Unit I

Lesson 1

Electromagnetic theory-1

Objective: To know about the laws of reflection, refraction and fresnel formulae with the help of electromagnetic theory.

Structure:

1.1.1 Reflection and refraction of electromagnetic waves

According to the electromagnetic theory of light, light consists of electromagnetic waves, propagated according to Maxwell's equations. We shall now consider the electromagnetic theory of the phenomena of reflection and refraction of monochromatic light wave at an infinite plane boundary between two semi-infinite media of different indices of refraction, for example a free space and a dielectric or two dielectrics or free space and a metallic conductor. We shall see that at the boundary between two dielectrics the electromagnetic wave also obeys the familiar laws of reflection and refraction which were established on the basis of the wave theory of light much before the electromagnetic theory. To discuss the behaviour of electromagnetic waves at the

boundary, we must first find the boundary conditions which the electric and magnetic fields must satisfy at the surface of discontinuity between the two media.

1.1.2. Boundary Conditions:

The electric and magnetic fields of the waves in the two media satisfy the following boundary conditions at an interface separating the two dielectric media:-

(i) The normal component of the magnetic induction B is continuous across a boundary. (A quantity is said to be continuous when its values at two neighbouring points, one on each side of the boundary, are equal.)

(ii) The normal component of electric displacement D is continuous across an uncharged boundary.

(iii) The tangential component of E is always continuous across a boundary.

(iv) The tangential component of H is continuous across a boundary that carries no surface current.

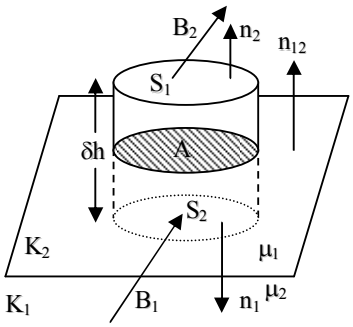


Fig.1. Derivation of the boundary condition for the normal component of **B**

These boundary conditions are deduced from Maxwell's equations as follows:-

(a) The magnetic induction B satisfies the second Maxwell's equation

Div **B** =0 …… (1.1.1a)

At the interface between two media we draw a small shallow pillbox like volume that encloses a portion of the boundary as shown in Fig.1. We apply Gauss divergence theorem to the integral of div B taken throughout the volume of the pillbox and obtain from Eq.(1.1.1a)

$$\int_v div\, B\, dv = \oint B.n\, ds = 0 \ldots\ldots\ldots (1.1.1b)$$

Applying Eq.(1.1.2) to the whole surface of pillbox, we get

$\int_{s_1} \mathbf{B_1}.\mathbf{n_1} ds + \int_{s_2} B_2.n_2 dS$+contribution from walls=0…….(1.1.1c)

If B is finite everywhere, then making the height of pillbox approaches zero, always keeping the boundary surface between its two flat faces, the last term in Eq. (1.1.1c) vanishes provided there is no surface flux of magnetic induction. $S_1$and S_2 approach the shaded area A. We therefore obtain in the limit as $\delta h \rightarrow O$

$$\int_A (\mathbf{B_1}.n_1 + \mathbf{B_2}.\mathbf{n_2}) dA = 0$$

Since the area A is quite arbitrary, the above equation will be true when

$\mathbf{B_1}.\mathbf{n_1} = -\mathbf{B_2}.\mathbf{n_2}$……..(1.1.1d)

If n_{12} is the unit normal vector pointing from the first into the second medium, then $n_1=-n_{12}$ and $n_2=n_{12}$and Eq. (1.1.1d) gives

$\mathbf{B_1}.\mathbf{n_{12}}=\mathbf{B_2}.\mathbf{n_{12}}$

$B_{1n}=B_{2n}$……(1.1.1e)

i.e. *the normal component of magnetic induction is continuous across the boundary.*

(b) The boundary condition for the electric displacement D can be obtained from Maxwell's equation in a similar way, but it will be slightly different if charges are present. Instead of Eq. (1.1.1c), we get in the present case

$$\int_{S^1} \mathbf{D_1}.n_1 dS + \int_{S^2} \mathbf{D_2}.\mathbf{n}_2 dS + contribution from walls = 4\pi \int_v \rho dV$$

As before, as $\delta h \rightarrow o$, the contribution from walls tends to zero. Instead of volume charge density ρ, the concept of surface charge density ρ_s defined by

$$\lim_{h\rightarrow 0} \int_v \rho dV = \int_A \rho_S dA$$

must be used in the preceeding equation, which now in the limit as $\delta h \rightarrow o$ becomes

$$\int_A (\mathbf{D_1}.\mathbf{n_1} + \mathbf{D_2}.\mathbf{n_2}) dA = 4\pi \int_A \rho_s dA \text{ …….(1.1.1f)}$$

Finally, we get

$\mathbf{n}_{12}.(\mathbf{D_2}-\mathbf{D_1})=4\pi\rho_S$……(1.1.1g)

i.e., *the normal component of electric displacement changes abruptly by an amount* $4\pi\rho_s$ *across the surface of charge density* ρ_s.

$D_{1n}=D_{2n}$……(1.1.1h)

i.e., the normal component of electric displacement is continuous across the boundary.

(c)The boundary condition which the tangential component of electric field must satisfy can be obtained from Maxwell's equation.

Curl $\mathbf{E}=-\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}$ ……(1.1.1i)

Integrating this equation over the surface bounded by rectangular loop A B C D shown in Fig.2 gives

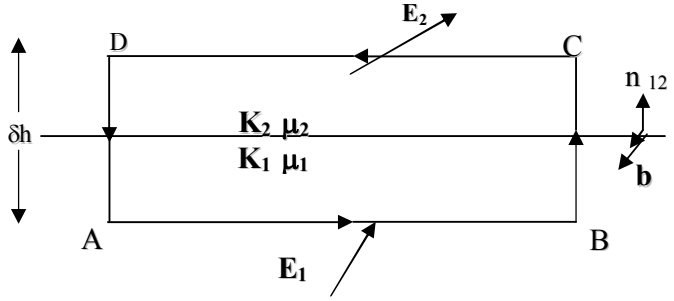


Fig. 2. Derivation of the boundary condition for the tangential component of **E.**

$$\int_S curl\mathbf{E}.\mathbf{b}\,ds = -\frac{1}{c}\int_S \frac{\partial \mathbf{B}}{\partial t}.\mathbf{b}\,d\mathbf{S}$$

where b is a unit vector perpendicular to the plane of rectangle. Applying the *stoke's theorem* to the integral on the left hand side we have

$$\oint \mathbf{E}.\mathbf{d_l} = -\frac{1}{c}\int_s \frac{\partial \mathbf{B}}{\partial \mathbf{t}}.\mathbf{b}dS$$

where the line integral is taken over the boundary A B C D . If the lengths AB and CD are small, E may be replaced by constant values E_1 and E_2 along each of these sides. The line integral along A B C D can be now easily written out.

lE_{1t}-lE_{2t}+contribution from sides BC and AD

$$=-\frac{1}{c}\int_s \frac{\partial \mathbf{B}}{\partial \mathbf{t}}.\mathbf{bdS}\ldots\ldots(1.1.1j)$$

Where $E_{1t}$and E_{2t} are the components of E tangent to the surface in respective media. If δh tends to zero but still keeping the two opposite sides AB and CD in different media, the contribution to the line integral from the sides BC and AD will tend to zero. Provided that $\partial \mathbf{B}/\partial t$ is everywhere finite, the right hand side of Eq.(1.1.1j) also vanishes. This is due to the fact that since the two paths AB and CD approach sufficiently close to the surface, the area of integration vanishes i.e. no flux can be enclosed. Thus, in the limit as δh→0, Eq. (1.1.1j) yields

$E_{1t} = E_{2t}$ ……..(1.1.1k)

i.e., *the component of E parallel to the surface of separation between the two media is continuous across this interface.*

(d) Finally consider the behaviour of the tangential component of the magnetic vector. The analysis is similar. For pure dielectrics, current density j=0. Instead of Eq.(1.1.1j), we have in this case

$$l\,H_{1t} - l\,H_{2t} + \text{contribution from sides BC and } AD = \frac{1}{c}\int_S \frac{\partial D}{\partial t}.bds$$

In the limit as δh→0,this equation yields

$$H_{1t} = H_{2t} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1.1.1l)$$

i.e. the tangential component is continuous across the surface separating two dielectrics.

At higher frequencies, the conduction current in a conductor travels only in the outer skin. In this case, the boundary condition for the tangential component of magnetic field becomes

$$H_{1t} - H_{2t} = (4\,\Pi/c)\,\boldsymbol{j_s} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1.1.1m)$$

It should be remarked that the continuity of normal components of **D** and **B** easily follows from the continuity of tangential components of **E** and **H** in passing through a boundary. Therefore, the continuity of the tangential components of **E** and **H** is sufficient in every problem involving the propagation of electromagnetic waves from one dielectric into the other. In general, it is not possible to satisfy the boundary conditions unless we postulate three distinct

electromagnetic waves-- an incident and a reflected wave in one medium and a refracted wave in the second medium.

1.1.3 Electromagnetic theory of dielectric reflection and refraction:

Let a plane polarized wave of monochromatic light in a medium of dielectric constant K_1 be incident at an angle Φ on the surface of a medium of dielectric constant K_2. It splits into two waves—a refracted wave proceeding into the second medium and a reflected wave propagated into the first medium. Let the surface of separation between the two media be chosen as the XY plane of a Cartesian frame of reference.

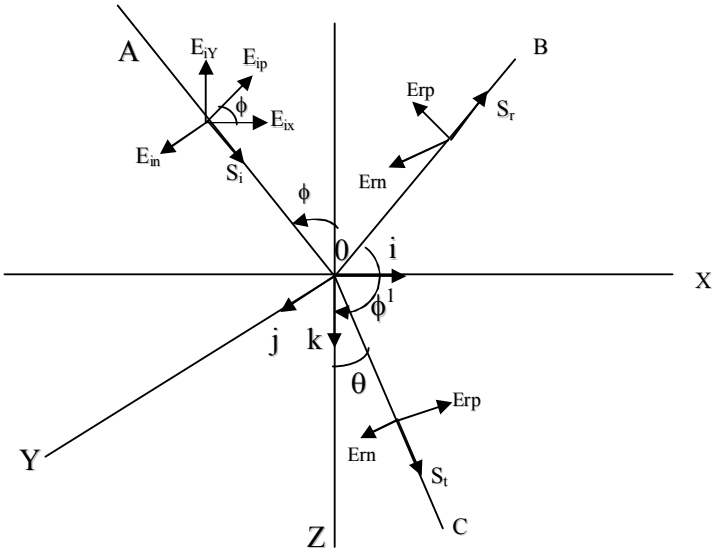


Fig. 3. Diagram for the electromagnetic theory of reflection and refraction at the boundary separating two dielectrics

We take the positive z-axis downwards and the x-axis in the plane containing the direction of propagation of the incident wave and normal to the interface i.e. the XZ plane is coincident with the plane of incidence Therefore, the x-, y-, z-, components of the unit wave normal vector s_i of the incident wave are

$$S_{ix}=\sin\Phi,\ S_{iy}=0, S_{iz}=\cos\Phi$$

We assume the incident wave to be plane polarized in an arbitrary plane. We can write the electric intensities $\mathbf{E}_i$, $\mathbf{E}_r$, $\mathbf{E}_t$, in the incident, reflected and transmitted waves in the following

form: -

$$\mathbf{E_i}=\mathbf{A}_i \exp[i\omega_i (t-\mathbf{r}\cdot\mathbf{s_i}/v_1)]\text{……..(1.1.2a)}$$

$$\mathbf{E}_r=\mathbf{A}_r\exp[i\omega_r(t-\mathbf{r}\cdot\mathbf{s}_r/v_1)]\text{……..(1.1.2b)}$$

$$\mathbf{E}_t=\mathbf{A}_t\exp[i\omega_t(t-\mathbf{r}\cdot\mathbf{s}_t/v_2)] \text{ …… (1.1.2c)}$$

Where $\mathbf{A}_i$, $\mathbf{A}_r$, $\mathbf{A}_t$ are the amplitude vectors. In writing these equations we have made no assumptions about the direction of propagation; the amplitude and even about the frequency of the reflected and refracted waves. Thus, we have not assumed the laws of reflection and refraction which we want to arrive at by the applications of principles of electromagnetic theory.

1.1.4 Laws of Reflection and Refraction:

The direction of propagation of the reflected and refracted waves can be obtained by the application of the boundary on condition expressed by Eq. (1.1.1k) to the electric components E_{ix}, E_{iy}, E_{iz} etc, The x-component of the electric field is $E_{ix}+E_{rx}$ in the first medium and E_{tx} in the second. Therefore, at the surface of separation of two media, Eq. (1.1.1k) gives

At z=0 $E_{ix} + E_{rx}=E_{tx}$……..(1.1.2d)

Also, at z=0 $E_{iy} +E_{ry}=E_{ty}$…… (1.1.2e)

Introducing the values of E_{ix}, E_{tx} etc. At z = 0 from Eqs.(1.1.2a),(1.1.2b) and (1.1.2c) in Eq.(1.1.2d) we get

$$A_{ix} \exp\{i\omega_i(t-\frac{xs_{ix}}{v_1})\}+A_{rx}\exp\{iw_r(t-\frac{xs_{rx}+ys_{ry}}{v_1})\}=A_{tx}\exp\{i\omega_t(t-\frac{xs_{tx}+ys_{ty}}{v_2})\}\text{……(1.1.2f)}$$

The only way in which Eq. (1.1.2f) remains valid for all values of t, x and y is if the three exponential factors are all the same i.e.

$$\omega_i(t-\frac{xs_{ix}}{v_1})=\omega_r(t-\frac{xs_{rx}+ys_{ry}}{v_1})=\omega_t(t-\frac{xs_{tx}+ys_{ty}}{v_2})\text{……(1.1.2g)}$$

It is essential that the exponential in Eq. (1.1.2f) must agree along the surface of separation z= 0, for otherwise, even if we were successful in satisfying the boundary condition at one point these would not hold at other points of the surface. Now, Eq. (1.1.2g) will be t rue at all points of the interface for all times only when the co-efficients of t, x and z are identical in each of these terms, since the variables are independent, We thus arrive at the following equations: -

(a) Equating the co-efficient of t gives

$\omega_i=\omega_r=\omega_t$……(1.1.2h)

Thus, on static boundary (z=0) the frequency of light on reflection and refraction does not vary. Therefore we shall suppress the suffix of ω in the subsequent theory.

(b) Equating the co-efficient of y in Eq. (1.1.2g) gives

$s_{ry}=s_{ty}=0$

i.e. The directions of propagation of the reflected and refracted waves lie in the **XZ** plane, which is the plane of incidence. Thus the reflected ray and the refracted ray lie in the plane of incidence.

(C) equating the co-efficients of x in eq (1.1.2g) gives

$s_{ix}/v_1=s_{rx}/v_1=s_{tx}/v_2$ ……(1.1.2j)

Let ϕ^1 be the angle of reflection and θ the angle of refraction, both the angles being measured with the positive direction of z-axis. Then, we have the following relations

$s_{rx}=\sin\phi^1$, $s_{ry}=0$, $s_{rz}=\cos\phi^1$

$s_{tx}=\sin\theta$, $s_{ty}=0$, $s_{tz}=\cos\theta$

Introducing these values in eq.(1.1.2j) we get

$\sin\phi/v_1=\sin\phi^1/v_1=\sin\theta/v_2$ ……(1.1.2k)

thus $\sin\phi=\sin\phi^1$

Since the incident ray and reflected ray have had opposite directions, the preceding equation yields

$\phi=\pi-\phi^1$

the angle of reflection equals the angle of incidence , which is the law of reflection .

also, from equation (1.1.2k) we have

$\sin\phi/\sin\theta=v_1/v_2$……(1.1.2 l)

If n_1 and n_2 are the refractive indices of the two media, then

$n_1=c/v_1$ and $n_2=c/v_2$

so that $v_1/v_2=n_2/n_1$

Eq.(1.1.2l) may be rewritten as

$$\frac{\sin\phi}{\sin\theta}=\frac{n_2}{n_1}={}_1n_2 \quad \text{.....(1.1.2m)}$$

since the refractive indices are constant for each medium for a given wavelength, the ratio of the sines of the angles of incidence and refraction is a constant ,which is snell's law of refraction . The electromagnetic wave thus obeys all the experimental laws of reflection and refraction at a surface separating two isotropic dielectric media.

1.1.5. Fresnel formulae:

we resolve the amplitude A_i of the electric vector of the incident incident wave into two components A_{ip} and A_{in} respectively parallel and normally to the plane of incidence .The parallel component can be further resolved along the x-axis and z-axis. Thus,the three components of A_i are

$A_{ix} = A_{ip}\cos\phi$, $A_{iy} = A_{in}$, $A_{iz} = -A_{ip}\sin\phi$

The choice of the positive direction for the parallel components is indicated in fig. 3. The convention of regarding displacement as positive is that looking against the light the positive direction of parallel component of electric vector is towards the right hand side of the observer. The perpendicular components marked in fig.3. are in reality the perpendicular to the plane of figure .

We can now write the components of the electric vector E_i of the incident wave by the help of eq. (1.2a).

$E_{ix} = A_{ip} \cos\phi \exp [i\omega \{t-(x \sin\phi + z \cos\phi)/v_1\}]$

$E_{iy} = A_{in} \exp [i\omega\{t-(x \sin\phi+ z \cos\phi)/v_1\}]$

$E_{iz} = -A_{ip} \sin\phi \exp [i\omega\{t-(x \sin\phi+z \cos\phi)/v_1\}]$

The components of the magnetic vector $\mathbf{H}_i$ of the incident wave are obtained by using

$Hi=\sqrt{K_1}s_i \times E_I$

The magnetic permeability of both media is equal to unity at optical frequencies. Therefore for the incident wave

$$H_i=\sqrt{K_1}\begin{vmatrix} i & j & k \\ S_{ix} & S_{iy} & S_{iz} \\ E_{ix} & E_{iy`} & E_{iz} \end{vmatrix}=\sqrt{K_1}\begin{vmatrix} i & j & k \\ \sin\phi & 0 & \cos\phi \\ E_{ix} & E_{iy`} & E_{iz} \end{vmatrix}$$

On equating the coefficients of i, j,k, of both sides separately, we get

$H_{ix}=-A_{in}\sqrt{K_1}\cos\phi \exp[i\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}]$

$H_{iy}=A_{ip}\sqrt{K_1}\exp i\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}]$

$H_{iz}=A_{in}\sqrt{K_1}\sin\phi \exp[i\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}]$

Similarly, we can write the components of the electric and magnetic vectors of the refracted and reflected waves.

Refracted wave:

$E_{ix}=A_{tp}\cos\theta \exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

$E_{ty}=A_{tn} \exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

$E_{tz}=A_{tp} \sin\theta \exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

$H_{ix}=-A_{tn}\sqrt{K_2}\cos\theta\exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

$H_{ty}=A_{tp}\sqrt{K_2} \exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

$H_{tz}=A_{tn}\sqrt{K_2}\sin\theta\exp[i\omega\{t-(x\sin\theta+z\cos\theta)/v_2\}]$

Reflected wave

$E_{rx}=A_{rp}\cos\phi^1\exp[i\omega\{t-x\sin\phi^1+z\cos\phi^1)/v_1\}]$

$E_{ry}=A_{rn}\exp[i\omega\{t-(x\sin\phi^1+z\cos\phi^1)/v_1\}]$

$E_{rz}=-A_{rp}\sin\phi^1\exp[i\omega\{t-(x\sin\phi^1+z\cos\phi^1)/v_1\}]$

$H_{rx}=-A_{rn}\sqrt{K_1}\cos\phi^1\exp[i\omega\{t-(x\sin\phi^1+z\cos\phi^1)/v_1\}]$

$H_{ry}=A_{rp}\sqrt{K_1}\exp[i\omega\{t-(x\sin\phi^1+z\cos\phi^1)/v_1\}]$

$H_{rz}=A_{rn}\sqrt{K_1}\sin\phi^1\exp[i\omega\{t-(x\sin\phi^1+z\cos\phi^1)/v_1\}]$

We now apply the boundary conditions to the components of the electric and magnetic intensities on the two sides of interface

$E_{ix}+E_{rx}=E_{tx}$ $E_{iy}+E_{ry}=E_{ty}$

$H_{ix}+H_{rx}=H_{tx}$ $H_{iy}+H_{ry}=H_{ty}$

On substituting in the above equations for all components and remembering that at the boundary i e. at z = 0 ,

$\omega(t-x \sin \phi/v_1)= \omega(t-x \sin \theta/v_2)=\omega(t-x \sin\phi^1/v_2)$

and also using the relation

$$\cos \phi^1=\cos (\pi-\phi) = - \cos\phi$$

we finally obtain the following four relations

$$(A_{ip}-A_{rp}) \cos \phi= A_{tp}\cos\theta \ldots\ldots\ldots\ldots\ldots \quad (1.1.2n)$$

$$A_{in}+A_{rn}=A_{tn} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.1.2o)$$

$$\sqrt{K_1}(A_{in}-A_{rn})\cos\phi=\sqrt{K_2}\, A_{tn}\cos\theta \ldots\ldots\ldots\ldots (1.1.2p)$$

$$\sqrt{K_1}(A_{ip}+A_{rp})=\sqrt{K_2}A_{tp} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.1.2q)$$

In deriving these equations we have not considered any possible change of phase on reflection or refraction since this will show up its effect in the magnitudes of A_{tp}, A_{tn}, A_{rp} and A_{rn} .we can solve Eqs.(1.1.2 n) and (1.1.2 q) for unknown quantities A_{rp} and A_{tp} Similarly Eqs(1.1.2 o) and (1.1.2 p) give A_{rn} and A_{tn} . Thus we obtain

$$A_{tp}=\frac{2\sqrt{K_1}\cos\phi}{\sqrt{K_2}\cos\phi+\sqrt{K_1}\cos\theta} A_{ip} \ldots\ldots\ldots \quad (1.1.2r)$$

$$A_{tn}=\frac{2\sqrt{K_1}\cos\phi}{\sqrt{K_1}\cos\phi+\sqrt{K_2}\cos\theta} A_{in} \ldots\ldots\ldots\ldots\ldots (1.1.2s)$$

$$A_{rp}=\frac{\sqrt{K_2}\cos\phi-\sqrt{K_1}\cos\theta}{\sqrt{K_2}\cos\phi+\sqrt{K_1}\cos\theta} A_{ip} \ldots\ldots\ldots\ldots\ldots (1.1.2t)$$

$$A_{rn}=\frac{\sqrt{K_1}\cos\phi-\sqrt{K_2}\cos\theta}{\sqrt{K_1}\cos\phi+\sqrt{K_2}\cos\theta} A_{in} \ldots\ldots\ldots\ldots\ldots (1.1.2u)$$

The law of refraction , given by Eqs (1.1.2 l), can be written as

$$\text{Sin } \phi/\sin\theta = \sqrt{K_2}/K_1$$

Since, when the magnetic permeability is unity, $v_1=c/\sqrt{K_1}$ and $v_2=c\sqrt{K_2}$. Using the new expression for the law of refraction we get for A_{tp}, A_{tn}, A_{rp} and A_{rn} the following expressions:-

$$A_{tp}= \frac{2\sin\theta\cos\phi}{\sin\phi\cos\phi+\sin\theta\cos\theta}A_{ip}$$

$$A_{tn}= \frac{2\sin\theta\cos\phi}{\sin\theta\cos\phi+\cos\theta\sin\phi}A_{in}$$

$$A_{rp}= \frac{\sin\phi\cos\phi-\sin\theta\cos\theta}{\sin\phi\cos\phi+\sin\theta\cos\theta}A_{ip}$$

$$A_{rn}= \frac{\sin\theta\cos\phi-\cos\theta\sin\phi}{\sin\theta\cos\phi+\cos\theta\sin\phi}A_{in}$$

This can be simplified by using the easily proved trigonometrical relations .

$\sin(\phi+\theta)\cos(\phi-\theta) = \sin\phi\cos\phi+\sin\theta\cos\theta$

$\sin(\phi-\theta)\cos(\phi+\theta)= \sin\phi\cos\phi-\sin\theta\cos\theta$

And we get the final expressions for A_{tp}, A_{tn}, A_{rp}, and A_{rn}.

$$A_{tp}= \frac{2\sin\theta\cos\phi}{\sin(\phi+\theta)\cos(\phi-\theta)}A_{ip} \quad \text{.....(1.1.2}v\text{)}$$

$$A_{tn}= \frac{2\sin\theta\cos\phi}{\sin(\phi+\theta)}A_{in} \quad \text{.....(1.1.2}w\text{)}$$

$$A_{rp}= \frac{\tan(\phi-\theta)}{\tan(\phi+\theta)}A_{ip} \quad \text{.....(1.1.2}x\text{)}$$

$$A_{rn}= -\frac{\sin(\phi-\theta)}{\sin(\phi+\theta)}A_{in} \quad \text{.....(1.1.2}y\text{)}$$

These equations were first obtained by Fresnel on the basis of the elastic theory of light. These relations are commonly called Fresnel's equations or Fresnel formulae for the reflection and refraction of light.

Summary: The electric and magnetic fields of the electromagnetic waves satisfy the boundary conditions at an interface separating the two dielectric media. Reflection and refraction phenomena explained on the basis of electromagnetic theory. Fresnel equations for the reflection and refraction of light obtained using electromagnetic theory.

Key terminology: Electric and magnetic fields – electromagnetic waves – reflection –refraction- Boundary conditions - Fresnel equations

Self assessment questions

1.State boundary conditions at the plane of separation between two dielectric media.

2.Deduce fresnel's laws of reflection and refraction from electromagnetic theory of light.

Reference books

1. Introduction to modern optics, B.K. Mathur

2. Optics, Born and Wolf

Unit 1

Lesson 2

Electromagnetic theory-II

Objective: To know about the production of polarization by reflection and refraction and the respective coefficients using electromagnetic theory. The behaviour of phase change on reflection and verification of Fresnel formulae along with total internal reflection is discussed.

Structure:

1.2.1. Polarization on reflection

The ratio of the reflected and incident amplitudes can be obtained from equations (1.1.2x) and (1.1.2y)

$$r_p = \frac{A_{rp}}{A_{ip}} = \frac{\tan(\phi - \theta)}{\tan(\phi + \theta)} \quad \text{.....(1.2.1a)}$$

$$r_n = \frac{A_{rn}}{A_{in}} = -\frac{\sin(\phi - \theta)}{\sin(\phi + \theta)} \quad \text{.....(1.2.1b)}$$

Now, r_n never vanishes but r_p is zero when $\tan(\Phi+\theta)=\infty$ or $\Phi+\theta=\Pi/2$ i.e, when the reflected ray is perpendicular to the incident ray. $r_p=0$ means that at this angle of incidence radiation with the electric vector parallel to the plane of incidence is not reflected.

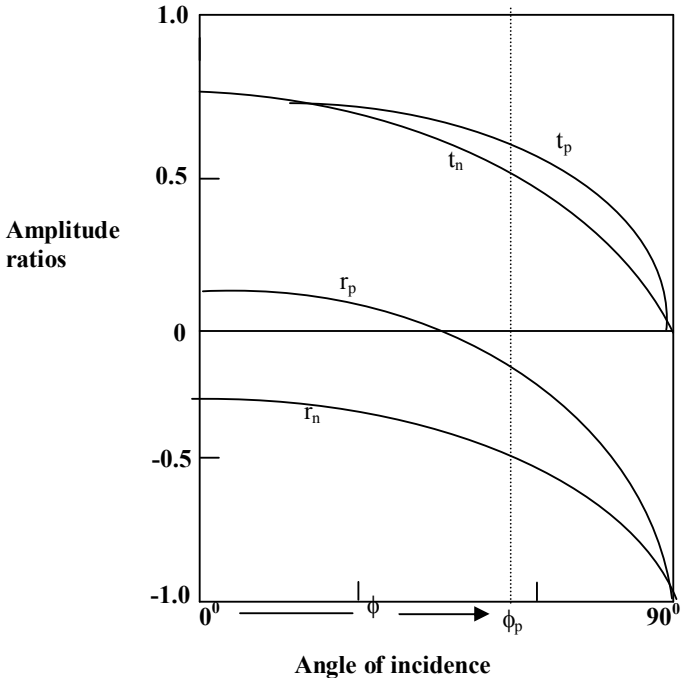


Fig .1. Graphs of r_p, r_n , t_p ,t_n against the angle of incidence for the n=1.5 for incidence from air to glass.

