified and returns to the position it has started from. Thus, in between the two mirrors, we have waves propagating along both directions. For resonance, it is necessary that when a wave returns after one round trip it is in phase with the existing wave. For this to happen, the total phase change suffered by the wave in one complete round trip must be an integral multiple of so that standing waves are formed in the cavity. Thus, if d represents the length of the cavity, then we may write

$$\frac{2\pi}{\lambda} 2d = 2m\pi \; ; \; m = 1, 2, 3, \dots \qquad \dots (12.3)$$

where $\lambda$ is the wavelength of the radiation in the medium enclosed by the cavity. If $n_0$ represents the refractive index of the medium enclosed by the cavity then

$$\lambda = \frac{\lambda_0}{n_0} \qquad \dots (12.4)$$

If we put $\lambda_0 = c/v$, (12.3) becomes

$$v = v_m = m\frac{c}{2n_0 d} \qquad \ldots (12.5)$$

which gives the discreate frequencies of oscillation of the modes. If we assume $n_0 = 1$ (like in a He-Ne laser), (12.5) simplifies to

$$v = v_m = m\frac{c}{2d} \qquad \ldots (12.6)$$

Different values of $m$ lead to different oscillation frequencies, which constitute the **longitudinal modes** of the cavity; for further details and for reasons why they are known as longitudinal modes the reader is referred to any textbook on lasers. The frequency difference between adjacent longitudinal modes is given by

$$\delta v = \frac{c}{2d} \qquad \ldots (12.7)$$

Returning to (12.5), we would like to mention that for a practical optical resonator, **$m$ is a very large number**. For example, for an optical resonator of length $d \approx 60$ cm operating at an optical frequency of $v \approx 5 \times 10^{14}$ Hz (corresponding to $\lambda \approx 6000$ Å), we obtain

$$m = \frac{5 \times 10^{14} \times 2 \times 60}{3 \times 10^{10}} = 2 \times 10^6$$

Equation (12.5) tells us that the cavity will support only those frequencies for which the round-trip phase shift is an integral multiple of $2\pi$. We may mention here that an **open resonator** consisting of two plane mirrors facing each other is nothing but the **Fabry-Perot interferometer** discussed in Unit 10 the main difference is that in a Fabry-Perot interferometer, the spacing between the mirrors is small compared to the transverse dimension of the mirrors while in an optical resonator, the converse is true.

**Resonator Configuration :**

An optical resonator as used in a laser can have different configurations. The simplest form of an optical resonator as used in a laser device is a pair of two plane parallel mirrors facing each other in between which is placed the active medium. These mirrors require very precise alignment. Such configuration is rarely used

nowadays. As an alternative configuration two identical concave mirrors separated by a distance equal to their radius are used. Diffraction losses are low in such a confocal configuration. Also, alignment problem is much less critical here.

A concentric or spherical resonator uses two concave mirrors of same radius separated by a distance equal to twice their radius. A hemispherical cavity uses a concave spherical mirror at one side and a plane mirror placed at the centre of curvature of the spherical mirror. Such a resonator is easier to align but power output is relatively low. A 'long-radius' cavity with two concave mirrors of equal and fairly long radius of curvature is found to be most common in commercial lasers.

**Note :**

1) The resonant modes of the cavity given by (12.5) are narrower in frequency than the bandwidth of single spontaneous emission line. So, several resonant cavity modes may lie within the width of a single emission line. Out of relatively broad range of frequencies available from an emission line the cavity selects and amplify only certain narrow bands. By adjusting the cavity modes separation, it is possible to generate only one cavity mode within the bandwidth of an emission line. This is the origin of extreme quasi-monochromaticity of laser light.

Broad emission line and narrow cavity modes
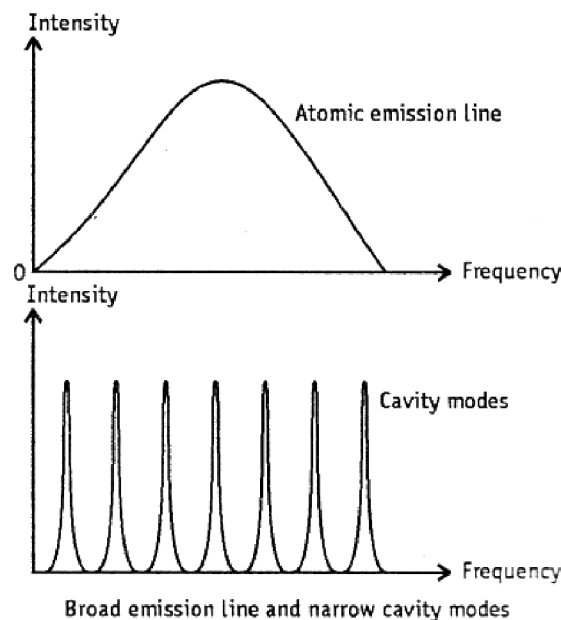
**Fig. 12.6**

2) In addition to longitudinal modes of oscillation there are a few **transverse modes**. Since the fields are nearly normal to the axis of cavity they are known as TEM (transverse electric and magnetic) mode. Lowest order transverse mode i.e., $TEM_{00}$ is the most widely used mode.

3) Laser cavities may be classified as **stable** and **unstable**. In *stable* type of resonator, the beam remains close to the axis and tends to bounce back and forth in between the mirrors without much loss. On the other hand, in an *unstable* resonator the beam goes away from the axis on each reflection and quickly leaves the resonator altogether.

4) The resonant mode of a resonator is not a sharp line but has a small but finite frequency spread. This spread is related to various losses (*scattering, absorption, diffraction*, etc.) associated with the modes in the cavity. A measure of these losses is often expressed in terms of the *quality factor* Q of the cavity, which is defined as

$$Q = 2\pi = \frac{\text{Maximum energy stored per cycle in the mode}}{\text{Energy dissipated per cycle in the mode}}$$

It can also be expressed as

$$Q = \frac{\omega}{\Delta\omega}$$

where $\Delta\omega$ is the linewidth. Thus, a high $Q$ cavity has low loss and narrow width of a mode.

5) One can visualize a mode as a wave having a well-defined transverse amplitude distribution which forms a standing wave pattern. In an actual laser, the modes that keep oscillating are those for which the gain provided by the laser medium compensates for the losses. When the laser oscillates in steady state, the losses are exactly compensated for by the gain. Since the gain provided by the medium depends on the extent of population inversion, for each mode there is a *critical value of population inversion (known as the threshold population inversion*) below which that particular mode would cease to oscillate in the laser.

6) **Threshold Condition for laser action :** It can be shown that the threshold population inversion for oscillation at the frequency $v$ is given by

$$\left(N_2 - N_1\right)_{th} = \frac{8\pi t_{sp} v_0^3}{c^3 g\left(v_0\right) t_c} \qquad \dots (12.8)$$

where $v_0$ is the central frequency, $t_c$ is the passive cavity lifetime and $t_{sp}$ is the spontaneous emission lifetime of upper state. In the above derivation we have assumed that the atom is capable of interacting with a radiation of a particular frequency.

Actually, the emission lines have some finite breadth caused by various factors such as *Doppler broadening, pressure broadening, natural breadth, Stark effect broadening* etc. As a result, an atom can interact with radiations over a range of frequencies. The strength of interaction is a function of frequency known as *line shape function* and is denoted by $g(v_0)$.

In order to get smallest possible threshold population inversion, we must have the following requirements :

(i) $t_c$ be as long as possible i.e., losses be minimum.

(ii) $t_{sp}$ out of the upper level be very short.

(iii) As $g(v_0) \propto 1/\Delta v$, the atomic linewidth be kept as small as possible. Temperature should be kept low to reduce $\Delta v$.

(iv) Lower value of $v_0$ is prospective for laser action.

(v) Laser material should be so selected that $N_1$ can be kept small.

## 12.7 Einstein's A, B coefficients

The consideration which led Einstein to the prediction of stimulated emission was the description of thermodynamic equilibrium between atoms and the radiation field. Consider an atom having two states. Let $N_1$ and $N_2$ be the number of atoms (per unit volume) in states 1 and 2 respectively; the levels correspond to energies $E_1$ and $E_2$ (see Fig. 12.5). As mentioned earlier, an atom in the lower energy level can absorb radiation and get excited to the level $E_2$. This excitation process can occur only in the presence of radiation. The rate of absorption depends on the density of radiation at the particular frequency corresponding to the energy separation of the two levels. Thus, if

$$E_2 - E_1 = \hbar\omega \qquad\qquad ... (12.9)$$

then the absorption process depends on the energy density of radiation at the frequency $\omega$; this energy density is denoted by $u(\omega)$ and is defined such that

$$u(\omega)d\omega = \text{Radiation energy per unit volume within the frequency}$$
$$\text{interval } \omega \text{ and } \omega + d\omega.$$

The rate of absorption is proportional to $N_1$ and also to $u(\omega)$. Thus, we may write

Number of absorptions per unit volume per unit time = $N_1 B_{12} u(\omega)$     ... (12.10)

where $B_{12}$ is the coefficient of proportionality and is a characteristic of the energy levels.

Let us now consider the reverse process namely the emission of radiation at a frequency $\omega$ when the atom de-excites from the level $E_2$ to $E_1$. As mentioned in Article 12.4, an atom in an excited level can make a radiative transition to a lower energy level either through spontaneous emission or through stimulated emission. In spontaneous emission, the probability per unit time of the atom making a downward transition is independent of the energy density of the radiation field and depends only on the levels involved in the transition. The rate of spontaneous transitions (per unit volume) from level $E_2$ to $E_1$ is proportional to $N_2$ and thus

$$\frac{dN_2}{dt} = -A_{21}N_2 = -\frac{N_2}{t_{sp}} \qquad \qquad \text{... (12.11)}$$

where $A_{21}$ represents the coefficient of proportionality and is known as the Einstein $A$ coefficient and depends on the energy level pair and

$$t_{sp} = \frac{1}{A_{21}} \qquad \qquad \text{... (12.12)}$$

represents the spontaneous emission lifetime of the upper level. The solution of (12.11) is given by

$$N_2(t) = N_2(0)e^{-\frac{t}{t_{sp}}} \qquad \qquad \text{... (12.13)}$$

implying that the population of level 2 reduces by 1/e in a time $t_{sp}$. For example, for the $2P \to 1S$ transition in hydrogen atom $A \approx 6 \times 10^8 S^{-1}$ giving a mean lifetime ($\approx 1/A$) of about $1.6 \times 10^{-9}$ s. In the case of stimulated emission, the rate of transition to the lower energy level is directly proportional to the number of atoms in the upper energy level as well as to the energy density of the radiation at the frequency $\omega$. Thus

Number of stimulated emissions (per unit volume per unit time) = $N_2 B_{21} u(\omega)$

with $B_{21}$ representing the corresponding proportionality constant. The quantities $A_{21}$, $B_{12}$ and $B_{21}$ are known as Einstein coefficients and are determined by the atomic system. At thermal equilibrium, the number of upward transitions must be equal to the number of downward transitions. Thus, we may write (at thermal equilibrium):

$$N_1 B_{12} u(\omega) = N_2 A_{21} + N_2 B_{21} u(\omega)$$

$$\text{or} \quad u(\omega) = \frac{A_{21}}{\dfrac{N_1}{N_2}B_{12} - B_{21}} \qquad \qquad \text{... (12.14)}$$

Now according to a fundamental principle in thermodynamics, at thermal equilibrium, we have the following expression for the ratio of the populations of two levels :

$$\frac{N_1}{N_2} = \exp\left[\frac{E_2 - E_1}{k_B T}\right] = \exp\left[\frac{\hbar \omega}{k_B T}\right] \qquad \text{... (12.15)}$$

where $k_B$ ( $= 1.38 \times 10^{-23}$ J/K) represents the Boltzmann constant and $T$ represents the absolute temperature and *we assume that degeneracy of each level is* **unity**. Equation (12.13) is known as the Boltzmann's law. Thus, we may write

$$u(\omega) = \frac{A_{21}}{B_{21} e^{\hbar\omega/(k_B T)} - B_{21}} \qquad \text{... (12.16)}$$

Now, at thermal equilibrium, the radiation energy density is given by Planck's law :

$$u(\omega) = \frac{\hbar \omega^3 n_0^3}{\pi^2 c^3} \frac{1}{e^{\hbar\omega/(k_B T)} - 1} \qquad \text{... (12.17)}$$

where $n_0$ represents the refractive index of the medium. Comparing (12.16) and (12.17), we obtain

$$B_{12} = B_{21} = B \text{ (say)} \qquad \text{(12.18)}$$

and 
$$\frac{A_{21}}{B_{21}} = \frac{\hbar \omega^3 n_0^3}{\pi^2 c^3} \qquad \text{... (12.19)}$$

These relations are known as **Einstein's relations**.

Notice that if we had not assumed the presence of stimulated emission, we would not have been able to arrive at an expression for $u(\omega)$; Einstein in 1917 had predicted the existence of stimulated emission which was later confirmed by rigorous quantum theory.

From (12.17) and (12.19), the ratio of the number of spontaneous to stimulated emissions under thermal equilibrium is given by

$$\frac{A_{21}}{B_{21} u(\omega)} = e^{\hbar\omega/(k_B T)} - 1 \qquad \text{... (12.20)}$$

We may note the following two important points :

1.  For normal optical source, T ~ $10^3$K with $\omega \approx 3 \times 10^{15} sec^{-1}$ (corresponding to $\lambda \approx 6000$Å) we have

$$\frac{\hbar \omega}{K_B T} = \frac{1.054 \times 10^{-34} \times 3 \times 10^{15}}{1.38 \times 10^{-23} \times 10^3} = 23$$

giving $\dfrac{A_{21}}{B_{21}u(\omega)} = 10^{10}$

Thus, *when the atoms are in thermal equilibrium, the emission (at optical frequencies) is predominantly due to spontaneous transitions and hence the emission from ordinary light sources is incoherent.*

2. From (12.19), one can see that the coefficient $B_{21}$ is inversely proportional to $\omega^3$ implying that laser action would become more difficult as we go to higher frequencies.

## 12.8 Ruby laser (three level laser)

In the first laser fabricated by Maiman in 1960, the population inversion was achieved in the following manner. It was made from a single cylindrical crystal of ruby whose ends were flat, with one of the ends completely silvered and the other partially silvered (see Fig.12.7). Ruby consists of $Al_2O_3$ as a host material doped with a small percentage (about 0.05% by weight) of $Cr_2O_3$ which acts as active material. The energy states of the chromium ion are shown in Fig. 12.8. The chief characteristic of the energy levels of a chromium ion is the fact that the bands labeled $E_1$ and $E_2$ have a lifetime of ~$10^{-8}$ sec whereas the state marked $M$ has a lifetime of ~$3 \times 10^{-3}$ sec— the lifetime represents the average time an atom spends in an excited state before making a transition to a lower energy state. A state characterized by such a long lifetime is termed a *metastable state.*
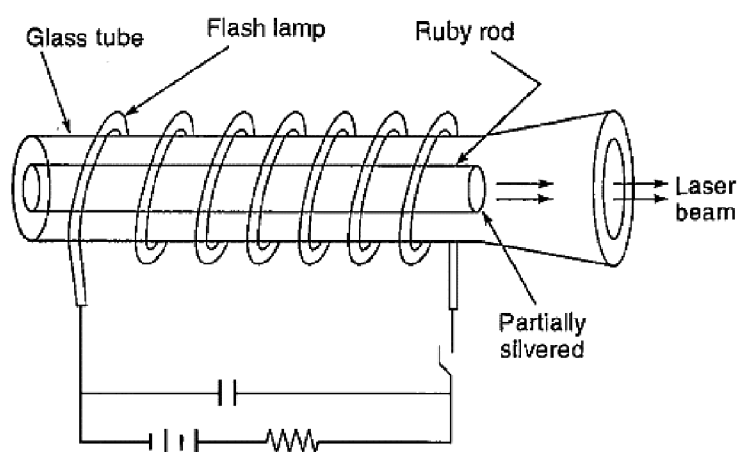


**Fig. 12.7**

The chromium ion in its ground state can absorb a photon (whose wavelength is around 6600 Å) and make a transition to one of the states in the band $E_1$; it could also absorb a photon of $\lambda \sim 4000$ Å and make a transition to one of the states in the band $E_2$—this is known as optical pumping and the photons which are absorbed by the chromium ions are produced by the flash lamp (see Fig. 12.7). In either case, it immediately makes a non-radiative transition (in a time $\sim 10^{-8}$ sec) to the metastable state $M$ – in a non-radiative transition, the excess energy is absorbed by the lattice and does not appear in the form of electromagnetic radiation. Also, since the state $M$ has a very long life, the number of atoms in this state keeps increasing and one may achieve population inversion between states $M$ and $G$. Thus, we may have a larger number of atoms in states $M$ and $G$. Once population inversion is achieved, light amplification can take place, with two reflecting ends of the ruby rod forming a cavity. The *ruby laser is an example of a three-level laser.*
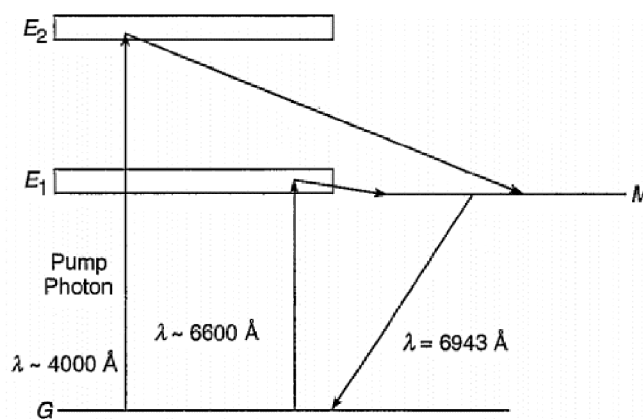


**Fig. 12.8**

In the original setup of Maiman, the flash lamp (filled with xenon gas) was connected to a capacitor (see Fig. 12.7) which was charged to a few kilovolts. The energy stored in the capacitor ($\sim$a few thousand joules) was discharged through the xenon lamp in a few milliseconds. This results in a power which is $\sim$ a few megawatts. Some of this energy is absorbed by the chromium ions resulting in their excitation and subsequent lasing action.

The state of inverted population is, however, not a stable one. One or two photons emitted spontaneously will induce stimulated emission and coherent laser beam is obtained through the partially silvered mirror. *Ruby laser is a pulsed laser* i.e., it emits laser light in pulses. The metastable state $M$ is in reality a doublet of separation 14Å. The output of the laser has two lines of wavelengths 6943Å and 6929Å. However, under lasing conditions the line 6943Å predominates.

### 12.8.1 Spiking in Ruby laser

The flash operation of the lamp leads to a pulsed output of the laser. Even in the short period of a few tens of microseconds in which the ruby is lasing, one finds that the emission is made up of spikes of high intensity emissions as shown in Fig. 12.9. This phenomenon is known as spiking and can be understood as follows.
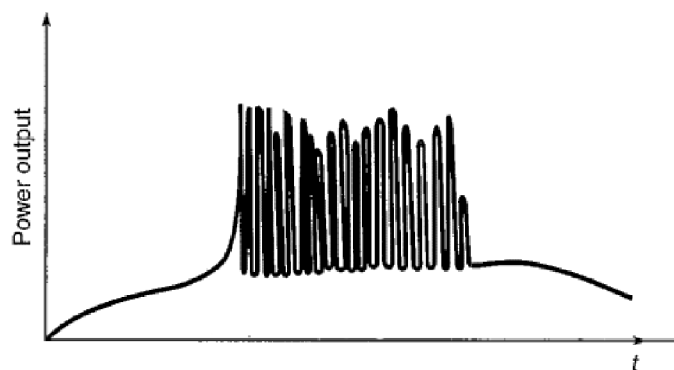


**Fig. 12.9**

When the pump is suddenly switched on to a value much above the threshold, the population inversion builds up and crosses the threshold value, as a consequence of which the photon number builds up rapidly to a value much higher that the steady state value. Since the photon number is higher than the steady state value, the rate at which the upper level depletes (because of stimulated transitions) is much higher than the pump rate. Consequently, the inversion becomes below threshold, and the laser action ceases. Thus, the emission stops for a few microseconds, within which time the flash lamp again pumps the ground state atoms to the upper level, and laser oscillations begin again. This process repeats itself till the flash lamp power falls below the threshold value and the lasing action stops (see Fig. 12.9).

## 12.9 He-Ne laser (four level laser)

We will now briefly discuss the He-Ne laser which was first fabricated by Ali Javan and his coworkers at Bell Telephone Laboratories in USA. This was also the first gas laser to be operated successfully.

The He-Ne laser consists of a mixture of He and Ne in a ratio of about 10:1, placed inside a long narrow discharge tube (see Fig. 12.10). The pressure inside the tube is about 1 Torr. The gas system is enclosed between a pair of plane mirrors or a pair of concave mirrors so that a resonator system is formed. One of the mirrors is of very

high reflectivity while the other is partially transparent so that energy may be coupled out of the system.
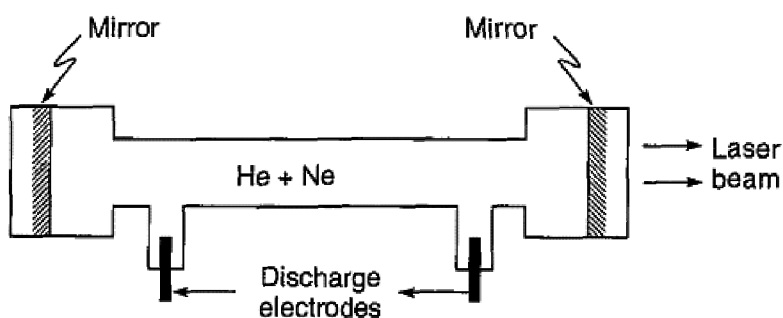


**Fig. 12.10**

The first few energy levels of He and Ne atoms are shown in Fig. 12.11. When an electric discharge is passed through the gas, the electrons traveling down the tube collide with the He atoms and excite them (from the ground state $F_1$) to the levels marked $F_2$ and $F_3$. These levels are metastable, i.e., He atoms excited to these states stay in these levels for a sufficiently long time before losing energy through collisions.
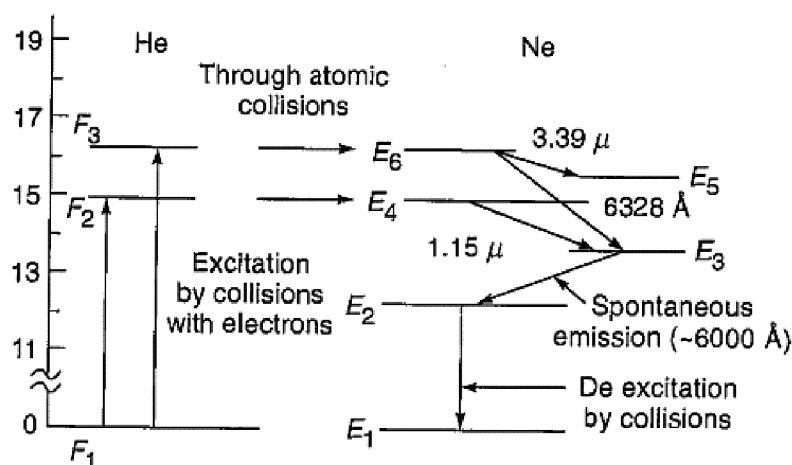


**Fig. 12.11**

Through these collisions, the Ne atoms are excited to the levels marked $E_4$ and $E_6$ which have nearly the same energy as the levels $F_2$ and $F_3$ of He. Thus, when the atoms in levels $F_2$ and $F_3$ collide with unexcited Ne atoms, they raise them to the levels $E_4$ and $E_6$ respectively. Thus, we have the following two step process :

1. Helium atom in the ground state $F_1$ + collision with electron →

Helium atom in the excited state ($F_2$ or $F_3$) + electron with lesser kinetic energy.

2. The spectroscopic states corresponding to the states $F_1$, $F_2$ and $F_3$ are $1^1S_0$, $2^3S_1$ and $2^1S_0$ respectively. The excited states of He ($F_2$ or $F_3$) are metastable - they would not readily lose energy through spontaneous emissions (the radioactive lifetime of these excited states would be about one hour). However, they can readily lose energy through collisions with Ne atoms :

He atom in the excited state $F_3$ + Ne atom in the ground state →

He atom in the ground state + Ne atom in the excited state $E_6$.

Similarly

He atom in the excited state $F_2$ + Ne atom in the ground state →

He atom in the ground state + Ne atom in the excited state $E_4$.

This results in a sizeable population of the levels $E_4$ and $E_6$. The population in these levels happens to be much more than those in the lower levels $E_3$ and $E_5$. Thus, a state of population inversion is achieved, and any spontaneously emitted photon can trigger laser action in any of the three transitions shown in Fig. 12.11. The Ne atoms then drop down from the lower laser levels to the level $E_2$ through spontaneous emission. From the level $E_2$ the Ne atoms are brought back to the ground state through collision with the walls. The transition from $E_6$ to $E_5$, $E_4$ to $E_3$ and $E_6$ to $E_3$ result in the emission of radiation having wavelengths 3.39 µm, 1.15 µm and 6328Å respectively. It may be noted that the laser transitions corresponding to 3.39 µm and 1.15 µm are not in the visible region. The 6328Å transition corresponds to the well-known red light of the He-Ne laser. A proper selection of different frequencies may be made by choosing end mirrors having high reflectivity over only the required wavelength range. The pressures of the two gases must be so chosen that the condition of population inversion is not quenched. Thus, the conditions must be such that there is an efficient transfer of energy from He to Ne atoms. Also, since the level marked $E_2$ is metastable, electrons colliding with atoms in level $E_2$ may excite them to level $E_3$, thus decreasing the population inversion. The tube containing the gaseous mixture is also made narrow so that He atoms in level $E_2$ can get de-excited by collision with the walls of the tube. Referring to Fig. 12.11, it may be mentioned that actually there are a large number of levels grouped around $E_2$, $E_3$, $E_4$, $E_5$ and $E_6$. Only those levels are shown in the figure which correspond to the important laser transitions.

Gas lasers are, in general, found to emit light, which is more directional and more monochromatic. This is because of the absence of such effects as crystalline imperfection, thermal distortion, and scattering, which are present in solid state lasers. Gas lasers are capable of operating continuously without need for cooling.

In an improved form of the He-Ne laser, the resonator mirrors $M_1$ and $M_2$ are used externally to the laser cavity, because the mirrors sealed inside the discharge tube are eroded by the gas discharge and have to be replaced. These are spherical in shape and the radius of curvature of each mirror is equal to the cavity length (see Fig. 12.12).

Further, to minimize the loss due to reflection at the ends of the discharge tube, Brewster windows $W_1$ and $W_2$ are used, as shown. These are fitted at an angle known as Brewster angle $\theta_B$ given by $\tan \theta_B = \mu$, where $\mu$ is the refractive index of the material of the window.
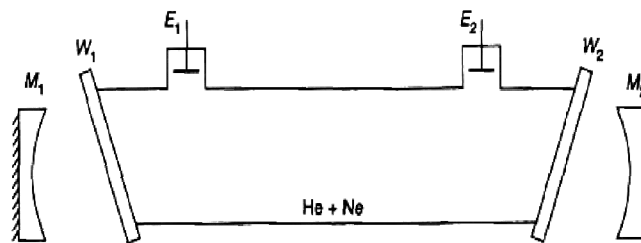


**Fig. 12.12**

To understand the action of Brewster's windows $W_1$ and $W_2$, we consider an unpolarized wave incident on a plate as a resultant of two superimposed plane polarized waves, one of which is polarized in a plane passing through the normal to the window and the tube axis and the other at right angle to this plane. The wave polarized normal to the plane of oscillation is completely reflected by the window plate and, therefore, does not play any part. The other wave, polarized in the plane of oscillation, is transmitted in the same direction, and is completely reflected by the mirrors $M_1$ and $M_2$ again and again, thus producing a strong plane polarized laser beam.

The power output of He-Ne Laser lies between 1 mW to 50 mW of continuous wave for inputs of about 5 W to 10 W. He-Ne laser gives a continuous operation.

## 12.10   Holography

Holography is a lens-less two-step imaging process which can produce three dimensional images. It was first suggested by D. Gabor in 1948. In ordinary photography we record only the intensity of light and no information about phase so that the three-dimensional character of the object is lost. But in holography we record information regarding boththe amplitudes and phases. The word 'hologram' is derived from Greek word 'holos' meaning 'the whole' i.e., the whole of information.

**Principle of holography :** A beam of coherent light from a laser is split into two by a semi-transparent mirror such that one beam (called object beam) can be scattered by the object to be photographed and the other beam (phase related reference beam) falls directly on the film. These two beams form interference patterns on the film. The fine speckled pattern on the film contains information regarding both the amplitude and phase. Thus, a hologram is produced. This step of holography is called **recording**
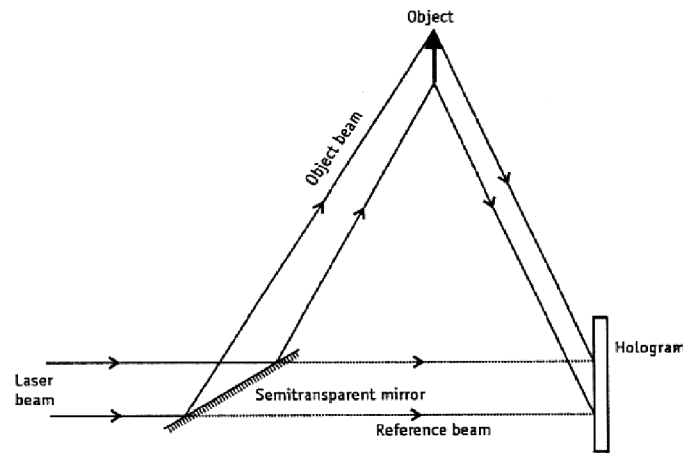


**Fig. 12.13 Recording**
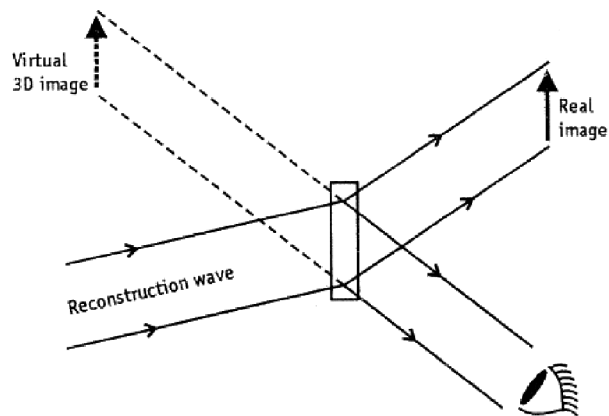


**Fig. 12.14 Recording**

(see Fig. 12.13). The next step is the **reconstruction step** (see Fig. 12.14). The hologram is illuminated with coherent light called reconstruction wave (usually of

same wavelength as the original beam). The hologram acts as a diffraction grating producing two sets of diffracted beams, one set forms a real image while the other forms a three-dimensional virtual image.

**Applications :** Holography gives a 3D idea of the object. A single hologram can record a large number of images. A small portion of the hologram can reproduce the entire image. All these properties make holography very popular in many applications. Most common applications are in holographic microscopy, holographic interferometry, 3D cinema or TV, espionage, non-destructive testing etc.

## 12.11  Summary

- Laser often possesses some very special characteristics - some of these are (a) **Directionality**, (b) **High power** and (c) **Spectral purity**.

- As put forward by Einstein, when an atom is in the excited state then, in addition to the spontaneous emission, it can also make a transition to a lower energy state by what is known as stimulated emission in which an incident signal of appropriate frequency triggers an atom in an excited state to emit radiation - this results in the amplification of the incident beam. If we are able to create a state of population inversion in which there are larger number of atoms in the upper state, then the number of stimulated emissions would exceed the number of stimulated absorptions resulting in the (optical) amplification of the beam.

- The three main components of any laser are

  (i) The active medium,

  (ii) The pumping mechanism, and

  (iii) The optical resonator.

- The first successful operation of a laser device ($\lambda \sim 0.684$ μm) was demonstrated by Theodore Mainman in 1960 using a ruby crystal. Within a few months of the operation of the ruby laser, Ali Javan and his associates constructed the first gas laser ($\lambda \sim 0.633$ μm), namely the helium - neon laser.

- The basic technique in **holography** has ben discussed.

## 12.12  Review Questions and Answers

**1.  Define meta-stable state.**

**Ans.** Normally an atom in the excited state comes back to the original or groundstate within a very short time of about $10^{-8}$ second by the emission of a photon.

*The average time for which an atom remains in an excited state is known as its mean life.* The mean life of an atom is characteristic of the energy state. Some energy states have comparatively longer mean life.

*Energy states having mean life of more than $10^{-3}$ sec. are known as meta-stable states.*

**2.    Name some important types of lasers.**

**Ans. Types of Lasers :** Various types of lasers in use are :

(i) **Solid lasers :** The most commonly used solid laser is a *Ruby laser.* The other solid laser used is $Y_2Al_5O_{12}$ doped with $Nd^{+3}$ ions in place of $Y$ (Yttrium). Its most useful operating wavelength is the infra-red light $\lambda = 1.064$ μm.

(ii) **Gas lasers :** Helium-Neon gas lasers (consisting of about 10 parts of Ne and 1 part of He is the best-known gas laser. It operates on three wavelength 6328 Å in the red and 1.15 and 3.3 μm in the infra-red region.

(iii) Other important types include semiconductor lasers, chemical laser, and dye lasers.

**In a semiconductor laser**, the conduction band population of electrons is increased by injecting free electrons into the negative side of an ordinary junction diode. The operating wavelength for GaAs laser is usually in the range 8400 - 8500 Å.

**In a chemical laser**, population inversion results from a chemical reaction.

**In a dye laser**, laser action is obtained by pumping certain dye solutions with micro second long flashes from high intensity lamps. The active material is an organic fluorescent dye dissolved in a common solvent. Optical excitation and emission in organic dyes takes place between the rotational vibrational levels of the *singlet ground state* $S_0$ and the *singlet excited state* $S_1$.

**3.    Explain the concept of negative temperature in connection with the population inversion in an active medium.**

**Ans.** See Article 12.5.

**4.    What are Einstein's *A,B* coefficients? Derive a relation between them.**

**Ans.** See Article 12.7.

**5.    What is optical resonator? Discuss the role played by it in a laser system.**

**Ans.** See Article 12.8.

**6.    Draw a neat sketch of a ruby laser. With the help of a simple energy level diagram describe the operation of a ruby laser.**

**Ans.** See Article 12.8.

7.  **Give the construction of a He-Ne laser. With the help of a simple energy level diagram show how population inversion is achieved here.**

**Ans.** See Article 12.9.

8.  **Mention some important applications of laser.**

**Ans.** Lasers find wide applications in different branches of science, technology and in surgery etc.

Laser beams can project intense energy into a small area of approximately 0.001 mm in diameter and produce, at the focal point, temperatures of the order of 10,000 to 16,000K. Therefore, laser has been used for drilling holes in hard materials like diamonds, vaporize metals, painless drilling and welding of teeth, retinal surgery i.e., welding the detached retina back into position and treatment of malignant tumours, specially removing eye tumours.

As the laser beam acts as a sharp knife and its localized heating causes all the blood vessels to be sealed instantaneously, it is used to perform 'blood less surgery' and for performing difficult operations of the scalp and abdominal skin.

In basic scientific research, laser has been used in exploring the molecular structure, nature of chemical reactions, Raman spectroscopy, precision measurement of length and accurate determination of velocity of light.

In high-speed photography, laser has been used to obtain very sharp images of moving objects like bullets in flight. Holography i.e., three-dimensional photography, has been made possible by the use of lasers.

Laser light, being highly coherent, can be modulated to transmit hundreds of messages at a time on radio and television. As laser beams are not readily absorbed by water, these have proved very useful for under water communication. Laser beam is also being used for automatic control and guidance of rockets and satellites, determining the distance of stars, planets, and moon, and ranging of the planes like Radar. A powerful laser beam can also be used to destroy aero planes and missiles, as well as damaging far-located tanks.

9.  **Briefly give the requirements for holography and mention various properties of a hologram.**

**Ans. Requirements :** The various requirements of holography are as given below:

1. **Monochromaticity**. The light source used should be monochromatic, as the fringe pattern produced is a function of wavelength of the illumination. If a broad spectrum is used, then each wavelength will give rise to its own fringe pattern, which will overlap and produce blurring.

2. **Spatial coherence**. If the source lacks spatial coherence i.e., it is broad, then each element of the source will produce interference fringes, which are displaced from those of other elements. The resultant effect of this fringe pattern will average to some uniform value and the fringe pattern will be missing.

The monochromaticity is obtained by passing the light through a monochromator or a narrow band colour filter. Spatial coherence is obtained by passing through a pin-hole. Such a traditional source is quite inefficient as illumination available is limited.

Thus, a laser beam, which is highly monochromatic and has extraordinary spatial coherence, is used for wavefront reconstruction photography.

**Properties.** 1. A hologram has remarkably three-dimensional image properties and a highly exact reproduction of the image is possible.