The angle of incidence satisfying this condition will be designated by Φ_p and the corresponding angle of refraction by θ_p. This angle Φ_p is called the polarizing angle because when unpolarized radiation is incident at this angle, the electric vector of the reflected light has no component in the plane of incidence. In other words, the reflected light is completely linearly polarized. We assume that the magnetic vector is in the plane of polarization so that the plane of incidence is the plane of polarization of the reflected light. The light vector is perpendicular to the plane of incidence since it is identical with the electric vector.

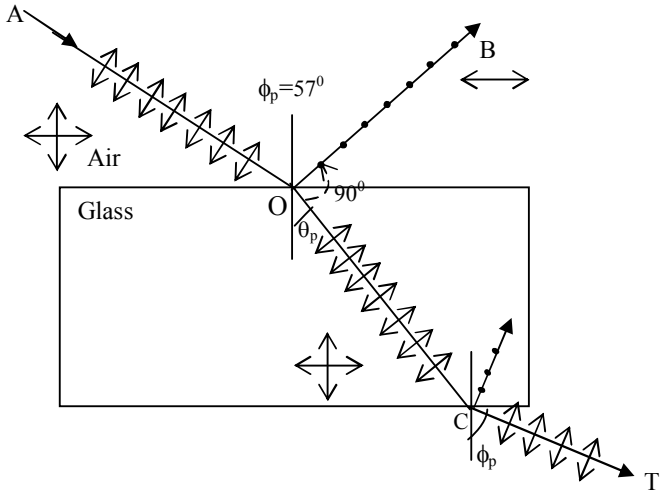


Fig.2. Illustrates Brewsters law at polarizing angle.

The polarizing angle is expressed by the relation

$$n = \frac{\sin\phi_p}{\sin\theta_p} = \frac{\sin\phi_p}{\sin(\frac{\pi}{2} - \phi_p)} = \tan\phi_p \text{ ……..(1.2.1c)}$$

This is called brewster's law. For glass interface, n =1.5 and $\Phi_p=56^0.3$. when the reflected and refracted rays are at right angles, which is brewster's experimental condition for complete polarization of the reflected wave.

This phenomenon can be employed as a means for the production of linearly polarized light but due to single reflection the degree of polarization and the intensity are poor.

1.2.2. Polarization by refraction(pile of glass plates.)

Let β_i be the angle between the plane of incidence and the plane of vibration of E_i in the linearly polarized incident light. This angle is called the azimuth of the incident wave. By the help of Fresnel's equations it is possible to write the azimuth of the refracted and reflected waves, β_t and β_r , in terms of the incident azimuth β_i as follows:-

$$\tan\beta_i = \frac{A_{in}}{A_{ip}},\ \tan\beta_t = \frac{A_{tn}}{A_{tp}},\ \tan\beta_r = \frac{A_{rn}}{A_{rp}}$$

Introducing the values of A_{rn}, A_{rp}, A_{tn} and A_{tp} from Fresnel's equations we have

$$\tan\beta_t = \cos(\phi-\theta)\tan\beta_i \text{ ……(1.2.2a)}$$

$$\tan\beta_r = -\frac{\cos(\phi-\theta)}{\cos(\phi+\theta)}\tan\beta_i \text{ ……(1.2.2b)}$$

Equation (1.2.2b) again indicates that for $\Phi+\theta=\Pi/2$, $\tan\beta_r=\infty$, $\beta_r=\Pi/2$ and hence $A_{rp}=0$ i.e. there is no E_r component parallel to the plane of incidence in the reflected wave. Whatever may be the values of θand Φ since $\cos(\phi-\theta) > \cos(\phi+\theta)$ it follows from eq(1.2.2b)) that the azimuth of the reflected wave is always greater than that of the incident wave.

$$|\tan\beta_r| > |\tan\beta_i|$$

Equation (1.2.2a) indicates that the plane of vibration of the electric vector of the refracted wave is always turned towards the plane of vibration in the incident wave. This fact is utilized for producing plane polarized light by refraction through pile of parallel glass plates. The unpolarized light is incident at polarizing angle Φ_p on the pile of parallel glass plates and it is easy to see that on each plate the angle of incidence is equal to the polarizing angle. The **E** component

perpendicular to the plane of incidence is gradually quenched in the refracted beam and the emergent beam is thus almost perfectly plane polarized and the **E** component parallel to the plane of incidence is predominantly present.

1.2.3. Reflection and Transmission coefficients:

The intensity I of radiation is given by

$$I= (c/8\pi)\sqrt{K}\, E_0^2$$

Again assuming that the permeability of medium is unity. According to this definition
We obtain for the intensity of incident light, reflected light and refracted light for the component parallel to the plane of incidence

$I_{ip}=(c/8\pi)\ \sqrt{K_1}\ A_{ip}^2$, $I_{rp}= (c/8\pi)\sqrt{K_1}\ A_{rp}^2$; $I_{tp}= (c/8\pi)\ \sqrt{K_2}\ A_{tp}^2$

We may substitute A_{rp} from Fresnel equations. Similar expressions can be written for the component perpendicular to the plane of incidence.

The reflection coefficients are defined by

$$R_p=\frac{I_{rp}}{I_{ip}} = \left(\frac{A_{rp}}{A_{ip}}\right)^2 = \frac{\tan^2(\varphi-\theta)}{\tan^2(\phi+\theta)} \quad \text{.....(1.2.3}a\text{)}$$

$$R_n= \frac{I_{rn}}{I_{in}} = \left(\frac{A_{rn}}{A_{in}}\right)^2 = \frac{\sin^2(\phi-\theta)}{\sin^2(\phi+\theta)} \quad \text{.....(1.2.3}b\text{)}$$

The transmission coefficients are defined by

$$\tau_P = \frac{I_{tp}}{I_{ip}} = \frac{2\sin\theta\sin 2\phi\cos\phi}{\sin^2(\phi+\theta)\cos^2(\phi-\theta)} \quad \text{.....(1.2.3}c\text{)}$$

$$\tau_n = \frac{I_{tn}}{I_{in}} = \frac{2\sin\theta\sin 2\phi\cos\phi}{\sin^2(\phi+\theta)} \quad \text{.....(1.2.3}d\text{)}$$

It is easy to see that $R_p+\tau_p$ is not equal to1 and $R_n+\tau_n$ is also not equal to one.

Thus the transmitted intensities are not complementary to the reflected ones. The intensity is defined as the energy crossing unit area per second but at oblique incidence the cross sectional areas of incident and reflected beams are different from that of the transmitted beams. Therefore the transmitted intensities are not complementary to the reflected ones. We will now show that the total energy in these beams is complementary. Suppose a polarized beam of finite cross section is incident on a unit area of the boundary. Then the cross-sectional areas of the incident, reflected and transmitted beams are respectively

$S_i=\cos\phi$, $S_r=\cos\phi$, $S_t=\cos\theta$

The amount of energy in the incident beam per second is

$J_i=\mathbf{P}_i\cdot\mathbf{S}_i=(c/4\pi)\ \sqrt{K_1}A_i^2\cos\phi$……(1.2.3e)

Where $\mathbf{p}_i$ is the poynting vector for the incident beam.

Similarly, the energies of the reflected and refracted beams leaving unit area of the boundary per second are given by

$J_r=\mathbf{P}_r\cdot\mathbf{S}_r=(c/4\pi)\sqrt{K_1}\,A_r^2\cos\phi$……..(1.2.3f)

$J_t=\mathbf{P}_t\cdot\mathbf{S}_t=(c/4\pi)\sqrt{K_2}\,A_t^2\cos\theta$……(1.2.3g)

The ratios

$$R=\frac{J_r}{J_i}=\frac{A_r^{\ 2}}{A_i^{\ 2}}\text{.....(1.2.3}h\text{)}$$

$$T=\frac{J_t}{J_i}=\sqrt{\frac{K_2}{k_1}}\frac{A_t^{\ 2}}{A_i^{\ 2}}\frac{\cos\theta}{\cos\phi}=n\frac{A_t^{\ 2}}{A_i^{\ 2}}\frac{\cos\theta}{\cos\phi}\text{.....(1.2.3}i\text{)}$$

are called **reflectivity** and **transmissivity** respectively.

Since $A_i^2=A_{ip}^2+A_{in}^2$, $A_t^2=A_{tp}^2+A_{tn}^2$ and $A_r^2=A_{rp}^2+A_{rn}^2$, by the application of Fresnel's equation it can be readily seen that

$R+T=1$…….. (1.2.3j)

1.2.4. Change of phase on reflection.

Equations (1.1.2v) and (1.1.2w) show that A_{tp} and A_{tn} have the same sign as A_{ip} and A_{in}. Hence electric fields of the incident and refracted waves are always in phase at the boundary separating the two media. In the case of reflected wave the phase will depend on the relative magnitudes of ϕ and θ. Two cases arise.

1.2.4.1. The second medium is optically denser than the first.

In this case $\phi>\theta$ hence it follows from equation (1.1.2 y) that the sign of A_{rn} and A_{in} are different. Therefore the phase of E_{ry} differs by Π from that of E_{iy} at the boundary surface. i.e. the component of electric vector perpendicular to the plane of incidence is reversed on reflection. we express it by writing $\delta_n=\Pi$. Also $\tan(\phi-\theta)$ is positive and for $\theta+\phi<\Pi/2$. That is $\phi<\phi_p$, $\tan(\phi+\theta)$ also positive. It follows from equation (1.1.2x) that A_{rp} has the same sign as A_{ip} i.e. E_{rp} has the same sign as E_{ip} indicating no phase change ie $\delta p=0$. But the directions of A_{rp} and A_{ip} are not the same but since both have the same sign ,this implies, as shown in fig.3., that A_{rx} is opposite in sign to A_{ix} and that A_{rz} has the same sign as A_{iz}.

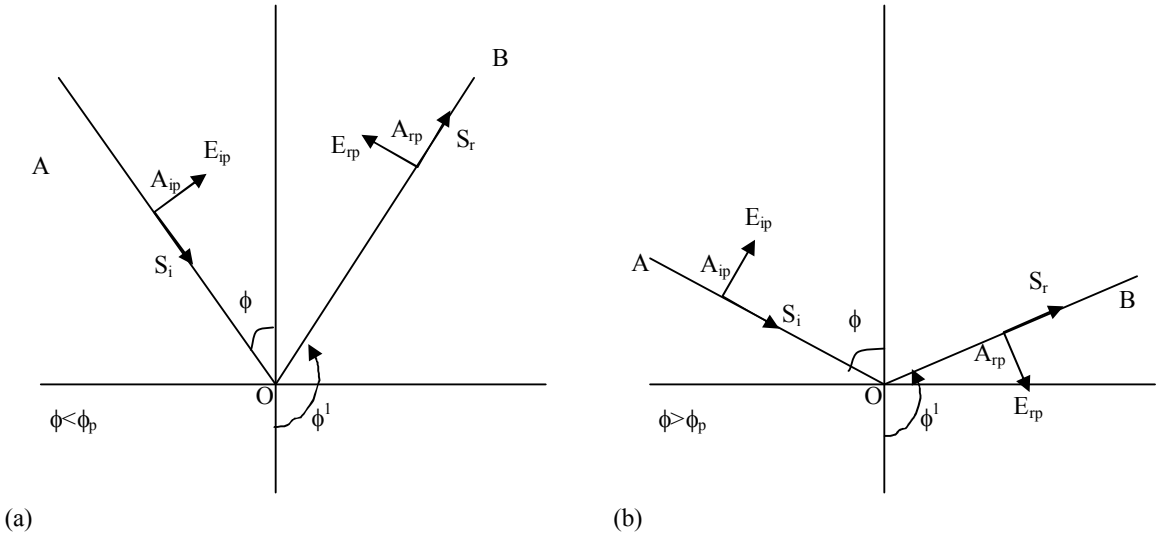


Fig.3, Direction of parallel component of the electric field at the boundary for the angles of incidence (a)less than (b) greater than Brewsters angle.

Thus any component of electric vector parallel to the reflecting surface is reversed on reflection when $\varphi<\varphi_p$ but the component normal to the reflecting surface is not reversed. The components of the magnetic vector H_i of the incident wave can be easily written as

$$H_{ix}=-\sqrt{K_1}\cos\phi E_{iy};\ H_{iy}=\sqrt{k_1}E_{ip};\ H_{iz}=\sqrt{k_1}\sin\phi\ E_{iy}$$

Thus, when E_{iy} changes sign on reflection H_{iz} also changes sign but H_{ix} does not. Similarly since E_{ip} does not changes sign hence H_{iy} also does not change sign on reflection. The important conclusion is that when the electric fields of incident and reflected waves have opposite phase, the magnetic fields are in phase .

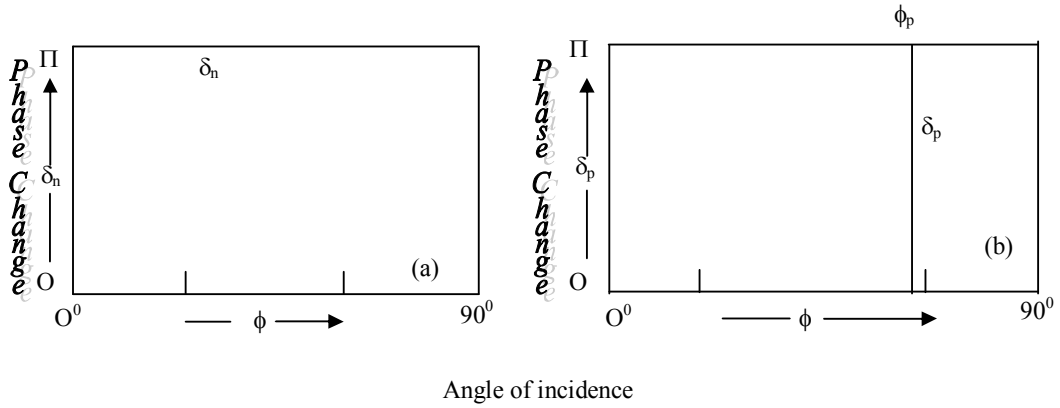


Fig.4. Phase change of the electric vector of plane polarized light externally reflected from a dielectric .

when reflection takes place at an angle greater than the polarizing angle φ_p then $\theta+\phi>\Pi/2$ and $\tan(\theta+\phi)$ become negative .it follows from Eq.(1.1.2x)that A_{rp} has a sign opposite to that of A_{ip}. The corresponding components of electric fields differ in phase by Π at the surface ie .A phase change of δ_p of Π occurs on reflection. Thus, we arrive at the following rule for the case in which the electric field is in the plane of incidence. The incident and reflected waves are in phase if their Z-components of electric fields are in the same direction and their X-components are opposite to each other just before and after reflection.

1.2.4.2. The second medium is optically rarer than the first.

In this case $\Phi < \theta$and it follows from Eq(1.1.2y) that E_{ry} and E_{iy} are always in phase at the boundary i.e. the electric component perpendicular to the plane of incidence is not reversed. The components of electric fields parallel to the plane of incidence differ in a phase by Π if $\theta+\Phi<\Pi/2$ and have the same phase if $\theta+\Phi>\Pi/2$. The phase changes up to the critical angle are exactly reverse of those at the corresponding angles of the previous case, which we may call as external reflection.

Perpendicular incidence.

In the case of normal incedence $\varphi=0$ and $\theta=0$. Therefore Fresnels formulae for A_{tp}, A_{tn}, A_{rp} and A_{rn} become indeterminate. But we can use eqs (1.2r),(1.2s),(1.2t),(1.2u) in which we substitute $\cos\theta = 1$and $\cos\varphi = 1$. Thus we get

$$A_{tp}=\frac{2\sqrt{k1}}{\sqrt{k1}+\sqrt{k2}}A_{ip}=\frac{2}{n+1}A_{ip} \quad \text{.....(1.2.4a)}$$

$$A_{tn}=\frac{2\sqrt{k1}}{\sqrt{k1}+\sqrt{k2}}A_{in}=\frac{2}{n+1}A_{in} \quad \text{.....(1.2.4b)}$$

$$A_{rp}=\frac{\sqrt{k2}-\sqrt{k1}}{\sqrt{k2}+\sqrt{k1}}A_{ip}=\frac{n-1}{n+1}A_{ip} \quad \text{..... (1.2.4c)}$$

$$A_{rn}=\frac{\sqrt{k1}-\sqrt{k2}}{\sqrt{k2}+\sqrt{k1}}A_{in}=-\frac{n-1}{n+1}A_{in} \quad \text{.....(1.2.4d)}$$

Thus for normal incidence the distinction between the parallel and perpendicular components disappear, since the plane of incidence is indeterminate.

12.5. Experimental verification of Fresnel Formulae

In the experimental verification of Fresnel formulae the polarizing nicol is set with its principle section at 45^0 to the plane of incidence. Thus in Eqs(1.4b) $\beta_I=45^0$,which now becomes

$$\tan\beta_t = -\frac{\cos(\phi-\theta)}{\cos(\phi+\theta)} \text{----------(1.2.5a)}$$

The angle, β_r, between the direction of vibration of the reflected light and the plane of incidence can be determined experimentally for different values of ϕ. The experimental values are compared with the theoretical values obtained from Eq.(1.8 a) since the angle of refraction can be known from snell's law of refraction

$$n = \sin\phi/\sin\theta$$

Provided n is known beforehand from some other experiment.

Experimental Procedure:

The index of refraction n of the material of prism for a monochromatic light is determined accurately with the help of an adjusted spectrometer. The polarizing angle ϕ_p is calculated from Brewster's law,

$$\tan\phi_p = n$$

The telescope is turned to a position to receive the direct image of the slit and then rotated through an angle $(180\text{-}2\phi_p)$ and clamped. A clean prism face is set on the prism table so that its plane is along the central axis and is also parallel to receive the light reflected from this face into the telescope. It is easy to see that the angle of incidence on the prism is equal to the polarizing angle ϕ_p. Hence the reflected light is plane polarized, the plane of vibration being perpendicular to the plane of incidence (horizontal plane). A biquartz plate is mounted infront of the telescope objective and a graduated circular scale carrying a nicol, capable of rotation, is clamped to the telescope objective. The nicol is rotated until the tint of passage appears. The telescope is then brought into initial position, prism is removed and another nicol is mounted on the collimator lens. This nicol is turned till the tint of passage reappears. In this setting of the collimator nicol, its principal section is vertical. This nicol is rotated through 45^o and thus the inclination of the plane of vibration of incident light is made 45^o to the plane of incidence. The telescope is now turned through a known angle α and the prism mounted in its original position and then rotated by rotating the prism table to a position to receive the reflected light into the telescope, the angle of incidence being $\phi = (180^o\text{-}\alpha)/2$. The telescope nicol is rotated from its initial setting through an

angle say Ψ to restore the tint of passage. Hence the inclination of the plane of vibration of the reflected light to the plane of incidence (horizontal plane) is

$$\beta_r = 90 - \psi \quad \text{.....(1.2.5b)}$$

The experiment may be repeated for various angles of incidence by repeating the above process. It is found that the experimental values of β_r agreed well with those given by eq.(1.2.5 a), thus confirming Fresnel's equations. The experimental values of β_r may be plotted against the corresponding values of ϕ. The point where the curve crosses the ϕ-axis gives ϕ_p. Again, the index of refraction n may be calculated from Brewster's law and compared with the value obtained with the spectrometer measurements.

1.2.6. Stationary Waves.

Electromagnetic waves reflected at a plane surface can interfere with the incident waves. We consider the case of normal incidence and n>1. From Eq.(1.2.4 d) it follows that A_{rn} is in a direction opposite to that of A_{in}. From Eq.(1.2.4c) it follows that A_{rp} has the same sign as that of A_{ip} but like signs actually denote opposite directions of their amplitudes, which is evident from the way in which A_{rp} and A_{ip} are marked positive in Fig 4. Thus the two components of electric vectors of normally incident wave, in the plane of reflecting surface, are reversed on reflection. Therefore, the reflecting surface must be a node for the electric vector. And other nodes will be formed at $\lambda/2$, λ, $3\lambda/2$ etc. from the surface. As the components of the magnetic vectors are not reversed on reflection i.e. the magnetic fields of the incident and reflected waves are in phase at the reflecting surface. The reflecting surface is therefore, the antinode of the magnetic vector. Thus in a stationary wave pattern the nodes of the electric vector coincide with the antinodes of the magnetic vector. In the Wiener's experiment, no blackening of the photographic film resulted along its line of contact with the reflecting surface. The first black band which corresponds to an antinode in the stationary wave was formed at a distance $\lambda/4$ from the reflecting surface. Hence the reflecting surface was a nodal plane so far as the photographic action of the light is concerned. Wiener, therefore, concluded that the photographic action is due to electric vector and not to magnetic vector of the electromagnetic light wave. In general the electric vector is effective in every optical phenomenon and so it is regarded as the light vector.

1.2.7. Total Internal Reflelcltion.

Suppose a plane wave of monochromatic light is incident in denser media, say glass, on the surface separating it from the rarer media say air. The angle of incidence ϕ is in the denser media and θ is the angle of refraction in a rarer media.

Hence $\sin\theta = n\sin\phi$ ……(1.2.6a)

And $\cos\theta = \sqrt{1-\sin^2\theta} = \sqrt{1-n^2\sin^2\phi}$ …… (1.2.6b)

Where n is the refractive index of the denser media with respect to the rarer media. The phenomenon of total reflection occurs for ϕ grater than the critical angle ϕ_c which corresponds to $\theta = \pi/2$. Therefore from Eq.(1.10a) we get

$$n \sin\phi_c = 1$$

When $\phi > \phi_c$, $n \sin\phi > 1$ and $\cos\theta$ as given by Eq.(1.2.6b) becomes imaginary i.e.

$$\cos\theta = \pm i\sqrt{n^2\sin^2\phi - 1} \quad \text{……(1.2.6c)}$$

We should re-examine Fresnel's equations by inserting the values of $\cos\theta$ and $\sin\theta$ given by Eqs. (1.2.6c) and (1.2.6a) respectively choosing only the positive sign in Eq. (1.2.6c). Thus the component A_{rn} of the amplitude of the electric vector E_r given by

$$A_{rn} = \frac{n\cos\phi - i\sqrt{(n^2\sin^2\phi - 1)}}{n\cos\phi + i\sqrt{(n^2\sin^2\phi - 1)}} A_{in} \quad \text{……(1.2.6d)}$$

The complex quantity by which A_{in} is multiplied is of the form

$$\frac{\rho\exp(-i\delta_n)}{\rho\exp(i\delta_n)} = \exp(-i2\delta_n)$$

where $\tan\delta_n = \sqrt{n^2\sin^2\phi - 1}\,/\,n\cos\phi$ ……(1.2.6e)

hence $A_{rn} = A_{in}\exp(-i2\delta_n)$… …… (1.2.6f)

In order to interpret this result we should write the component of the electric vector E_i of the incident wave in a direction normally to the plane of incidence i.e. the y-component of E_i .

$$E_{in} = A_{in}\exp[i\omega\{t-(x\sin\phi + z\cos\phi)/v_1\}] \quad \text{…….. (1.2.6g)}$$

Similarly for the reflected wave we write

$$E_{rn} = A_{rn}\exp[i\omega\{t-(x\sin\phi + z\cos\phi)/v_1\}]$$

$$E_{rn} = A_{in}\exp(-i2\delta_n)\exp[i\omega\{t-(x\sin\phi + z\cos\phi)/v_1\}]$$

$$E_{rn} = A_{in}\exp[i\omega\{t-(x\sin\phi + z\cos\phi)/v_1\} - 2\delta_n] \quad \text{…… .(1.2.6h)}$$

A comparison of Eqs. (1.2.6g) and (1.2.6h) indicates that the amplitude of E_{rn} is the same as that of E_{in} but both differ in phase by $-2\delta_n$ which varies with the angle of incidence ϕ.

We can similarly examine the expression for A_{rp} given by Eq.(1.1.2x) and we get

$$A_{rp} = A_{ip}\exp(-i2\delta_p)\ldots\ldots \quad \ldots\ldots(1.2.6i)$$

Where $\tan\delta_p = n\sqrt{n^2 \sin^2\phi - 1}/\cos\phi$……(1.2.6j)

The component of electric vector E_i parallel to the plane of incidence is

$$E_{ip}=A_{ip}\exp[i\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}]\ldots\ldots..(1.2.6k)$$

And the component of E_r parallel to the plane of incidence, viz.

$$E_{rp}=A_{rp}\exp[i\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}]$$

Becomes

$$E_{rp}=A_{ip}\exp i[\omega\{t-(x\sin\phi+z\cos\phi)/v_1\}-2\delta_p]\ldots\ldots..(1.2.6l)$$

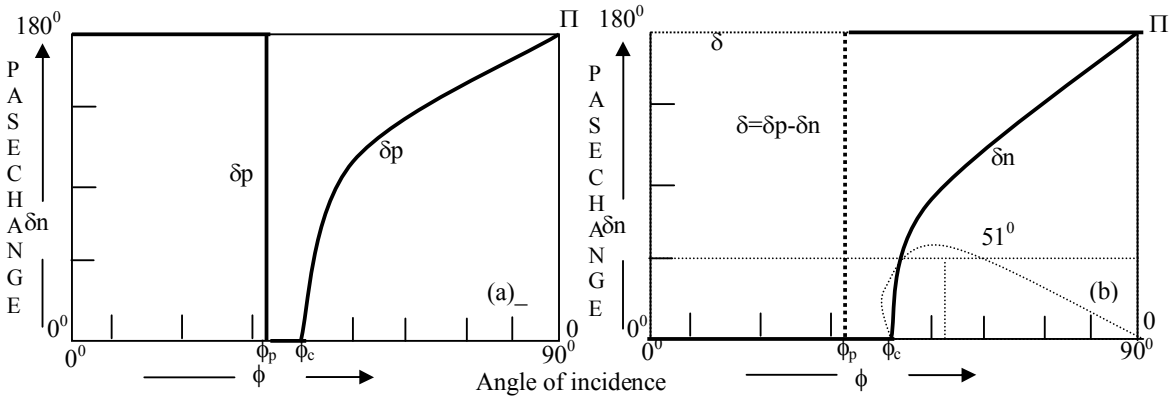


Fig.5. Phase changes of the electric vector for internal reflection(n=1.51).

Thus the components of electric vector parallel and perpendicular to the plane of incidence on reflection experience phase retardations of $2\delta_p$ *and* $2\delta_n$ respectively but their amplitudes remain unaltered.

$|A_{rp}|=|A_{ip}|$ and $|A_{rn}|=|A_{in}|$

Thus, for each component, the intensity of light which is totally reflected is equal to the intensity of the incident light. Consequently we saythat the incident wave is totally reflected.

The components of the electric vector of the reflected wave parallel and perpendicular to the plane of incidence have a relative phase difference δ is given by

$\delta=2(\delta p.\delta n)$

$$\tan\frac{1}{2}\delta=\tan(\delta p.\delta_n)=\frac{\tan\delta_p-\tan\delta_p}{1+\tan\delta_p\tan\delta_p}$$

$$\tan\frac{1}{2}\delta=\cos\phi\sqrt{\sin^2\phi-(1/n^2)}\,/\sin^2\phi\ldots\ldots\ldots\ldots\ldots(1.2.6m)$$

The phase difference depends upon the angle of incidence greater than the critical angle. The plane polarized light on reflection becomes elliptically polarized . Any desired phase difference δ can be introduced between the two components of reflected light by the proper choice of ϕ. Therefore any desired type of elliptically polarized light can be conveniently produced by the method of total internal reflection. It is easy to see from Eq (1.2.6 m) that for n =1.5, δ=π/4 when $\phi=53^0 15^1$ or $\phi=50^0 14^1$ and so after another internal reflection at the same angle the phase difference between the reflected components becomes π/2. The reflected light is , of course , elliptically polarized, for in general the components of electric vector of the incident light parallel and perpendicular to the plane of incidence are of unequal amplitudes.

Circularly polarized light can also be produced by two internal reflections as follows:

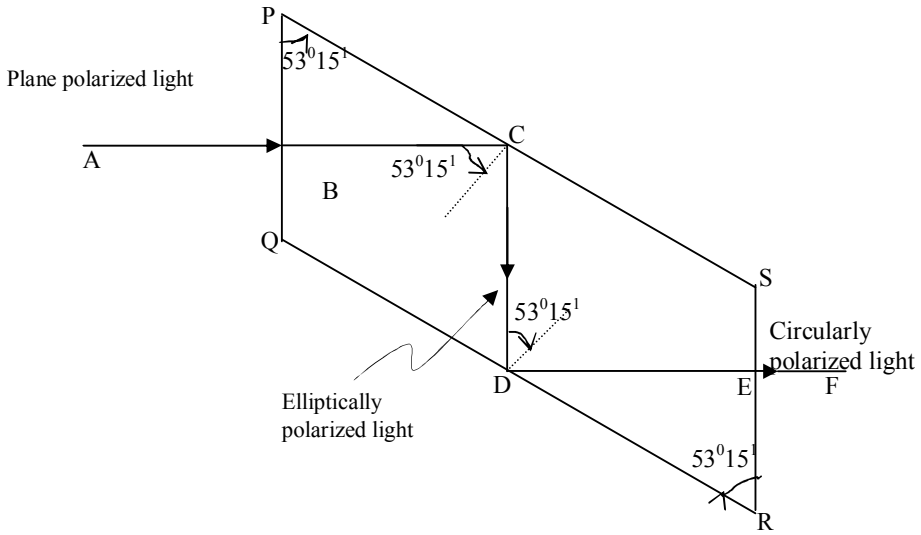


Fig.6. Production of circularly polarized light by fresnels rhomb .

A glass block (n = 1 ·55), with opposite surfaces mutually parallel and the end faces PQ and RS as squares, is formed with angels of transverse section equal to53^0 15' and $(180^0 - 53\ ^0 15')$. A plane polarized monochromatic light is incident normally on the rhomb from air as shown in Fig.21.9,with its plane of vibration inclined at 45^o to the plane of incidence. At it

suffers total reflection at an angle $53^o15'$ and so phase difference of $\prod/4$ is introduced between the two perpendicular components. The reflected light CD is, therefore, elliptically polarized. On the second internal reflection at D at the same angle a further phase difference of $\prod/4$ is introduced so that the total phase difference is $\prod/2$ between the two perpendicular components .Thus ,the light after two internal reflections i.e. DE is circularly polarized. The glass block constructed and used in this way is known as Fresnel's rhomb. The rate of variation of δ with the angle of incidence is greater when $\phi = 50^o14'$. Due to this reason it is desirable to construct a rhomb with larger angle $\phi = 53^o15'$ when it is to be used for producing circularly polarized light.

1.2.8. Disturbance in the second Medium:

At total internal reflection, the phase difference between the incident and reflected waves is not $\prod$. Therefore, there is some resultant electric field in the denser medium at the boundary. As a consequence, in order to satisfy the boundary condition there must also be a disturbance in the rarer medium. Therefore, we should again analyze the expression for the refracted wave by substituting in it the values of $\sin\theta$ and $\cos\theta$ from eqs. (1.2.6 a) and (1.2.6c). The component of electric vector of the refracted wave perpendicular to the plane of incidence is

$E_{tn} = A_{tn} \exp [i\omega\{t-(x \sin\theta + z \cos\theta)/v_2\}]$……..(1.2.7a)

Now, by the help of eqs.(1.1.2d) and (1.2.6f)

$A_{tn} = 2A_{in}\cos\delta_n \exp(-i\delta_n)$……..(1.2.7b)

Hence

$$E_{tn}=2A_{in}\cos\delta_n \exp[\pm\frac{2\pi z}{\lambda_2}\sqrt{n^2\sin^2\phi-1}] \times \exp[i\{\omega(t-\frac{xn\sin\phi}{v_2})-\delta_n\}]\text{……..(1.2.7c)}$$

In order that this result may correspond to physical situation we must take the negative value of the root in the exponential, since otherwise, the amplitude would tend to infinity with the depth z of penetration

$$E_{tn}=2A_{in}\cos\delta_n\exp[-\frac{2\pi z}{\lambda_2}\sqrt{n^2\sin^2\phi-1}] \times \exp[i\{\omega(t-\frac{xn\sin\phi}{v_2})-\delta_n\}]\text{……..(1.2.7d)}$$

Thus E_{tn} is periodic in x and not in z. consequently, it corresponds to non-homogeneous wave disturbance in the rarer media close up to the boundary and moving parallel to the x axis without attenuation with a velocity less then v_2, since $n \sin\phi > 1$. The amplitude of the disturbance decays exponentially with the depth z of penetration, the effective depth of penetration being of the order of $\lambda_2/2\prod$. The disturbance ordinarily becomes negligibly small after a distance of few wavelengths

along the z axis from the boundary. The periodicity in x implies that the energy flow is along the boundary and not down into the rarer medium. This can be explained by evaluating the x and z components of the poynting vector. It will be found that the z component has in general, finite value but its time average will be found to vanish, meaning thereby that there are places where energy enters alternating with places where energy leaves. Thus there is no permanent transmission of energy in the second medium.

The boundary wave has another peculiarity that it is not transverse, since $E_{tx,}$ the x component of electric intensity in the refracted wave is not equal to zero , and of course the wave is propagated in the x-direction.

The electromagnetic Theory of light predicts that the electromagnetic field penetrates for a very small distance in to the rarer medium at total reflection . Various experiments have been designed to demonstrate the existence of this disturbance . In one experiment a very fine layer of lamp black is deposited on that face of totally reflecting glass prism at which a beam of light is totally reflected . the fine carbon particles, viewed through a microscope , are seen to be illuminated. It is claimed that each carbon particle being in the oscillatory electromagnetic field of the penetrating wave becomes earlier, however, does not apply exactly to this experiment, since the presence of carbon particles disturbs the boundary condition in their own neighbourhood and hence disturbs the penetrating electromagnetic field.

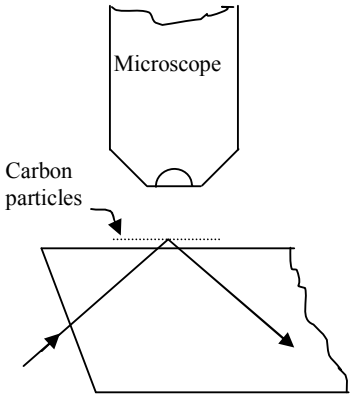


Fig 7. Arrangement for observing light scattered from fine particles on the upper side of a surface when light is incident at an angle$\phi > \phi_c$ on the lower-side

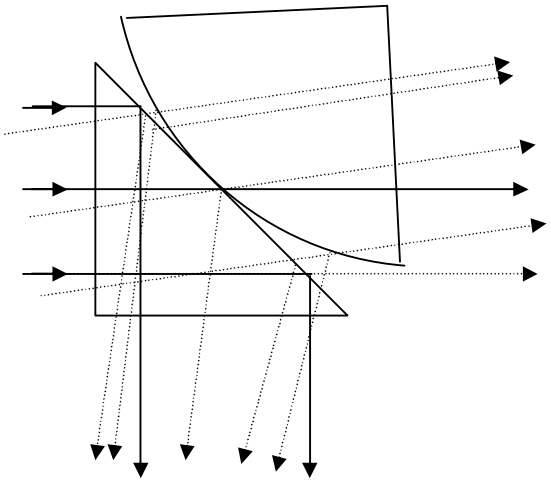


Fig 8.Hall's experiment –demonstrating that light does not undergo total reflection near the point of contact between the hypotenuse surfaces of the teo points.

Another experiment to demonstrate this penetration was performed by E.E.Hall in 1902. His apparatus, shown in Fig.8 essentially consists of two total reflecting prisms, one of which has the hypotenuse faces mutually touch at one point. If a broad source of monochromatic light is viewed through two prisms, one can observe a bright central disc surrounded by several faintly

illuminated interference rings. The central disc corresponds to the point of contact where the light is transmitted without any attenuation and the interference rings to the narrow region where the thickness of the air film is comparable to wave length surrounding the actual point of contact. The rings get weaker as the thickness of air film increases being quite invisible if the air gap exceeds a few wavelengths. The rings correspond to angle of incidence just below the critical angle ϕ_c , the corresponding rays shown by dotted lines. It is argued that where the thickness of the air film is comparable to the wavelength , the electromagnetic field that penetrates in the air film has appreciable intensity on reaching the curved face of the second Prism. Here it changes again into ordinary light wave which gives rise to production of Newton's rings by the usual process of partial reflection and refraction at the two surfaces of air film.

It should be remarked that no experiment can be designed to demonstrate the disturbance in the rarer medium without withdrawing energy. Therefore, the boundary condition must be modified and the preceding theory does not apply exactly to any experiment.

1.2.9. Summary

Production of polarization of light by reflection and refraction phenomena is explained. The total energy of the wave is conserved. Showed the change of phases on reflection in different conditions. Experimental verification of Fresnel formula and total internal reflection has been explained.

1.2.10. Key terminology

Polarization – Reflection – refraction – Transmission coefficients – Phase change – Total internal reflection – Fresnel formula

1.2.11. Self assessment questions

1. Verify the Fresnel formula. Show that Brewester law of polarizing angle is a direct consequence of Fresnels laws of reflection.
2. On Electromagnetic theory , explain total internal reflection of light and show that on total reflection there is a penetration of light in the other medium. Give its experimental verification.
3. Explain the action of Fresnels rhomb.

1.2.12. Reference books

1. Introduction to modern optics, B.K. Mathur

2. Optics, Born and Wolf

Unit I

Lesson 3

ELECTROMAGNETIC THEORY OF ABSORPTION

Objective: To know about the absorption in the medium while electromagnetic wave incidenting at different angles and production of elliptically polarized light due to metallic reflection.

Structure

1.3.1 Introduction:

Absorbing media is defined as the media in which the intensity of light diminishes as it penetrates deeper and deeper within the medium. Exponentially strong absorbing and high reflecting power characterize the metals. According to the electromagnetic theory of absorption outline below all media, which are not perfect dielectric, should exhibit absorption of light. This is to be expected since the electric current from conduction of free electrons under the influence of electric field of the incident radiation produces heat energy as a result of the electrical resistance of the medium. By the principle of conservation of energy, this heat energy must come from the radiant energy of the light wave, which originated the current.

Consider a homogeneous isotropic conducting medium of dielectric constant K, permeability μ and conductivity σ. The theory of propagation of electromagnetic waves in conducting material is based on Maxwell's field equations, which may be written by using the material. **D**=K **E**, **B**=μ **H**, **j**=σ**E**, in the following form:-

$$curlH = \frac{K}{c}\frac{\partial E}{\partial t} + \frac{4\pi}{c}\sigma E \quad \text{……(1.3.1a)}$$

$$curlE = -\frac{\mu}{c}\frac{\partial H}{\partial t} \quad \text{……(1.3.1b)}$$

$$divE = {4\pi\rho}/{K} \quad \text{……(1.3.1c)}$$

$$divH = 0 \quad \text{…… (1.3.1.d)}$$

1.3.2 Reflection and refraction of electromagnetic waves:

In the case of an electromagnetic wave incident from outside on a conductor we can replace eq (1.3.1c) by div **E** =0. For if we take the divergence of eq.(1.3.1a) and using eq.(1.3.1c) we find

$$\frac{\partial \rho}{\partial t} + \frac{4\pi\sigma}{K}\rho = 0$$

which on integration gives

$$\rho = \rho_0 \exp(-t/\tau) \quad \text{……(1.3.1e)}$$

where t = K/4πσ. This indicates that ρ→0 as t→∞ the relaxation time T is extremely small compared with the periodic time of the light wave so that in a good conductor any initial distribution of charge is quickly dispersed to the surface. Hence we may regard the interior of a conductor to be unchanged that is in a conducting medium there can be no permanent charge density. Therefore we may put the right hand side of eq (1.3.1 c) equal to zero.

div **E**=0 ……(1.3.1f)

To study the propagation of electromagnetic waves in a conductor we shall form the differential equation of electric vector **E** as follows.

$$curlcurlE = -\frac{\mu}{c} curl \frac{\partial H}{\partial t} = -\frac{\mu}{c}\frac{\partial}{\partial t} curlH$$

$$curlcurlE = -\frac{\mu}{c}\left[\frac{K}{c}\frac{\partial^2 E}{\partial t^2} + \frac{4\pi\sigma}{c}\frac{\partial E}{\partial t}\right]$$

But $curlcurlE = graddivE - \nabla^2 E = -\nabla^2 E$

Therefore E satisfies the differential equation of the wave.

$$\nabla^2 E = \frac{K\mu}{c^2}\frac{\partial^2 E}{\partial t^2} + \frac{4\pi\sigma\mu}{c^2}\frac{\partial E}{\partial t} \quad \text{.....(1.3.1g)}$$

The presence of the term $\partial E/\partial t$ in eq.(1.3.1g) implies that the wave is damped that is its amplitude and energy suffer a progressive attenuation as it penetrates deeper and into the medium.

1.3.3 Electromagnetic waves propagated normally to conducting surface:

We consider the XY-plane as the surface of the conducting media, the Z-axis being normal to the surface. We consider the electromagnetic waves incident normally on the conducting surface and so they are propagated in the direction of the positive Z-axis.

Hence $\frac{\partial}{\partial y} = 0$ and $\frac{\partial}{\partial x} = 0$

Now from eq (1.3.1f) it follows that $\partial E_z/\partial z = 0$ which yields E_z=0, since we are not interested in constant field. Similarly eq (1.3.1d) gives H_z=0. Thus **E** has only two components E_x and E_y. To simplify further, we suppose that the wave is plane polarized in the XZ plane then E_x=0 and **E** = **j**E_y. The wave equation viz eq (1.3.1g)

now reduces to $\frac{\partial^2 E_y}{\partial z^2} - \frac{K\mu}{c^2}\frac{\partial^2 E}{\partial t^2} - \frac{4\pi\sigma\mu}{c^2}\frac{\partial E_y}{\partial t} = 0$ .....(1.3.1h)

We are interested in the special case of electromagnetic waves in which the field is a simple periodic function of time . Therefore we assume that E_y is expressed by

$E_y = f(z)\exp(i\omega t)$.....(1.3.1i)

By substituting this in eq(1.3.1 h) we obtain the following differential equation for the space function f(Z):-

$$\frac{d^2 f}{dz^2} + \left(\frac{K\mu}{c^2}\omega^2 - i\frac{4\pi\sigma\mu\omega}{c^2} \right) f = 0$$

The solution if this differential equation is

$$f(z) = E_{oy} \exp\left(\pm i\sqrt{K\mu\omega^2 - i4\pi\sigma\mu\omega}\; z/c \right)$$

We choose the negative sign, since the wave is propagated in the positive Z direction.

$$f(z) = E_{oy} \exp\left(-i\sqrt{K\mu\omega^2 - i4\pi\sigma\mu\omega}\; z/c \right) \ldots\ldots\ldots\ldots\ldots\ldots (1.3.1j)$$

To simplify this expression we introduce the following notation.

$$\sqrt{K\mu\omega^2 - i4\pi\sigma\mu\omega} = \omega\pi\left(1 - i\kappa\right) \ldots\ldots\ldots\ldots\ldots\ldots (1.3.1k)$$

Where n and κ are real and we call κ the *extinction coefficient* . Squaring and equating real and imaginary parts separately we obtain the following equations:

$$n^2\left(1-\kappa^2\right) = K\mu \ldots\ldots\ldots\ldots (1.3.1l)$$

$$n^2\kappa = 2\pi\mu\sigma/\nu \ldots\ldots\ldots\ldots\ldots\ldots (1.3.1m)$$

It should be emphasized that equation(1.3.1l) and (1.3.1m)apply only to normal incidence.

Now, using equations (1.3.1j) and (1.3.1k) we can write the solution, equation (1.3.1i), of the differential equation of wave motion in the following form

$$E_y = E_{oy} \exp(-\omega n\kappa z/c) \exp \prec i\omega(t - nz/c) \succ$$

Multiplying both sides by unit vector j, gives

$$E = E_o \exp(-\omega n\kappa z/c) \exp \prec i\omega(t - nz/c) \succ \ldots\ldots\ldots\ldots\ldots (1.3.1n)$$

Now

$$\frac{n\omega}{c} = \frac{2\pi\upsilon}{c} = \frac{2\pi}{\lambda}$$

Where is the wave length in the free space of refractive index n. The real part of equation (1.3.1n) viz.

$$E = E_0 \exp(-2\pi\kappa z/\lambda)\cos\omega(t - nz/c) \ldots\ldots\ldots\ldots\ldots\ldots (1.3.1o)$$

gives the electric vector in the plane polarized wave propagated along positive z axis, the amplitude $E_0\exp(-2\pi\kappa z/\lambda)$ diminishing exponentially as wave advances. The planes of equal phase are identical with planes of equal amplitude. Therefore the wave is homogeneous. It should be remarked that homogeneous waves exist only for normal incidence of light on metallic boundary. For oblique incidence refracted wave is inhomogeneous. This is due to fact that various parts of plane front (plane of equal phase) have traversed different thickness of absorbing material and thus suffered unequal absorption and therefore amplitude is not constant over wave front.

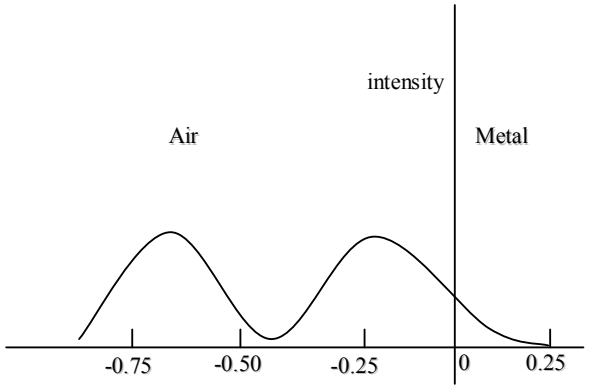


Fig.1. Intensity of wave in metal compared to the intensity of the stationary wave on the side of the incident and reflected waves.

The intensity I of light is proportional to square of amplitude of electric vector. Therefore it follows that intensity I of light transmitted through a metal film of thickness z decreases in accordance with relation

$$I = I_0\exp(-4\pi\kappa z/\lambda).....(1.3.1p)$$

Thus intensity falls to 1/e value after wave has advanced a distance d given by

$$d = \lambda/4\pi\kappa.....(1.3.1q)$$

This quantity is usually a very small fraction of wavelength. Greater value of κ for a given the smaller penetration of wave with in metal i.e. greater the absorption or attenuation of energy. Due to this reason we have called κ as Extinction coefficient. It is also called Attenuation index. The coefficient of z in equation (1.3.1p) viz.

$$\chi = 4\pi\kappa/\lambda = 1/d.....(1.3.1r)$$

is called the *absorption coefficient*, where is wavelength in free space of index n. The absorption coefficient is therefore defined by condition that energy in a wave falls by in a distance. Therefore and κ can be directly determined by measurements of transmission of different thickness.

We now calculate H but only its X component remains. Using equations (1.3.1 b) and (1.3.1 n) H_x given by

$$\frac{\partial H_x}{\partial t} = \frac{c}{\mu}\frac{\partial E_y}{\partial z}$$

$$\frac{\partial H_x}{\partial t} = -\frac{\omega}{\mu}(n\kappa + in)E_{oy}\exp(-\omega n\kappa z/c)\exp\{i\omega(t - nz/c)\}$$

$$i.e. H_x = -(n/\mu)(1 - i\kappa)E_y$$

Now we write $1\text{-}i\kappa = \sqrt{1+\kappa^2}\exp(-i\gamma)$ where $\tan\gamma = \kappa$, hence

$$H_x = (n/\mu)\sqrt{1+\kappa^2}E_{oy}\exp(-\omega n\kappa z/c)\exp[i\omega(t - nz/c) - i\gamma] \quad \text{.....(1.3.1}s\text{)}$$

Thus in a conducting media there is a phase difference between electric and magnetic vectors H lags behind E. It will be recalled that in a pure dielectric E and H are in phase.

In an isotropic dielectric Y component of electric vector E for a plane wave traveling along positive Z axis is expressed by

$$E_y = E_{oy}\exp\{i\omega(t - z/v)\} \quad \text{.....(1.3.1}t\text{)}$$

But in a conducting media, equation(1.3.1 o) gives

$$E_y = E_{oy}\exp[i\omega\{t - n(1 - i\kappa)z/c\}]$$

$$\text{i.e. } E_y = E_{oy}\exp[i\omega \prec t - n^* z/c \succ] \quad \text{.....(1.3.1}u\text{)}$$

where $n^* = n(1\text{-}i\kappa)$ .....(1.3.1v)

It is clear that Eq;(1.3.1u) is formally similar to Eq.(1.3.1t). Duo to this analogy n* is Often called the complex refractive index of the conducting or absorbing media.
To complete the analogy with the non-conducting media we also introduce complex Phase velocity v* and the complex dielectric constant k* by the relations analogous to That for isotropic dielectrics

$$V^* = \frac{c}{\sqrt{\mu K^*}} \qquad n^* = \sqrt{\mu K^*}$$

Hence $K^*=\frac{n^{*2}}{\mu}=\frac{n^2(1-i\kappa)^2}{\mu}$.....(1.3.1*w*)

Now,using eq.(1.3.1k),k* is given bys

$K^*=K-i\,(4\pi\sigma/\omega)$.....(1.3.1*x*)

It should be remarked that is the conductivity at optical frequency concerned and is not generally equal to the direct current or low frequency conductivity. According to the theory outlined above, only difference between isotropic transparent medium and isotropic absorbing media is that the constants n and k which are real for transparent media, become complex n* and k* given by Eq(1.3.1v) and (1.3.1x) for absorbing media.

According to theory outlined above, every conductor should absorb electromagnetic Waves. However NaCl solution is perfectly transparent although its electrical conductivity is high. Absolutely perfect conductor is characterised by infinitely large conductivity.

From Eqs.(1.3.1l)and (1.3.1m)

$K/\sigma = 2\pi(1-\kappa^2)/\omega\kappa$

And so, thus a perfect conductor would not allow the electromagnetic wave to Penetrate through any depth with in it; On the other hand it would reflect all the incident Waves.

1.3.4 Electro magnetic waves propagated obliquely to conducting surface.

We take the equation of vector wave of electric field strength in the general form.

$E=E_0\exp[i\omega\prec t-(r.s)/v]$.....(1.3.2*a*)

Hence $\frac{\partial E}{\partial t}=i\omega E$

And $\sigma E=\frac{\sigma}{i\omega}\frac{\partial E}{\partial t}=-\frac{i\sigma}{\omega}\frac{\partial E}{\partial t}$

On introducing this in Eq.(1.3.1a) we can write it in the form

$\text{Curl}H=\frac{1}{c}(K-i\frac{4\pi\sigma}{\omega})\frac{\partial E}{\partial t}=\frac{K^*}{c}\frac{\partial E}{\partial t}$.....(1.3.2*b*)

Where K* is a complex dielectric constant

$K^*=K-i(4\pi\sigma/\omega)$…….(1.3.2c)

Equation(1.3.2 b) is representing isotropic dielectric constant except that for dielectrics the dielectric constant is real while for metals it is a complex number. In order to simplify the problem of oblique propagation of wave with in the conducting medium the XZ plane is chosen as plane of incidence of light. Then the harmonic solution of the vector wave equation (1.3.1g) is written as

$$E=E_0\exp[i\omega \prec t-(a^*x+b^*z)/v \succ] \quad \text{.....(1.3.2}d\text{)}$$

Where a* ,0 ,b*are the direction cosines of the unit wave normal with in the conducting media. Substituting this solution in Eq.(1.3.1g) we get the result

$$\frac{a^{*2}+b^{*2}}{v^2}=\frac{1}{c^2}(K-i\frac{4\pi\sigma}{\omega})=\frac{K^*}{c^2} \quad \text{.....(1.3.2}e\text{)}$$

It is clear that if is not zero, the direction cosines a* and b* of the unit wave normal are complex and therefore they may be written in the form

$$A^*=\sin\theta-ik_x \quad \text{.....(1.3.2}f\text{)}$$

$$B^*=\cos\theta-ik_z \quad \text{.....(1.3.2}g\text{)}$$

Now ,the harmonic solution, Eq(1.3.1d), of the vector wave equation can be written as

$$E=E_0e^{-(2\pi/\lambda)(k_x+k_zz)}e^{i\omega\prec t-(x\sin\theta+z\cos\theta)/v\succ} \quad \text{.....(1.3.2}h\text{)}$$

Where λ is the wavelength in a medium of refractive index n It is clear from eq.(1.3.1i) that the equation of constant phase is.

$$X\sin\theta+z\cos\theta=C \quad \text{.....(1.3.2}i\text{)}$$

While the amplitude is constant in the plane

$$x\,k_x + z\,k_z = C' \qquad (1.3.2\ j\ )$$

Where C and C ' are constants. If we put

$$k_x = \kappa\sin\alpha \qquad \text{and} \qquad k_z = \kappa\cos\alpha \qquad (1.3.2\ k)$$

we can rewrite eq. (1.3.2 j) as

$$\kappa(x\sin\alpha + z\cos\alpha) = C' \qquad (1.3.2\ l\ )$$

Where κ is called the *extinction co-efficient.* In the preceding equation θ is the angle between the z-axis and normal to plane of equal phase and α is the angle between the z-axis and normal to the plane of equal amplitude. In highly absorbing materials we can regard the planes of equal amplitude as parallel to the XϒZ plane and α may be put equal to zero.

Now Eq. (1.3.2h) becomes

$$E=E_0 \exp(-2\pi\kappa z/\lambda) \quad \exp[i\omega\{t-(x\sin\theta + z\cos\theta)/v\}] \qquad (1.3.2\ m)$$

Thus, the angle between the planes of equal phase and equal amplitude is θ and the wave is said to be non-homogeneous. When equal amplitude planes coincide with equal phase plane the wave is called homogeneous and it is only for normal incidence of light on the metallic boundary that the wave is homogeneous.

The amplitude $E_0 \exp(-2\pi\kappa Z/\lambda)$ of the non-homogeneous wave also decreases exponentially. The amplitude of the wave is attenuated in the ratio 1: $\exp(-2\pi\kappa)$ after traversing a distance whose projection on the Z-axis is $Z=\lambda$, where λ is the wavelength within the conducting medium.

The intensity of light transmitted through a metal film of thickness Z is proportional to the square of the amplitude and so it can be written as

$$I=I_0 \exp(-4\pi\kappa Z/\lambda)\ldots\ldots\ldots\ldots\ldots\ldots(1.3.2n)$$

The intensity diminishes exponentially with the depth of penetration of the light wave.

Now, from Eqs.(1.3.2 f) and (1.3.2 g) we calculate

$$a^*/v = 1/v\,(\sin\theta - i\kappa\sin\alpha) = 1/c\,(n\sin\theta - i n\kappa\sin\alpha) \qquad (1.3.2\ o)$$

$$b^*/v = 1/v\,(\cos\theta - i\kappa\cos\alpha) = 1/c\,(n\cos\theta - i n\kappa\cos\alpha) \qquad (1.3.2\ p)$$

Substituting in Eq. (1.3.2 e) we get

$$\kappa - i\,(4\pi\sigma/\omega) = (n\sin\theta - i\,n\kappa\sin\alpha)^2 + (n\cos\theta - i n\kappa\cos\alpha)^2$$

$$\kappa - i\,(4\pi\sigma/\omega) = n^2(1-\kappa^2) - i\,2n^2 k\cos(\theta-\alpha)$$

But to a close approximation $\alpha=0$, the above equation reduces to

$$\kappa - i\,(4\pi\sigma/\omega) = n^2(1-\kappa^2) - i\,2n^2 k\cos\theta$$

By comparing the real and imaginary parts we obtain

$$n^2 (1- \kappa^2) = \kappa \qquad (1.3.2\ q)$$

$$n^2 k \cos \theta = 2\pi\sigma/\omega = \tau \qquad (1.3.2r)$$

Thus, the refractive index n and extinction co-efficient κ depend upon the direction of propagation of the wave within the conduction medium. For normal incidence $\theta=0$, these equations are identical with Eqs. (1.3.1 k) and (1.3.1 l) respectively.

The permeability μ of the conducting medium is unity at optical frequency. Therefore the complex refractive index of the metal is

$$n^{*2} = k^* = n^2 (1-k^2) - i2n^2 \kappa \cos \theta \qquad (1.3.2\ s)$$

For normal incidence this relation gives

$$n^* = n (1 - i \kappa) \qquad (1.3.1t)$$

We shall use the expression for complex refractive index given by (1.3.2 t) even when light is incident obliquely on the metal surface.