2. With pulsed laser sources, instant three dimensional volume holograms have been prepared which contain a fog like suspension of particles. Once image has been reconstructed, the particles can be studied and counted with a microscope.

3. Multicolour holograms have also been prepared. These utilize the fact that photographic emulsion has thickness. Successive layers in emulsion can reflect a particular colour strongly if their spacing is proper for that colour and angle of illumination.

4. Each part of the hologram, no matter how small, can reproduce the entire image. Thus, a hologram can be broken into small fragments, each of which can give a complete image. It is because each part of the hologram receives light from all parts of the object and hence contains the entire image in an encoded form.

5. The hologram constructed is regarded as a negative, but the image it produces is positive. If a contact print is obtained, the hologram would be reversed in the sense that the opaque areas would now become transparent and vice versa. The image produced from a copy will, however, remain a positive and is indistinguishable from the image produced by the original.

**10.** **Indicate one major disadvantage of a three level laser.**

**Ans.** One major disadvantage of a three level laser, such as a ruby laser, is that no significant stimulated emission occurs until at least half the ground state population has been excited to the meta-stable state. That is, population inversion exists only when the meta-stable state is more densely populated than the ground state. This makes a high threshold for optical pumping and a considerable waste of energy. Better performance is expected from a four level laser in which population inversion begins between the meta-stable state and laser terminal state with very little pumping.

# Unit-13 : Polarization of Light

**Structure**

## 13.1 Objectives

In going through this unit, you will learn:
1.   Linear, Circular and Elliptical Polarization
2.   Propagation of E.M. Waves in Anisotropic Media

## 13.2  Introduction

The phenomena of interference and diffraction of light show that the light is a wave motion which may be longitudinal or transverse in nature. It is the phenomenon of **polarization** which for its explanation requires that the light must be a **transverse wave**. In fact, in the electromagnetic theory (Article 8.4) light is considered as a transverse electromagnetic wave consisting of vibrating electric and magnetic vectors at right angles to each other and also at right angles to the direction of propagation of light.

The electric vector is called **light vector** because it is responsible for the sensation of vision. In an ordinary beam of light with millions of waves the light vectors of component waves will remain in all possible directions on a plane drawn at right angles to the direction of propagation. This happens so due to random orientations of excited atoms or molecules in the source. Such an ordinary beam of light with the electric vectors arranged symmetrically about the direction of propagation is called an **unpolarized light**. We may assume all these vibrations to be resolved into two rectangular components of equal amplitude but having a relative phase difference which varies rapidly and randomly. Now, if by some means one of these rectangular vibrations is cut off, we get vibrations of all the component waves confined in one definite direction. Such a light is said to be plane polarized light (also called p-state

light). The pictorial representations of unpolarized and plane polarized lights are shown in Fig. 13.1.



(a) Unpolarised light

(b) Plane polarised light with its vibration in the plane of the paper

(c) Plane polarised light with its vibration in the plane perpendicular to the plane of the paper
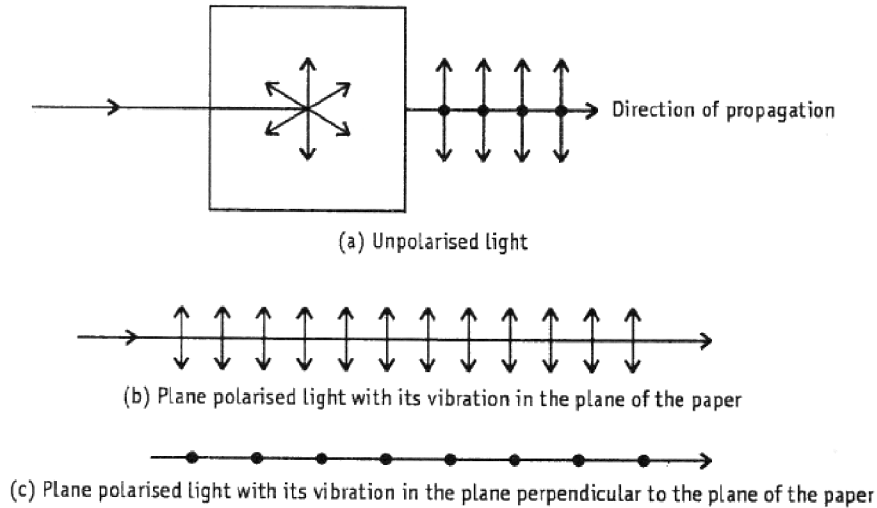
**Fig. 13.1**

The electric vector of a plane polarized light is confined to a particular plane, known as the plane of vibration. This plane contains the $\vec{E}$ - vector and the propagation vector $\vec{k}$. The plane perpendicular to the plane of vibration is known as **plane of polarization** (see Fig. 13.2).
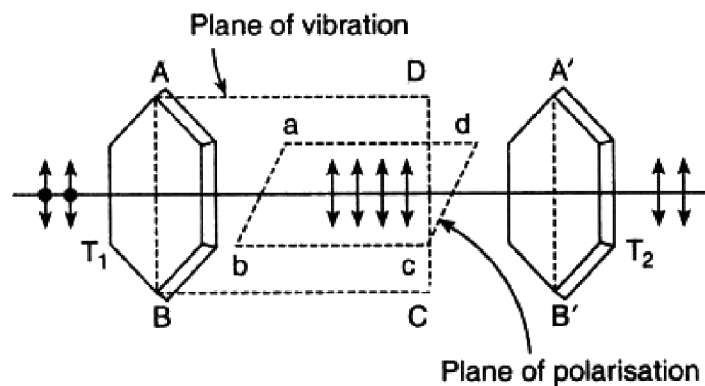


**Fig. 13.2**

By the superposition of two plane polarized waves under suitable conditions the resultant light vector may be made to rotate in a plane perpendicular to the direction of propagation. If the magnitude of the resulting light vector remains constant the tip

of the light vector appears to trace out a circle at a fixed space. Such a light is said to be circularly polarized. If, on looking towards the incoming light the resultant light vector appears to rotate clockwise, then the light is said to be right circularly polarized light. If the light vector rotates counter-clockwise, then the light is said to be left circularly polarized light. On the other hand, if the magnitude of the resulting light vector varies periodically between a maximum and a minimum value then the tip of the light vector appears to trace out an elliptic path. Such a light is said to be elliptically polarized.

A mixture of polarized and unpolarized lights is called partially polarized light.

## 13.3 Propagation of EM waves in anisotropic media

In this Article, we will discuss the plane wave solutions of Maxwell's equations in an anisotropic medium and we will completely restrict ourselves to uniaxial crystals for which the refractive indices are given by

$$n_x = n_y = n_o \text{ and } n_z = n_e \qquad \qquad \text{... (13.1)}$$

The difference between an isotropic and an anisotropic medium is in the relationship between the displacement vector **D** and the electric vector **E**; the displacement vector **D** is defined elsewhere (see any textbook on Electricity and Magnetism). In an isotropic medium, **D** is in the same direction as **E** and one can write

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \qquad \qquad \text{... (13.2)}$$

where $\varepsilon$ is the dielectric permittivity of the medium. On the other hand, in an anisotropic medium **D** is not, in general, in the direction **E** and the relation between **D** and **E** can be written in the form

$$\left.\begin{array}{l} D_x = \varepsilon_{xx}E_x + \varepsilon_{xy}E_y + \varepsilon_{xz}E_z \\ D_y = \varepsilon_{xy}E_x + \varepsilon_{yy}E_y + \varepsilon_{yz}E_z \\ D_x = \varepsilon_{zx}E_x + \varepsilon_{zy}E_y + \varepsilon_{zz}E_z \end{array}\right\}$$

where $\varepsilon_{xx}$, $\varepsilon_{xy}$, ... are constant. One can show that

$$\varepsilon_{xy} = \varepsilon_{yx} \text{ ; } \varepsilon_{xz} = \varepsilon_{zx} \text{ and } \varepsilon_{yz} = \varepsilon_{zy} \qquad \qquad \text{... (13.4)}$$

Further, one can always choose a coordinate system (i.e., one can always choose approximately the directions of $x$, $y$, and $z$ axes inside the crystal) such that

$$\left.\begin{array}{l} D_x = \varepsilon_x E_x \\ D_y = \varepsilon_y E_y \\ D_z = \varepsilon_z E_z \end{array}\right\} \qquad \qquad \text{... (13.5)}$$

This coordinate system is known as the principal axis system and the quantities $\varepsilon_x$, $\varepsilon_y$ and $\varepsilon_z$ are known as the principal dielectric permittivities of the medium. If

$$\varepsilon_x \neq \varepsilon_y \neq \varepsilon_z \quad \text{(biaxial)} \qquad \qquad \text{... (13.6)}$$

we have what is known as a biaxial medium and the quantities

$$n_x = \sqrt{\frac{\varepsilon_x}{\varepsilon_0}}, \quad n_y = \sqrt{\frac{\varepsilon_y}{\varepsilon_0}}, \quad n_z = \sqrt{\frac{\varepsilon_z}{\varepsilon_0}} \qquad \qquad \text{... (13.7)}$$

are said to be the principal refractive indices of the medium; in the above equation $\varepsilon_0$ represents the dielectric permittivity of free space ($= 8.8542 \times 10^{-12}$ C²/N–$m^2$). If

$$\varepsilon_x = \varepsilon_y \neq \varepsilon_z \quad \text{(uniaxial)} \qquad \qquad \text{... (13.8)}$$

we have what is known as a uniaxial medium with the z-axis representing the optic axis of the medium. The quantities

$$n_0 = \sqrt{\frac{\varepsilon_x}{\varepsilon_0}} = \sqrt{\frac{\varepsilon_y}{\varepsilon_0}} \quad \text{and} \quad n_e = \sqrt{\frac{\varepsilon_z}{\varepsilon_0}} \qquad \qquad \text{... (13.9)}$$

are known as ordinary and extraordinary refractive indices. For a uniaxial medium, since $\varepsilon_x = \varepsilon_y$ the $x$ and $y$ directions can be arbitrarily chosen as long as they are perpendicular to the optic axis, i.e., *any* two mutually perpendicular axes (which are also perpendicular to the z-axis) can be taken as the principal axes of the medium. On the other hand, if

$$\varepsilon_x = \varepsilon_y = \varepsilon_x \quad \text{(isotopic)} \qquad \qquad \text{... (13.10)}$$

we have an isotropic medium and can choose any three mutually perpendicular axes as the principal axis system.

We will assume the anisotropic medium to be non-magnetic so that

$$\boldsymbol{B} = \mu_0 \boldsymbol{H}$$

where $\mu_0$ is the free space magnetic permeability.

Let us consider the propagation of a plane electromagnetic wave (EM). For such a wave the vectors **E**, **H**, **D** and **B** would be proportional to exp[$i(\boldsymbol{k}.\boldsymbol{r}-\omega t)$ ]. Thus

$$\left.\begin{array}{l} \boldsymbol{E} = \boldsymbol{E}_0 e^{i(k.r-\omega t)} \boldsymbol{H} = \boldsymbol{H}_0 e^{i(k.r-\omega t)} \\ \boldsymbol{D} = \boldsymbol{D}_0 e^{i(k.r-\omega t)} \boldsymbol{B} = \boldsymbol{B}_0 e^{i(k.r-\omega t)} \end{array}\right\} \qquad \qquad \text{... (13.11)}$$

where the vector $\boldsymbol{E}_0$, $\boldsymbol{H}_0$, $\boldsymbol{D}_0$ and $\boldsymbol{B}_0$ are independent of space and time; **k** represents the propagation vector of the wave and $\omega$ the angular frequency. The wave velocity

$v_w$ (also known as the phase velocity) and the wave refractive index $n_w$ are defined through the following equation :

$$v_w = \frac{\omega}{k} = \frac{c}{n_w} \qquad \qquad ... (13.12)$$

$$\therefore \quad |k| = k = \frac{\omega}{c} n_w \qquad \qquad ... (13.13)$$

In the present article, it is our objective the possible values of $n_w$ when a plane wave propagates through an anisotropic dielectric. Now, in a dielectric medium

$$div\ D = 0 \qquad \qquad ... (13.14)$$

For a plane wave given by (6.11) the above equation becomes

$$i(k_x Dx + k_y D_y + k_z D_z) = 0$$

$$\text{or} \quad D.k = 0 \qquad \qquad ... (13.15)$$

implying that $D$ is always at right angles to $k$.

Similarly, since in a non-magnetic medium $div\ H = 0$,

$H$ will always be right angles to $k$. $\qquad \qquad (13.16)$

Now, in the absence of any currents (i.e., $J = 0$) Maxwell's curl equations become

$$\nabla \times E = -\frac{\partial B}{\partial t} - i\omega B = i\omega\mu_0 H \qquad \qquad ... (13.17)$$

$$\text{and} \quad \nabla \times H = \frac{\partial D}{\partial t} = i\omega D \qquad \qquad ... (13.18)$$

where we have assumed the medium to be non-magnetic (i.e., $B = \mu_0 H$). Now it can be shown that

$$H = \frac{1}{\omega\mu_0}(k \times E) \qquad \qquad ... (13.19)$$

$$D = \frac{1}{\omega}(H \times k) \qquad \qquad ... (13.20)$$

Equations (13.19) and (13.20) show that

$H$ is at right angles to $k$, $E$ and $D$ $\qquad \qquad ... (13.21)$

implying

$k$, $E$ and $D$ will always be in the same plane.

Further [see (13.15)]

$D$ is at right angles to $k$ $\qquad \qquad ... (13.22)$

Substituting for $H$ in (13.20), we get

$$D = \frac{1}{\omega^2 \mu_0}\left[(k.k)E - (k.E)k\right] \qquad \text{... (13.23)}$$

where we have used the vector identity

$$(A \times B) \times C = (A.C)B - (B.C)A$$

Thus $\qquad D = \frac{k^2}{\omega^2 \mu_0}\left[E - (\hat{k}.E)\hat{k}\right] = \frac{n_w^2}{c^2 \mu_0}\left[E - (\hat{k}.E)\hat{k}\right] \qquad \text{... (13.24)}$

where $\hat{k} = \dfrac{k}{k}$

represents the unit vector along $k$. Since

$$D_x = \varepsilon_x E_x = \varepsilon_0\, n_x^2 E_x$$

We have for the x-component of (13.24)

$$\frac{\varepsilon_0 \mu_0 c^2 n_x^2}{n_w^2} E_x = E_x - k_x\left(k_x E_x + k_y E_y + k_z E_z\right)$$

Since $c^2 = 1/(\varepsilon_0\mu_0)$, we have

$$\left(\frac{n_x^2}{n_w^2} - k_y^2 - k_z^2\right)E_x + k_x k_y E_y + k_x k_z E_z = 0 \qquad \text{...(13.25)}$$

where we have used the relation $k_x^2 + k_y^2 + k_z^2 = 1$ (since $\hat{k}$ is a unit vector).

Similarly, $\qquad k_x k_y E_x + \left(\dfrac{n_y^2}{n_w^2} - k_x^2 - k_z^2\right)E_y + k_y k_z E_z = 0 \qquad \text{... (13.26)}$

$$k_x k_z E_x + k_y k_z E_y + \left(\frac{n_z^2}{n_w^2} - k_x^2 - k_y^2\right)E_z \qquad \text{... (13.27)}$$

Since the above equations form a set of three homogeneous equations, for non-trivial solutions, we have

$$
\begin{vmatrix}
\dfrac{n_x^2}{n_w^2} - k_y^2 - k_z^2 & k_x k_y & k_x k_z \\[3mm]
k_x k_y & \dfrac{n_y^2}{n_w^2} - k_x^2 - k_z^2 & k_y k_z \\[3mm]
k_x k_z & k_y k_z & \dfrac{n_z^2}{n_w^2} - k_x^2 - k_y^2
\end{vmatrix} = 0
$$

... (13.28)

We should remember that we still do not know the possible values of $n_w$. Indeed, for a given direction of propagation (i.e., for given values of $k_x$, $k_y$ and $k_z$) the solutions of (13.28) gives us the two allowed values of $n_w$. It may be mentioned that from (13.28) it appears as if we will have a cubic equation in $n_w^2$ which would give us three roots of $n_w^2$: however, the coefficient of $n_w^6$ will *always* be zero and hence there will be always two roots. We illustrate the general procedure by considering propagation through a uniaxialmedium.

## 13.4 Fresnel's formula

The simple laws of reflection and refraction give no indication in regard to what fraction of the amplitude or energy of the incident wave is reflected and what fraction is transmitted. This is what we shall now inquire. The equations that relate the amplitudes of reflected and transmitted (refracted) waves with that of the incident wave are called Fresnel's equations of formula.

We will now derive expressions for the reflection and transmission coefficients when a plane polarized wave is incident on an interface of two dielectrics. We will first consider the case when the electric vector lies in the plane of incidence which will be followed by the case when the electric vector is at right angles to the plane of incidence.

**Case 1. E Parallel to the plane of incidence :** We will assume the electric vector to lie in the plane of incidence as shown in Fig. 13.3.
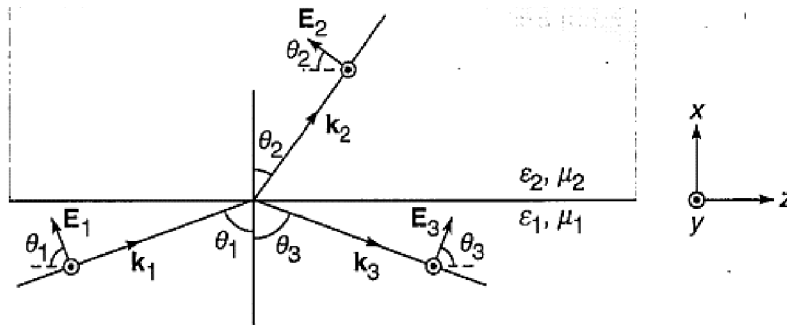
**Fig. 13.3**

The magnetic vectors are along the y-axis. Clearly, the z-component of the electric field represents a tangential component which should be continuous across the surface. Thus

$$E_{1z} + E_{3z} = E_{2z}$$

$$\text{or } -E_1 \cos \theta_1 + E_3 \cos \theta_1 = -E_2 \cos \theta_2 \qquad \qquad ... (13.29)$$

$$[-E_{10} \, exp\{i(\boldsymbol{k}_1.\boldsymbol{r} - \omega t)\} + E_{30} \, exp\{i(\boldsymbol{k}_3.\boldsymbol{r} - \omega t)\}]_{x=0} \cos \theta_1$$

$$= [-E_{20} exp\{i(\boldsymbol{k}_2.\text{r} - \omega t)\}]_{x=0} \cos \theta_2 \qquad (13.30)$$

where $E_{10}$, $E_{20}$ and $E_{30}$ are the peak values of $E_1$, $E_2$ and $E_3$ respectively and independent of space and time but may, in general, be complex. The vectors $k_1$, $k_2$ and $k_3$ represent the propagation vectors associated with the incident, refracted, and reflected waves respectively. We must notice that the fields must satisfy Maxwell's equations. We also remember that $\boldsymbol{k}_{1x}$, $\boldsymbol{k}_{1y}$ and $\boldsymbol{k}_{1z}$ represent the $x$, $y$, and z-components of $\boldsymbol{k}_1$; similarly, for $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$.