1.3.5 Metallic Reflection at Oblique Incidence

When a plane monochromatic wave in a dielectric medium (say air) is incident on the surface of a conductor, reflection and refraction occur but quite differently from the case of incidence on a dielectric – dielectric boundary. We may follow the procedure given in sec. 1.1.3 but the results apply to the present case also if we replace n by the complex refractive index n^* for as shown in sec.1.3.1, the propagation of plane wave in a conducting medium is the same as in transparent dielectric except that n^* takes the place of n. For example, the law of refraction is obtained by replacing n_1 n_2 in Eq. (21.2 m) by n^* and we get

$$\sin\theta = 1/n^* \sin \phi = 1/n(1-\kappa) \quad \sin\phi \qquad (1.3.3\ a)$$

$$\text{And } \cos\theta = \frac{1}{n(1-i\kappa)}\sqrt{n^2(1-i\kappa)^2 - \sin^2\phi} \quad \text{.....(1.3.3}b\text{)}$$

Thus the angle θ is complex and so it has no simple interpretation of the angle of refraction.

We may calculated the amplitudes and phases of the reflected and refracted waves by substituting in fresnel's formulae the value of θ given by eq.(1.3.3a). We shall only consider the problem of determining the optical constants of the metal from the observation on the reflected wave .the metallic index of refraction n and the extinction coefficient κ defined eqs. (1.3.1l) and (1.3.1m) are the optical constants of the metal. We shall first consider the nature of the reflected wave. The amplitude components A_{ip} and A_{in} of the incident wave and the corresponding components A_{rp} and A_{rn} of the reflected wave are related by Fresnel's Eqs. (1.1.2 x) and (1.1.2 y), viz.

$$A_{rp} = \tan(\phi-\theta)/\tan(\phi+\theta) \,.\, A_{ip} \quad \text{----------} \quad (1.3.3\text{ c})$$

$$A_{rn} = \sin(\phi-\theta)/\sin(\phi+\theta) \,.\, A_{in} \quad \text{----------} \quad (1.3.3\text{d})$$

Since θ is now complex A_{rp} and A_{rn} are complex quantities and so the ratios A_{rp}/A_{ip} and A_{rn}/A_{in} are also complex. Therefore, as in the case of total internal reflection, here also characteristic phase changes δ_p and δ_n occur on reflection in parallel and perpendicular components of the electric vector of the incident wave. Let r be the absolute value of A_{rp}/A_{rn} and δ the phase difference between the two components of reflected light i.e.

$\delta = \delta_p - \delta_n$. Then from Eqs. (1.3.3 c) and (1.3.3 d) we have

$$r\exp(i\delta) = A_{rp}/A_{rn} = -A_{ip}/A_{in}\cos(\phi+\theta)/\cos(\phi-\theta) \quad \text{----------(1.3.3 e)}$$

Now consider the case when the electric vector in the incident plane polarized light is at 45° to the plane of incidence. Then $A_{ip} = A_{in}$ and Eq. (1.3.3 e) becomes

$$r\exp(i\delta) = -\cos(\phi+\theta)/\cos(\phi-\theta) = \tan\phi\tan\theta - 1/\tan\phi\tan\theta + 1 \quad \text{----------} \quad (1.3.3\text{ f})$$

$$\frac{1+r\exp(i\delta)}{1-r\exp(i\delta)} = \tan\phi\tan\theta$$

And by eliminating θ with the aid of Eq. (1.3.3 a) we finally get

$$\frac{1+r\exp(i\delta)}{1-r\exp(i\delta)} = \frac{\tan\phi\sin\phi}{\sqrt{n^2(1-i\kappa)^2 - \sin^2\phi}} \quad \text{.....(1.3.3}g\text{)}$$

Since the right hand side of eq. (1.3.3 g) is complex, δ must differ from zero or ∏. There is a phase difference other than ∏ between the two complaints of reflected light. Therefore a *plane polarized light on metallic reflection, in general, becomes elliptically polarized.*

We will now study the variation of the phase difference and the nature of reflected light with the angle of incidence ϕ.

(i)At normal incidence ϕ =90^o,hence r exp(iδ) =-1 i.e. δ =o and r = -1. Thus the wave is reflected with a sudden phase change of ∏ but with no phase difference between the components of the electric vector. The plane polarized light remains plane polarized on reflection at normal incidence from the metallic mirror.

(ii)At grazing incidence ϕ =90^o, hence r exp(iδ) =1 i.e. δ =0 and r =1. The reflection occurs without any phase change and the reflected light is again plane polarized.

(iii)For other angles of incidence the reflected light is elliptically polarized and the axes of ellipse have, in general, no special relation to the plane of incidence as shown in Fig..2 and fig.3

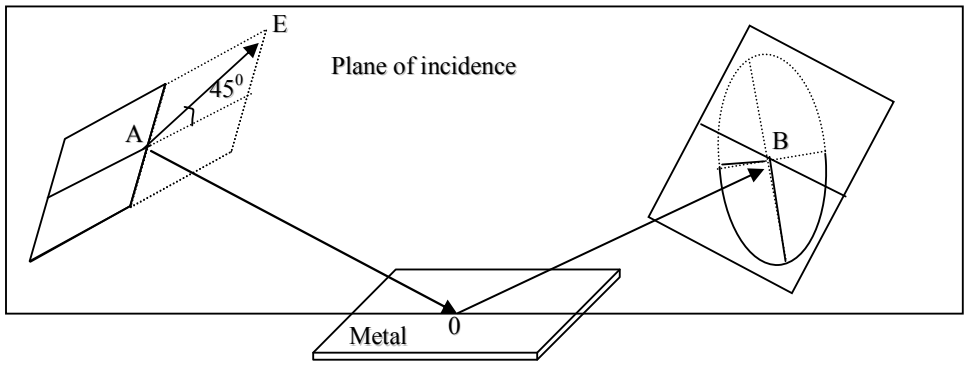


Fig 2 Reflection of plane polarised light from a metal surface to give elliptically polarised light

However there is an angle of incidence ϕ_p called the principal angle of incidence for which δ = ∏/2. Then the axes of vibration ellipse are parallel and perpendicular to the plane of incidence as shown in the end on view Fig.3 (b).

Since exp (i δ) = i when δ= π/2, Eq .(1.3.3 g) for principal angle of incidence becomes

$$\frac{1+ir}{1-ir} = \frac{\tan\phi_p \sin\phi_p}{\sqrt{n^2(1-i\kappa)^2 - \sin^2\phi_p}} \quad \text{.....(1.3.3}h\text{)}$$

The elliptically polarised light can always be converted into plane polarized light by allowing it to pass through a crystal plate of suitable thickness and orientation which introduces opposite difference between the components of elliptically polarised light.this can be easily accomplished by the use of Babinet compensator.

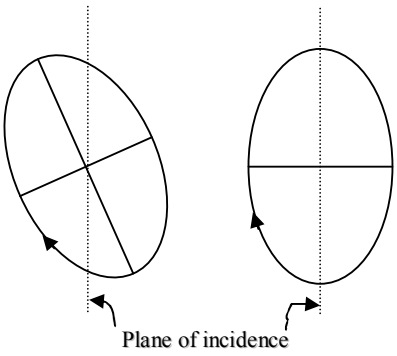


Fig.3. End on view of the reflected light from metallic mirror at
(a) an angle of incidence ϕ;
(b) at principal angle of incidence ϕ_i

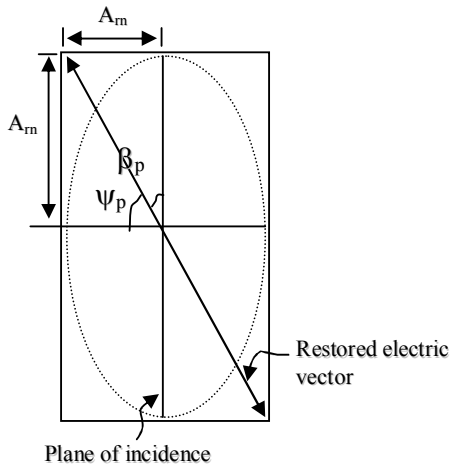


Fig4. Vibration ellipse of reflected from a metal at the principal angle of incidence and restored electric vector

The plane of vibration of the "restored" electric vector is at right angles to the shorter diagonal of the end face of the nicol when the latter is adjusted for complete extinction of "restored" linearly polarized light .the angle between the plane incidence and the direction of vibration of "restored" electric vector is called the azimuth of the restored electric vector .the principal azimuth corresponds to the principal angle of incidence ϕ_p .

$$\tan\beta_p = \left|\frac{A_{rn}}{A_{rp}}\right| = \frac{1}{r} = \cot\psi_p$$

$$r = \tan\psi_p \quad \text{.....(1.3.3}i\text{)}$$

Where $\psi_p = \pi/2 - \beta_p$ is the complement of the principal azimuth. Introducing the value of r from Eq. (1.3.3 i) in Eq.(1.3.3 h)we get

$$\frac{\tan\phi_p si\phi_p}{\sqrt{n^2(1-i\kappa)^2 - in^2\phi_p}} = \frac{1+i\tan\psi_p}{1-i\tan\psi_p} = \exp(i2\psi_p) \quad \text{.....(1.3.3}j\text{)}$$

Now, if, as is usually the case in the visible region

$$n^2(1+\kappa^2) >> 1$$

May be neglected in comparison to $n^2(1-i\kappa)^2$ and Eq. (1.3.3 j) becomes

$$\frac{n(1-i\kappa)}{\tan\phi_p \sin\phi_p} = \cos 2\psi_p - i\sin 2\psi_p$$

Equating real parts, we obtain

$$n = \tan\phi_p \sin\phi_p \cos 2\psi_p \text{.....(1.3.3}k\text{)}$$

Equating the real parts and using the value of n given by Eq. (1.3.3 k) we obtain

$$\kappa = \tan 2\psi_p \text{.....(1.3.3}l\text{)}$$

Thus, one can use these expressions to evaluate the optical constants n and κ of metals after measuring the principal angle of incidence and the principal azimuth or it's complement. We give below the optical constants for some metals determined by Minor.

Optical constants for various metals for sodium light =5893A^0.

Metal	ϕ_p	ψ_p	N	κ	ρ,%
Cobalt	$78^0 5^1$	$31^0 40^1$	2.120	1.900	67.5
Copper	$71^0 34^1$	$39^0 5^1$	0.617	4.258	74.1
Silver	$75^0 35^1$	$43^0 47^1$	0.177	20.554	95
Gold	$72^0 18^1$	$41^0 39^1$	0.37	7.62	85.1
Sodium	$71^0 19^1$	$44^0 58^1$	0.005	522	99.7

Kundt was successful in measuring the index of refraction n by measuring the deviation produced by a metallic prism of very small refracting angle (a fraction of a minute of arc). When light is incident approximately normally on the prism deviation is given by the same expression as in the case of transparent medium , The metal prism was prepared by the electrolytic deposition upon platinised glass. In general, the results were in agreement with those obtained by reflection experiment. In particular, his measurements also confirmed the curious result that for some metals n is less than 1. The small indices of silver, gold, copper and sodium mean that light travels faster in these metals than in air.

We note from the table that in all cases $n < n\kappa$, hence it follows from Eq.(1.3.1l) that K is negative, the permeability μ being unity at optical frequencies. The negative value of dielectric constant K can be explained by the electron theory of dispersion in metals.

The values of n and k given in the table are not in agreement with that given by Eq.(1.3.1m). For example for copper $\sigma = 5.14*10^{17} sec^{-1}$, υ of sodium light is of the order of $5*10^{14}$sec., hence $\sigma/\upsilon \sim 10^3$ ie $n^2\kappa \sim 10^3$ but from the table $n^2\kappa = 1.57$.This indicates that our theory is inadequate for the visible region of the electromagnetic spectrum . This is due to the fact that K , μ and σ are not true and constants but depend on the Wavelength of radiation and so the refractive index n and the extinction coefficient k also depend on the wavelength. This can explained on the basis of electron theory of dispersion in metals. Our theory is true only when K and σ are regarded as absolute constants which will be true only in the infrared region, since the vibrations of free electrons are sufficiently slow and electron theory of dispersion in metals is not applicable.

1.3.6. Reflection at normal incidence:

We have explained in fig 4 that at normal incidence, the distinction between A_{rp} and A_{rn} disappears , the plane of incidence being undetermined. We may therefore omit the indices p and n. The reflection co-efficient or the amplitude reflectivity at normal incidence on the dielectrics is given by Eq.(1.2.4 c) or (1.2.4d),which is rewritten.

$$\frac{A_r}{A_i} = \frac{n-1}{n+1} \quad (1.3.4a)$$

The amplitude reflectivity at normal incidence on a metallic surface is contained on replacing n in Eq.(1.3.4a) by the complex refractive index.

$$\left|\frac{A_r}{A_i}\right| \exp(-i\delta) = \frac{n^* - 1}{n^* + 1} = \frac{n(1-i\kappa)-1}{n(1-i\kappa)+1} \quad \text{.....(1.3.4b)}$$

Where A_r and A_i are the amplitudes of the reflected and incident fields and δ the phase change on reflection. The intensity reflectivity or reflecting power ρ of the metal is the ratio of the intensity of the reflected light to that of the incident light when the angle of incidence ϕ is zero.

$$\rho = \frac{I_r}{I_i} = \left|\frac{A_r}{A_i}\right|^2 \quad \text{.....(1.3.4c)}$$

The reflecting power is obtained on multiplying Eq.(1.3.4 b) by the complex conjugate expression $|A_r/A_i| \exp(i\delta)$ and thus we get

$$\rho = \frac{(n-1)-in\kappa}{(n+1)-in\kappa} \times \frac{(n-1)+in\kappa}{(n+1)+in\kappa}$$

$$\rho = \frac{(n-1)^2 + n^2\kappa^2}{(n+1)^2 + n^2\kappa^2} = \frac{n^2(1+\kappa^2)+1-2n}{n^2(1+\kappa^2)+1+2n} \quad \text{.....(1.3.4.}d\text{)}$$

The phase shift δ is calculated by writing eq.(1.3.4b) with real denominator. Then a comparison of the real and imaginary parts leads to the result

$$\tan\delta = \frac{2n\kappa}{n^2(1+\kappa^2)-1} \quad \text{.....(1.3.4}e\text{)}$$

When the light is reflected from the boundary between a dielectric and a metal, the terms $(n-1)^2$ and $(n+1)^2$ in eq.(1.3.4d) are small compared to $n^2\kappa^2$-in reality both the terms are regarded as only a correction to $n^2\kappa^2$. The intensity reflectivity or reflecting power of metals is therefore enormously large. Silver reflects 95% of the incident light i.e. ρ=0.95. A substance which exhibits this strong reflecting power characteristic of the metal is said to have metallic lustre. Moreover it is also seen from Eq.(1.3.4d) that greater the value of κ for given λ, the nearer the value of reflecting power ρ to unity for that wavelength. Thus gold and copper appear yellow since κ is greatest for this colour which is therefore reflected more strongly than other colours. In sec. 1.3.1, eq.(1.3.1 q) means that greater the value of κ for a given λ, the smaller the penetration of the wave within the metal. Thus all wavelengths which are strongly absorbed are also strongly reflected by metals- all good conductors are good reflectors. Hence colours of metals by transmitted and by reflected light are approximately complementary. For example, a thin film of gold appears green by transmitted light as the gold colour yellow is absorbed and reflected. In order to observe this effect it is necessary to prepare sheets of metals of few thousandth of millimeter thickness.

We can easily co- relate the reflecting power p with the electrical conductivity of the Metal by the help of eq(1.3.1m).Now, Eq.(1.3.4d) may be written in the form

$$\rho = 1 - \frac{4n}{(n+1)^2 + n^2\kappa^2} \quad \text{.....(1.3.4f)}$$

For metals the absorption is very nearly unity. Therefore we put $\kappa=1$ in Eq.(1.3.4f) And get

$$\rho = 1 - \frac{4n}{2n^2 + 2n + 1} \quad \text{.....(1.3.4g)}$$

In the long wavelength region it is found that for metals $n >> 1$.

Hence we may ignore all terms in the denominator of Eq.(1.3.4g) other than $2n^2$ and so it becomes $\rho = 1 - \left(2/n\right)$ .....(1.3.4h)

Also putting $\kappa=1$ in Eq.(1.3.1m) and also $\mu=1$, since at optical frequencies permeability of all metals is unity, we have $n^2 = \sigma/\nu$

Hence $\rho = 1 - 2\sqrt{\nu/\sigma}$.....(1.3.4*i*)

This approximate relationship is, however, not valid below $\lambda=0.5\times10^{-8}$cm. Hagen and Rubens using infra-red radiation of wavelengths $\lambda=12\times10^{-8}$ cm. found experimentally that for copper $1-\rho=1.6*10^{-2}$ while from conductivity the calculated value is $1.4*10^{-2}$. They showed that the reflecting power of most metals increases rapidly with wavelength and approaches the value given by Eq.(1.3.4i).

1.3.7 Summary

In absorbing media the electromagnetic wave intensity diminishes as it penetrates into the media. Strong absorbing and high reflecting power are characteristics of metals. Reflection and refraction of electromagnetic waves in the absorbing media at normal and oblique incidence are explained . On metallic reflection, plane polarized light becomes elliptically polarized light and also good absorbers are good reflectors.

1.3.8 Key terminology

Absorbing media - normal and oblique incidence - metallic reflection - complex refractive index - complex dielectric constant - extinction coefficient

1.3.9 Self assessment questions

1. Show on the basis of electromagnetic theory, how a good absorber is also a good reflector.
2. Show that when κ approaches zero, the principle angle approaches the polarizing angle.
3. Explain on the basis of electromagnetic theory how a plane polarized light is converted to the elliptically polarized light.

1.3.10 Reference books

1. Introduction to modern optics, B.K. Mathur
2. Optics, Born and Wolf

Unit I

Lesson 4

DISPERSION

Objective: To know about different theories of normal and anomalous dispersion and explanation of these dispersions through electromagnetic theory.

Structure of the lesson

1.4.1 Introduction:

The variation of the refractive index of a medium with wavelength (or frequency) constitutes the phenomenon of dispersion. It is represented as $dn/d\lambda$. The index of refraction is found to decrease with the increase of wavelength, but over small wavelength ranges there is increase of index of refraction accompanied by an increased absorption of the passing radiation

through the medium. The term anomalous dispersion is used to describe this effect. Lorentz gave an adequate theory of dispersion based on the electromagnetic theory of light and Cauchy, Selleimer and Helmholtz made earlier attempts.

1.4.2 Normal Dispersion:

If the values of the refractive index of the material of a prism are plotted against wavelength, a curve like one of those shown below is obtained.

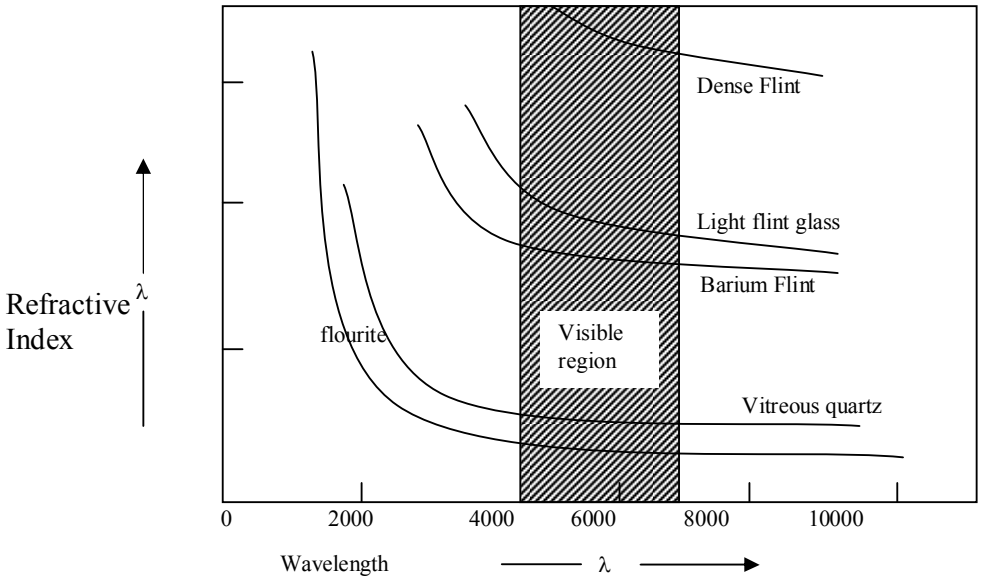


Fig.1. Normal dispersion

From these curves the following important conclusions can be drawn.

* The index of refraction decreases as the wavelength increases.
* The range of increase $dn/d\lambda$, i.e., the slope of the curve is greater at shorter wavelengths. The violet end of the prismatic spectrum is spread out on a larger scale than the red end.
* At any given wavelength the curve is steeper the larger the index of refraction of the substance.

The dispersion is greater for a medium of higher index of refraction.

The dispersion of a medium with the above said characteristics is called the normal dispersion. All non-coloured transparent substances exhibit normal dispersion in the visible region of the electromagnetic spectrum.

1.4.3 Cauchy's Equation:

Cauchy found that in the region of normal dispersion the dependence of refractive index on wavelength for a given medium is represented with reasonable accuracy by an empirical equation

$$n=A+\frac{B}{\lambda^2}+\frac{C}{\lambda^4} \quad \text{.....(1.4.1a)}$$

where A,B and C are constants

depending on the medium, diminishing rapidly in magnitude as proceed to higher order of terms. For some purposes the variation of refractive index with wavelength can be represented sufficiently accurate by including only the first two terms of the power series.

i.e. $n = A + B/\lambda^2$ .....(1.4.1b)

where A and B are positive and n is maximum at the violet end and least at the red end of the visible region.

By differentiating the equation

$$\frac{dn}{d\lambda} = -\frac{2B}{\lambda^3} \quad \text{.....(1.4.1c)}$$

Thus the dispersion varies inversely as the cube of the wavelength. The minus sign indicates that the slope of the dispersion curve is negative and its magnitude decreases as λ increases.

1.4.4 Hartmann Formula:

Hartmann gave another empirical expression as given below

$$n = n_o + \frac{b}{(\lambda-\lambda_o)} \quad \text{.....(1.4.2a)}$$

which gives an accurate variation of n with λ if the range of wavelength is not too large and where n_o ,b, and λ_o are constants to be determined from the observations.

Anomalous Dispersion:

Cauchy's equation fits the curve of normal dispersion of many transparent substances in the visible region with reasonable accuracy. All refracting substances exhibit selective absorption in some wavelength regions. If measurements of the index of refraction are extended into infra-red region, a marked deviation of results from Cauchy's equation are observed as one approaches the spectral region of the absorption lines of the refracting medium. It is found that the index of refraction decreases more rapidly than given by Cauchy or Hartmann formulae, as the absorption band in the neighbourhood of absorption wavelength is approached from the short wavelength side and jumps to large values on the long wavelength side as shown in the figure below for a substance having two absorption bands. MNO,PQR, ST are experimental curves

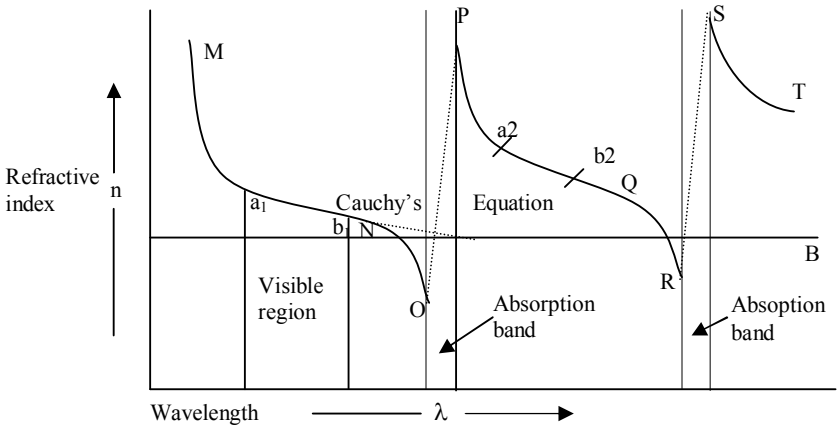


Fig..2. Anomalous dispersion of a transparent substance having two absorption bands

while the dotted curve beyond N is given by Cauchy's formula. The regions $a_1 b_1$ and $a_2 b_2$ are normal in form being closely represented by Cauchy's equation but with a different set of values of constants A and B. It is difficult to measure index of refraction in the absorption band since the light is absorbed strongly. But measurements of index of refraction with thin layers and employing Michelson interferometer have shown that the curves MNO, PQR and ST are connected up within the absorption band by smooth continuous curves OP and RS as shown in the above figure.

The dispersion in the region of the absorption band is called anomalous dispersion since historically this was the first observation that within the absorption band the refractive index increases with increase in wavelength. In the neighbourhood of the absorption band, the refractive index is greater for longer wavelength and so longer wavelengths are more refracted than certain shorter ones. For example, Iodine vapour has an absorption band in the visible region and so a prism made of iodine vapour deviates the red rays more than the violet.

1.4.5 Sellmeier's formula:

The first success in deriving a dispersion formula of general applications was achieved by Sellmeier on the basis of the elastic solid theory of light. He assumed that every medium consists of elastically bound particles capable of vibrating with a natural frequency of vibration say 'ν_0' in the absence of any periodic force. He regarded the particles as frictionless oscillators. A passage of light wave through the medium exerts a periodic oscillatory force on the particles, which are thus forced to vibrate with relatively small amplitude with a frequency 'ν' equal to that of the incident light wave. The amplitude of forced oscillation increases as 'ν' approaches 'ν_0' and when $\nu = \nu_0$ the particle resonates and a very large amplitude is built up. These vibrations cause a change in the velocity of light wave. Sellmeier carried out the mathematical investigation of the above mechanism and arrived at the following equation.

$$n^2 = 1 + \frac{A_0 \lambda^2}{(\lambda^2 - \lambda_0^2)} \quad \text{.....(1.4.3a)}$$

For a medium in which all particles have the same natural frequency corresponding to the light of wavelength 'λ_o' . This equation contains two constants A_0 and λ_0. This existence of several absorption bands is explained by postulating the existence of particles with several different natural frequencies. In this case Eq.(1.4.3a) can be written with a series of terms

$$n^2=1+\frac{A_0\lambda^2}{(\lambda^2-\lambda_0^2)}+\frac{A_1\lambda^2}{(\lambda^2-n\lambda_1^2)}+.....(1.4.3b)$$

$$n^2=1+\sum_k\frac{A_k\lambda^2}{(\lambda^2-\lambda_k^2)}.....(1.4.4c)$$

Where A_k is proportional to the number of particles per unit volume capable of vibrating with a natural frequency corresponding to wavelength 'λ_k' in vacuum . A plot of Eq . (1.4.3b) for the simple case of two oscillators is shown in Fig.3 .

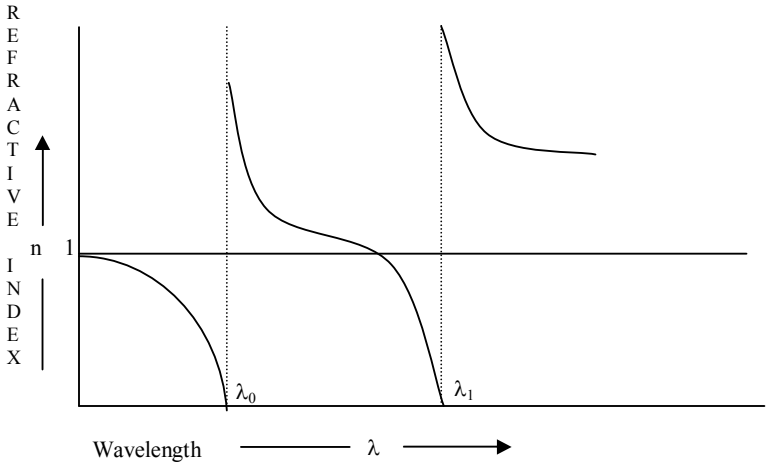


Fig.3. Theoretical dispersion curve given by sellemier's equation for a medium having two natural frequencies

For this case Sellmeier's equation is

$$n^2=1+\frac{A_0\lambda^2}{(\lambda^2-\lambda_0^2)}+\frac{A_1\lambda^2}{(\lambda^2-\lambda_1^2)} \quad .....(1.4.3d)$$

As λ approaches λ_o or λ_1 from the short wavelength side n tends to ∞ at resonance ($\lambda=\lambda_o$ or $\lambda = \lambda_1$) and as λ approaches λ_o or λ_1 from long wavelength side n tends to $+\infty$ at resonants wavelenghths. Between two resonances the curve passes through a inflection point and index of refraction becomes less than unity again. As λ approaches zero n tends to 1.

Rewriting eq 1.4.3b in the form

$$n^2=1+\frac{A_0}{(1-\lambda_0^2/\lambda^2)}+\frac{A_1}{(1-\lambda_1^2/\lambda^2)} \quad \text{.....(1.4.3e)}$$

it follows that in the microwave region on the long wavelengths and side of resonance λ becomes much greater than any of λ_k values so that the above equation reduces to

$$n^2=1+\sum_k A_k \quad \text{.....(1.4.3f)}$$

Sellmeier's equation is a great improvement over that of Cauchy in as much as it not only represented the experimental data not to close to the absorption lines but also agreed with the normal dispersion portion of the experimental curve more accurately than Cauchy's equation. In fact the Cauchy equation is nearly an approximation to Sellmeier's equation. However it can not represent the dispersion in the regions extremely close to the absorption band since n becomes $-\infty$ or $+\infty$ according as λ approaches λ_k from the short or long wavelength sides not only this physically impossible but the form of the curve near λ_k does not agree with the experiment

1.4.6 Helmholtz mechanical theory:

The dispersion curve in the region of absorption band is entirely different from that required by Sellmerier equation. Helmholtz pointed out that this discrepancy arises from the fact that in deriving Sellmerier equation the absorption of energy of the wave by the medium is not taken into account. Helmholtz postulated an oscillator which experiences a damping force of frictional character. This kind of resistance is essential if energy is to be continuously absorbed from the wave by the oscillator. He assumed the frictional force directly proportional to the velocity of an oscillator derived the following equation for dependence of n on λ which therefore takes into account of absorption

$$n^2(1-\kappa^2)=1+\sum_k \frac{A_k\lambda^2(\lambda^2-\lambda_k^2)}{(\lambda^2-\lambda^2{}_k)^2+g_k{}^2\lambda^2} \quad \ldots\ldots\ldots\ldots(1.4.4a)$$

The constant g_k is a measure of the strength of the frictional force. This equation is true for all wavelengths including those within an absorption band. This formula is identical with that obtained by electromagnetic theory. κ is called the extinction coefficient of the medium.

1.4.7 Experimental demonstration of anomalous dispersion

The anomalous dispersion of sodium vapour in the neighbourhood of the yellow lines was demonstrated by means of the experimental arrangement is shown in fig .4. Pieces of sodium are placed along the bottom of a steel tube T which is fitted with water cooled glass window at each end . The tube is partially evacuated to a pressure of about 1 to 3 cm. of Hg and on heating it at the bottom sodium vapour diffuses upward across the tube.

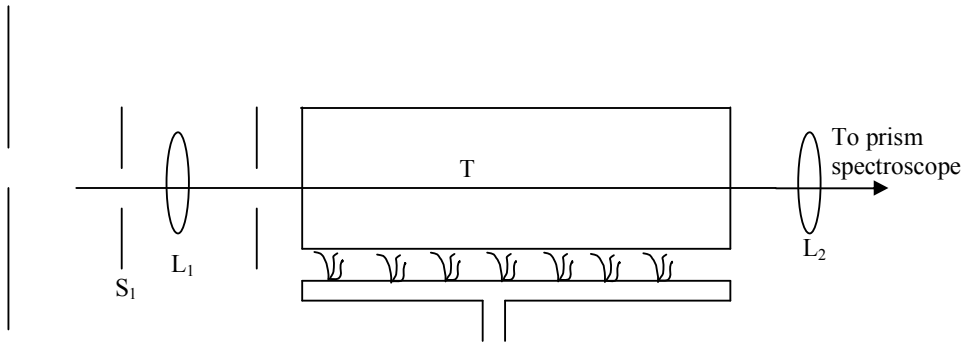


Fig..4, Experimental arrangement for observing the anomalous dispersion of sodium vapour.

Thus the density of vapour gradually increases from top to bottom and hence it is optically equivalent to a prism whose thickness increases down words resting upon its base, the refracting edge being perpendicular to the axis of tube. White light from the arc is focussed on a horizontal slit S_1 rendered parallel by the lens L_1 and then refracted on passing through the sodium vapour prism and finally brought to focus on the vertical slit S_2 of the prism spectroscope by means of the lens L_2. The refracting edge of the prism in the spectroscope is vertical. When tube is cold, the

gas is homogeneous and a white image of S_1 is formed on S_2 illuminating only one point and this is spread out into a narrow horizantal continuous spectrum in the focal plane of the spectroscope camera. When the sodium is vapourised, due its prismatic action the images of S_1 corresponding to different wavelengths are dispersed across the length of the slit S_2. For the spectral region remote from the absorption lines, the D lines, the index of refraction of the sodium vapour is nearly unity and so the horizantal continuous spectrum in the spectroscope remains constant . But as the D lines are approached from the short wavelength side, the light is deviated upwards in passing through the sodium vapour. Since the spectroscope lens inverts the image of the slit, the spectrum on the green side of the D lines is bent down in the spectroscope. Similarly, the spectrum on the long wavelength side of the D lines is bent up. The spectrum is depicted in the Fig.5.

Fig 5. Anomalous dispersion of sodium vapour.

which illustrates the anomalous dispersion for two neighbouring wavelengths 5896A , 5890A. When the density of vapour is too high the absorption lines broaden and it is not possible to observe the spectral region that lies between the two curved portions.

1.4.8 Electromagnetic theory of dispersion

An adequate theory of dispersion based on the electromagnetic theory of light was given by Lorentz by assuming that the atomic dipoles in the material medium execute forced vibrations under the action of the oscillatory electric field of the passing light wave. The atomic dipoles are created in the medium due to the relative displacements of the electrons and nuclei of the neutral atoms by the electric field of the light wave.

The total electric intensity on a electron of charge e and mass m at any point with in a polarized isotropic medium is $[\mathbf{E}+(4\pi/3)\mathbf{P}]$ and so the force on the electron is

e [**E**+(4π/3)**P**]. Here **E** denotes the electric intensity in the electromagnetic wave and (4π/3)**P** is the local electric intensity due to the molecular dipoles. On fundamental grounds, the local field correction (4π/3)**P** should be omitted in highly polarisable solids. It is also true for gases since the molecules are far apart. Therefore we shall take the electric force on the electron as **E**e. The magnetic force on the electron on account of its motion can be neglected. In addition to the electric forces the electron is also acted upon by two mechanical forces

(1) an elastic restoring force proportional to its displacement **r** from its equilibrium position i.e equal to -γ **r** where γ is a certain positive constant which depends on the structure and the properties of the particle.

(2) a damping frictional force against the motion of electron , which is proportional to its velocity i.e equal to - α(d**r**/dt) where α is a positive constant of proportionality, called the resistance constant. Thus the total resultant force on the electron is

$$[eE-\alpha(\frac{dr}{dt})-\gamma r] \quad .$$

By the Newton's second law the differential form of the motion of electron is

$$m\frac{d^2r}{dt^2}=eE-\alpha(\frac{dr}{dt})-\gamma r$$

$$\frac{d^2r}{dt^2}+\left(\frac{\alpha}{m}\right)\frac{dr}{dt}+\left(\frac{\gamma}{m}\right)r=eE \quad \ldots\ldots\ldots\ldots (1.4.5a)$$

In an isotropic medium, **E** and the displacement **r** in the same direction. We may therefore use the magnitude of vectors and rewrite Eq (1.4.5a) as

$$\frac{d^2r}{dt^2}+\left(\frac{\alpha}{m}\right)\frac{dr}{dt}+\left(\frac{\gamma}{m}\right)r=\left(\frac{e}{m}\right)E \quad \ldots\ldots\ldots\ldots (1.4.5b)$$

We first consider the free motion of electron in the absence of electric field.

Putting E=0 in (22.B) the equation of damped oscillation of electron is obtained.

$$\frac{d^2r}{dt^2}+\left(\frac{\alpha}{m}\right)\frac{dr}{dt}+\left(\frac{\gamma}{m}\right)r=0$$

Its solution is

$$r = r_0 \exp((-1/2)g_0 t + i\omega_0 t)$$ where $g_0=\alpha/m$ and $w_0^2 = (g/m) - (a^2/4m^2)$

If damping is low which we shall treat the resonance frequency of the system is

$$v_0 = 1/2\pi\sqrt{\gamma/m}$$

and the angular frequency is

$$\omega_0 = 2\pi v_0 = \sqrt{\gamma/m}$$

i.e we assume that the free period of electron is equals to that of undamped oscillator. Introducing the constants ω_0 and g_0 in eq. (1.4.5b) we obtain standard form of the differential equation of the forced oscillation of the electron.

$$\frac{d^2r}{dt^2}+g_0\left(\frac{dr}{dt}\right)+\omega_0^2 r=\left(\frac{e}{m}\right)E \quad \text{.....(1.4.5c)}$$

The polarization P is equal to the dipole moment per unit volume so that if there are N dispersion electrons per unit volume we may write

$$P = Ner \quad \text{.....(1.4.5d)}$$

Multiplying Eq. (1.4.5c) by Ne and using Eq. (1.4.5d) we get

$$\frac{d^2P}{dt^2}+g_0\frac{dP}{dt}+\omega_0^2P=\frac{Ne^2}{m}E \quad \text{.....(1.4.5e)}$$

Taking a monochromatic wave $E=E_0\exp(i\omega t)$ and using the symbolic differential operator D for differentiation with respect to t we have

$$[D^2+g_0D+\omega_0^2]P=\frac{Ne^2}{m}E_0\exp(i\omega t)$$

But $D=i\omega$ and $D^2=-\omega^2$

hence

$$P=\left(\frac{Ne^2}{m}\right)\frac{E}{(\omega_0^2-\omega^2+ig_0\omega)} \quad \text{.....(1.4.5f)}$$

$$\frac{4\pi P}{E}=4\pi\beta=\frac{4\pi Ne^2}{m(\omega_0^2-\omega^2+ig_0\omega)} \quad \text{.....(1.4.5g)}$$

The ratio $(4\pi P/E)$ is a complex number. Hence it follows that the dielectric constant is complex and so we conclude that the refractive index is also complex .

$$n^{*2}-1=4\pi\beta=\frac{4\pi Ne^2}{m(\omega_0^2-\omega^2+ig_0\omega)} \quad \text{.....(1.4.5h)}$$

In deriving Eq (1.4.5 h) we have assumed that the medium contains oscillators all of which have the same natural frequency. This simply means that the medium absorbs only one spectral line. To take into account the possibility of existence of other absorption lines we suppose that there are N molecules per unit volume and in each molecule there are f_1 oscillators whose constants are ω_1 and g_1; f_2 oscillators whose constants are ω_2 and g_2 and so on. Then Eq (1.4.5 d) becomes

$$P=\sum P_k = Ne\sum f_k r_k \quad \text{.....(1.4.5i)}$$

The differential equation of forced oscillation of electron characterized by constants ω_k and g_k may be written by replacing r by r_k in Eq (1.4.5 c) and then it is easy to get

$$\frac{d^2P_k}{dt^2}+g_k\frac{dP_k}{dt}+\omega_k{}^2P_k=\frac{Ne^2f_k}{m}E \qquad \text{.....(1.4.5j)}$$

hence

$$P_k=\left(\frac{Ne^2f_k}{m}\right)\frac{E}{(\omega_k{}^2-\omega^2+ig_k\omega)} \quad \text{.....(1.4.5k)}$$

Dividing Eq (1.4.5 k) by E and summing over all k we get

$$\beta=\sum\beta=\frac{Ne^2}{m}\sum_k\frac{f_k}{(\omega_k{}^2-\omega^2+ig_k\omega)} \quad \text{.....(1.4.5l)}$$

Thus, instead of Eq (1.4.5 h) we now get

$$n^{*2} - 1 = (4\pi Ne^2 / m)\sum_k f_k /(\omega_k^2 - \omega^2 + ig_k\omega) \ldots\ldots.(1.4.5\text{m})$$

This is called dispersion formula. The significance of the complex refractive index is that there is absorption of energy in the medium. The index of refraction has become frequency dependent. This is what is meant by dispersion.

1.4. 9 Normal dispersion. In the case of transparent substance there is no appreciable absorption. In the region remote from the natural frequencies of oscillators i.e. the absorption frequencies of the medium, the term $g_k\omega$ in the denominator of Eq(1.4.5 m) is so small that it may be neglected in comparison to $\omega_k^2-\omega^2$. The right hand side is also real i.e.κ =0 and Eq(1.4.5 m) simplifies to

$$n^2 - 1 = (4pNe^2 / m)\sum_k f_k / (w_k^2 - w^2) \ldots\ldots.(1.4.5\text{n})$$

The refractive index is real. Since

$$\omega_{red} < \omega < \omega_{violet} < \omega_k$$

The denominator $\omega_k^2-\omega^2$ is positive in the visible spectrum and is large at the red end than at the violet end. Red light is refracted less than blue light. This is normal dispersion.

The wavelength in vacuum is given by

$$\lambda = 2\pi c / \omega$$

and if

$$\lambda_k = 2\pi c / \omega_k \qquad \text{and} \quad A_k = (Ne^2 / \pi c^2 m) f_k \lambda_k^2$$

then Eq(1.4.5 n) becomes

$$n^2 = 1 + \sum_k A_k l^2 / (l^2 - l_k^2) \quad \text{.....(1.4.5o)}$$

Which is **Sellmeier's dispersion formula.**

Let us suppose that λ_k lies in the ultra violet. Assuming that the medium has oscillators of one kind only, we can rewrite Eq (1.4.5o) as

$$n^2 = 1 + A_k \left[1 - (l_k^2 / l^2)\right]^{-1}$$

$$n^2 = 1 + A_k + \left[A_k l_k^2 / l^2\right] + \left[A_k l_k^4 / l^4\right] + \text{.....} \quad \text{.....(1.4.5p)}$$

Stopping at the third term in the binomial expression.

$$A = 1 + A_k, B = A_k l_k^2 \text{ and}$$
$$C = A_k l_k^4$$

Eq (1.4.5p) Becomes

$$n^2 = A + B/\lambda^2 + C/\lambda^4 \quad \text{.....(1.4.5q)}$$

Which is Cauchy's dispersion formula. The dispersion of most transparent substances can be represented fairly well by the above formula.

Dispersion of most transparent substances can be expressed by assuming that the medium contained electrons with free period in the ultra-violet and also electrons with free period in the infrared. The dispersion formula may be written as

$$n^2 = 1 + \frac{A_v \lambda^2}{(\lambda^2 - \lambda_v^{\,2})} + \frac{A_r \lambda^2}{(\lambda^2 - \lambda_r^{\,2})} \quad \text{.....(1.4.5r)}$$

The index of refraction increases with decreasing wavelength, the normal behavior.

For gases n is only slightly greater than 1. We may therefore put

$n^2-1= (n+1)(n-1)=2(n-1)$.

Then Eq. (1.4.5n) reduces to

$$n - 1 = 2p \frac{Ne^2}{m} \sum_k \frac{f_k}{(w_k^{\,2} - w^2)} \quad \text{.....(1.4.5s)}$$

i.e. at low pressures the refractive index of a given gas should be proportional to N
i.e. to the density.

It should be remarked that in the visible region Hartmann's empirical formula represents the refractive of transparent substances much better than the theoretical formula, Eq. (1.4.5o), for n^2. But close to absorption bands Hartmann's formula fails. The theoretical formula is much better when we consider simultaneously ultra-violet as well as visible region of the spectrum.

1.4.10 Anomalous Dispersion:

The dispersion is always normal as long as the impressed frequencies do not include a natural frequency of electrons or of ions. In an extremely narrow spectral region of an absorption line like a D-line of sodium when the impressed frequencies contain the natural frequency of a D line, the normal course of dispersion is disturbed. In this narrow spectral region the effect of other absorption frequencies on the index of refraction is practically constant. Therefore in Eq. (1.4.5m)

only variable term is that which corresponds to one natural frequency of electron say ω_o. We may easily represent in Eq. (1.4.5m) the effect of all terms except the **0**th, together with the term unity by a constant n_0^2

$$n^{*2} = n_0^2 + \frac{4pNe^2}{m}\frac{f_0}{(w_0^2 - w^2 + ig_0w)}$$

$$n^2(1-ik)^2 = n_0^2 + \frac{4pNe^2}{m}\frac{f_0}{(w_0^2 - w^2 + ig_0w)}$$

Equating the real and imaginary parts we obtain

$$n^2(1-k^2) = n_o^2 + (\frac{4pNe^2}{m})\frac{f_0(w_0^2 - w^2)}{((w_0^2 - w^2)^2 + g_0^2w^2)} \quad \text{.....(1.4.5}t\text{)}$$

$$2n^2k = (\frac{4pNe^2}{m})\frac{f_0g_0w}{((w_0^2 - w^2)^2 + g_0w^2)} \quad \text{.....(1.4.5}u\text{)}$$

In the extremely narrow spectral region of the absorption line the range of ω is also narrow and we may replace $\omega_0 + \omega$ by $2\omega_0$ and the constant g_0 by another constant $h_0\omega_0$ so that Eqs. (1.4.5t) And (1.4.5u) become

$$n^2(1-k^2) = n_0^2 + A_0\frac{(2w_0(w_0 - w)}{(4(w_0 - w)^2 + h_0^2w^2)} \quad \text{.....(1.4.5}v\text{)}$$

$$2n^2k = A_0\frac{(h_0w_0w)}{(4(w_0 - w)^2 + h_0^2w^2)} \quad \text{.....(1.4.5}w\text{)}$$

In the case of gases and vapours at extremely low pressure $\kappa << 1$ so that we may neglect the term $n^2\kappa^2$ in Eq.(1.4.5v). Moreover, the refractive index n is so close to n_o that we may write

$n^2 - n_o^2 = 2n_o(n - n_o)$ in Eq. (1.4.5v).

$$n = n_0 + \left(\frac{A_0}{n_0}\right)\frac{[\omega_0(\omega_0 - \omega)]}{(4(\omega_0 - \omega)^2 + h_0^2\omega^2)} \quad \text{.....(1.4.5}x\text{)}$$

Equating dn/dω to zero, we easily obtain the conditions for the maximum and minimum values of the index of refraction n.

$$h_0^2\omega\omega_0 = 4(\omega_0 - \omega)^2$$

To a good approximation we may replace $\omega\omega_0$ by ω^2

$$h_0^2\omega^2 = 4(\omega_0 - \omega)^2$$

Then, the maximum in the index of refraction lies at

$$\omega/\omega_0 = 1 - (h_0/2) \quad \text{.....(1.4.5y)}$$

And the minimum at

$$\omega/\omega_0 = 1 + (h_0/2) \quad \text{.....(1.4.5z)}$$

The extreme values of n on the scale of ω/ω_0 are separated by h_0 and

$$n_{max} = n_0 + A_0/4n_0h_0$$

$$n_{min} = n_0 - A_0/4n_0h_0$$

At the maximum positions and minimum positions of n on the scale of ω/ω_0

$$2n^2k = A_0/2h_0$$

At resonance when $\omega = \omega_0$ we have from eq(1.4.5w)

$$A_0/2h_0 = (1/2)(2n^2k)_{max}$$

Hence the half width of resonance plot of against ω/ω_0 is equal to the separation of the maximum and minimum of n on the frequency scale . Fig. 6. is a plot of n and $2n^2\kappa^2$ against ω/ω_0 assuming n_0= 1, h_0 = 0.001 and A_0= 0.0001. In this particular case

$w/\ w_0 = 0.9995\ \ and\ n_{max} = 1.025$

$w/\ w_0 = 1.0005\ and\ n_{min} = 0.975$

$w/\ w_0 = 1\ \ and\ n = 1$

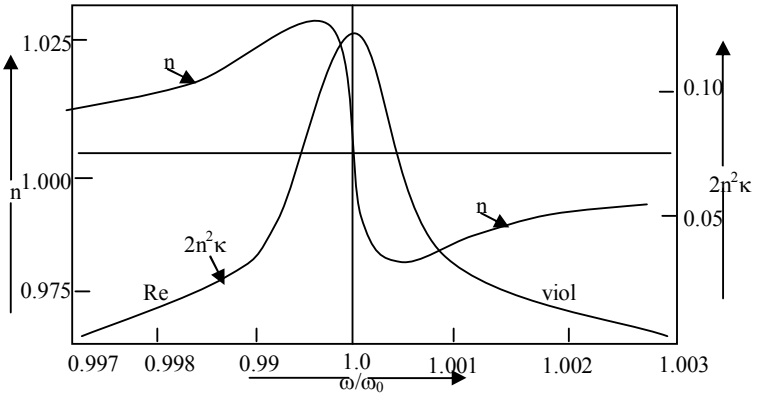


Fig..6. Plots of the index of refraction n and of the product $2n^2\kappa$ against ϖ/ϖ_0

Thus as the frequency of the light wave increases gradually from a value slightly less than ω_0, n increases rapidly pass through a maximum then drops to a minimum and finally again increases . Thus in the immediate neighbourhood of the absorption frequency ω_0 the index of refraction decreases with increasing frequency contrary to the normal behaviour viz n increasing with increase in ω. For similar plots of n and $2n^2\kappa$ are shown in Fig. 7.and would be in a direction opposite to that of ω/ω_0

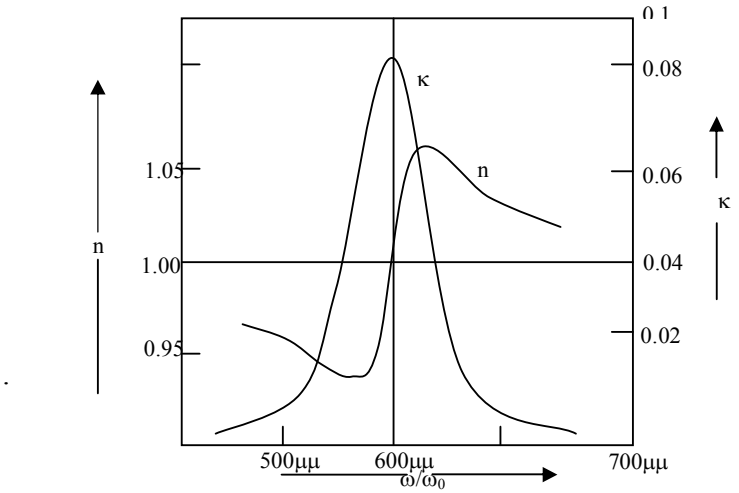


Fig.7. Plots of n and κagainst λ.

1.4.11 Summary: We came to know about the definition of dispersion and types of dispersions. The different theories of normal dispersion and complete explanation of the behaviour of normal and anomalous dispersion through the electromagnetic theory is given

1.4.12 Key words: Dispersion - normal dispersion – anomalous dispersion – Cauchy's, Hartmann's –Sellmeier's - Helmholtz – Lorentz theories - electromagnetic theory of dispersion .

1.4.13 Self assessment questions:

1. What do you mean by dispersion. Discuss it in the case of gaseous medium having both real and complex refractive index
2. Give the theory of normal and anomalous dispersion and describe how the latter has been studied in the case of sodium vapour.
3. Explain what is meant by anomalous dispersion. Give a simple theoretical explanation of the phenomenon .
4. Discuss the phenomenon of dispersion in absorbing media on the basis of electromagnetic theory.
5. What is Cauchy's dispersion formula. What are its limitations and under what conditions it fails. Discuss the necessary modification of the formula .
6. Write an essay on theories of dispersion

1.4.14 Reference books

1. Introduction to modern optics, B.K. Mathur
2. Optics, Born and Wolf

UNIT II -
LESSON-1

Characteristics of Lasers

Objective: - To describe the extraordinary properties of laser radiation and the determination of characteristic parameters.

Structure:

2.1.1 Introduction: -

During 20^{th} century spectroscopy has been enriched by the development of many new branches to understand the structure of matter through interaction studies with light .In the middle of the century (1954), the demonstration of MASER principle by C.H. Townes, MIT, Colombia, in microwave region using Ammonia, has made the scientific community to provide the experimental demonstration of LASER in1960. The LASER stands as an acronym for Light Amplification by Stimulated Emission of Radiation. This principle of stimulated emission of radiation was given by Einstein as early as in 1917. The special characteristics of this spectacular

source made the scientists to call in 1960 itself, as a tool of this (20^{th}) century. This is a source which can emit a kind of light of unrivalled purity and intensity not found in any of the previously known sources of electromagnetic radiation. In course of time the lasing action has been experimentally demonstrated and developed using various systems involving atoms, ions, molecules in liquids, gases and semiconductors etc., operating over major portion of EM spectrum.

The most striking features of LASER are

1. directionality,
2. brightness (light intensity)
3. extraordinary monochromaticity,
4. high degree of coherence.

These four are the distinguished properties responsible for the development of many new lines of studies in Raman spectroscopy, Holography, Non-linear optics, optical communication, military warfare and its utility in medical diagnostics and treatment.

2.1.2 Directionality:

It can be seen as a thin beam of light coming out from small power (1mw) laser in one direction with high intensity. The laser radiation comes out from a cavity formed by two highly polished plane parallel mirrors separated through a distance. The curvatures of the waves coming out are nearly planar. The directionality of the laser beam is expressed in terms of the full angle beam divergence which is twice the angle that outer edges of the beam makes with the center of the beam. The divergence is a measure of rate of expansion of the beam with the distance traveled. The divergence is measured in radians (2π radians equals to 360^0). Usually the laser beam divergence is about 1milli radian (10^{-3} radians). It is shown as the beam increases in size about 1mm for every meter distance of beam's travel. It travels d^2/λ distance as a parallel beam and then begins to spread linearly with distance due to unavoidable effects of diffraction (λ = wavelength of radiation). The angular spread is $\Delta\theta = \lambda/d$. Where 'd' is aperture diameter.

For a searchlight, the beam spreads about a kilometer for every kilometer distance of its travel. Laser beam reaching the surface of moon would be just a few kilometers wide, enabled the scientists to determine the distance between the surfaces of earth and moon with reasonable accuracy.

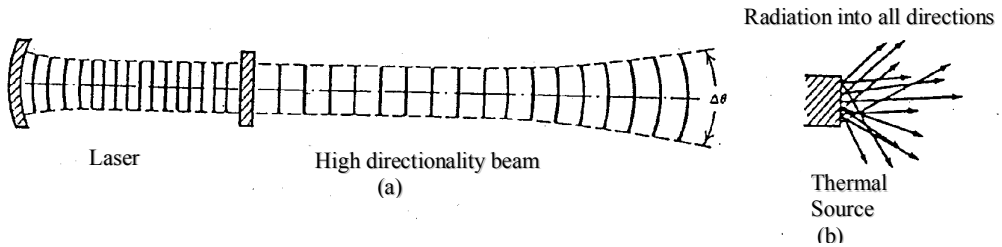


Fig.:2.1.1.

2.1.3 Monochromaticity:

Wave like nature of electromagnetic radiation is characterized by combination of time varying electric and magnetic fields propagating through space. The frequency of which these fields oscillate 'ν' and their wavelength in vacuum 'λ' are related by

$$\lambda\nu = c$$

where 'c' is the velocity of light in vacuum $\approx 3X10^8$ m/sec

In any medium $\lambda\nu = c/\eta = \upsilon$

where 'η' is the refractive index of the medium and 'υ' is velocity of light in the medium.

Conventional sources emit discrete electromagnetic radiations and some times broad band emissions. The discrete radiations are called lines with a particular value of 'λ'corresponding to the central portion of the line profile.

Though it is not seen with human eye the intensity profile of line varies with frequency/wavelength as shown in the figure.2.1.2.

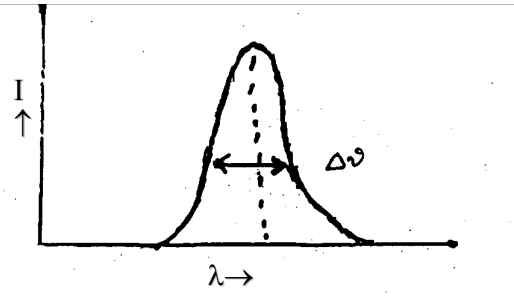


Fig: 2.1.2

The line width $\Delta\nu$ varies from line to line and from source to source. In particular, lasers produce the spectral emissions with less line width compared to conventional sources.

Even the lasers operating in single mode emit radiations with a minimum line width ($\Delta\nu$ = 500 Hz) at 6000A^o. So far, the source that can emit a single frequency or wavelength has not yet been designed. Hence the absolute monochromatic emission of radiation with $\Delta\nu = 0$ is an unattainable goal.