Once again, since this condition has to be satisfied at all space points in the plane $x = 0$ and at all times, the exponents must be identically equal which leads to the following equations

$$\omega = \omega_2 = \omega_3 \qquad \qquad ... (13.31)$$

$$k_{1y} = k_{2y} = k_{3y} \qquad \qquad ... (13.32)$$

$$k_{1z} = k_{2z} = k_{3z} \qquad \qquad ... (13.33)$$

Thus

$$\left[E_{10} - E_{30}\right]\cos\theta_1 = E_{20}\cos\theta_2 \qquad \dots (13.34)$$

Further, the normal component of **D** must also be continuous and since $D = \varepsilon E$, we must have

$$\varepsilon_1 E_{1x} + \varepsilon_1 E_{3x} = \varepsilon_2 E_{2x}$$

or $\quad \varepsilon_1\left[E_{10} + E_{30}\right]\sin\theta_1 = \varepsilon_2 E_{20}\sin\theta_2 \qquad \dots (13.35)$

Substituting for $E_{20}$ from (13.34), we get

$$\varepsilon_1\left[E_{10} + E_{30}\right]\sin\theta_1 = \varepsilon_2\sin\theta_2\frac{\left[E_{10} - E_{30}\right]}{\cos\theta_2}\cos\theta_1$$

or, $\quad \left[\varepsilon_2\sin\theta_2\cos\theta_1 + \varepsilon_1\sin\theta_1\cos\theta_2\right]E_{30}$

$$= \left[\varepsilon_2\sin\theta_2\cos\theta_1 - \varepsilon_1\sin\theta_1\cos\theta_2\right]E_{10}$$

Thus $\quad r_{\parallel} = \dfrac{E_{30}}{E_{10}} = \dfrac{\varepsilon_2\sin\theta_2\cos\theta_1 - \varepsilon_1\sin\theta_1\cos\theta_2}{\varepsilon_2\sin\theta_2\cos\theta_1 + \varepsilon_1\sin\theta_1\cos\theta_2} \qquad \dots (13.36)$

where $r_{\parallel}$ denotes the amplitude reflection coefficient, the subscript $\parallel$ refers to the fact that we are referring to parallel polarization. If we now divide (13.34) by $E_{10}$ and substitute the expression for $E_{30}/E_{10}$ from (13.36), we will get

$$\left[1 - \frac{\varepsilon_2\sin\theta_2\cos\theta_1 - \varepsilon_1\sin\theta_1\cos\theta_2}{\varepsilon_2\sin\theta_2\cos\theta_1 + \varepsilon_1\sin\theta_1\cos\theta_2}\right]\cos\theta_1 = \frac{E_{20}}{E_{10}}\cos\theta_2$$

or, $\quad t_{\parallel} = \dfrac{E_{30}}{E_{10}} = \dfrac{2\varepsilon_2\sin\theta_1\cos\theta_1}{\varepsilon_2\sin\theta_2\cos\theta_1 + \varepsilon_1\sin\theta_1\cos\theta_2} \qquad \dots (13.37)$

where $t_{\parallel}$ denotes the amplitude transmission coefficient.

For non-magnetic media, $\mu_1 = \mu_0 = 4\mu \times 10^{-7}$ $N/amp^2$ and the expression for the amplitude reflection coefficient (13.36) simplifies to

$$r_{\parallel} = \frac{n_2^2\sin\theta_2\cos\theta_1 - n_1^2\sin\theta_1\cos\theta_2}{n_2^2\sin\theta_2\cos\theta_1 + n_1^2\sin\theta_1\cos\theta_2}$$

using the fact that for non-magnetic media $n = c/v = \sqrt{\varepsilon\mu/\varepsilon_0\mu_0} = \sqrt{\varepsilon/\varepsilon_0}$ , thus $n^2 = \varepsilon/\varepsilon_0$)

Since $n_1 \sin \theta_1 = n_2 \sin \theta_2$      ... (13.38)

We get

$$r_\parallel = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \qquad ... (13.39a)$$

$$= \frac{\sin \theta_1 \cos \theta_1 - \sin \theta_2 \cos \theta_2}{\sin \theta_1 \cos \theta_1 + \sin \theta_2 \cos \theta_2} \qquad ... (13.39b)$$

$$= \frac{\sin 2\theta_1 - \sin 2\theta_2}{\sin 2\theta_1 - \sin 2\theta_2} = \frac{2\cos(\theta_1 + \theta_2)\sin(\theta_1 - \theta_2)}{2\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)}$$

$$= \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)} \qquad ... (13.39c)$$

Similarly, starting from (13.37) one easily obtain

$$t_\parallel = \frac{2\cos \theta_1 \sin \theta_2}{\sin(\theta_1 + \theta_2)\cos(\theta_1 - \theta_2)} \qquad ... (13.40)$$

Equations (13.39a) – (13.39c) and (13.40) are known as *Fresnel's equations.*
From (13.39) and (13.40) we may deduce the following :

(a) **No Reflection when $n_2 = n_1$**

When $n_2 = n_1$, $\theta_2 = \theta_1$ and we get

$$r_\parallel = 0 \text{ and } t_\parallel = 1$$

Thus, there is no reflection when the second medium has the same refractive index as the first medium. Thus, if we have a transparent solid immersed in a liquid of the same refractive index, the solid would not be seen!

(b) **Polarization by Reflection :**

**Brewster's Law :** If the angle of incidence is such that

$$\theta_1 + \theta_2 = \frac{\pi}{2}, \text{ then } r_\parallel = 0$$

i.e., there is no reflected beam. Thus, if an unpolarized beam is incident at an angle such that $\theta_1 + \theta_2 = \frac{\pi}{2}$, then the parallel component of the E-vector will not be reflected,

and the reflected light will be polarized with its E-vector perpendicular to the plane of incidence (see Fig 13.4).
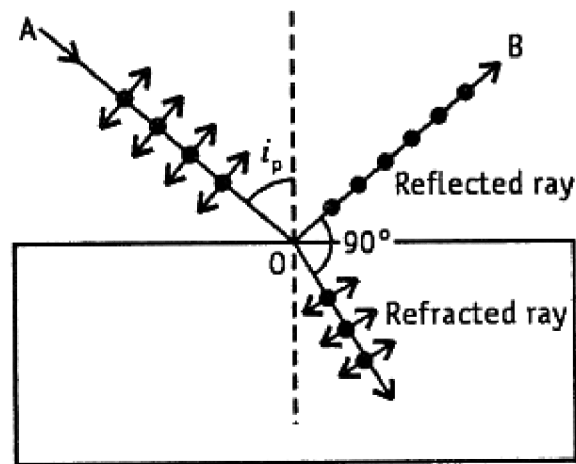


**Fig. 13.4**

This is the famous Brewster's law. The corresponding angle of incidence is known as the Brewster angle (or polarizing angle) and is usually denoted by $i_p$.

Notice that the angle of refraction will be $\frac{\pi}{2} - i_p$ and therefore Snell's law takes the form

$$\frac{n_2}{n_1} = \frac{\sin\theta_1}{\sin\theta_2} = \frac{\sin i_p}{\sin\left(\frac{\pi}{2} - i_p\right)} - \tan i_p \qquad \text{... (13.41)}$$

$$i_p = \tan^{-1}\left(\frac{n_2}{n_1}\right) \qquad \text{... (13.42)}$$

Thus, when the angle of incidence is equal to $\tan^{-1}\left(\frac{n_2}{n_1}\right)$ then the reflected beam is plane polarized. Further, the transmitted beam is partially polarized. It is easily seen that at the polarizing angle, the reflected ray is at right angles to the refracted ray.

Further, the following facts can easilybe explained from Fresnel's equations:

(c) Phase Change on Reflection and Stokes' Relations

(d) Reflection at Grazing Incidence and

(e) Total internal reflection

**Case 2. E Perpendicular to the plane of incidence :** Let us next consider the reflection and refraction of a linearly polarized plane wave with its electric vector

perpendicular to the plane of incidence; the reflection is assumed to occur at the interface of two dielectrics. Thus, the electric vectors will be along the y-axis (see Fig. 13.5).
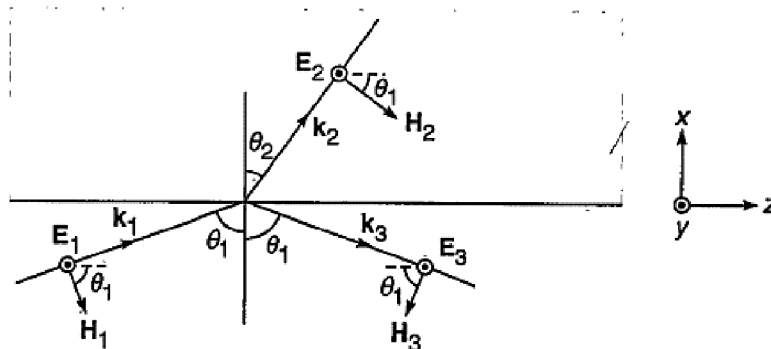


**Fig. 13.5**

Since the y-axis is tangential to the interface, the y-component of **E** must be continuous across the interface; consequently

$$E_{10} + E_{30} = E_{20} \qquad \qquad ... (13.43)$$

The directions of the magnetic fields are also shown in Fig. 6.5; they lie in the plane of incidence and are given by

$$
\left.
\begin{array}{l}
\boldsymbol{H}_1 = \boldsymbol{H}_{10} \exp\left[i\left(\boldsymbol{k}_1.r - \omega t\right)\right] = \dfrac{\boldsymbol{k}_1 \times \boldsymbol{E}_{10}}{\omega\mu_1} \exp\left[i\left(\boldsymbol{k}_1.r - \omega t\right)\right] \\[4mm]
\boldsymbol{H}_2 = \boldsymbol{H}_{20} \exp\left[i\left(\boldsymbol{k}_2.r - \omega t\right)\right] = \dfrac{\boldsymbol{k}_2 \times \boldsymbol{E}_{20}}{\omega\mu_2} \exp\left[i\left(\boldsymbol{k}_2.r - \omega t\right)\right] \\[4mm]
\boldsymbol{H}_3 = \boldsymbol{H}_{30} \exp\left[i\left(\boldsymbol{k}_3.r - \omega t\right)\right] = \dfrac{\boldsymbol{k}_3 \times \boldsymbol{E}_{30}}{\omega\mu_1} \exp\left[i\left(\boldsymbol{k}_3.r - \omega t\right)\right]
\end{array}
\right\} \qquad ... (13.44)
$$

(Notice that H lies in the plane of incidence). Since $k_1$ is at right angles to $E_{10}$, the magnitude of $H_{10}$ is simply $\dfrac{k_1 E_{10}}{\omega\mu_1}$; similarly, for $H_{20}$ and $H_{30}$. It is obvious from Fig. 13.5 that for the z-component of the magnetic field to be continuous, we must have

$$H_{10} \cos\theta_1 - H_{30} \cos\theta_1 = H_{20} \cos\theta_2 \qquad ... (13.45)$$

or, $\quad \dfrac{k_1}{\omega\mu_1}\left(E_{10} - E_{30}\right)\cos\theta_1 = \dfrac{k_2}{\omega\mu_2} E_{20} \cos\theta_2 \qquad ... (13.46)$

Substituting the expression for $E_{20}$ from (13.43) we get

$$\frac{k_1}{\omega\mu_1}\left(E_{10} - E_{30}\right)\cos\theta_1 = \frac{k_2}{\omega\mu_2}\left(E_{20} + E_{30}\right)\cos\theta_2$$

Rearranging, we get

$$r_\perp = \frac{E_{30}}{E_{10}} = \frac{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 - \dfrac{k_2}{\omega\mu_2}\cos\theta_2}{\dfrac{k_1}{\omega\mu_1}\cos\theta_1 + \dfrac{k_2}{\omega\mu_2}\cos\theta_2} \qquad \text{... (13.47)}$$

$$= \frac{\sqrt{\varepsilon_1/\mu_1}\cos\theta_1 - \sqrt{\varepsilon_2/\mu_2}\cos\theta_2}{\sqrt{\varepsilon_1/\mu_1}\cos\theta_1 + \sqrt{\varepsilon_2/\mu_2}\cos\theta_2} \qquad \text{... (13.48)}$$

$$= \frac{\sin\theta_2\cos\theta_1 - \sin\theta_1\cos\theta_2}{\sin\theta_2\cos\theta_1 - \sin\theta_1\cos\theta_2}$$

$$= -\frac{\sin\left(\theta_1 - \theta_2\right)}{\sin\left(\theta_1 + \theta_2\right)} \qquad \text{... (13.49)}$$

Further

$$t_\perp = \frac{E_{20}}{E_{10}} = 1 + \frac{E_{30}}{E_{10}}$$

$$= \frac{2\sqrt{\varepsilon_1/\mu_1}\cos\theta_1}{\sqrt{\varepsilon_1/\mu_1}\cos\theta_1 + \sqrt{\varepsilon_2/\mu_2}\cos\theta_2} \qquad \text{... (13.50)}$$

$$t_\perp = \frac{2\sin\theta_2\cos\theta_1}{\sin\left(\theta_1 + \theta_2\right)} \qquad \text{... (13.51)}$$

where the subscript $\perp$ on $r$ and $t$ refers to the fact that we are referring to the state of polarization in which $E$-vector is perpendicular to the plane of incidence. It may be mentioned that (13.47), (13.48) and (13.50) are exact, whereas (13.49) and (13.51) are valid for non-magnetic media. Once again, we can show that when $\theta_1 > \theta_c$, total internal reflection will occur and for grazing incidence the reflection is complete.

Equations (13.49) and (13.51) are known as Fresnel's equations.

## 13.5 Uniaxial and Biaxial crystals

As light propagates through a transparent material it excites the atoms within the medium. The electrons are driven by the **E**-field, and then reradiate; these secondary wavelets recombine, and the resultant refracted wave moves on. Thus, optical properties depends on the electrical properties of a transparent material, i.e., on $\varepsilon$ and $\mu$. Here we shall consider only non-magnetic crystalline materials so that the effect of $\mu$ may be neglected (i.e., $\mu = 1$). Transparent crystals are divided structurally into *seven* categories :

Cubic, trigonal, tetragonal, hexagonal, orthorhombic, monoclinic, and triclinic.

As regards optical properties these transparent crystals fall into three distinct groups :

**Cubic :** Crystals in which three crystallographically-equivalent, mutually orthogonal directions may be chosen. Here $\varepsilon_x = \varepsilon_y = \varepsilon_z$.

**Uniaxial :** These are crystals of the trigonal, tetragonal, and hexagonal systems. For example, calcite, tourmaline, quartz, etc. Here $\varepsilon_x = \varepsilon_y \neq \varepsilon_z$, where $\varepsilon_z$ is said to be the distinguished direction.

**Biaxial :** These are crystals belonging to orthorhombic, monoclinic, and triclinic systems. Examples are mica, topaz, borax, etc. $\varepsilon_x \neq \varepsilon_y \neq \varepsilon_z$ i.e., crystals in which no two crystallographically-equivalent directions may be chosen.

**Optic Axis :** In doubly refracting crystals, there is a direction fixed relative to the crystal geometry along which no separation of the *o*-ray and e-ray take place. This direction is called the *optic axis*. In other words, both the rays travel with the same velocity along the *optic axis*. It should be noted that an optic axis is just a direction and not a particular line. However, optic axis is an axis of symmetry with respect to both the crystal form and the arrangement of atoms. In *uniaxial crystals* there is only *one* optic axis and in *biaxial crystals* there are *two* optic axes.

**Principal Plane :** A plane that contains the optic axis and is perpendicular to the opposite faces of the crystal is called a principal section. As shown in Fig. 13.6, any plane parallel to the plane BDEG is a *principal section* of the calcite crystal, the line DG being the optic axis of the crystal through the point D.

As is seen the principal section cuts the calcite crystal in a parallelogram, the angles of the parallelogram being $71°$ and $109°$.
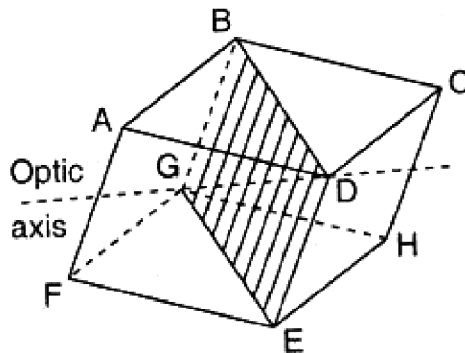
**Fig. 13.6**

A plane that contains the optic axis and o-ray or e-ray is called a **principal plane**. In general, there are two principal planes. When the plane of incidence is a principal section, it will contain both the principal planes.

## 13.6 Propagation in uniaxial crystals

As discussed earlier, for uniaxial crystals, $n_x = n_y = n_o$ and $n_z = n_e$ and the $x$ and $y$ directions can be arbitrarily chosen as long as they are perpendicular to the optic axis. Now, for a wave propagating along any direction **k**, we choose our y-axis in such a way that it is at right angles to **k**, i.e., the y-axis is normal to the plane defined by **k** and the z-axis; obviously, the x-axis will lie in the same plane (see Fig. 13.7).
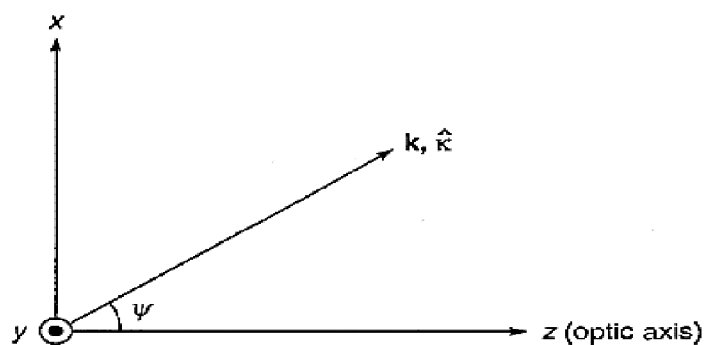


**Fig. 13.7**

Thus, we may write

$$k_x = \sin \psi, \ k_y = 0 \text{ and } k_z = \cos \psi$$

where $\psi$ is the angle that the **k** vector makes with the optic axis (see Fig. 13.7). Equations (13.25) – (13.27) therefore become

$$\left(\frac{n_0^2}{n_w^2} - \cos^2\psi\right)E_x + \sin\psi\cos\psi\; E_z = 0 \qquad \dots\dots (13.52)$$

$$\left(\frac{n_0^2}{n_w^2} - 1\right)E_y = 0 \qquad \dots (13.53)$$

$$\sin\psi\cos\psi\; E_x + \left(\frac{n_e^2}{n_w^2} - \sin^2\psi\right)E_z = 0 \qquad \dots (13.54)$$

Once again, we have a set of three homogenous equations and for non-trivial solutions, the determinant must be zero. However, since two equations involve only $E_x$ and $E_z$ and one equation involves only $E_y$ we have the following two independent solutions :

**First Solution :** We assume $E_y \neq 0$ then $E_x = 0 = E_z$. From (13.53) one obtains the solution

$$n_w = n_{wo} = n_o \text{ (ordinary wave)} \qquad \dots (13.55)$$

The corresponding wave velocity is

$$v_w = v_{wo} = \frac{c}{n_0} \text{ (y – polarized o – wave)} \qquad \dots (13.56)$$

Since the wave velocity is independent of the direction of the wave, it is referred to as the ordinary wave (usually abbreviated as the o-wave) and hence the subscript 'o' on $n_w$ and $v_w$.