When the emission sources are compared in this respective $\Delta\nu$, the laser sources produce a high degree of monochromatic sine wave with small line width. So this exceedingly narrow line width contains small spread of frequencies. Hence even the single mode operating laser can not produce the absolute monochromatic radiation with $\Delta\nu = 0$.Monochromaticity is a relative parameter to the emission characteristics of various sources in spectroscopy. The qualitative comparison is shown in fig2.1.3 (a) and (b)

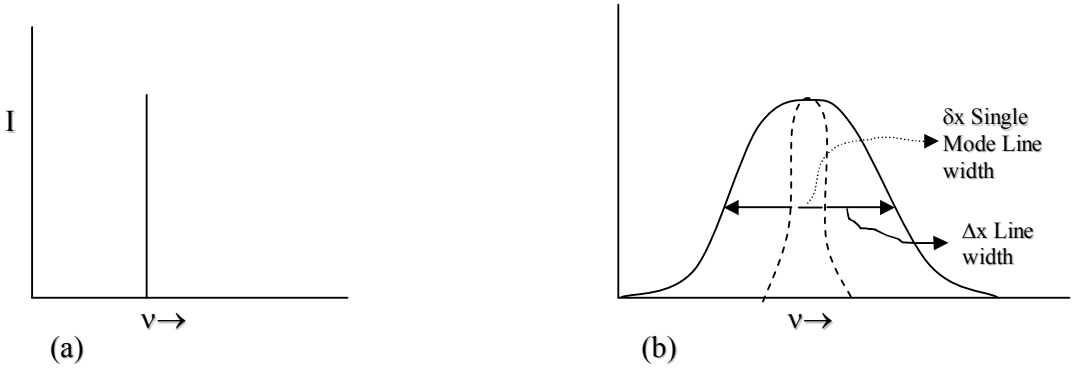


Fig : 2.2.3

I = intensity
ν = frequency
a) theoretical monochromaticity
b) $\Delta\nu$ line-width – 1500 MHz (FWHM)} for lasers
$\delta\nu$ single mode linewidth ≈ 500 Hz

2.1.4.Intensity: -

General sources emit radiation in all directions uniformly. So a small portion of power of the lamp enters into eye. In a laser beam the energy concentrated in small region specifically and spatially accounts for its large intensity. Thus even 1mW laser would appear more intense than 100 watt lamp.

The measurement of electromagnetic radiation is known as radiometry treating all wavelengths equally. Radiant energy and power are measured in joules and watts respectively. A subdivision of radiometry known as photometry is confined to the measurement of visible radiation. In most of the cases, radiometric units are used for comparing intensities of various sources. Terms, symbols units are given in the **Table 2.1.1**

Table 2.1.1

Radiometry			**Photometry**		
Term	**Symbol**	**Units**	**Term**	**Symbol**	**Units**
Radiant energy	Q_e	Joule	Luminous energy	Q_υ	Talbot
Radiant power (or)flux	ϕ_e	Watt (= 680 lumens at 555 nm)	Luminous power (or) flux	ϕ_ν	Lumen
Irradiance	E_e, I	Watt/m^2	Illuminance	E_ν	Lumen/m^2
Radiant intensity	I_e	Watts/sterad	Luminous intensity	I_ν	Lumens/ sterad
Radiance	L_e	Watts/m^2-sterad	Luminance (brightness)	L_ν	Lumens/ m^2-sterad

The radiometric and photometric units are linked by a standard luminosity curve drawn on the basis of human eye response to light of various wavelengths. The systems are connected by an

equality 680 lumens=1 Watt at 555 nm. The infrared and ultraviolet laser output is measured in radiometric units.

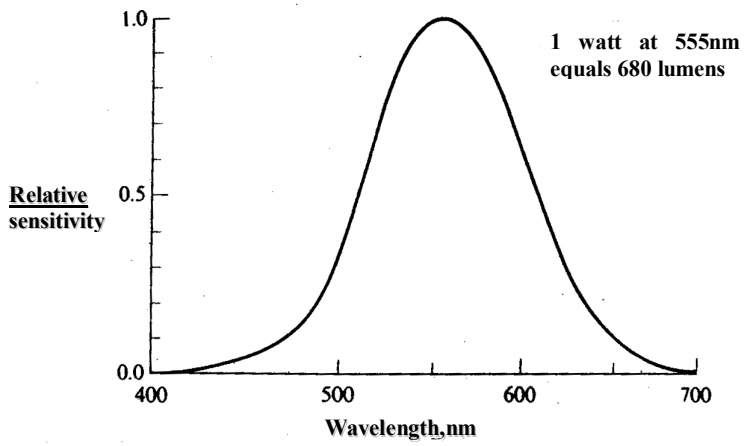


Fig:2.1.4

The power incident on a unit area, popularly known as intensity of light, is called irradiance in radiometry (watts/m^2) and illuminance in photometry (lumens/m^2).

The brightness of sources are compared in quantities radiance (watts/m^2 - sterad), the radiometric term and luminance (lumen/m^2 – sterad). The brightness is the luminous power per unit area of source per unit solid angle into which source is radiating.

The spectral brightness of sun (500 lumens/m^2–sterad–nm) is compared with 1mw Helium–Neon laser brightness (10^8 lumens/m^2–sterad – nm) at 633 nm.

2.1.5. Coherence: -

A high degree of order exhibited by laser radiation in respect of electric field is named as coherence. In a propagating electromagnetic wave there is complete correlation between electric field variation at any two points in space in the direction of propagation. It is shown that monochromatic wave can be coherent in both time and space. The electric field amplitude has the form

$$E(x,y,z,t) = A(x,y,z,t)\,[\text{Cos } wt + \theta(x,y,z)]$$

where $\theta_{(xyz)}$ is space–varying part of the phase enables to specify the variations in the field at any point of time 't' from the knowledge of A and θ.

Measurement of Coherence:

In general the light sources are classified into

1) Coherent
2) Partially coherent
3) Incoherent

based on the degree of coherence studied by using interferometers. The measurement of coherence is made from young's experiment, in which the two pinholes are illuminated and the interference pattern consisting of dark and bright bands is observed on a screen at some distance from pinholes. Fringe contrast is measured in terms of visibility factor (V.F.)

Where $$VF = \frac{I_{Max} - I_{Min}}{I_{Max} + I_{Min}}$$

I_{Max} = Maximum intensity of bright band

I_{Min} = Minimum intensity of dark band

VF assumes a maximum value of unity and a minimum value of zero. Since irradiance is a non-negative quantity, VF is a direct measure of degree of coherence of the light used to illuminate the two pinholes. Let I_1 and I_2 be the irradiances that result on the screen from each of the pinholes separately. If $I_1=I_2$,the fringe VF is equal to the degree of coherence of light at the two pin holes.

The light used, is classified on the basis of the value of VF when ($I_1=I_2$)

If VF=1, sources are perfectly coherent – electric fields are completely correlated,

=0, sources are incoherent, electric fields are completely incoherent,

=1>x>0, sources are partially coherent.

Physical reality is the existence of partially coherent sources only. (i.e. VF assumes only a value between 0 and 1)

The laser beam exhibits two types of coherence:

1) Spatial coherence or transverse coherence
2) Temporal coherence or longitudinal coherence.

2.1.6 Spatial Coherence:

When a wavefront falls on the two holes S_1 and S_2 whose separation can be varied, produces an interference pattern on the screen placed at a small distance. By varying the separation of S_1 and S_2 the visibility factor of the pattern also varies accordingly. The area of the wavefront, over which the pinholes are moved to from the fringes, is called as coherence area of the beam. This area is a measure of the spatial coherence or transverse coherence. This produces the spatial variation of coherence across the wavefront in the direction traverse to the propagation direction.

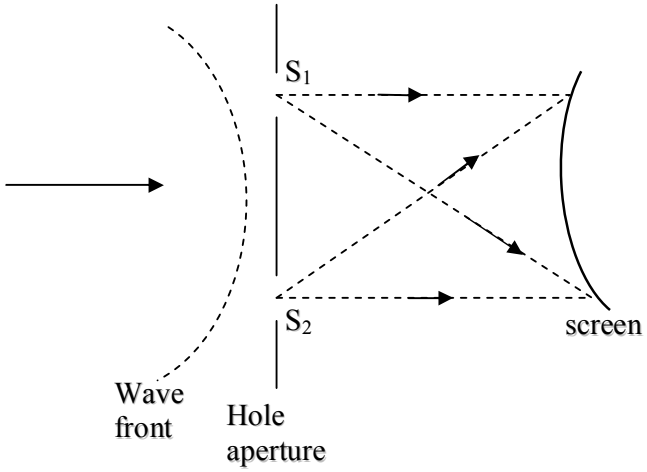


Fig:2.1.5

2.1.7. Temporal Coherence:

Temporal Coherence or longitudinal coherence is the property maintained (retained) along the length of the beam. This is studied by considering the field E of the beam at a point at time t_A (E x,y,z, t_A) and the field after some time t_B is E(x,y,z, t_B) at the same point. This is studied by arranging the experimental setup as shown in fig.2.1.6.

f

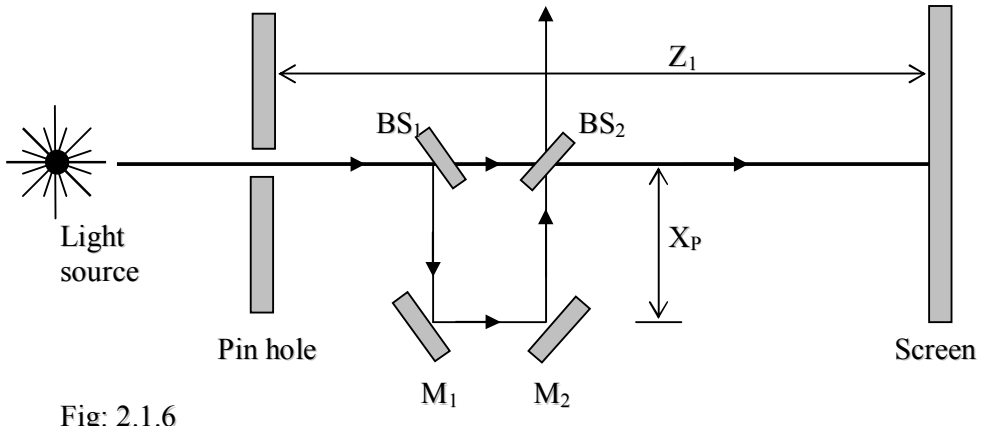


Fig: 2.1.6

Separation of time ($t_B - t_A$) in the wave field corresponds to a distance traveled by beam during that time in the direction of propagation. Let 'c' be the velocity of light and the distance traveled $\Delta Z = (t_B - t_A)c$.

Source beam from a pinhole is split into two equal beams by a beam splitter BS_1 and made the beam to travel in two different perpendicular directions. One path A passes straight in Z direction through BS_2 on to the screen. The other part of the beam takes route BS_1, M_1, M_2 BS_2 and then on to the screen.

The path A takes (Z_1/ e) time to reach the screen. The other path takes time ($Z_1 + 2x_D$)/e. By varying x_D through M_1, M_2 movement as shown, the correlation of the beam pattern is seen at different time delays. It is observed that for small change in x_D the visibility factor (VF) remains constant. As x_D increases, the fringe visibility decreases and ultimately reaches zero visibility at a particular value. This extra path distance $2x_D$ is known as coherence length l_C and the corresponding time is known as $t_c = l_c/c$.

$$t_c = \frac{l_c}{c}$$

However, there is a direct relationship between coherence length and monochromaticity as

$$l_c = \frac{c}{\Delta v}$$

where $\Delta\nu$ is the linewidth of the source radiation and

$$t_c = \frac{1}{\Delta\nu}$$

usually the measurement of coherence is evaluated by using interferometers.

Coherence is the manifestation of great regularity. Hence the amplitude of the field and polarization of EM wave produced by laser, can be predicted in time and space. In summary, all the properties of laser radiation are closely coupled with coherence property, the only aspect most clearly differentiates the laser from all other sources of light.

Zermike defined that the degree of coherence is equal to the visibility of fringes when path difference between beams is small and amplitudes $I_1=I_2$.The degree of coherence of light, exhibited at the two pin holes is taken as visibility. If the visibility is >0.85, the light at the two secondary sources is monochromatic and highly coherent in time and space. The monochromatic radiation is characterized by centre frequency ν_0 and linewidth $\Delta\nu$ in the interval between $(\nu_0 - \Delta\nu/2)$ to $(\nu_0 + \Delta\nu/2)$. The average time during which the sinusoidal emission exists is called coherence time t_c and the corresponding length of the wave is called coherence length $l_c = ct_c$., After t_c the correlation between the phase of the waves ceases.

2.1.8. Relation between the coherence of the field and size of the source: -

Let S be an extended source, on which S_1 and S_2 are two points radiates incoherently ($S_1S_2 = x$). The disturbances at S1 and S2 are represented by.

$$f(S_1) = \exp [i\omega t+\phi(S_1)]$$
$$f(S_2) = \exp [i\omega t+\phi(S_2)]$$

Where $\phi(S_1)$ and $\phi(S_2)$ fluctuate irregularly with time

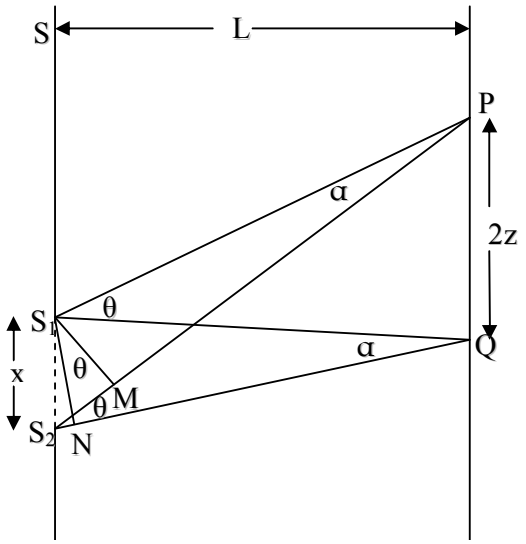


FIG. 2.1.7:Dependence of coherence on the size of the source

The disturbance E_p at p on observable plane due to S_1 and S_2 is

$$E_P = f(S_1)\exp -ik(S_1P) + f(S_2)\exp -ik(S_2P)$$

$$= \exp i\omega t\{ \exp [\phi(S_1)-ik(S_1P)]+\exp [\phi(S_2) - ik(S_2P)] \}$$

The corresponding expression for disturbance at Q is E_Q in the plane

$$E_Q = \exp i\omega t\{ \exp [\phi(S_1) -ik(S_1Q)] + \exp [\phi(S_2) -ik(S_2Q)] \}$$

If P and Q coincide, E_P and E_Q are equal and coherent.

$\therefore$ $ik(S_1)P = ik(S_1Q)$ and $ik(S_2P) = ik(S_2Q)$

Let us see how far P and Q can be separated before coherence disappears.

Let PQ = 2Z and L the separation of source and screen is the large compared to PQ and X.

M and N are feet of the perpendiculars from S_1 on S_2 P and S_2 Q respectively.

$$S_1P - S_2P \approx S_2M \qquad S_1Q - S_2Q \approx S_2N$$

$$S_2M - S_2N \approx X\theta \qquad X \approx L\alpha\ ; \quad \frac{2Z}{L} \approx \theta$$

θ and α are as shown in the figure

$$S_2M - S_2N \approx L\alpha \frac{2Z}{L} = 2Z\alpha$$

For a clear fringe system to appear, the path difference must be less than λ.

Coherence disappears when

$2\alpha Z = \lambda$

or $PQ = \frac{\lambda}{\alpha}$

It is a good measure of spatial coherence.

wave the electric field in space and time points in particular direction approximately described by a vector E perpendicular to both propagation Vector 'K' which describe the direction of travel of the wave and to the instantaneous direction of magnetic field of the wave 'H'. The direction of the electric field vector is the direction of polarization of light.

The light from general sources generally referred, as unpolarized or randomly polarized, since electric field vector has no preferred direction of orientation. Certain sources like some lasers, produce light waves of highly direction oriented electric fields. Such sources are referred as polarized sources.

The output of many lasers is linearly polarized as results of Brewster surface with-in the laser. Cutting the ends of laser rod or mounting windows of gas laser tube at the Brewster angle, assures that light of one polarization direction is transmitted out of the lasing medium to the reflecting mirrors and back into lasing medium with no loss. Then the angled-cut surfaces and windows are called Brewster cuts or Brewster windows. For light polarized perpendicular to the plane of incidence, there is a large loss at Brewster surface due to refraction out of the lasing medium. Usually this loss makes it impossible for light of this polarization to lase. As the preferred polarization can only lase, the output is of high degree of polarization.

2.1.10. Summary: -

The characteristic properties of laser radiation like directionality; monochromaticity, intensity and spatial and temporal coherence are described in a detailed manner. Experimental techniques to determine coherence parameters like coherence length, coherence area are described.

2.1.11. Key terminology: -

Laser-linewidth-directionality-monochromaticity-spatialcoherence-temporalcoherence - visibility factor-radiometry and photometry-polarization-Brewster angle windows and cuts.

2.1.12.Self assessment questions: -

1.Distinguish between conventional sources and lasers.

2.Describe the characteristic properties of laser.

3.Explain the technique to determine coherence length.

4.Describe units used in radiometry and photometry.

2.1.13.Reference books:-

1.Introduction to lasers and their applications D.C. Oshea, W.R.Callen and W.T. Rhodes.

2.Lasers and Non-linear optics by B.B.Laud

3.Lasers Theory and applications by K.Thyagarajan and A.K.Ghatak.

Unit II
LESSON-2

Basic principles of lasers

Objective: - To present the description of basic requirements for lasing and the importance of Einstein coefficients. To obtain the necessary condition for amplification of a signal in a medium. To study of various laser schemes and to discuss their power requirements.

Structure:

2.2.1. **Basic requirements for lasing:**

The basic requirements for stimulated emission of radiation are

1) Active Medium
2) Population inversion
3) Optical feedback system (Resonator).

The principles involved in these conditions are discussed.

2.2.2. **Active Medium:**

It was found that gas discharge tubes emit light at discrete wavelengths in the form of line spectra, as characteristics of an emitter. It was also found the gasses absorb discrete wavelengths on irradiation from white light or a continuous spectral source. It was shown that each atomic element had its signatures in the form of characteristic identifiable line spectra in both emission and absorption. Later it was shown that molecules also had similar characteristic band spectra in microwave, infrared and visible regions.

Neil's Bohr developed a theory that atoms can exist in certain discrete energy states and postulated that emission or absorption of light occurs due to certain allowed atomic transitions in between the discrete energy states of the same system. The frequency of the emitted or absorbed radiation is related to the energy difference of the two participating states in the transition.

$$E_2 - E_1 = \Delta E = h\nu$$

where 'h' is the Planks constant = 6.625×10^{-34} joule-sec
' ν' is the frequency of radiation.
and 'E_1' and E_2' are the energies of respective states.

The concepts of stable energy states and radiative transitions are valid as per the highly developed quantum theory. An atom in some arbitrary energy state E_i can be induced by radiation at

frequency 'ν_{ij}' to undergo a transition to higher energy state E_j if $\nu_{ij} = (E_j - E_i)/ h$, in absorption process. The emission is the reverse of this process in which a photon of energy '$h\nu_{ij}$' is emitted.

2.2.3.**Quantum numbers**:

Each quantum state for one electron atomic system is characterized by a unique set of quantum numbers.

n –principle quantum number represents size of the electron orbit.
l-orbital quantum number represents angular momentum (l h)
m-magnetic quantum number represents z-component of angular momentum (mh)
s-spin quantum number represents electron spin angular momentum(sh).

The electron spin designated by 's' can contribute to the energy of the atom when the electron is in a magnetic field. Atoms with same values of 'n' and 'l' with different values of 'm' and 's' often have the same total energy. Such quantum states are described as degenerate energy states. When atoms in these states in a magnetic field, the energy of each degenerate state increases or decreases by an amount that depends on the values of 'm' and 's'. The degeneracy is removed by the presence of the field. In multi-electron atoms each electron is associated with a set of quantum numbers according to Pauli's exclusion principle.

The change of atomic energy state during absorption or emission is called an atomic transition.The allowed transition as per selection rules are called as allowed transition while the highly improbable transitions are called forbidden transitions.

The selection rules $\Delta S=0$ and $\Delta L=0,+1,-1$ tell that transitions between singlet to singlet or triplet to triplet states are allowed but not from a singlet to triplet states.

The transition life time is a measure of probability of a transition i.e. the time during which the transition (either emission or absorption) takes place. The allowed transitions take place over a time varying from microsecond to nanoseconds. Some atomic systems remains in upper state of a forbidden transition much longer periods i.e., milliseconds. Such states are called metastable states. The

metastable states found in atoms and molecules are of great importance in the development of various laser systems.

2.2.4.**Absorption:**

Absorption and emission processes occur in a collection of atoms or molecules by virtue of allowed transitions in between the quantum states.

Consider an absorption spectrometer consisting of a cell with windows, to hold the atomic gas characterized by its energy in ground state E_o and an exited state E_1 ($E_1 > E_o$). A collimated beam from a source of continuous spectrum is allowed to pass through the cell and dispersed by a prism. Spectra are taken with out gas and with gas in the absorption cell. Two distinct types of continuous spectra with a smooth variation in colours with out gas in the cell and the other with dark lines at discrete wavelengths on bright background with gas in the cell, are recorded. The frequency of the absorption lines is related by famous quantum conditions.

The experimental observation of dark lines at certain frequencies like ν_{10} etc, is an indication that certain frequencies are totally absorbed under suitable quantum conditions. Hence the intensity loss at certain frequencies corresponds to characteristic, absorption spectra of the atomic gas. The excited atoms will emit at the same frequencies partially sometimes or the excited energy may be dissipated.

The Fraunhoffer line spectrum is an ideal example of absorption spectra.

$(E_1-E_o)/h = \nu_{10}$ (frequency in Hz).

The dark lines at discrete wavelengths are called characteristic absorption spectrum of the atomic gas in the cell. This is due to absorption of the quantum of radiation ($h\nu_{10}$). The ground state atoms are stimulated to absorb this discrete quantum of energy from incident radiation. Hence this transition induced by the incident radiation is often referred as stimulated absorption. This process of stimulated absorption is not allowed endlessly till the ground state population is exhausted, since the emission processes occur from the excited species resulting the excited atom to come to the ground state. The emission process makes the excited atoms to come to the ground state in a fraction of nanoseconds time resulting a photon of energy ($h\nu_{10}$).

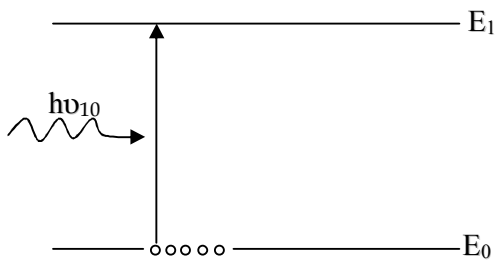


Fig: 2.2.1(a) Absorption

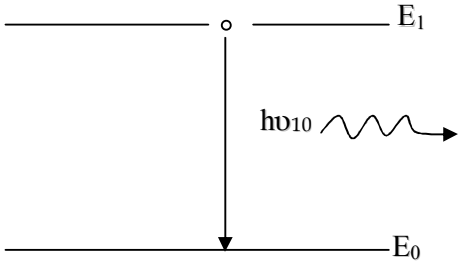


Fig 2.2.1(b) Emission

2.2.5.**Emission:**

Emission process is of two types.

1) Spontaneous emission
2) Stimulated emission.

The photon of energy $h\nu_{10}$ is emitted from the excited atoms spontaneously in random direction. The photons are emitted from various excited atoms are independent in direction and phase. Under certain conditions these excited species are also induced to undergo induced emission or stimulated emission in which the emitted photons are in phase and unidirectional with the stimulated photon. This stimulated emission occurs when the medium is provided with:

1) Population Inversion
2) Stimulating Radiation
3) Feedback mechanism

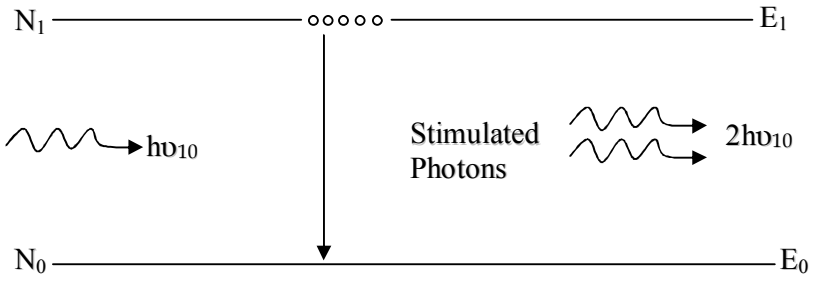


Fig 2.2.2. Stimulated Emission

This stimulated emission process is essentially part of operation of LASER (Light Amplification by Stimulated Emission of Radiation). The photons emitted in phase increase the amplitude on constructive basis resulting high intensity of LASER radiation ($I = N^2a^2$) where

I = intensity, N= number of photons, a = amplitude of the field.

2.2.6. **Population inversion:**

The population of the excited state (N_1) should be more than the population of the final state of the transition (ground state in thIs case) i.e. $N_1 > N_0$, This is created in the medium by means of pumping mechanism using optical, electrical, chemical etc. techniques.

2.2.7. **Stimulating Radiation:**

Some of the excited species emit spontaneous emission of radiation of energy $h\nu_{10}$ which is available in the medium itself. Thus a spontaneous photon of energy $h\nu_{10}$ will act as a stimulating photon in lasing action in an inverted medium.

The important condition is that the energy of the stimulating photon must be equal to the energy difference of the levels participating in the transition resulting stimulated emission.

A comparison is made between spontaneous emission and stimulated emission of radiation.

2.2.8.**Distinction between spontaneous emission and stimulated emission**

Spontaneous emission	Stimulated emission
1. Proportional to population of the upper level participating in a particular transition (N_U)	1. Proportional to the population of the level participating in stimulated process 'N'
2. No need of external stimulant (except the excitation source)	2.External stimulating photon of having same energy ($h\nu_{ul}$)as that of the E_u-E_l, is essential.
3. Proportional to spontaneous emission rate/ probability. (A_{ul})	3. Proportional to stimulated coefficient B_{ul} or B_{lu}
4. No need of population inversion.	4. Population inversion is required essentially. $N_u - N_l$ = positive (ΔN_{ul})

5. No need of external agency.	5. Proportional to the density of stimulating energy density. $\{\omega(\nu)\}$
6. It is a multi-directional process.	6. It is unidirectional emission.
7. No phase relation between emitted photons.	7. Stimulating and stimulated photons will be in phase.
8. Occurs from the levels excited spontaneously (nano-sec process).	8. Levels with long life-times are the best candidates.
9. Simple excitation is sufficient.	9. Pumping technique is required to create population inversion.
10. Photons of short duration are emitted (incoherent radiation).	10. Long coherent beams are obtained with temporal coherence.
11. No need of feedback mechanism.	11. Feedback mechanism (resonator) is needed.
12. Two levels are sufficient always.	12. Generally more levels are needed (at least) for an efficient operation. In certain cases two levels are sufficient.

2.2.9. **Feedback Mechanism**:-

A combination of two mirrors plane or curved or combination with different reflectivities (100% and 95%) placed at the end of the cell containing the medium to provide feedback mechanism, This combination is called as “resonator”.

2.2.10.**Einstein Relations**:

In 1917, Albert Einstein showed that the three processes – stimulated absorption, spontaneous emission and stimulated emission occur in a collection of atoms in an interaction with suitable electromagnetic radiation. In a quantified study, mathematical relation has been established between these three processes.

Considering only an excited state with E_1 energy and the other ground state at energy E_0 and the respective population in these states N_1 & N_0, the rate at which the atoms spontaneously drops to the lower state is designated by a constant A_{10}.

This is called as Einstein coefficient for spontaneous emission and the reciprocal of A_{10} is the time for the transition $1 \rightarrow 0$ to occur. Hence this is the transition life time of the state T_{10} ($T_{10}=1/A_{10}$).

The number of atoms undergo spontaneous emission is given as N_1A_{10}.

For stimulated absorption and emission induced radiation should be available. Let this spectral density of radiation energy be ρ_ν (joules-sec/m^3) at the transition frequency ν_{10}.

The stimulated absorption transition rate = $N_0 \rho_\nu B_{01}$.

Where B_{10} is the constant of proportionality called 'Einstein coefficient for stimulated absorption'.

Similarly the stimulated emission transitions take place per second i.e.

Stimulated emission transition rate.= $N_1 \rho_\nu B_{10}$ (where B_{10} is stimulated emission coefficient for level '1' to ground state '0').

At conditions of thermal equilibrium

1. The number of upward transitions is equal to downward transitions.
2. The coefficient of stimulated absorption and stimulated emission are equal ($B_{01} = B_{10}$)

The effective spectral radiation energy density associated with light beam is related to spectral irradiance of the beam by $\rho_\nu=I_\nu/c$.

On substitution it can be shown that

$N_0 (I_\nu/c) B_{01}=N_1 A_{10}+N_1(I_\nu/c) B_{10}$.