**Second Solution :** The second solution of (13.52) – (13.54) will corresponds to

$$E_y = 0; \;\; E_x, E_z \neq 0 \qquad \dots (13.57)$$

We use (13.52) – (13.54) to obtain

$$\frac{E_z}{E_x} = \frac{\dfrac{n_0^2}{n_w^2} - \cos^2\psi}{\sin\psi\cos\psi} = -\frac{\sin\psi\cos\psi}{\dfrac{n_e^2}{n_w^2} - \sin^2\psi}$$

Simple manipulations would give us

$$\frac{1}{n_w^2} = \frac{1}{n_{we}^2} = \frac{\cos^2\psi}{n_o^2} + \frac{\sin^2\psi}{n_e^2} \qquad \dots (13.58)$$

where the subscript $e$ refer to the fact that the wave refractive index corresponds to the extra ordinary wave. The corresponding wave velocity would be given by

$$v_{we}^2 = \frac{c^2}{n_{we}^2} = \frac{c^2}{n_o^2}\cos^2\psi + \frac{c^2}{n_e^2}\sin^2\psi \qquad \dots (13.59)$$

Since the wave velocity is dependent on the direction of the wave, it is referred to as the extraordinary wave and hence the subscript $e$.

## 13.7 Double refraction

When an unpolarized light beam is incident on a uniaxial crystal (e.g., calcite), it usually splits up into two linearly polarized beams as shown in Fig. 13.8. The beam which travels undeviated is known as the ordinary ray (o-ray) and obeys Snell's laws of refraction. On the other hand, the second beam, which in general, does not obey Snell's laws, is known as the extraordinary ray (e-ray). Both of these o- and e-rays are plane polarized whose vibrations are along and at right angles to the principal section. This phenomenon in which a single incident ray is refracted into two rays is called **double refraction** or **birefringence** and the crystals which exhibit the phenomenon are called **doubly refracting crystals or birefringent**. For a *negative crystal* $n_e < n_o$ (e.g., calcite) and for a positive crystal $n_e < n_o$ (e.g., quartz).
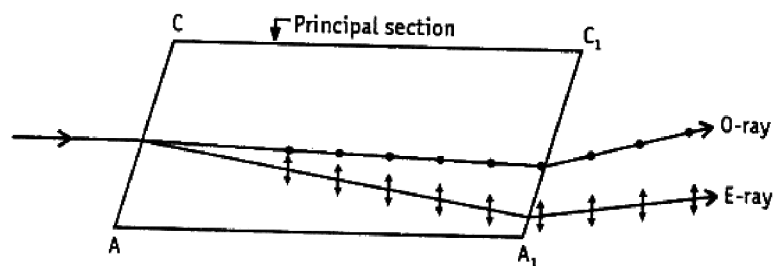


**Fig. 13.8**

### 13.7.1 Polarization by double refraction

When an unpolarized beam enters a dichroic crystal - like tourmaline, it splits up into two linearly polarized components. It has different coefficients of absorption for the two linearly polarized beams. Consequently, one of the beams gets

absorbed quickly and the other component passes through without much attenuation (see Fig. 13.9).
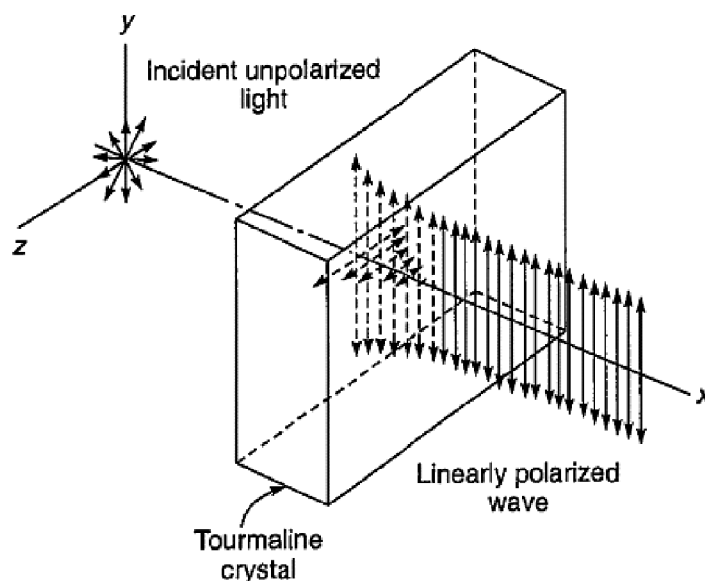


**Fig. 13.9**

**Nicol Prism :** Another method for eliminating one of the polarized beams is through total internal reflection. We noticed in Article 13.6 that the two beams have different velocities and as such the corresponding refractive indices are different. If one can sandwich a layer of a material whose refractive index lies between the two, then for one of the beams, the incidence will be at a rarer medium and for the other it will be at a denser medium. This principle is used in a Nichol prism which consists of a calcite crystal cut in such a way that for the beam, for which the sandwiched material is a rarer medium, the angle of incidence is greater than the critical angle. Thus, this particular beam will be eliminated by total internal reflection. Figure 13.10 shows a properly cut calcite crystal in which a layer of Canada Balsam has been introduced so that the ordinary ray undergoes total internal reflection. The extra ordinary component passes through and the beam emerging from the crystal is linearly polarized.
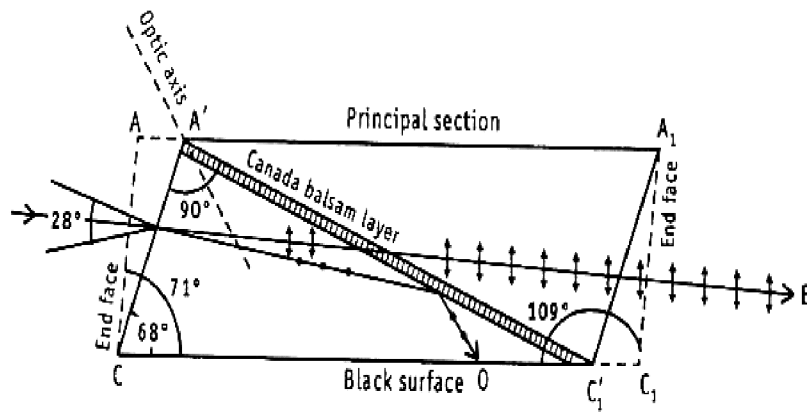
**Fig. 13.10**

---

## 13.8 Malus's law

Malus experimentally obtain a relation which shows how the intensity of light transmitted by the analyzer varies with the angle that its plane of transmission makes withthat of the polarizer. Suppose that the angle between the two planes of transmission be θ at any instant, as shown in Fig. 13.11.
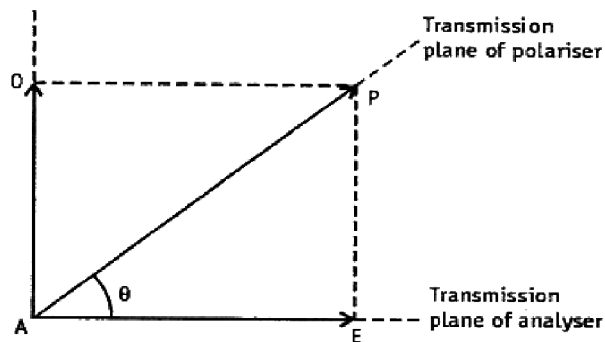


**Fig. 13.11**

The light vector $AP = E_0$ in the plane polarized light emerging from the polarizer may be resolved into two components $AE = E_0 \cos θ$, $AO = E_0 \sin θ$ respectively along and perpendicular to the plane of transmission of theanalyzer. The perpendicular component $E_0 \sin θ$ is eliminated while the parallel component is freely transmitted

through the analyzer. Hence the intensity of light that emerges from the analyzeris given by,

$$I = E_0^2 \cos^2 \theta = I_0 \cos^2 \theta \qquad \text{... (13.60)}$$

where $I_0$ is the intensity of the polarized light incident on the analyzer. This is, of course, one-half of the intensity of the unpolarized light incident on the polarizer. (13.60) shows that the transmitted intensity varies as the square of the cosine of the angle between the planes of transmission of polarizer and analyzer. This is known as **Malus's law**.

## 13.9 Retardation plates

A plate made of doubly refracting crystal with its refracting faces cut parallel to the optic axis and employed to introduce a given phase difference between $e$-ray and $o$-ray is called a *retardation plate*. If $d$ be the thickness of the plate, $n_o$ and $n_e$ be the refractive indices for $o$-ray and $e$-ray then the path difference introduced by the plate is given by

$$(n_o \sim n_e).d \qquad \text{... (13.61)}$$

Corresponding phase difference is given by

$$\delta = \frac{2\pi}{\lambda}\left(n_o \sim n_e\right).d \qquad \text{... (13.62)}$$

If the thickness $d$ of the plate is such that a path difference of $\lambda/4$ or a phase difference of $\pi/2$ is introduced between $o$-ray and $e$-ray then the plate is called a **quarter-wave plate**. Thus, the thickness of a quarter-wave plate is given by

$$(n_o \sim n_e).d = \lambda/4; \quad \text{or,} \quad d = \frac{\lambda}{4\left(n_o \sim n_e\right)} \qquad \text{... (13.63)}$$

On the other hand, If the thickness d of the plate is such that a path difference of $\lambda/2$ or a phase difference of $\pi$ is introduced between $o$-ray and $e$-ray then the plate is called a **half-wave plate**. Thus, the thickness of a half-wave plate is given by

$$(n_o \sim n_e).d = \lambda/2; \quad \text{or,} \quad d = \frac{\lambda}{2\left(n_o \sim n_e\right)} \qquad \text{... (13.64)}$$

## 13.10  Production of elliptically and circularly polarized light

**Production of elliptically polarized light :** An elliptically polarized light can be produced by superposing two mutually perpendicular coherent linear vibrations of unequal amplitudes but differing in phase by $\lambda/2$. For example, the linear optical vibrations (at $z = 0$)

$$x = a \, \cos(\omega t + \pi/2), \ y = b \, \cos \omega t$$

yield  an elliptic vibration given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Two such linear vibrations can be obtained by allowing a beam of plane polarized light to be incident normally on a quarter-wave plate with the direction of vibration making an angle $\theta$ other than 45° (about 30°) with the optic axis of the plate. The incident light of amplitude A can be resolved into two components, one along the optic axis forming e-wave of amplitude $a = A \cos \theta$ and the other normal to the optic axis forming o-wave of amplitude $b = A \sin \theta$. Let these two components (at $z = 0$) be represented by

$$x = a \cos \omega t \text{ and } y = b \cos \omega t$$

On passing through the $\lambda/4$ plate a relative phase difference of $\lambda/2$ will be introduced between these two waves. As a result, the emergent light will be elliptically polarized as discussed above.

**Production ofcircularly polarized light :** A circularly polarized light can be produced by superposing two mutually perpendicular coherent linear vibrations of equal amplitudes but differing in phase by $\pi/2$. For example, the linear optical vibrations (at $z = 0$)

$$x = a \, \cos(\omega t + \pi/2), \ y = a \, \cos \omega t$$

yield  a circular optical vibration $x^2 + y^2 = a^2$

Two such linear vibrations can be obtained by allowing a beam of plane polarized light to be incident normally on a $\lambda/4$ plate with the direction of vibration making an angle $\theta = 45°$ to the optic axis of the plate. The incident light of amplitude $A$ can be resolved into two components, one along the optic axis forming e-wave of amplitude $A \cos 45° = A / \sqrt{2}$ and the other normal to the optic axis forming o-wave of amplitude $A \sin 45° = A / \sqrt{2}$. Let these two components (at $z = 0$) be represented as

$$x = \frac{A}{\sqrt{2}} \cos \omega t \text{ and } y = \frac{A}{\sqrt{2} \cos} \omega t$$

Now on passing through the $\lambda/4$ plate a relative phase difference of $\pi/2$ will be introduced between these two waves. As a result, the emergent light will be circularly polarized as discussed above.

## 13.11  Detection of elliptically and circularly polarized light

**Detection of elliptically polarized light :** To confirm if the light coming out of the quarter-wave plate is elliptically polarized, first examine it with a rotating Nicol and adjust the position of the Nicol for maximum intensity. Place another quarter-wave plate between the first quarter-wave plate and the analyzing Nicol such that the optic axis of the second quarter-wave plate is parallel to the principal section of the Nicol adjusted for maximum intensity. If the light was elliptically polarized it will now become plane polarized. If the Nicol prism is now rotated, variation of intensity from maximum to zero will be observed.

**Detection of circularly polarized light :** To confirm whether light emerging out of the quarter-wave plate is circularly polarized, it is first passed through another quarter wave plate with optic axis in any position. The light on emergence from the second quarter-wave plate will become plane polarized. If it is now examined through a Nicol prism, it will show variation of intensity withzero minimum.

## 13.12  Analysis of circularly polarized light

A circular vibration consists of two linear vibrations of equal amplitudes mutually perpendicular to each other. When circularly polarized light is examined through a Nicol prism the intensity of the transmitted beam is not altered as the Nicol is rotated. *The behaviour is similar to that of unpolarized beam of light.*

To distinguish between the two, allow the light to fall normally on a quarter-wave plate and then examine it through the Nicol prism. If the intensity varies between maximum and zero, it is circularly polarized light. It is because the circular vibration on entering the quarter-wave plate is broken up into two linear vibrations of the same amplitude mutually perpendicular to each other. The quarter-wave plate further introduces a phase change of $\pi/2$ between the two. The resultant phase difference becomes 0 or $\pi$ and the outgoing light is *plane polarized.* On examination through the rotating Nicol, maximum intensity is observed when the principal section of the Nicol is parallel to the vibration of the plane polarized light. Light will be cut off completely when the principal section of the Nicol is perpendicular to the direction of the resultant plane polarized vibration.

The unpolarized light after passing through a quarter-wave plate remains unpolarized and hence no change in intensity will be observed when passed through a Nicol prism which is rotated gradually.
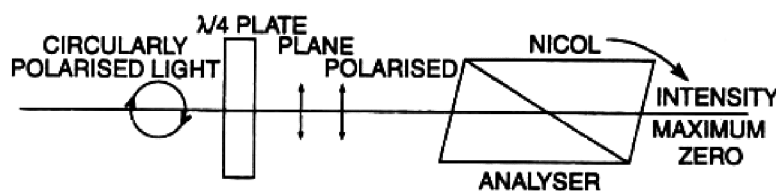


**Fig. 13.12**

Thus, we conclude that if the beam after passing through the quarter wave plate is extinguished twice in each rotation of the Nicol prism, it is circularly polarized (see Fig. 13.12).

## 13.13 Analysis of elliptically polarized light

When a beam of elliptically polarized light is examined through a rotating Nicol prism about the ray as an axis, its intensity varies in magnitude, but is never zero. It is because an elliptic vibration consists of a combination of two unequal rectilinear vibrations in two directions perpendicular to each other along the major and minor axis of the ellipse. When the principal plane of the Nicol is parallel to the vibration along the major axis, the transmitted light has a maximum intensity and when the principal plane is parallel to vibrations along the minor axis it has a minimum intensity.

The behaviour is similar to a mixture of unpolarized and plane polarized light when examined through a Nicol prism.

To distinguish between the two :

(i) Allow the light to pass through the rotating Nicol alone and adjust its position for maximum intensity.

(ii) Allow the light under examination to pass through the quarter-wave plate normally and the optic axis of the plate parallel to the principal plane of the Nicol adjusted for maximum intensity.

The elliptic vibration is resolved into two rectangular vibrations of unequal amplitudes with a phase difference of $\pi/2$ between them. By passage through the quarter wave plate a further phase difference of $\pi/2$ is introduced between the two rectangular components along the two axes of the ellipse, so that the resultant phase

difference becomes π or 0. This light on emergence becomes plane polarized. If the light is now examined through the Nicol prism, the intensity is reduced to zero twice in one rotation of the Nicol prism when the principal section of the Nicol is perpendicular to the plane of vibration of the emergent light. The incident light is thus **elliptically polarized**.

## 13.14 Babinet's compensator

In the analysis of polarized light, it is desirable to have a device which can produce a phase difference of any arbitrary value between two plane polarized waves in mutually perpendicular planes. A single crystal, such as a quarter-wave plate, can only produce one particular phase difference, depending upon its thickness and the wavelength of light used. Babinet devised an arrangement called Babinet's compensator by means of which a measurable and variable pathdifference can be introduced.

**Construction :** It consists of two slender right-angled quartz prisms *ABD* and *ACD* placed together with their hypotenuse *AD* in contact with each otherso that they together form a rectangular slab (Fig. 13.13). The axis of *ABD* is on the face *AB* and parallel to *AB* of the section *ABD* whilethe axis of *ACD* is on the face *CD* but perpendicular to the section *ACD*. Thus, the axes of the two prisms are at right angles to each other so that the ordinary ray in one will behave as the extraordinary ray when travelling in the other and vice versa. The thickness of the prisms are very small so that the separation between the ordinary and extraordinary rays is practically nil.

In the Jamin's modified form one of the prisms is shifted by a micrometer screw (Fig. 13.13).
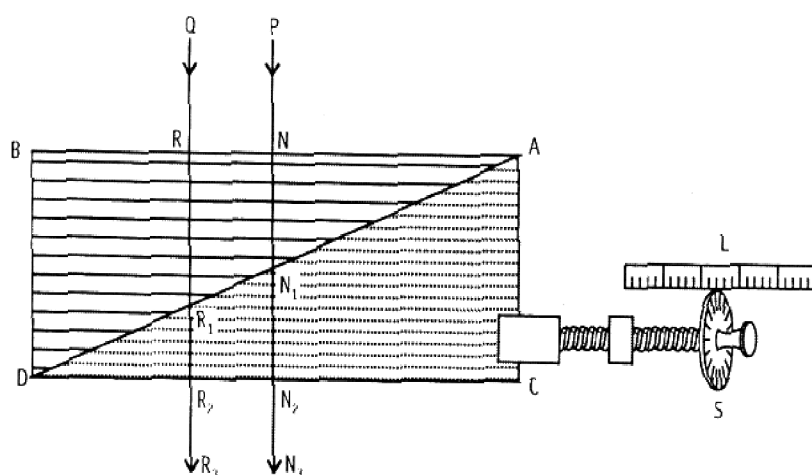


**Fig. 13.13**

**Action :** Let a ray *PN* of plane polarized light be made incident on the face *AB* normally so that its plane of vibration is making an angle 45° with the optic axes. This ray will be broken up into extraordinary and ordinary rays whose vibration will respectively be parallel and perpendicular to the axis of *ABD*. If these two rays traverse the thickness d of the prism *ABD*, then the path retardation introduced between these two rays by the prism *ABD* is $d(n_e - n_o)$ for quartz has a greater refractive index $(n_e)$ for extraordinary ray than that $(n_o)$ for ordinary ray. These two rays on entering the prism *ACD* will change their character, i.e., the ordinary and the extraordinary rays in the prism *ABD* will respectively behave as the extraordinary and the ordinary rays in the prism *ACD*. If these two rays traverse the thickness $d'$ of the prism *ACD* then the path retardation introduced between these two rays by the prism *ACD* will be, $-d' (n_e - n_o)$, for here opposite phase will be introduced. Thus, the resultant phase difference between the rays emerging from the two prisms is

$$(d - d') (n_e - n_o) \qquad\qquad ... (13.65)$$

and phase difference is

$$\delta = \frac{2\pi}{\lambda}(d - d')(n_e - n_o) \qquad\qquad ... (13.66)$$

For the central region, $d$ $(=NN_1)$ becomes equal to $d'$ $(=N_1N_2)$ and hence the resultant phase difference is zero. The emergent lightwill, therefore, remain plane polarized whose direction of vibration will be parallel to that of the incident light.

If the compensator is placed between two crossed Nicols we get a dark band in this position. This is the central dark band. Keeping the positions of crossed Nicols fixed if the micrometer screw is rotated the phase difference will go on changing and when $\delta$ becomes $2\pi$ again a dark band appears. If the shift of the micrometer for this is $2b$ then a shift $x$ of the micrometer will introduce a phase difference

$$\frac{2\pi}{2b} \times x = \frac{\pi x}{b}$$

Thus, by shifting one prism with help of a micrometer screw any amount of phase difference can be introduced between *e*-rays and *o*-rays. On the other hand, if any phase difference exists between e-ray and o-ray (for example, due to transmission through any doubly refracting crystal plate), it can be neutralized or compensated by the Babinet's compensator. This is why it is called a compensator.

**Uses :**

(a) To produce elliptically and circularly polarized light and

(b) To analyze an elliptically polarized light

## 13.15  Optical activity

If a plane polarized light is passed through some crystals like quartz along the optic axis the plane of vibration gradually undergoes rotation about the optic axis. The angle of rotation is found to depend on the thickness and nature of the crystal and also on the wavelength of the light employed. This phenomenon is called **optical activity** or **rotatory polarization** and the substance which rotates the direction of vibration of the incident polarized light is called optically active substance.

There are two classes of active substances. One class of substance rotates the line of vibration of the incident light towards right and the substances belonging to this class are called **dextro-rotatory** substances. Another class of substance rotates the line of vibration towards left and the substances which belong to this class are called **laevo-rotatory** substances.

The optical activity of crystals is associated with the structural dissymmetry of the crystals. Optically active crystals are found to possess one or more screw axis of symmetry. Here the atoms or molecules constituting the crystal are arranged on a spiral which may be left-handed or right-handed. Some crystals occur in both the varieties, one of which is the mirror image of the other producing rotation in opposite directions. Some substances in solution show this property. In this case optical activity is associated with the asymmetrical structure of the molecules themselves. There may be a preferential direction of rotation associated with a molecule. Though molecules in a solution are randomly oriented the rotations do not cancel out because the sense of rotation is same for molecules having same type of dissymmetry. Thus, optical activity is caused by structural dissymmetry of crystals or by molecular dissymmetry. There are substances in which optical activity is associated with both kinds of dissymmetry.

## 13.16  Biot'slaws for rotatory polarization

Biot made a systematic study of the natural rotation of the plane of polarization by different substances, and established the following laws :

(1) The angle ($\theta$) of rotation produced by an active substance is proportional to the length ($l$) of the substance traversed by the ray and is not affected by turning the substance around, consequently the rotation will be annulled if the ray is reflected back through the substance.

(2) The combined rotation produced by two different substances having different thickness is the algebraic sum of the rotation caused by each separately.

(3) The amount of rotation ($\theta$) in the case of a solution, is proportional to the amount of active substance ($m$) present per c.c. of the solution.

(4) The amount of rotation varies with the wavelength of the incident light and with temperature of the substance. The rotation is approximately proportional to the inverse square of the wavelength.

In case of a solution, the first and third laws may be combined to have the relation, $\theta = slm$, where $s$ is a constant and is called **rotatory power** or **specific rotation** of the solution. Thus, specific rotation of a solution is *defined* as the amount of rotation produced by a 10 cm column of liquid containing lgm of the active substance per c.c. of the solution. Thus, for a solution containing mmg. of active substance per c.c. the rotation ($\theta$) for a path length lcm is given by

$$\theta = s.m. \ (l/10) \hspace{3cm} ... (13.67)$$

where s is the specific rotation. For a pure liquid $m$ is simply the density.

In the case of optically active crystals specific rotation is defined as the rotation produced by a crystal plate of thickness 1mm.

Molecular rotationis obtained by taking the product of the specific rotation and the molecular weight of the substance. According to the fourth law the angle of rotation of the plane of vibration of a plane polarized light produced by an optically active substance is found to be approximately proportional to the inverse square of the wavelength. So, if a plane polarized white light is passed through an optically active substance it will decompose into constituent colours. This phenomenon is known as **rotatory dispersion**.

## 13.17  Fresnel's theory of optical rotation

According to Fresnel, a plane polarized light may be considered as the superposition of two equal and oppositely described circularly polarized light. In case of optically active quartz the secondary wavelets due to $o$- and $e$-rays do not touch each other along the optic axis. As a result, the two equal and opposite circular vibrations travel along the axis with unequal speed. On passing through the crystal some amount of phase difference (say $\delta$) will be introduced between them. So emergent circular vibrations will combine to form a plane polarized light whose plane of vibration is rotated by certain amount ($\delta/2$) with respect to the incident vibration. This can be explained as follows :

Let the incident linear optical vibration at $z = 0$ be represented by

$$x = 2a \cos \omega t \hspace{3cm} ... (13.68)$$

This can be resolved into two circular motions :

$$\left.\begin{array}{l} x_1 = a\ \cos\omega t \\ y_1 = a\ \sin\omega t \end{array}\right\} \text{left circular} \qquad \left.\begin{array}{l} x_2 = a\ \cos\omega t \\ y_2 = -a\ \sin\omega t \end{array}\right\} \text{right circular}$$
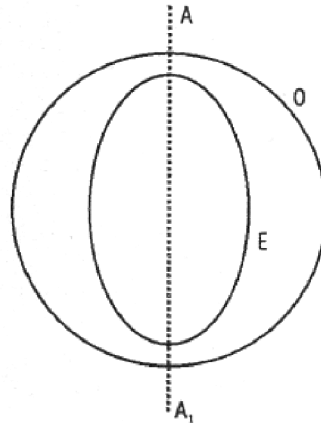


**Fig. 13.14**

These vibrations will travel with different velocities and as a result some phase difference will be introduced between them. Let the right circular motion travels faster than the left circular motion and there is a relative phase advancement of right circular motion by $\delta$. Since only relative phases matter the emergent light can be represented by

$$\left.\begin{array}{l} x_1 = a\ \cos\omega t \\ y_1 = a\ \sin\omega t \end{array}\right\} \text{and} \left.\begin{array}{l} x_2 = a\ \cos(\omega t + \delta) \\ y_2 = -a\ \sin(\omega t + \delta) \end{array}\right\}$$

Thus, the resultant vibrations along the $x$ and $y$ directions, on emergence are respectively given by

$$x = x_1 + x_2 = a[\cos\omega t + a\ \cos(\omega t + \delta)]$$

$$= 2a\cos\frac{\delta}{2}.\cos\left(\omega t + \frac{\delta}{2}\right)$$

$$y = y_1 + y_2 = a[\cos\omega t - a\ \sin(\omega t + \delta)]$$

$$= -2a\sin\frac{\delta}{2}.\cos\left(\omega t + \frac{\delta}{2}\right)$$

Taking the ratio of $x$ and $y$ we get

$$y = \tan\left(-\frac{\delta}{2}\right)x \qquad\qquad \text{... (13.69)}$$

So, the emergent light is plane polarized whose direction of vibration makes an angle $-\delta/2$ with the x-axis (which is the direction of vibration of incident plane polarized light).

If $v_L$ and $v_R$ be respectively the velocities of the left- and right-handed circular motion within the active substance, $v_R > v_L$ then their time difference $= t = (l/v_L - l/v_R)$ (where $l$ = thickness of active substance).

$\therefore$ Phase difference between the two circular vibrations is given by

$$\delta = \frac{2\pi}{T}\left(\frac{1}{v_L} - \frac{1}{v_R}\right)$$

$[\because$ for periodic time difference $T$, the phase difference is $2\pi.]$

$$\text{or,} \quad \delta = \frac{2\pi l}{cT}\left(\frac{c}{v_L} - \frac{c}{v_R}\right) = \frac{2\pi l}{\lambda}\left(n_L - n_R\right) \qquad \ldots (13.70)$$

where $n_L$ and $n_R$ are respectively the refractive indices of the active substance for left and right circularly polarized light along the direction of optic axis.

Again

$$\delta = 2\pi l\left(\frac{1}{\lambda_L} - \frac{1}{\lambda_R}\right) \quad [\because v_L T = \lambda_L; v_R T = \lambda_R]$$

$$\text{or,} \quad \delta = \frac{2\pi l \Delta\lambda}{\lambda^2} \text{ (approx.)} \qquad \ldots (13.71)$$

Taking $\lambda_L = \lambda$ and $\lambda_R = \lambda + \Delta\lambda$

Thus, the rotation of the plane of vibration is given by

$$\phi = \delta/2 = \frac{nl}{\lambda}\left(n_L - n_R\right) \qquad \text{[from (13.70)]}$$

$$\text{or,} \quad \phi = \delta/2 = \frac{nl \Delta\lambda}{\lambda^2} \qquad \text{[from (13.71)]}$$

Thus, the rotation varies as the thickness $l$ and inversely as the square of the wavelength $\lambda$.

## 13.18  Polarimeter

Polarimeter is an optical device designed for accurate determination of the angle of rotation of the plane of vibration produced by an optically active substance. When

a polarimeter is calibrated to directly read the percentage of cane sugar in the solution, then it is called a saccharimeter.

A polarimeter in its simplest form consists of two Nicols capable of rotation about a common axis. The active substance is placed between them. Such an arrangement is not sensitive enough for accurate measurement. For these various types of polarimeters have been designed for accurate measurement of the angle of rotation.

### 13.18.1 Laurent half-shade polarimeter

The construction of a Laurent half-shade polarimeter is shown in Fig. 13.15. $S'$ is a slit kept at the focal plane of the convex lens $L$ so that the rays emerging from $L$ will be parallel. The polarizing Nicol $P$ polarizes this parallel light. This polarized light then falls on the half-shade plate $H$. $T$ is a tube containing the solution of active substance.
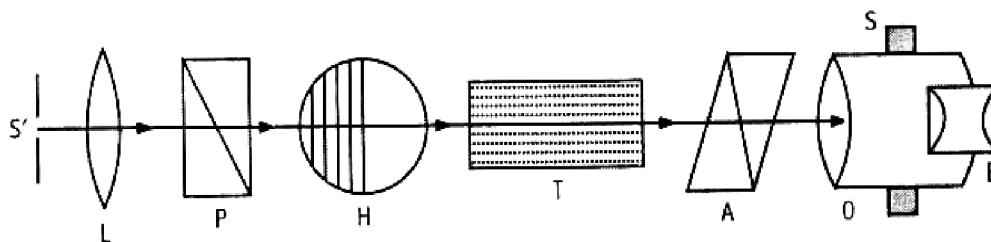


**Fig. 13.15**

The analyzing Nicol $A$, the objective $O$ and an achromatic eye-piece $E$ are kept in a tube which can be rotated about a horizontal axis and this angle of rotation can be recorded by two verniers rotating with the tube over a fixed circular scale $S$ graduated in degrees.

**Working of the apparatus :** The slit $S'$ is illuminated by monochromatic light suitable for half-shade plate. Light from $S'$ will be rendered parallel by the lens $L$ and after being polarized by the Nicol $P$, will pass out of the half-shade plate $H$. This light from $H$ is received by the analyzer $A$ and the telescope in the absence of the tube $T$. The tube containing the analyzeris rotated until the two halves of the half-shade plate are equally bright. The readings $R_1$ of the verniers are noted from the scale $S$.

The tube $T$, containing the active solution is now put in its position when the two halves of the plate $H$ will be unequally bright due to the rotation of the plane of polarization by the solution. The tube containing the analyzeris again rotated either towards right or towards left until the two halves of $H$ again become equally bright.

The readings $R_2$ of the verniers are again noted. Then $\theta = (R_2 \sim R_1)$ will be the angle of rotation of the plane of polarization by the solution.

**Action of half-shade plate :** One half of the half-shade plate is made of glass while the other half is made of quartz whose axis is on the surface and parallel to theline of junction of glass and quartz. Let the directions of vibration of the polarized rays of light on the half-shade plate be parallel to RS (Fig. 13.16). If $\theta$ be the angle which the direction of vibration $OS$ of a ray of light on quartz $Q$, makes with $OX$ (a line perpendicular to the axis of the quartz which is parallel to $YY'$) then the vibration of this ray of light on the quartz half will be resolved into two vibrations. One parallel to the axis of the quartz known as extraordinary ray and another perpendicular to the axis known as ordinary ray. These two rays travel within the quartz with unequal speed and hence a path difference will be created.
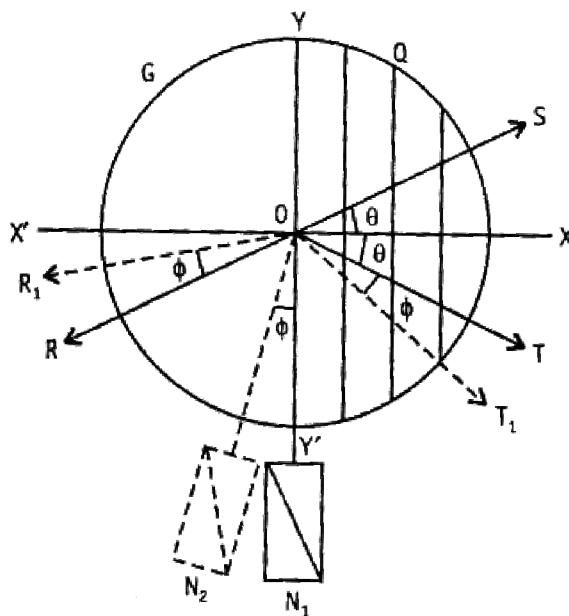


**Fig. 13.16**

The thickness of the quartz is such that a path difference of $\lambda/2$ or a phase difference of $\pi$ will be introduced between the emergent ordinary and extraordinary rays and they will combine to form a linear vibration whose direction will be parallel to $OT$ and inclined with $OX$ by the same angle $\theta$, but on the other side of $OX$. The directions of vibration of the rays of light. incident on the glass half ($G$), will remain unchanged on emergence i.e., the vibrations of the rays of light from the glass half

will remain parallel to $OR$. If now the principal section of the analyzer be kept in position $N_1$ so that it bisects the $\angle ROT$, then the amplitudes of light received by the analyzer at $N_1$ from both the glass half and the quartz half will be equal and both halves will appear equally bright.

If now the tube $T$ containing the active solution be placed in its position, the vibrations parallel to $OR$ and $OT$ from glass half and quartz half respectively will be equally rotated in the same direction by the same angle $\phi$ and the resolved parts of their amplitudes in a direction parallel to the principal section of the analyzer at $N_1$ will be unequal and hence the two halves $G$ and $Q$ will appear unequally bright. When the tube containing the analyzeris rotated by an angle $\phi$ to bring the analyzing Nicol at $N_2$ so that its principal section may bisect the angle between the new directions of $OR_1$ and $OT_1$ the two halves $G$ and $Q$ will again become equally bright.

**Determination of the strength of solution :** The length $l$ of the solution is measured and expressed in decimetre. If the specific rotation $s$ of the substance is known, we can find the mass $m$ of the active substance per c.c. of the solution from the relation $\theta = slm$. If $m$ is known, we can find $s$, the specific rotation of the substance.

## 13.19 Summary

■ Linearly polarized light can be produced by various methods : e.g.,

(a) by allowing an unpolarized light to pass through a Polaroid

(b) by allowing an unpolarized light to fall on a dielectric surface at the Brewster angle $i_p \left( = \tan^{-1} \dfrac{n_2}{n_1} \right)$.

(c) by passing through a Nicol prism

■ If an unpolarized plane wave is incident on a uniaxial crystal, the plane wave will split into two plane waves. One is referred to as the ordinary wave (usually abbreviated as the o-wave) and the other is referred to as the extraordinary wave (usually abbreviated as the $e$-wave). For both waves, the space and time dependence of the vectors **E**, **D**, **B** and **H** can be assumed to be of the form $e^{i(k.r-\omega t)}$ where **k** denotes the propagation vector and represents the direction normal to the phase fronts. In general, **k** vector for the o- and e-waves will be different. Further,

(a) Both ordinary and extraordinary waves are linearly polarized.

(b) $\boldsymbol{D.k} = 0$ for both $o$- and $e$-waves

(c) For the $o$-wave, the **D** vector is at right angles to the optic axis as well as to **k**.