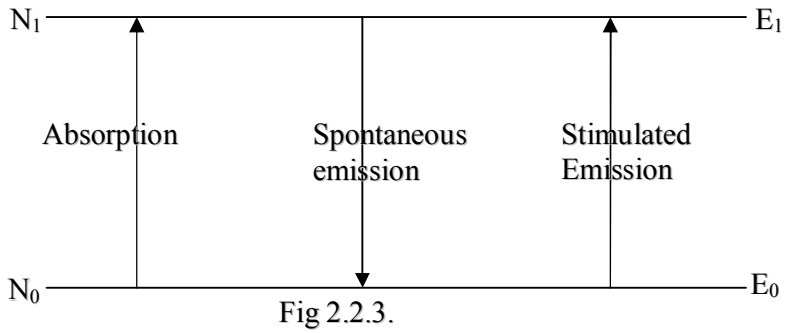


Fig 2.2.3.

The Einstein coefficient A_{10} and B_{10} are related as

$$A_{10} = (8\pi h\nu_{10}{}^3/ c^3) B_{10}.$$

The coefficient of spontaneous emission radiation A_{10} is proportional to ν^3 times B_{10}. This cubic dependency on ν accounts for the difficulty of attaining laser action (stimulated emission) at high frequencies of electromagnetic spectrum i.e. at X- rays and Υ -rays. Hence an attempt has been made by C.H.Townes from MIT in 1954 to demonstrate stimulated emission from ammonia maser successfully at microwave frequencies.

2.2.11 **Amplification in a medium**

In an electromagnetic interaction in a medium, the amplification or absorption of a signal depends upon the gain coefficients.

Consider only a single radiative transition between two energy states E_i and E_j (E_j>E_i) and the incident radiation be monochromatic at the transition frequency $\nu_{ij} = E_j$-E_i/h. The irradiance in the beam is a function of position along z axis .Consider the change I(z) after traversing a small distance dz in between the planes at z and z+Δz in the gas.

$$\Delta I_z(z) = I_{(z+\Delta z)} - I_{(z)} \quad \text{--------(1)}$$

Let the constant of proportionality is absorption coefficient

$$\Delta I_z = -\alpha\, I_z\, \Delta z \text{---------- (2)}$$

The negative sign indicates the reduction in irradiance due to absorption since 'α' is a positive quantity

In differential equation form

$$dI_{(z)} / dz = -\alpha\, I_{(z)} \text{ ----------(3)}$$

Solving the equation,

$$I_{(z)} = I_0\, e^{-\alpha z} \quad \text{----------(4)}$$

This is known as Beer's law.

$I_{(0)}$ is the irradiance of the beam at the input plane.

ΔI is usually negative due to normal attenuation of the incident beam. But under certain circumstances ΔI may be positive i.e. the incident beam at certain frequency can be amplified as it travels through the medium. This is possible only when more number of atoms in the element are in the excited state E_j. Hence no absorption but stimulated emission be possible when the incident radiation is of the frequency ν_{ij}. Due to stimulated emission the energy is radiated at the same frequency ν_{ij} and also in phase with the incident radiation. Thus the amplification is responsible in lasing action.

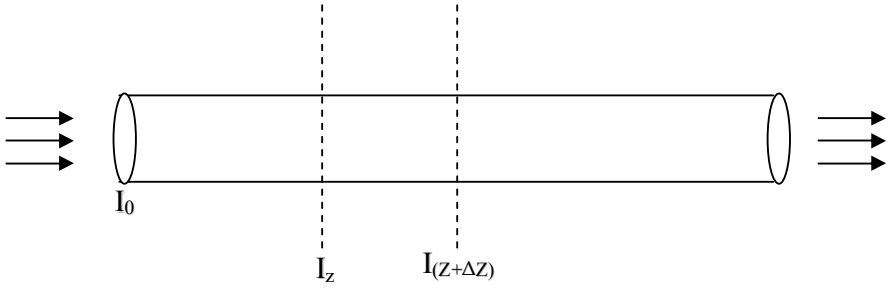


Fig: 2.2.4. Irradiance change

Let n_{ij} of energy quanta (photons) lost per second to the collimated beam as it travels a volume element of thickness Δz

$$-\frac{dn_{ij}}{dt} = N_i \frac{I_Z}{c} B_{ij} - N_j \frac{I_Z}{c} B_{ji} \quad \text{----------(5)}$$

$$= B_{ij}(N_i - N_j)\frac{I_Z}{c} \quad \text{---- ----------(6)}$$

where N_i and N_j are the number of atoms in the lower and upper energy states respectively in the volume element. B_{ij} and B_{ji} are stimulated absorption and emission coefficients respectively.

Ignoring scattering and spontaneous emission loss, the photon loss in terms of the beam irradiance at Z and Z+ΔZ is equated to the difference of populations and the absorption coefficients as

$$- dn_{ij} / dt = [I_{(z)} - I_{(z+\Delta z)}] A / h\nu_{ij} = B_{ij} [N_i - N_j] I_{(z)} / c \text{ ----------(7)}$$

A = area of cross-section of the beam, $I_{(z)}$ and $I_{(z+\Delta z)}$ are beam irradiances at Z and Z+ΔZ positions.

If Δz is sufficiently small, the above equation can be written as

$$(- dI_{(z)} / dz) \Delta z [A/ h\nu_{ij}] = B_{ij} (N_i - N_j) I_{(z)} / c \text{ ----------(8)}$$

Substituting for α from equ. (3) and dividing by I (z)

$$\alpha = [B_{ij} (N_i - N_j) / A \Delta z] [h\nu_{ij} / c] \approx B_{ij} (N_i - N_j) h\nu_{ij} / c \text{ ---------- (9)}$$

Let $(N_i / A \Delta z)$ and $(N_j / A \Delta z)$ represent number of atoms per unit volume or atomic population densities of two energy states.

The equation provides the link between absorption coefficient 'α' and population difference of two states. Normally N_i is greater than N_j and α is positive quantity, absorption is expected.
In case $N_j > N_i$, then α is negative, the quantity (-αz) i.e. exponent in the equation is positive. Hence the irradiance of the beam grows with distance as per the relation

$$I_{(z)} = I_o e^{\beta z} \text{ ----------(10)}$$

Where β is so called small signal gain coefficient.

$$\beta = -\alpha \quad , \alpha < 0$$

$$\beta = \beta_{ij} (N_j - N_i) h\nu_{ij} / c \text{ ----------(11)}$$

Under this condition $(N_j - N_i)$ the beam is amplified instead of the expected attenuation.
This is achieved by a process known as 'pumping' to reverse the situation. This attainment of this condition, known as 'population inversion' is required for light amplification.

2.2.12. Laser pumping: Pumping is the technique used to create population inversion among the energy levels participating in laser transition in a medium. Usually the distribution of population of atoms among various energy levels is governed by Boltzmann's principle in thermal equilibrium.

2.2.13. Boltzmann's principle and the population of energy levels:

At room temperature the atoms of a gas in thermal equilibrium, majority of atoms remain in the ground state configuration and a small fraction will be in the excited states due to the available thermal energy. By increasing temperature some more atoms are made to occupy higher excited energy states.

Boltzmann's principle states that the population of atoms in an higher energy level N_i is equal to

$$N_i = N_0 e^{-Ei / kT}$$

Where

N_0 is the number of atoms in the ground state.
E_i = energy measured relative to ground state
T = temperature in absolute scale (Kelvin) in degrees
k = Boltzmann's constant 1.38 x 10^{-23} joule/k.

This relationship is called Boltzmann's ratio.

$$N_j / N_i = e^{-(Ej - Ei)/kT} = e^{-\Delta Eji / kT}$$

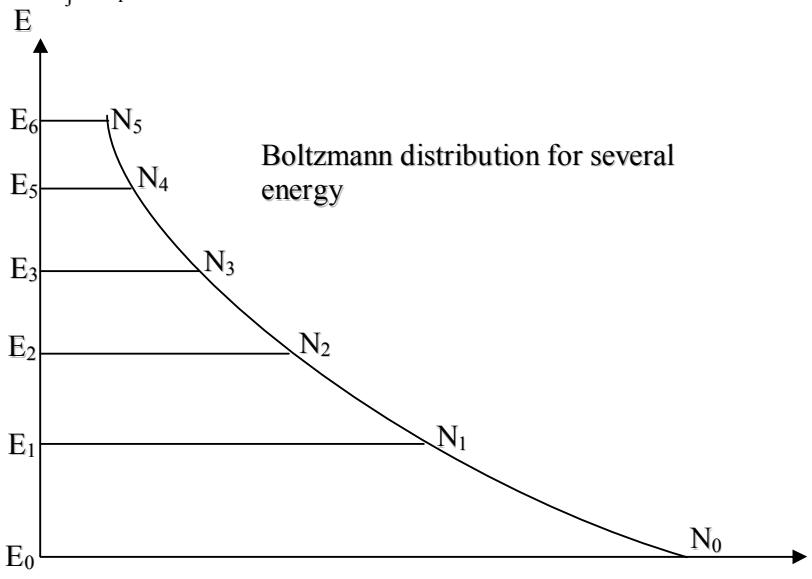


Fig: 2.2.5

The E_0, E_1, E_2, E_3, E_4, E_5 are energy levels with respective populations N_0, N_1,N_2, N_3, N_4, N_5. Where $\Delta E = kT$, the Boltzmann's ratio is 1/e.

$h\nu_{ij} = kT$

At room temperature T = 300k, the corresponding frequency is $6x10^{12}$ Hz in far infra red part of the spectrum ($\lambda = 50$ μm). As the temperature increases to high value, the greater occupancy of higher energy levels is responsible for new lines in the absorption spectrum.

2.2.14. Attainment of population inversion

Under thermal equilibrium conditions the population of higher energy states can never exceed the lower energy state (i.e $N_j \leq N_i$). The essential condition for lasing is population inversion obtained in the medium through application of external source of energy. Various pumping techniques are in vogue to create population inversion among the energy states responsible for stimulating emission of radiation. The three general schemes adopted are

1. Two-level pumping schemes of ammonia maser

2. Three or four level optical pumping used in solid state lasers (Ruby and ND:YAG laser)

2.2.15. Two level pumping:

This scheme is unique to ammonia (NH_3) beam maser. Maser is an acronym for Microwave Amplification by Stimulated Emission of Radiation. This was the first successful demonstration of amplification of electromagnetic radiation (microwaves) by stimulated emission of radiation(1954) by Gordon, Zeiger and Townes at Columbia University.

In NH_3 maser, the resonance transition occurs between two quantized levels of particular mode of vibration of the molecule at frequency 24 GHz(24×10^9 Hz) in microwave region. At room temperature ,$h\upsilon/kT<<1$ for υ=24 GHz. The Boltzmann ratio is near unity ie the number of molecules in higher energy states nearly equals but never exceed the number in the lower state. The population

inversion on this transition is achieved by physical separation of higher energy molecules from lower energy molecules in an electric field gradient produced by using a quadrupole focusing system as shown in the figure 2.2.6.

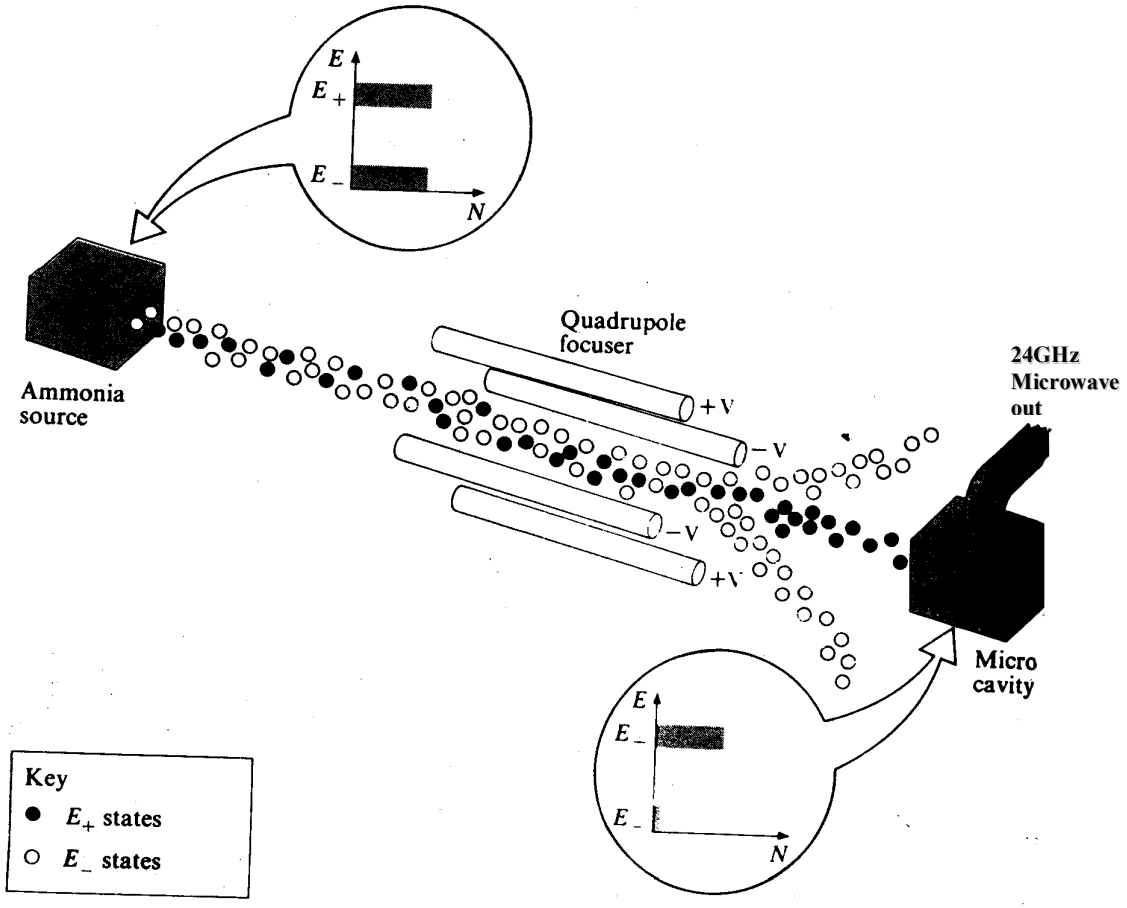


Fig:2.2.6.NH_3 laser pumping

The molecules in higher energy state experience a net force in the direction of zero electric field ie towards the axis of focuser. The molecules in lower energy state experience a net force away from the axis. Thus the lower energy molecules condense on the cool walls of the focuser and are effectively removed. The majority of molecules remain in the beam are of higher energy state enter the microwave

cavity resonant at transition frequency. As population inversion exists in the cavity, any small signal at transition frequency is amplified by stimulated emission. This amplified disturbance builds to a sustained oscillation with a narrow linewidth of NH_3 transition (6000Hz). Hence this is used as a frequency standard in microwave spectrum.

In this two-level experiment population inversion is possible by physical separation of molecules on the basis of energy content. By conventional technique of stimulated absorption using 24GHz (optical pumping), it is not possible to achieve population inversion since the gas can absorb only till the population in the two states are equal.

2.2.16 Optical pumping: three and four level schemes

The three level and four level schemes are used in optically-pumped, doped solid state lasers to create population inversion in the active medium indirectly to the upper state of the transition involved.

Three-level scheme:- This was proposed by Bloembergen in 1956 at Harvard University. Consider a three level system consisting of energy levels E_1, E_2 and E_3 (1,2,3) with the respective populations N_1, N_2, and N_3. The pump is applied between 1 and 3(1→3 transition) and lasing transition is $2 \rightarrow 1$. The pump lifts from level 1 to level 3 from which they decay to level 2through non-radiative fast decay. Thus pump effectively transfers atoms from level 1 to level 2 via level 3. Since level 1 is the ground state , more than 50% of the atoms of level 1 have to be lifted to level 2 to create inversion.

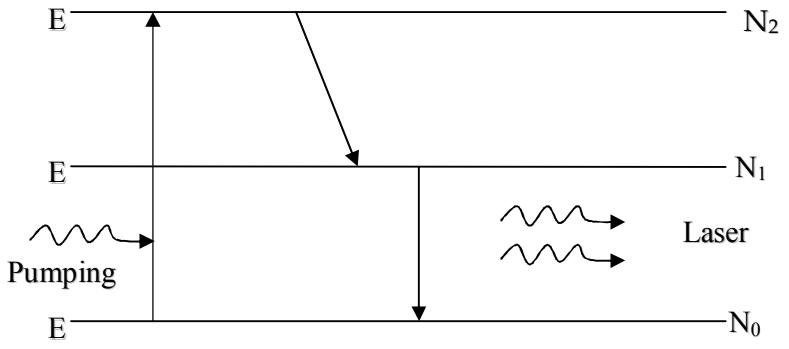


Fig : 2.2.7 Three level system

Let $N = N_1 + N_2 + N_3$ ------(1)

be the number of atoms per unit volume in levels E_1, E_2, and E_3.

Change in the population of level 3 is described by the equation

$$\frac{dN_3}{dt} = Wp\,(N_1 - N_3) - N_3\,T_{32} \text{ ---------- (2)}$$

where W_pN_1 represents the number of induced transitions per unit time per unit volume resulting $1 \rightarrow 3$ transition. Similarly W_pN_3 represents number of stimulated transitions per unit time with $3 \rightarrow 1$ transition. N_3T_{32} term represents extremely fast non-radiative process.

Further

$$T_{32} = A_{32} + S_{32} \text{ ---------- (3)}$$

Where A_{32} is Einstein Coefficient corresponding to radiative transition and

S_{32} is non-radiative transition rate from 3 to level 2

The rate of change of population of level 2

$$\frac{dN_2}{dt} = W_l\,(N_1 - N_2) - N_3\,T_{32} - N_2T_{21} \text{ ---------- (4)}$$

First term represents stimulated transitions between 1 and 2, second term represents spontaneous transitions ($3 \rightarrow 2$) and third term represents spontaneous transitions ($2 \rightarrow 1$) W_l is proportional to B_{21}Einstein coefficient for stimulated emission.

$A_{21} = T_{21}$ ---------- (5)

Similarly

$$\frac{dN_1}{dt} = W_p\,(N_3 - N_1) + W_l\,(N_2 - N_1) + N_2T_{21} \text{ ---------- (6)}$$

$$\frac{dN_1}{dt}+\frac{dN_2}{dt}+\frac{dN_3}{dt}=0$$ consistent with the equation (1) -------- (7)

on simplification

$$\frac{N_2-N_1}{N}=\frac{Wp-T_{21}}{Wp+T_{21}}$$ ---------- (8)

$W_p > T_{21}$ and population inversion is independent of laser transition energy

$$W_p > T_{21} = A_{21} = \frac{1}{t_{sp}}$$ ---------- (9)

The minimum power P to be spent is $\frac{N_2 h\nu_p}{t_{sp}}$ per unit volume

$$P=\frac{N_2 h\nu_p}{t_{sp}}$$

$$N_2 \approx \frac{N}{2} \qquad P_t=\frac{N}{2}\left[\frac{h\nu p}{t_{sp}}\right]$$ ---------- (10)

For Ruby laser $\rightarrow P_t \approx 1100$ W/cm^3

Advantages of Ruby :-

Absorption band of ruby crystal is very well matched to pump lamps. All most all the atoms pumped to 3 drop down to level for radiative transition. Line width of radiation is very narrow. This scheme is used in the operation of ruby laser. This requires very high pump powers since the terminal level of lasing transition involves the ground state, more than half of the ground state atoms must be pumped to the upper state to achieve population inversion. However, this intense pumping requirement is greatly reduced in four level pumping scheme.

The Four level System : Atoms from level 1 ground state are excited to level 4 by pump light. From level 4 excited atoms decay to level 3 throught non-radiative transition. Level 3 being metastable (life time 10^{-3}sec) forms the upper level and level 2 forms the lower (final) level of laser transition with very short life time so that atoms relax down immediately to level 1. If the rate of relaxation from 2 to 1 is faster, population can easily be developed with small pumping rate. If level 4 is broad, large amount of energy from pump light can be utilized for lasing action.

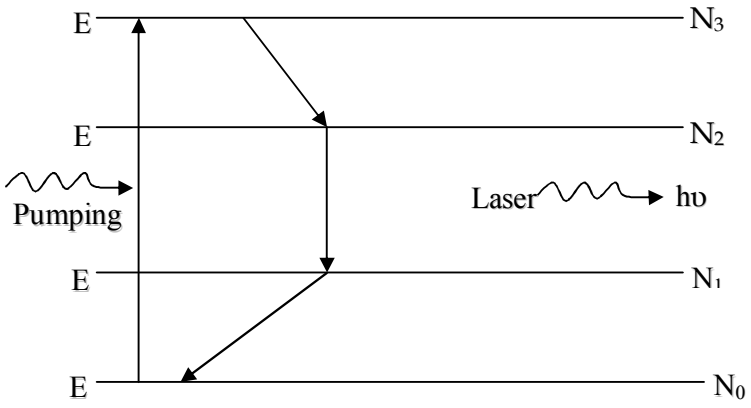


Fig 2.2.8 Four level

The upper laser level (3) must be narrow and laser level (2) must also be sufficiently above level (1) so that the population in level (2) is negligible at operating temperature.

The rate equations for the four level system are

$$\frac{dN_4}{dt} = W_p(N_1 - N_4) - T_4 N_4 \quad \text{----------(1)}$$

Where W_p is pump rate ; T's represent total relaxation time.

$$T_4 = T_{43} + T_{42} + T_{41} \quad \text{----------(2)}$$

$$\frac{dN_3}{dt} = T_{43}N_4 + W_l(N_2 - N_3) - T_3 N_3 \quad \text{----------(3)}$$

$$T_3 = T_{32} + T_{31} \quad \text{----------(4)}$$

$$\frac{dN_2}{dt} = T_{42}N_4 + W_l(N_3 - N_2) - T_{21}N_2 + T_{32}N_3 \quad \text{---------(5)}$$

$$\frac{dN_1}{dt} = W_p(N_4 - N_1) + T_{41}N_4 + T_{31}N_3 + T_{21}N_2 \quad \text{----------(6)}$$

$$\frac{dN_1}{dt} + \frac{dN_2}{dt} + \frac{dN_3}{dt} + \frac{dN_4}{dt} = 0 \quad \text{----------(7)}$$

which is consistent with requirement that total number of atoms

$$N = N_1 + N_2 + N_3 + N_4 \text{ ----------(8)}$$

must be constant.

At steady state $\frac{dN_1}{dt} = \frac{dN_2}{dt} = \frac{dN_3}{dt} = \frac{dN_4}{dt} = 0$ ----------(9)

For good laser action : $T_3 << T_{43}$ and $T_{21} >> T_{32}$ and also $T_4 = T_{43}$

$$\frac{N_3 - N_2}{N} = \frac{W_p / T_3}{1 + W_p / T_3} \quad \text{----------(10)}$$

The threshold population inversion density is much less than the total number of atoms per unit volume

$$\because (N_3 - N_2)_t << N \quad \text{----------(11)}$$

In the case of Nd:YAG Laser $W_{pt} \cong \frac{(N_3 - N_2)_t}{N} T_3$ --------(12)

This is much smaller than that is required in the case of ruby laser (three level system)

$P_t \approx 5$ W / cm^3

In the case of He-Ne laser $P_t = 1.1 \times 10^{-3}$ W /cm^3

2.2.17. Summary of lesson :- Basic phenomenon like absorption, spontaneous emission and stimulated emission and the requirements for lasing action like active medium, population inversion and feedback mechanism are described. The relation between the Einstein coefficients and their importance have been discussed. The three-level and the four-level laser schemes have been discussed including the parameters responsible for their efficiency.

2.2.18. Key Terminology :– Absorption-Spontaneous emission-Stimulated emission-Active medium-Population inversion-Resonator-Einstein coefficients- Boltzmann ratio- Three-level and Four-level laser systems.

2.2.19. Self assessment questions:-

1.Distinguish between Spontaneous and Stimulated emission.

2.Obtain the relation between Einstein coefficients and discuss their importance.

3.What is Boltzmann Ratio? Discuss the situation when $\Delta E = kT$.

4.Compare the power requirements in a three-level and four-level laser systems.

5.Obtain the condition for an amplification of the signal in a medium.

2.2.20. Reference books :-

1. Lasers theory and applications- K.Thyagarajan, A.K.Ghatak.

2. Introduction to lasers and their applications - D.C. Oshea etal

3. Laser and non-linear optics – B.B.Laud.

Unit II
Lesson 3

LASER APPLICATIONS

Objective: The characteristic features of laser radiation like high degree of monochromaticity, temporal and spatial coherence, very high intensity, directionality and above all the availability of laser sources at required wavelength, are responsible for an exciting applications in scientific, industrial, engineering and medical fields. Variety of new lines of research and study have been developed and grown to help the mankind in their daily life. Some of the worth mentioning topics are given below very briefly for a glance.

Structure:

2.3.1. Applications at a glance:

I. In spatial frequency filtering and Holography, the applications are:

1. To filter out high frequency and low frequency components in the object.
2. To enhance contrast in the image
3. To detect non-periodic errors in a periodic structure.
4. In character recognition problems in an optical image.
5. To observe the depth concept in an image (Third dimension)
6. To analyze strain patterns in a model by using interferometric and speckle techniques.
7. To use in non-destructive testing techniques.

II. To design a thermonuclear reactor based on laser induced fusion.

III. Light wave communications through fiber optics

IV. To generate harmonics, stimulated Raman emission and self focusing

V. To trigger chemical and photochemical phenomenon reactions.

VI. To separate isotopes.

VII. To study absolute rotation of earth and ether drift.

VIII. Lasers in industry for

1. Laser welding
2. Hole drilling
3. Laser cutting
4. Vaporization and subsequent deposition on a substance.
5. Laser Tracking- LIDAR –development
6. Developing precision in length measurement

7.Velocity measurement

IX. To produce the best available clock or time standard.

X.To develop nonlinear optical techniques

XI. In medicine-diagnostics and treatment and cosmetic tool.

2.3.2.Lasers are also used in :

Phototherapy of the eye, laser printers, cell cytometry stereo lithography, fluorescence analysis, inspection of electronic circuit boards compact disk, laser disk, C.D-Rom mastering, large image projection television, material processing, stroboscopic illumination of various rapidly moving objects like bullets, Ablation of materials photodynamic therapy in treating some types of cancer.

Annealing, hardening, cutting and cauterization, laser surgery and corneal scalping. Plasma diagnostics, heterodyne sources for far infrared astronomy, to shatter kidney stones or gall stones by directing the beam through an optical fiber catheter, to remove tattoos and also skin lesions resulting from excess melanin, selective photothermlysis, pollution detection, remote sensing, cryogenic cooling. Sources of local area networks, avionics, and satellite networks, high definition television etc.

2.3.3. Therapeutic Applications of Lasers.

Specialty	**Procedure/purpose**	**Lasers used**
Burn therapy	Narcotic and infected tissue Debridgement Burn Escher Excision	CO_2, Ar+, Nd: YAG
Dermatology	protwine stain, strawberry marks Varicose vein excision Melanomas and basal cell Carcinomas	Ar+, Nd: YAG
Gastroenterology	Gastric bleeding hemostasis Hepatetectomy Gallstone removal	CO_2 , Ar+, Nd: YAG
Gynaecology	Fallopian tube reconstruction Fertility micro surgery Hyper plastic destroy Vaginal adenoids	Ar+, CO_2 , Nd: YAG

Neurosurgery	Spinal and brain tumor excision Photo radiation of tumors Debulk Large exophytic tumors	CO_2, Ar^+ Nd: YAG
Ophthalmology	Photocoagulation of retinal bleeding Retinal detachment ,lens capsule surgery	CO_2
Orthopedics	Joint surgery ligament chips Calcification deposits, Bone Tumor excision	CO_2
Otolaryngology	Polyp excision Turbinectomy	CO_2
Plastic & reconstructive Surgery	Eye lid repair, breast reconstruction, fatty tissue Removal, Maxillo facial Surgery	CO_2
Thoracic surgery	Lung Cancer photo radiation (Treatment & Diagnosis) Heart revasculization	Ar+, Nd: YAG CO_2
Urology	Kidney stone excision Bladder tumor excision s	CO_2, Ar+ CO_2, Ar+ CO_2

After having gone through the multifarious applications of lasers in scientific disciplines such as physics, chemistry, Biology, medicine etc one may not wonder, why the scientists called LASER as a tool of 20^{th} century.