(d) On the other hand, for the e-wave, **D** lies in the plane containing **k** and the optic axis and of course, **D.k** = 0

- Consider a uniaxial crystal whose optic axis lies on the surface of the crystal. If the thickness of the crystal is such that a phase difference of $\pi/2$ is introduced between the ordinary and extraordinary waves, then the plate is said to be a quarter wave plate (usually abbreviated as a QWP). If a linearly polarized wave (with its **E** making 45° with the optic axis) is allowed to fall normally on a QWP, the output is a circularly polarized wave.

- When a linearly polarized light beam propagates through an "optically active" medium like sugar solution then—as the beam propagates—its plane of polarization rotates.

- In a uniaxial crystal, the wave velocities associated with the ordinary and extra ordinary waves are given by

$$v_w = v_{wo} = \frac{c}{n_o} \quad (\, o - \text{wave})$$

$$v_{we}^2 = \frac{c^2}{n_{we}^2} = \frac{c^2}{n_0^2}\cos^2\psi + \frac{c^2}{n_e^2}\sin^2\psi \quad (e - \text{wave})$$

where $\psi$ is the angle that **k** makes with the optic axis; $n_o$ and $n_e$ are constants of the crystal.

- For $E_1$ (incident wave) lying in the x-y plane (which is the plane of incidence)

$$r_\| = \frac{n_2\cos\theta_1 - n_1\cos\theta_2}{n_2\cos\theta_1 + n_1\cos\theta_2} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)}$$

$$t_\| = \frac{2n_1\cos\theta_1}{n_2\cos\theta_1 + n_1\cos\theta_2} = \frac{2\cos\theta_1\sin\theta_2}{\sin(\theta_1 + \theta_2)\cos(\theta_1 + \theta_2)}$$

Notice that $r_\| = 0$ when $\theta_1 + \theta_2 = \frac{\pi}{2}$ implying

$$\theta_1 = i_p = \tan^{-1}\left(\frac{n_2}{n_1}\right)$$

This is the Brewster's angle.

■ For $E_1$ (incident wave) perpendicular to the plane of incidence (i.e., along y-axis)

$$r_\perp = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} = -\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)}$$

$$t_\perp = \frac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} = \frac{2\sin\theta_2 \cos\theta_1}{\sin(\theta_1 + \theta_2)}$$

■ In both cases, if $n_2 < n_1$ and $\theta_1 > \theta_c = \sin^{-1}(n_2/n_1)$ we have total internal reflection. We can still use the above expressions for $r_\parallel$, $t_\parallel$, $r_\perp$ and $t_\perp$ but we must remember that

$\sin\theta_2 = n_1/n_2 \sin\theta_1$ will be greater than 1.

## 13.20    Review Questions and Answers

1. **What is the difference between negative and positive crystal?**

**Ans. Negative crystal** The wavefronts surrounding a point source $S$ in such a crystal are shown in Fig. 13.17(a). It is clear that

(i) ordinary wave surface lies inside the extra-ordinary wave surface.

(ii) Velocity of o-ray is constant in all direction.

(iii) The velocity of e-ray varies with the direction.It is minimum and equal to the velocity of the ordinary ray along the optic axis. It is *maximum* in a direction perpendicular to the direction of the optic axis.

(iv) The refractive index for the *e*-ray is less than the refractive index for the *o*-ray i.e. , $n_e < n_o$.
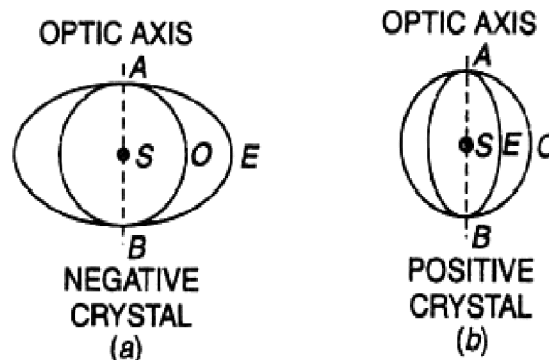


Fig. 13.17

Positive crystal : The wavefronts surrounding a point source $S$ in such a crystal are shown in Fig. 13.17(b). It is clear that

(i) The ordinary wave surface lies outside the extraordinary wave-surface.

(ii) The velocity of o-ray is constant in all directions.

(iii) The velocity of $e$-ray varies with the direction. It is *maximum* and equal to the velocity of the o-ray along the optic axis. It has a minimum value in a direction perpendicular to the optic axis.

(iv) The refractive index for the e-ray is more than the refractive index for the $o$-ray i.e., $n_e > n_o$.

2. **State and explain Malus law.**

**Ans.** Statement of Malus law :It states that the intensity of the polarized light transmitted through the analyzer varies as the square of the cosine of the angle between the plane of transmission of the analyzer and the plane of polarized.

For explanation, see Article 13.8.

3. **Give theory and construction of a half-wave plate and a quarter-wave plate.**

**Ans.** See Article 13.9.

4. **What is the nature of emergent polarized light when circularly polarized light is passed through (i) a quarter-wave plate (ii) a half-wave plate.**

**Ans.** (i) When circularly polarized light is passed through a *quarter-wave plate* such that the vibrations make an angle of 45° with the direction of optic axis, the outgoing light will be plane polarized as the phase difference between the $e$-component and o-component becomes $\pi/2 + \pi/2 = \pi$.

(ii) When left-handed circularly polarized light is passed through a *half-wave plate* such that the vibrations make an angle of 45° with the direction of optic axis, the outgoing light is right-handed circularly polarized light as $\delta = \pi/2 + \pi = 3\pi/2$.

5. **Plane polarized light is incident normally on a quarter-wave plate. Determine the nature of emergent light in each case when the vibrations in the incident light make angles of 0°, 30°,45° and 90° with the axis of the plate.**

**Ans.** When plane polarized vibration of amplitude $A$ is incident normally on a quarter-wave plate, it is split up into the $e$ and $o$ vibrations. If $\theta$ is the angle

which the incident vibration makes with the direction of optic axis, then on emergence the resultant vibration is given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

where $a = A \cos \theta$ and $b = A \sin \theta$

(i) When $\theta = 0°$, $a = A$ and $b = 0$

This shows that the amplitude of the vibration perpendicular to the optic axis is zero which therefore vanishes. The emergent light is *plane polarized* having vibrations in a direction parallel to the optic axis.

(ii) When $\theta = 30°$, $a = A \cos 30°$ and $b = A \sin 30°$ i.e., $a$ is not equal to $b$. The resultant vibration is given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

This shows that the resultant light is elliptically polarized, the major and the minor axes of the ellipse coincide with the co-ordinate axes.

(iii) When $\theta = 45°$, $a = A \cos 45°$ and $b = A \sin 45°$, then $a = b$ and we have
$$x^2 + y^2 = a^2$$

This shows that the emergent light is circularly polarized.

(iv) When $\theta = 90°$, $a = 0$ and $b = A$

The amplitude of the vibration along the optic axis vanishes. The outgoing light, therefore, is plane polarized having vibrations perpendicular to the optic axis.

6. **In a uniaxial crystal, the wave velocities associated with the ordinary and extra-ordinary waves are given by**

$$v_w = v_{wo} = \frac{c}{n_o} \quad (o - \text{wave})$$

$$v_{we}^2 = \frac{c^2}{n_{we}^2} - \frac{c^2}{n_0^2} \cos^2 \psi + \frac{c^2}{n_e^2} \sin^2 \psi \quad (e - \text{wave})$$

where $\psi$ is the angle that **k** makes with the optic axis; $n_o$ and $n_e$ are constants of the crystal.

Ans. See Article 13.6.

7. **Deduce Fresnel's laws of reflection and refraction from EM theory of light. Show that Brewster's law of polarization is a direct consequence of Fresnel's laws of reflection.**

**Ans.** See Article 13.4.

## 13.21　Illustrated Examples

**Example 1 :** By Brewster's law show that light incident on a transparent substance at polarizing angle gives reflected and refracted rays at right angle to each other.

**Solution :** According to Brewster's law $n = \tan i = \sin i / \cos i$

According to Snell's law $n = \sin i / \sin r$

Comparing above two equations, we get

$$\cos i = \sin r = \cos(90° - r)$$

$$i = 90° - r$$

$$\therefore \quad i + r = 90°$$

Angle between reflected and refracted $= 180° - (i + r) = 180° - 90° = 90°$.

Hence at the polarizing angle the reflected and refracted beams are at right angle to each other.

**Example 2 :** If the refractive indices for the ordinary ray in the case of calcite and Canada balsam are 1.658 and 1.550 respectively, calculate the maximum possible inclination with the Canada balsam surface so that the ordinary ray is still quenched.

**Solution :** Here $n_o = 1.658$, $n_b = 1.550$

If C is the critical angle when light travels from calcite to balsam layer, then

$$\sin C = \frac{1.550}{1.658} = 0.9348$$

$$\therefore \quad C = 69.2°$$

Thus, the $o$-ray incident on the balsam layer at an angle less than 69.2° will not be quenched. Hence maximum angle with the balsam layer for the $o$-ray to be quenched

$$= 90° - 69.2° = 20.8°.$$

**Example 3 :** Two polaroids are adjusted so as to obtain maximum intensity. Through what angle should one polaroid be rotated to reduce the intensity to (i) half and (ii) one fourth.

**Solution :** According to Malus law

$$I_1 = I_o \cos^2\theta$$

or,      $I_1/I_0 = \cos^2\theta$

Now      (i) $I_1/I_0 = 1/2$

∴      $\cos^2\theta = 0.5$   or  $\cos\theta = 0.7071$

Hence      $\theta = \mathbf{45°}$

Now      (ii) $I_1/I_0 = 1/4 = 0.25$

∴      $\cos^2\theta = 0.25$   or  $\cos\theta = 0.5$

Hence      $\theta = \mathbf{60°}$

**Example 4 :** Calculate the thickness of a quarter-wave plate for light of wavelength 5893 Å, given refractive indices for ordinary ray and extra-ordinary ray 1.544 and 1.553 respectively.

**Solution :** Here $n_o = 1.544$, $n_e = 1.553$ and $\lambda = 5893$ Å $= 5893 \times 10^{-10}$ m

As $\mu_e > \mu_0$, the quarter-wave plate is made from a uniaxial positive crystal, the thickness $d$ of which is given by

$$d = \frac{\lambda}{4(n_e - n_0)} = \frac{5893 \times 10^{-10}}{4(1.553 - 1.544)}$$

$$= \frac{5893 \times 10^{-10}}{4 \times 0.009} = \mathbf{0.1637 \times 10^{-4}} \; \boldsymbol{m}$$

**Example 5 :** Calculate the thickness of a half-wave plate for light of wavelength 5000 Å, given refractive indices for ordinary ray and extra-ordinary ray 1.544 and 1.553 respectively.

**Solution :** Here $n_o = 1.544$, $n_e = 1.533$ and $\lambda = 5000$ Å $= 5000 \times 10^{-10} \; m$

$$d = \frac{\lambda}{2(n_e - n_0)} = \frac{5000 \times 10^{-10}}{2(1.544 - 1.533)} = \mathbf{0.227 \times 10^{-4}} \; \boldsymbol{m}$$

**Example 6 :** The adjoining diagram shows the section of an extraordinary wave surface in a positive uniaxial crystal with optic axis along OX. Show that the ray refractive index of the medium at an angle θ to the optic axis is given by

$$n_\theta = \sqrt{n_o^2 \cos^2\theta + n_e^2 \sin^2\theta}$$

where $n_o$ and $n_e$ are respectively the refractive index for the o-ray and the principal refractive index of e-ray.
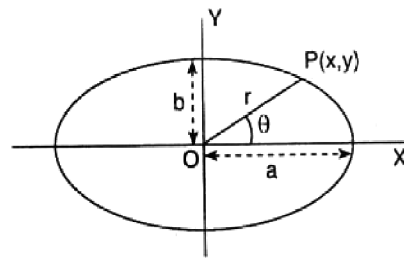
**Solution :** The equation of the wave surface is given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

But $x = r\cos\theta$, $y = r\sin\theta$, $b = v_e t$, $v_e$ being the wave velocity for e-ray along OY (perpendicular to optic axis) and $r = OP = v_\theta t$, where $v_\theta$ is the ray velocity of e-ray along OP. Again, since the wave surface for o-ray is spherical, touching the wave surface for e-ray along optic axis OX, $a = v_o t$, where $v_o$ is the wave velocity of o-ray, being the same in every direction.

Thus, the above equation may be written as

$$\frac{v_\theta^2 t^2 \cos^2\theta}{v_o^2 t^2} + \frac{v_\theta^2 t^2 \sin^2\theta}{v_e^2 t^2} = 1$$

or, $$\frac{\cos^2\theta}{v_o^2} + \frac{\sin^2\theta}{v_e^2} = 1$$

or, $$\frac{c^2\cos^2\theta}{v_o^2} + \frac{c^2\sin^2\theta}{v_e^2} = \frac{c^2}{v_\theta^2}$$

or, $$n_o^2 \cos^2\theta + n_e^2 \sin^2\theta = n_\theta^2$$

$$n_\theta = \sqrt{n_o^2 \cos^2\theta + n_e^2 \sin^2\theta}$$

where $n_\theta = c/v_\theta$, the ray refractive index for e-ray at an angle θ to the optic axis.

**Example 7 :** The rotation in the plane of polarization ($\lambda = 5893\text{Å}$) in a certain substance is 10° per cm. Calculate the difference between the refractive indices for right-handed and left-handed circularly polarized light in the substance.

**Solution :** Let $t$ cm be the thickness of the substance and $n_R$ and $n_L$ the indices for right and left circularly polarized light in it. Then the rotation $\theta$ in the plane of polarizationis given by

$$\theta = \frac{\pi}{\lambda}\left(n_R \sim n_L\right)t \quad \Rightarrow n_R \sim n_L = \frac{\theta}{t}\cdot\frac{\lambda}{\pi}$$

Here $\theta/t = (10 \times 2\pi/360)$ $rad/cm$ and $\lambda = 5893$ Å $= 5893 \times 10^{-8}$ $cm$

$$\therefore \quad n_R \sim n_L = \frac{10 \times 2\pi}{360} \times \frac{5893 \times 10^{-8}}{3.142} = \mathbf{4.09 \times 10^{-9}}$$

**Example 8 :** A certain length of 5% solution causes an optical rotation of 20°. How much length of 10% solution of the same substance will cause rotation of 35° ?

**Solution :** Let $S$ be the specific rotation of the solution, $l_1$ the length of 5% solution causing 20° rotation and $l_2$ the length of 10% solution causing 35° rotation.

$$\therefore \quad S = \frac{\theta_1}{l_1.c_1} = \frac{\theta_2}{l_2.c_2} \quad \Rightarrow \frac{20°}{l_1 \times 5\%} = \frac{35°}{l_2 \times 10\%}$$

$$\therefore \quad l_2 = \frac{5 \times 35}{10 \times 20}l_1 = \frac{\mathbf{7}}{\mathbf{8}}l_1$$

**Example 9 :** A 20 cm length of a certain solution causes right-handed rotation of 38° and 30 cm of another solution causes left-handed rotation of 24°. What optical rotation will be caused by 40 cm length of a mixture of the above solutions in the ratio 1:2 by volume?

**Solution :** Optical rotation, $\theta \propto l$, length of substance traversed for a given $\lambda$.

Rotation by 20 cm of sol. 1 = +38° (right-handed)

Rotation by 30 cm of sol. 2 = –24° (left-handed)

$$\Rightarrow \text{ Rotation by 40 cm of sol. 2 } = -24° \times \frac{40}{30} = -32°$$

Since rotation of a mixture of active substances is the algebraic sum of individual rotations, we write as

Rotation by (20 cm of 1 + 40 cm of 2) = + 38° – 32° = 6° i.e., mixture in the ratio 1 : 2

$\therefore$ rotation by 30 cm of the mixture = **+3° (right-handed)**.

# Unit-14 : Optical Fibres

## Structure

## 14.1 Objectives

In going through this unit, you will learn :

1. Propagation of light in fibre
2. Numerical Aperture
3. Step and Graded Indices
4. Concepts of Single and Multiple Mode Fibres

## 14.2 Introduction

In recent times transparent dielectric fibres have been developed to carry efficiently optical signals from one place to another. These are known as optical fibres. The study of the properties of such optical fibres is known as fibre optics. Prior to the introduction of fibre optics, EM waves down to microwave region were used for the purpose of communication. But since the amount of information carried by EM waves is proportional to its frequency, only about 100 pulses per second at best could be transmitted by a wave of frequency 100. Now, by using laser light through fibre,

347

million pulses per second can be transmitted. Moreover, because of low cost, high reliability and extremely low losses optical fibres are gradually being used more and more.

The diameter of optical fibres is usually large compared to the wavelength of optical signals. The wave nature of light may thus be neglected to study the properties of optical fibres and the known laws of geometrical optics may be applied. In fibres of thin variety, the diameter of the fibre becomes of the order of the wavelength and the wave nature of light cannot be neglected. In such cases, light propagates as EM wave along the fibre, similar tomicrowave propagation along a wave guide.

## 14.3 Basic principle of fibre optics

An optical fibre is a very thin cylindrical fibre of glass or any transparent dielectric medium such as plastic, having a radius ~ a micrometer ($10^{-6}$ m). A bundle of such thin fibres forms what is called **light pipe**, consisting of many thousands of very long fine-quality glass or quartz fibres. the fibres are coated with a layer of material of lower refractive index.



Coating (n = 1.5)

Fiber material
(n = 1.7)
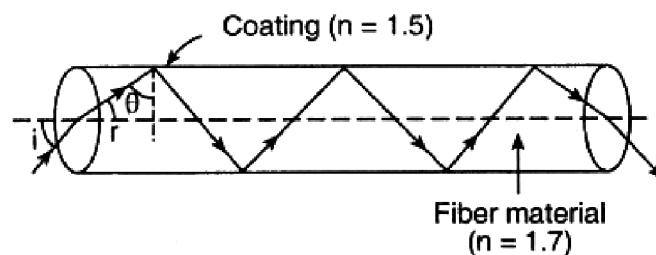
**Fig. 14.1**

*Total internal reflection* is the basic principle of optical fibres. When light is incident at one end of the fibre wall, it makes a nearly glancing incidence because of the small radius of the fibre. Let i be the angle of incidence of the ray with the axis (Fig. 14.1) and $r$ the angle of refraction. If $\theta$ be the angle at which the ray is incident on the fibre boundary, then $\theta = 90° - r$. Let $n_1$ be the refractive index of the fibre. If $\theta > \theta_C$, the critical angle, where $\theta_C = \sin(-1)1/n_1$, then the ray will undergo total internal reflection. The ray is thus transmitted along the fibre by repeated internal reflections until its emergence by the other end. Even if the fibre is bent in a complicated shape, the transmission occurs without any appreciable loss.

Thus, the light ray is guided through the fibre from end to the other without any energy being lost due to refraction.

It is important to note that there must be very little *absorption* of light during transit over a long distance inside the fibre. This is achieved by *purification* and *special preparation* of the material used.

## 14.4 Structure and classification

**Structure :** Structurally, an optical fibre consists of an inner cylinder usually made of glass and is known as the core. It is the core that carries the light. The dia of core varies from about 5μm to 100 μm.

The core is surrounded by another coaxial cylindrical shell of refractive index lower than that of the core. This is called the **cladding** whose diameter is usually about 125 μm. The function of cladding is to retain the light within the core through total internal reflections (Fig. 14.2).
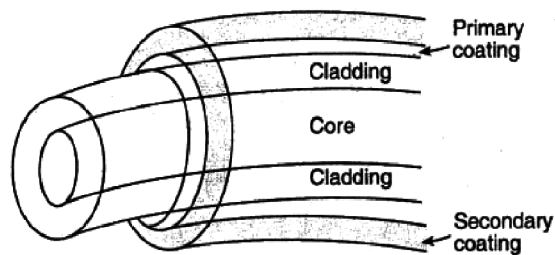


**Fig. 14.2**

To provide greater strength and protection, a layer of soft plastic coating of dia about 250 μm surrounds the cladding. This coating is the *primary coating*. A second layer of coating, called secondary coating, is often used outside the primary.

**Classification :** There are mainly two types of optical fibres, based on refractive index profile of the core, namely (i) **step-index fibre** and (ii) **graded-index fibre**.

### 14.4.1 Step-index fibre

In step-index fibre, the refractive index $n_1$ of the core is uniform throughout the core. Similarly, the refractive index $n_2$ of the cladding is also uniform. Necessarily, $n_1 > n_2$. Let a and $b$ be the radius of the core and cladding respectively so that $b > a$. The index profile and the radii of the core and the cladding of a step-index fibre is illustrated in Fig. 14.3.
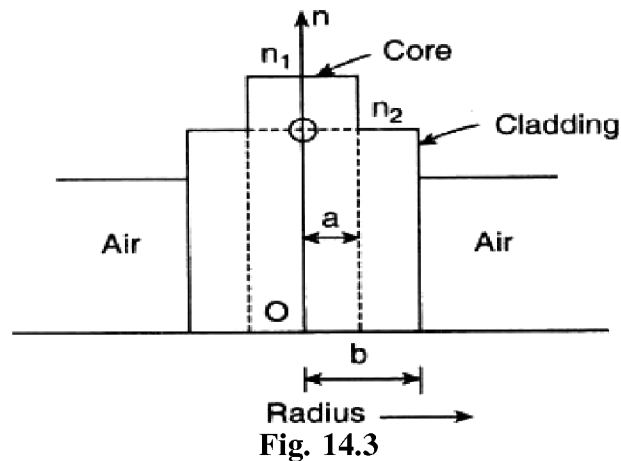
Fig. 14.3

Depending on the launch angle of the incident light into the fibre, there may a large Limber of different zig-zag paths of the rays or modes by which energy travels down the core. In Fig. 14.4, we have restricted to two rays only, entering at different angles of incidence with the fibre-axis and emerging out of the fibre at different times. This is then called a **step-index multimode** fibre. The rays travelling along different paths take different times to reach the other end of the fibre leading to a widening of input pulse as it traverses the fibre. This is called **intermodal dispersion** that distorts the signal to be transmitted.
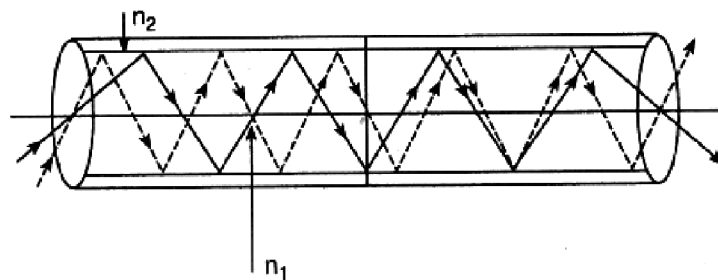


Fig. 14.4

### 14.4.2 Graded-index fibre

In a graded-index optical fibre, unlike the step-index one, the core has a non-uniform refractive index that gradually decreases from the centre towards the core-cladding interface. The cladding however has a uniform index The refractive index profile of a graded-index fibre is illustrated in Fig. 14.5 and the paths of the rays in Fig. 14.6.
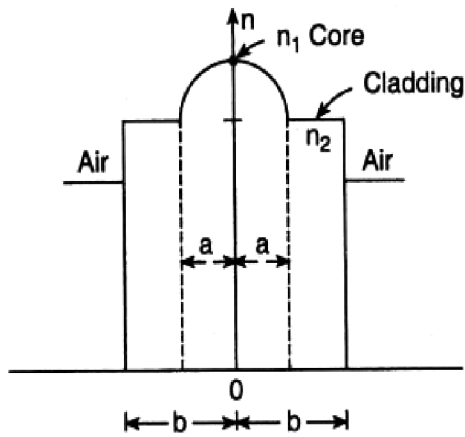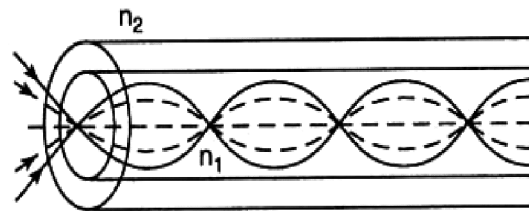
Fig. 14.5                                                Fig. 14.6

Here in multimode graded-index fibre, the ray-paths are not zig-zag; they are refracted gradually. At the central region, the index is higher implying that the light velocity is smaller; but near the cladding the situation is reverse, the index is smaller, and the light velocity is greater. The rays thus spiral around near the axis, take shorter paths, and travel slowly. But the rays spiraling around the vicinity of cladding take longer paths and travel faster. As a result, all the rays tend to recross the axis at almost the same instant as shown in Fig. 14.6.

### 14.4.3 Single mode fibre

There is another type of optical fibres, called *single or monomode optical fibres* having *narrow* cores (~ 10 μm) which *allow only one mode to propagate*, as shown in Fig. 14.7. Here the rays travel only parallel to the axis and the problem of intermodal dispersion is practically absent.
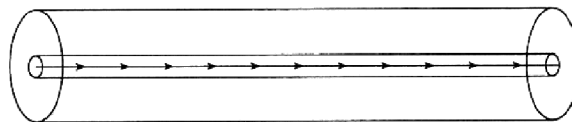


Fig. 14.7

### 14.4.4 Comparison among different fibres

Of the different types of optical fibres, the *multimode step-index* fibre is the oldest. Its advantages are that it is least expensive, rugged and can be infused easily with

light. But its *serious drawback* is its intermodal dispersion and is unsuitable for long distance communication.

The *multimode graded-index* fibre is medium-priced and intermodal dispersion is not serious. It is thus widely used in medium distance communication.

The *monomode* fibres are relatively expensive and require laser sources. There is no intermodal dispersion. They are most suitable for long distance communication.

**Note :** (1) If many uncladded fibres are packed to form a bundle, then light may leak through from one fibre to the other. This is termed **cross-talk**. In modern fibres, therefore, cladding is used. Cross-talk may also occur if a fibre is bent too sharply for then the angle of incidence at core-cladding interface may be less than the critical angle and a part of light may escape.

(2) If fibres are properly aligned in a bundle, it is called a **coherent** bundle; if not, the bundle is an **incoherent** one.

(3) The **reason for using glass** for optical fibres are : (i) It has a range of accessible temperatures where its viscosity is variable; it does not solidify at a fixed freezing point but gradually becomes harder and harder and in the transition region can be easily drawn into a thin fibre, (ii) Pure silica is characterized by extremely low 'loss' e.g., in commercial silica fibres 96% of power gets transmitted through 1 km of optical fibre, (iii) Its intrinsic strength is as high as $2 \times 10^{16}$ lb/inch$^2$. Further, silica is very cheap and abundantly available.

## 14.5 Acceptance angle and Numerical aperture

We consider a step-index optical fibre. Let $n_1$, $n_2$ be the indices of the core and cladding respectively and $n_0$ be that of the outside medium.

Let a ray be incident on the entrance aperture of the fibre at an angle $\theta_1$ with its axis and $\theta_2$ be the corresponding angle of refraction (Fig. 14.8).
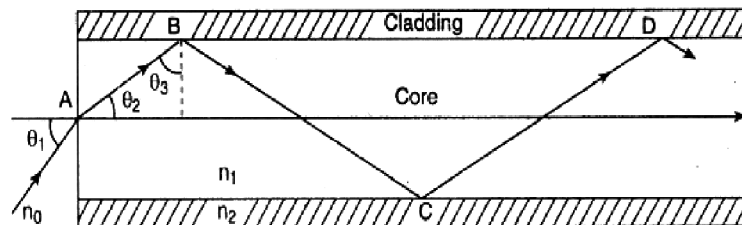


**Fig. 14.8**

$$\therefore \quad n_0 \sin \theta_1 = n_1 \sin \theta_2 \qquad \qquad ... (14.1)$$

To retain the ray inside the core, the angle of incidence $\theta_3$ at the core-cladding interface must not be less than the critical angle $\theta_C$.

But $\theta_C = 90° - \theta_2$        ... (14.2)

Now, if $\theta_1$ is increased, $\theta_2$ also increases and so $\theta_3$ decreases. There is thus a maximum value of $\theta_1$ for which $\theta_3$ is not less than $\theta_C$ and internal reflection at core-cladding interface occurs. This angle is called the *acceptance angle*, $\theta_A$ or $\theta_{max}$.

The **acceptance angle** is thus *defined as the maximum value of launch angle incident light into the fibre for which the ray is totally internally reflected at core-cladding interface and gets transmitted without loss.*

The acceptance angle is a measure of the light gathering power of the fibre.

A cone of light of semi-angle $\theta_A$ is called the **acceptance cone**. *Maximum acceptance angle* is $2 \times \theta_A$.

**Determination of $\theta_A$ :** We have from (14.1) and (14.2),

$$n_0 \sin \theta_1 = n_1 \sin \theta_2 = n_1 \sin(90 - \theta_3) = n_1 \cos \theta_3$$

$$\Rightarrow n_0 \sin \theta_A = n_1 \cos \theta_C \qquad ... (14.3)$$

for $\theta_1 = (\theta_1)_{max} = \theta_A$ and $\theta_3 = \theta_C$, the critical angle.

$$\therefore \quad n_0 \sin \theta_A = n_1 \sqrt{1 - \sin^2 \theta_C} = n_1 \sqrt{1 - n_2^2/n_1^2} = \sqrt{n_1^2 - n_2^2} \qquad ... (14.4)$$

Since, $\sin \theta_C = n_2 / n_1$.

The quantity $n_0 \sin \theta_A$ or $\sin \theta_A$ (when the outside medium is *air*) is defined as the **numerical aperture** (*NA*) of the fibre. *It is a measure of the light gathering power of the fibre*. If *NA* is large, it becomes easier to launch power into the fibre. Since for all practical purposes $\left(n_1^2 - n_2^2\right) > 1$, we can define the numerical aperture as

$$NA = \sqrt{n_1^2 - n_2^2} \qquad ... (14.5)$$

### 14.5.1 Fractional index change

Another important parameter, in this context, is the *fractional refractive index change*.

It is *defined* as the fractional difference $\Delta$ between the indices of the core and the cladding of the fibre.

$$\therefore \quad \Delta = \frac{n_1 - n_2}{n_1} \qquad ... (14.6)$$

The parameter $\Delta$ is always positive since $n_1$ is always greater than $n_2$ to satisfy one of the conditions for total internal reflection. For guiding the light rayseffectively through a fibre $\Delta \ll 1$ and typically $\Delta \sim 0.01$.

**Relation between $\Delta$ and *NA* :** We have from (14.5)

$$NA = \sqrt{n_1^2 - n_2^2}$$

Now, $\quad n_1^2 - n_2^2 = (n_1 + n_2)(n_1 - n_2)$

$$= \frac{1}{2}(n_1 + n_2)\frac{(n_1 - n_2)}{n_1}2n_1$$

$$\simeq (n_1)\Delta 2n_1 \left[\because \frac{1}{2}(n_1 + n_2) \simeq n_1\right]$$

$$= 2n_1^2\Delta$$

$$\therefore \quad NA = \sqrt{2n_1^2\Delta} = n_1\sqrt{2\Delta} \qquad\qquad \dots (14.7)$$

## 14.6 Summary

- We have explaind optical fibre and its working principle and described the stucture of a typical optical fibre giving the necessary diagrams.

- We have discured the step-index and graded-index optical fibre with necessary diagram.

- We have mentioned different classes of fibres and discured their relative merits and demerits.

- Relation between the numerical apearture (*NA*) of an optical fibre and fractional index change $\Delta$ is $NA = n_1\sqrt{2\Delta}$, where $n_1$ is the refractive index of the core.

## 14.7 Review Questions and Answers

1. **What is an optical fibre? Explain the working principle of optical fibre as a waveguide for light.**

**Ans.** See Articles 14.2 and 14.3.

2.  **Describe the structure of a typical optical fibre giving the necessary diagrams.**

**Ans.** See Article 14.4.

3.  **Describe, with necessary diagram, the step-index and graded-index optical fibre.**

**Ans.** See Articles 14.4.1 and 14.4.2.

4.  **What are meant by acceptance angle and numerical aperture ($NA$) of a fibre. Find their expressions and significance for a step-index fibre.**

**Ans.** See Article 14.5.

5.  **What are the different classes of fibres? Discuss their relative merits and demerits.**

**Ans.** See Article 14.4.4.

6.  **Give the expression for the numerical aperture ($NA$) of an optical fibre and show that it is related to fractional index change $\Delta$, by the equation**

$$NA = n_1\sqrt{2\Delta}$$

**Ans.** See Article 14.5.1.

## 14.8 Illustrated Examples

**Example 1 :** Find the numerical aperture ($NA$) of a step-index fibre when the refractive index of the core is 1.51 and that of the material used for cladding is 1.47.

**Solution :** Here, $n_1 = 1.51$, $n_2 = 1.47$

$$NA = \sqrt{n_1^2 - n_2^2} = \sqrt{1.51^2 - 1.47^2}$$

$$= \sqrt{(1.51 + 1.47)(1.51 - 1.47)}$$

$$= \sqrt{2.98 \times 0.04} = \mathbf{0.3452}$$

**Example 2 :** Calculate the acceptance angle of the fibre referred to in Ex. 1 when it is placed in air.

**Solution :** Let $\theta_A$ be the required acceptance angle. Then we have

$$n_0 \sin\theta_A = \sqrt{n_1^2 - n_2^2} = NA = 0.3452 \quad (\text{Ex.1}) \text{ and } n_0 = 1 \text{ (air)}.$$

So, we obtain

$$\sin \theta_A = 0.3452 \Rightarrow \theta_A = \sin^{-1}(0.3452) = \mathbf{20.19°}$$

**Example 3 :** The numerical aperture (*NA*) of an optical fibre is 0.5 and the core refractive index is 1.54. Find the refractive index of the cladding

**Solution :** Refractive index of cladding is

$$n_2 = \sqrt{n_1^2 - (NA)^2} = \sqrt{11.54^2 - 0.5^2} = \mathbf{1.46}$$

**Example 4 :** In a graded-index fibre the radial distribution of index is given by

$$n(r) = 1.52 - 2r^2$$

With *r* in mm up to a radius $r_0 = 0.2$ mm. Calculate the acceptance angle of the fibre.

**Solution :** Here, $n_1 = 1.52$ and $n_2 = 1.52 - 2 \times (0.2)^2 = 1.52 - 0.08 = 1.44$

∴ Acceptance angle,

$$\theta_A = \sin^{-1} \sqrt{n_1^2 - n_2^2}$$

$$= \sin^{-1} \sqrt{1.52^2 - 1.44^2}$$

$$= \sin^{-1} \sqrt{2.96 \times 0.08} = \mathbf{29.12°}$$

**Example 5 :** For a typical step-index (multimode) fibre with $n_1 = 1.45$ and $\Delta = 0.01$. Calculate acceptance angle.

**Solution :** $\sin \theta_A = NA = n_1 \sqrt{2\Delta} = 1.45 \times \sqrt{2 \times 0.01} = 0.205 = \sin 12°$

$$\therefore \quad \theta_A = \mathbf{12°}$$

# REFERENCES AND FURTHER READINGS

## Waves and Vibrations

1. The Phytsics of Vibrations and Waves, Sixth Edition, . J. Pain, John Wiley 7 Sons.

2. Waves & Oscillation by Pramod Kumar L. P. Verma and G. N. Lohani, Published from S. J. Publication.

3. Waves & Oscillation by J. C. Upadhyay, Kedar Nath Publication.

4. Elementary book on waves & Oscillations, IGNOU, New Delhi.

5. Objective Physics, Satya Prakash, A S prakashan, Meerut.

6. Fundamentals of Physics, David Halliday, Robert Resnick, Jearl Walker, Hohn Wiley & Sons.

7. Concepts of Physics, Part-I, HC Verma, Bharati Bharati Bhawan, Patna.

8. Fundamental of waves and Oscillations, K. U. Ingard, Cambridge Univ. Press.

9. Waves & Oscillations by R. B. Singh, Pragati Prakashan.

10. The physics of Waves & Oscillations, N. K. Bajaj, Tata Mc Graw Hill.

11. Principles of Accoustics, B. Ghosh, Sreedhar publishers. Kolkata.


## Waves and Optics

1. A.B. Gupta, Modern Optics, Books and Allied (P) Ltd., Kolkata, 2008.

2. Ajoy Ghatak, Optics, 4th Edition, McGraw-Hill Co., New Delhi, 2005.

3. B. Ghosh and K.G. Mazumdar, A Textbook on Light, Sheedhar Publishers, Kolkata, 2003.

4. E. Hecht, Optics, 5th Edition, Pearson Education, London, 2017.

5. F.A. Jenkins and H.E. White, Fundamentals of Optics, 4th Edition, McGraw-Hill, 2001.

6. M. Born and E. Wolf, Principle of Optics, 7th Edition, Cambridge University Press, Cambridge, 2019.

7. R.K. Kar, Optics (Classical & Quantum), Books and Allied (P) Ltd., Kolkata, 2008.

8. R.S. Longhurst, Geometrical and Physical Optics, 3rd Edition, Longman Group Ltd., London, 1973.

# NOTE

..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................
..........................................................................................................................................

# NOTE

..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................
..............................................................................................................................

# NOTE