APPLICATION OF LASERS

2.3.4. Some experiments of Fundamental importance:

2.3.4.1 Study of Brownian Motion:

An experiment has been described by Clark etal for the study of Brownian motion using light scattering with 1m.w., power of He-Ne laser. The laser light is focused on to a cell containing a dilute aqueous solution of polystyrene spheres of uniform diameter. The scattered light by each particle in a small angle θ is collected on to a photo surface of RCA 931 photo multiplier tube. As the particles move the intensity is constantly fluctuating around its average value, the frequency spectrum of the scattered light provides the conditional probability.

$$P(r, t)=(4\Pi Dt)^{-3/2} \exp(-r^2/4Dt)$$

The diffusion constant 'D' is determined accurately to obtain quantitative information about the motion of the particles.

2.3.4.2. Absolute Rotation of the Earth: To detect the absolute rotation of earth, light is made to pass in both clockwise and anti clock wise directions around a square in a fixed path. In a square frame at rest, the light will take same time to travel in both directions in both beams will have some frequency. When the square frame is made to rotate about an axis perpendicular to the plane containing the system, the two beams will have slightly different frequencies. On recombination detected beats indicate the rotation of the square. In a square of 3m side, the difference in path length create a shift of hundred thousandth of a wavelength, is produced.

The same experiment was conducted in 1914 by Sagnae and then by Michelson and Gale in 1925 using ordinary light. With four He-Ne lasers arranged in the sides of square as shown in fig2.3.1. The experiment can be done with much more precision and sensitivity,

In New York at a latitude of (110^0 40^1 N) the effective speed of rotation is about 1/6 of a degree/minute, would correspond to beat frequency of 40Hz. However, new designs are suggestions to increase difference between the beams by using low loss optical fibers to increase the path length and area.

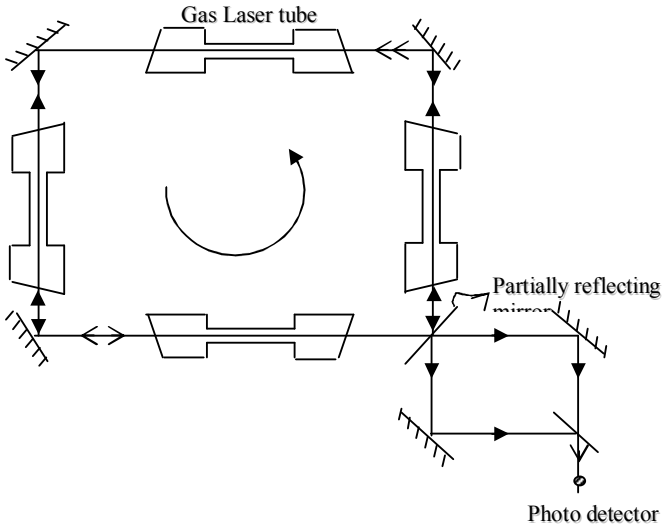


Fig:2.3.1 A ring laser for detecting the absolute rotation of the earth when the system at rest , both the clockwise and anti-clockwise rotating beams have the same frequency . When the system is rotated about an axis normal to the plane containing the system, the clock wise rotating and the anticlockwise rotating beams have slightly different frequencies. When they are combined, beats are

2.3.4.3. Lasers in Isotope Separation:-

By using tunable lasers (Dye lasers) an efficient and economical separations of isotopes can be achieved. This laser isotope separation (LIS) is used for large scale enrichment of uranium (U^{235}) for nuclear power reactors.

Isotope separation using laser beam is fundamentally a different technique, where one makes use of the slight differences in the energy levels of atoms of the isotopes due to difference in nuclear mass. This difference is termed as the isotope shift. As the energy levels of the isotopes of the same element are slightly different, precise photon energy (hν) from a laser of high degree of monochromaticity, can be absorbed by one isotope and excited while the other atoms of the second isotope are not at all effected. Thus shine of laser light on a mixture of two isotopes can excite one of the isotopes and earmarking it for subsequent separations as shown in the block diagrams.

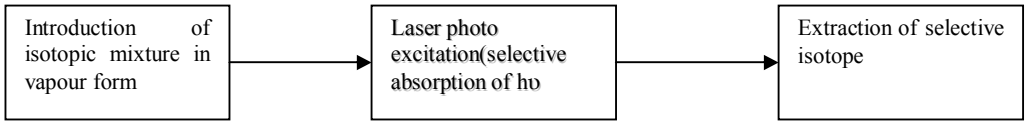


Fig 2.3.2 : Block diagram of a typical laser isotope separation process. A laser excites one of the isotopes from the isotopic mixture through selective absorption and the excited isotope atoms are separated using one of the many techniques.

2.3.4.4. LIS – Process

LIS involves the selective excitation of atoms of an isotope by irradiation. Then the excited atoms are separated from the mixture using radiation pressure.

On absorption of photon hν, the isotope atom acquires momentum hν/c as per conservation law. Due to absorption the atom, is pushed in the direction of travel of the incident photon. The excited atom will come to ground state by emitting a photon. In emission also it acquires equal and opposite momentum to that it acquired during absorption. Since emission occurs in all random directions, the net effect of many absorptions and emissions is to push the atoms of that isotope in the direction of the laser beam. Thus the isotopic atoms participated in absorption are deflected by laser beam. Scheme of this sort has been used to separate the isotopes of barium.

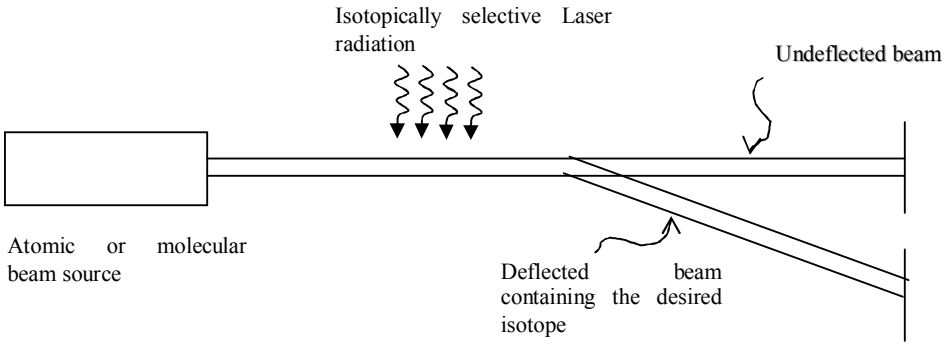


Fig 2.3.3 Laser isotope separation-Deflection method

2.3.4.5. Separation by selective photo ionization or photo dissociation:-

A most popular and universally applicable scheme of isotope separation is the two-step photo ionization of atoms or the two step dissociation of molecules. The first step causes the selective excitation followed by second excitation which ionizes the excited atoms or dissociates the molecules. The ionized atoms are separated by means of electric fields. In the case of molecules the dissociation products are separated from other molecules by means of chemical reactions. The two step photo ionization was used to demonstrate the applicability of LIS for uranium at Lawrence Livermore laboratory in USA.

Atomic uranium beam from a furnace at 2100 ^{o}C. temperature was excited by a tunable dye laser beam and then ionized by the light of high pressure mercury lamp. The ^{235}U isotope present in natural uranium at a concentration of 0.71% was enriched to 60%.

Two-step excitation to molecules was used in separating the two isotopes ^{10}B and ^{11}B. In this experiment $^{11}BC_3$ isotope was selectively excited by CO_2 laser radiation corresponding to vibrational transitions of $11BC_3$. The molecules thus excited were dissociated by light with a wavelength 2130A^{o} and 2150A^{o}. The fragments generated by this dissociation originating mainly from $^{11}BC_3$, were bound by reaction with O_2. It was found that with five light pulses of CO_2 laser radiation, 14% isotopic enrichment could be obtained.

LIS has been used in nuclear power industry which requires uranium enriched with isotope mass number 235(^{235}U). LIS was also used to obtain isotopes used as traces in medicine, agriculture, research and industry.

2.3.4.6. Laser-Induced Fusion:-

In nature sun and stars are releasing enormous amounts of energy continuously due to thermonuclear fusion reactions. During the last five decades scientists are working on various schemes to generate fusion energy in a controlled manner, so as to have an almost inexhaustible supply of pollution free energy. A thermonuclear reactor based on laser induced fusion is of great promise for humanity.

It is known that nucleons are held together in the nucleus by short-range attractive forces operating at short distances (<=10^{-13} cm). To split the nucleus into its constituent nucleons certain amount of energy has to be supplied. This is known as binding energy and is calculated using the famous Einstein mass-energy relation $E=mc^2$. Considering the nuclear reaction in which two deuterons $D={}^1H_2$ react to form a tritium $T={}^1H_3$ nucleus and a proton.

$$D+D \rightarrow T(1.01\ Mev) + H\ (3.01\ Mev)$$

In this fusion reaction, two loosely bound light nuclei produce a heavier tightly bound nucleus giving a net gain in binding energy of about 4.02 Mev. The total binding energy of nucleus is calculated from

$$\Delta = (Z\ m_p + N\ m_n - Ma)\ c^2$$

where m_p, m_n and Ma represent the masses of proton, neutron and atomic nucleus. The another nuclear fusion reaction of deuterium and tritium.

$$D+T \rightarrow \alpha\ (3.5Mev) + neutron\ (14.1\ Mev)$$

is also of considerable importance as a possible source of thermonuclear energy of about 17.6 Mev. As tritium is not available naturally, this has to be produced by allowing the neutron produced in D-T reaction, to interact with lithium.

$${}_3Li^6 + {}_0n^1 \rightarrow T\ ({}_1H^3) + ({}_2He^4) + 4.79\ Mev$$

$${}_3Li^7 + {}_0n^1 \rightarrow T\ ({}_1H^3) + ({}_2He^4) + {}_0n^1 - 2.46\ Mev.$$

The energy released in D-T reaction is about four times that of a D-D reaction.

Very high temperatures (100 million ^{o}k) are required for fusion reactions so that high K.E of nuclei can dominate over coulomb repulsion. At such high temperatures the matter will be in a fully ionized state in plasma. The two major problems 1) heating plasma to very high temperature. 2) confinement of plasmas for sufficiently long time for fusion reaction to occur, have been solved with a new idea of fusion using intense laser pulses. This idea is essentially compressing, heating and confining thermonuclear material by way of interacting with intense laser pulse of energy ~ 1.2 X 10^5 J. The expected fusion energy yield is ~ 14 MJ.

2.3.4.7. Laser - induced Fusion reactor:-

Intense laser pulses are made to strike on a D-T pellet in the form of cryogenic solid of particle densities 4 X 10^{22} cm^{-3} from all directions. Within a short time the outer surface of the pellet is heated and gets converted into hot plasma ($T \approx 100$ million K). The ablation layer expands into vacuum the reaction of which is a push to the pellet towards the center of the pellet. Thus irradiation from all sides as shown in the fig 2.3.4. by the pulses of intensity 10^{17} w/cm^2 an impulsion front traveling towards the core with an inward pressure 10^{12} atmospheres. Such shock waves lead to very high compression of core and fusion energy is released. It is shown that 1.06 μm radiation from Nd:glass laser can heat D-T pellet five times the temperature in one fourth time compared to 10.6 μm radiation from CO_2 laser.

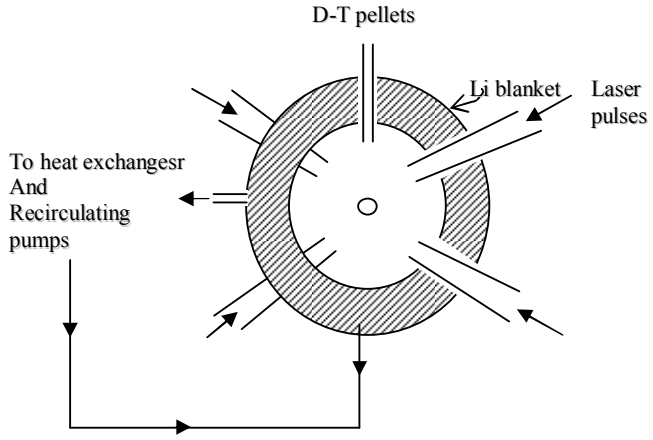


Fig 2.3.4: D-t fusion reactor

Limitations :-

Building of high power lasers to a stage of getting more output energy then input energy. Exact time delivery of laser energy onto the pellet production of targets with extremely good surface finish to avoid unstable compressions.

2.3.4.8. Lasers in Industry:-

The extraordinary properties of laser radiation are used in industry for various applications due to feasibility of focusing into a region of 'λ' dimensions resulting very high field strength of the order 10^7 V/cm. Let λ be the wavelength, 'a' is the radius of the beam and ' f ' is focal length of the lens, then the incoming beam can be focused into a region 'b'

$$b = \lambda \ f/a$$

If 'p' represents the power of the laser beam, then the intensity obtained at the focused region be

$$I = \frac{P}{\pi b^2} = \frac{P\,a^2}{\pi \lambda^2 f^2}$$

By using beam expander and using lasers of short focal length, very high intensities are produced in a small region. At such smaller focused laser spots, the beam has large divergence. Hence the beam expands again within a short distance. This distance is called the depth of focus. Thus smaller focused spots lead to a smaller depth of focus.

2.3.4.8.1. Laser Welding:-

High power laser CO_2 having an output of 3.5 Kw can weld of ¼" thick stainless steel at a speed of 2cm/sec in the focal length 25cm. Laser welding has been used in the field of electronics and microelectronics which require precise welding of very thin wires. Further, welding in inaccessible areas like inside the glass envelope can be done using laser beam. One can have effective weld even without the removal of the insulation. Laser weld not only achieves welding between dissimilar metals but also allows precise location of the weld. Nd: YAG and CO_2 lasers are two important kinds of lasers used for wide- ranging applications in welding.

2.3.4.8.2. Hole drilling:- Drilling of holes in various substances like 1mm thick steel plate of hole of radius 0.1mm and for the drilling of diamond dies used for drawing wires, holes of 10μm

diameter through hardest substances, to drill ruby stones used in time pieces can be made by using Nd:YAG and ruby lasers.

2.3.4.8.3. Laser cutting:- High power CO_2 in the most common laser used in cutting process with as small as possible width of the cut. Laser cutting of stainless steel, nickel alloys and other metals finds widespread application in aircraft and automobile industries. Laser cutting has also been used in the textile industry and apparel manufacturing for cutting cloth.

2.3.4.9. Laser Tracking:- Tracking means either the determination of the trajectory of a moving object like air craft or rocket or determining the daily positions of heavenly object like moon or artificial satellite.

Principle:- This technique involves the measurement of time taken to travel to and fro for a sharp laser pulse sent and received back by the observer using a retro reflector on the target. In a retro reflector the incident and reflected rays are parallel and travel in opposite directions. Laser tracking system is much smaller in size compared to microwave radar system. When fog and snow are present in the atmosphere it is difficult to work at optical frequencies in laser systems.

NASA ,USA launched an aluminium sphere called laser geodynamic satellite into orbit at an altitude of 5800 km for studying the movements in earth's surface, which would be of great help in predicting the earthquakes.

2.3.4.10. LIDAR:-

Lidar is an acronym for light detection and ranging. Laser systems are used to study the laser beam scattered from atmosphere. Laser pulses are sent from systems base and the scattered radiation by various particles in the atmosphere is analyzed with sensitivity far superior than microwave radar technique. In particular this technique is used to study the nature of aerosols in the atmosphere by measuring the back scattered laser light using a photo detector. Time variation is considered into height from which the laser beam is back scattered.

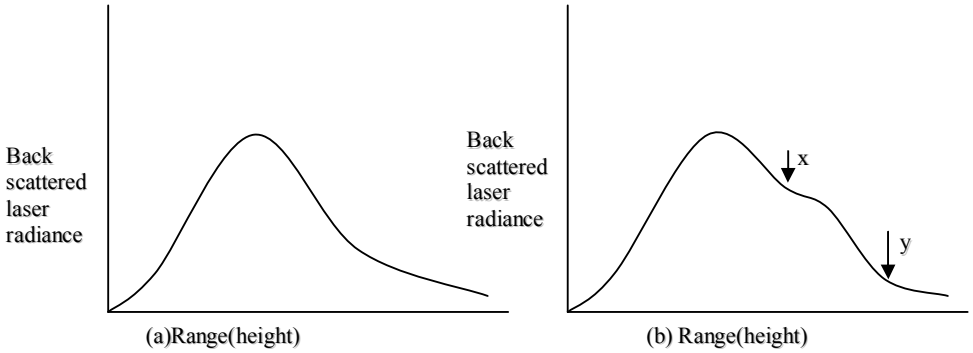


Fig.

The presence of aerosols between the heights h1 and h2 are responsible for greater intensity of back scattered radiation compared to that of aerosol free atmosphere. The size and concentration of various particles important in pollution studies are also determined. Lidar technique can also be used to determine visibility of atmosphere, the diffusion of particular materials and also the study of clouds, fog etc., The study of turbulence and winds and probing of stratosphere have also been carried out by lidar systems.

2.3.4.11. Heat Treatment: It is common in tooling and automotive industry where in a portion of a larger item is to be treated for hardening. This is achieved by focusing laser beam using mirror arrangement precisely at the required portion without producing undesirable changes over the entire item. This process is highly effective and economical in the case of hardening a well-defined portion above the ring of an automobile piston. This can be done even intricate portions that are not accessible by usual means.

This heat treatment is also used to form an alloy entirely different from the bulk material, by adding proper elements to the molten layer. Thus alloying process enables to selectively harden the areas of greatest wear.

2.3.4.12. Velocity Measurement:- The frequency of scattered radiation from a moving object is different from that of incident frequency. Let ν be the frequency and V represents velocity of a

moving object moving at an angle with respect to incident radiation. The change in frequency ν between the incident and scattered beams is given by

$$\frac{\Delta\nu}{\nu} = \frac{2V}{c}\cos\theta$$

Where 'c' represents velocity of light in free space. The change in frequency is directly proportional to velocity V of the moving object is known as Doppler shift. Using this principle portable velocity measuring meters have been fabricated to measure the speeds of automobiles on highways and fluid flow rates in industry.

The beam from a CW He-Ne laser is made to split from a beam splitter. One of the components undergoes scattering from a moving object. The scattered radiation and the second component of the beam (ν) are made to interfere. The resulting beat frequency is a direct measure of the velocity of the object.

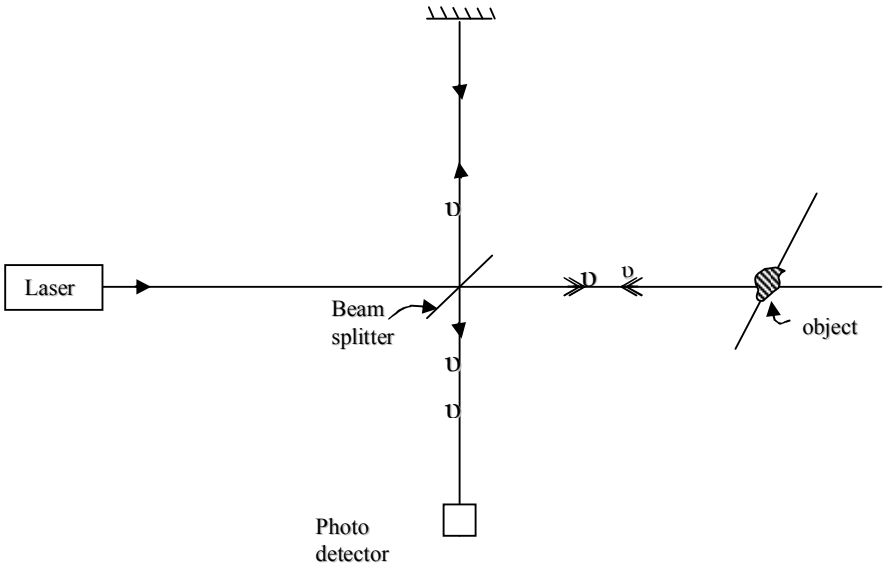


Fig 2.3.6. Schematic of an arrangement for measuring the velocity of a moving object using Doppler shift

2.3.4.13. Lasers in chemistry and Biology:- The production capability of high temperature and electric fields of the order 10^9 v/cm at the focus of laser beam is considered to be an excellent tool for triggering chemical and photochemical reactions. The analysis of trace metals present in various tissues is carried out through spectral microanalysis using laser attached microscopes. The extremely high monochromaticity of laser is used for selective excitation of bands in a molecule to produce new chemical species in certain chemical reactions. The picosecond laser pulses enabled to study ultra fast chemical process with unprecedented time resolution.

Brillouin spectral measurements of biological fibers, collagen and keratin, synthetic polypeptides and material from the lens and cornea of the eye, coupled with density measurements, values of elastic moduli are determined.

Lasers are used in eye surgery to reattach a detached retina. Lasers are used for treatment of Glucoma in the eye. New types of surgery with ultraviolet excimer lasers and high power pulsed neodymium laser energy through optical fiber is now used in the treatment of liver cancer.

2.3.4.14. Low power applications:- The first laser-based device to be used by public at large, is the point of sale device (POSD) to price the items in super markets. This is by reading the label printed in the form of parallel bars of varying width used to encode numerical information in universal product code UPC by the manufacturer. The device scans a focused laser beam back and forth across the label many times each second in an intricate design. Detectors sensitive only to laser wavelength, detects the variations in the reflected light from dark and bright bands of the label and converts them into electrical signals. The price of the item, posted at the shelf and stored in the computer is added to the bill. Thus man power costs are reduced and increase check out times and provides upto date inventory control.

Laser is introduced in videodisc system. The readout using a focused laser beam holds several advantages since there is no contact with disc surface, the wear and tear problems are totally eliminated.

2.3.5. Summary : Laser applications in variety fields are given for a glance study. Detailed experiments of fundamental importance in physics, industry and chemistry are described.

2.3.6. Key terminology: Laser induced fusion – LIDAR – Harmonics – Stimulated Raman emission – Laser isotope separation – Velocity meter – Medical diagnostics and treatment – Stereo lithography – Laser-disk – Pollutants detection Aerosol Study – Avionics etc.

2.3.7. Self assessment questions:-

1.Describe briefly the applications of laser.

2.Explain laser isotope separation.

3.Discuss laser induced fusion.

4.Mention the applications of laser in medical field.

2.3.8. Reference Books:-

1.Lasers and non-linear optics – B.B Laud, Wiley Eastern Ltd.

2.Lasers theory and applications – K. Thyagarajan, A.K. Ghatak, Macmillan India ltd.

3.Lasers fundamentals - William T. Silfvast, Cambridge university Press.

4. Introduction to lasers and their applications D.C. Oshea, W.R.Callen and W.T. Rhodes

Unit II

LESSON – 4

OPTICAL FEEDBACK

Objective:

To explain the mechanism of feedback in lasing action and to discuss the necessary conditions for stability of a resonator configuration.

Structure:

THE OPTICAL RESONATOR

2.4.1.**Introduction :**

A combination of two mirrors M_1 & M_2 with 100% and 95% reflectivities (R_1 and R_2) respectively placed on one axis at the two ends of active medium is called an optical resonator. Photons generated by stimulated emission of radiation will have a long path due to multiple reflections in between the

two mirrors through the inverted medium so as to reach sufficient intensity by virtue of continuous amplification. Laser systems operate with only inverted medium without mirrors for multiple reflections, are referred as super-radiant. Such systems are referred as light amplifiers rather than lasers.

2.4.2.**Round trip power gain and threshold condition**

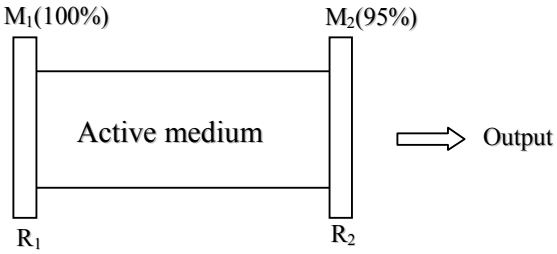


Fig: 2.4.1: Resonator

Lasers with highly reflecting mirrors at the ends are not only amplifiers but also oscillators Optical resonator supports continuous Electromagnetic oscillations at laser transition frequency. A small signal provided by spontaneous emission undergoes continuous amplification by stimulated emission until some steady state level of oscillation is reached. Though the amplification is continuous, simultaneously the photons lose part of their energy to overcome various lossy mechanisms such as scattering, absorption reflections and the output of the beam etc.,. In order to overcome these losses, minimum gain coefficient called threshold gain β_{th} is required to initiate and sustain laser oscillation . This threshold gain in turn specifies the minimum population inversion

required for lasing .

Assuming that active medium is completely filled the volume in between the mirrors and the pumping excitation is uniform, the irradiance of the beam I is given by

$$I = I_0 e^{(\beta - \alpha_1) L}$$

Where L is the separation of mirrors,

β = small signal gain coefficient

α_1 = loss per unit distance

After a round trip of the beam due to reflections at M_1 & M_2,

the beam irradiance is given by

$$G = R_1 R_2 \exp[2(\beta - \alpha_1) L]$$

$$= \frac{Beam\ irradiance\ at\ the\ end\ of\ round\ trip}{Beam\ irradiance\ at\ the\ start\ of\ round\ trip}$$

where G = net round trip power gain.

If G is greater than unity, the disturbance at resonance frequency of laser, experiences a net round trip growth in magnitude and cavity oscillations sustain and increase.

If G is less than unity, the oscillations die out.

Therefore the threshold condition for laser oscillation is,

$$G = R_1 R_2 \exp [2 (\beta - \alpha_1) L] = 1$$

In a continuous operation laser, the steady state value of small signal-gain must be equal to the threshold value β_{th} . This steady state gain is referred as gain saturation. The small signal gain to support steady-state operation depends on the parameters of the laser medium α_1, L, R_1, and R_2.

Taking logarithm on both sides of the above equation.

$$\beta_{th} = \alpha_1 + \frac{1}{2L} \ln\left(\frac{1}{R_1 R_2} \right) = \alpha_1 + \alpha_0$$

where α_1 = losses of cavity radiation due to absorption and scattering

α_0 = loss of radiation as an useful output from the cavity.

The gain must be equal to all types of losses from the cavity thus β can assume various values not only due to N_J but also the intrinsic properties of the medium and reflectivities of the mirrors and their alignment.

2.4.3. **The efficiency of a laser**

$$\text{The efficiency of a laser} = \frac{output\ power}{input\ power}$$

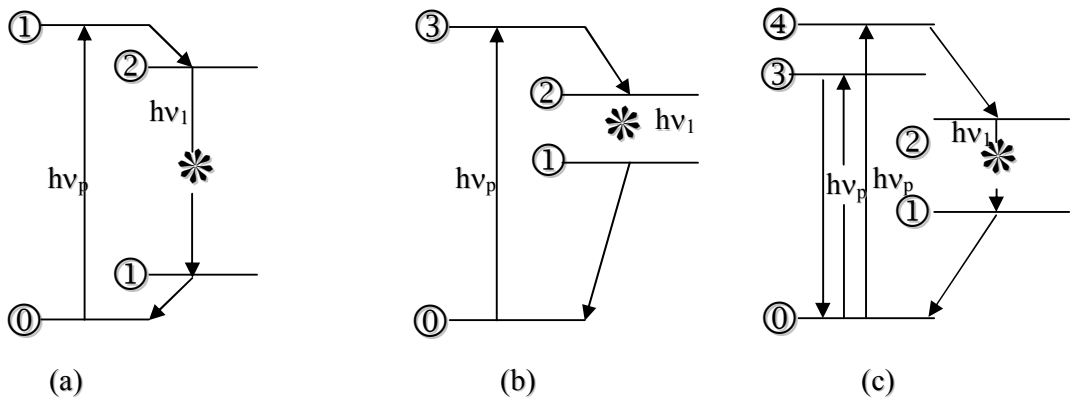


FIG 2.4.2 :four level systems

Hence it depends upon the laser transition probabilities and populations of the respective levels. Though the four level schemes are considered to be efficient, the system's efficiency varies through wide range as seen from the following schemes.

a) Highly efficient system since most of the pump energy is converted to output energy ($h\nu_1$).
b) Less efficient system than (a) scheme since small fraction of pump energy is converted as output ($h\nu_2$).
c) This is also less efficient system than scheme (a) since much of the pump energy is channeled into non-lasing transitions.

Thus the efficiency of the system depends on ($h\nu_1 > h\nu_3 > h\nu_2$) laser output.

2.4.4.**Confinement of the beam within the Resonator : The stability condition**.

The resonator is a combination of mirrors M_1 and M_2 (plane, curved and plane-curved combinations) used to strictly confine the beam into the cavity even after undergoing multiple reflections. The role played by mirrors depends upon the characteristics of mirrors namely radius of curvature r_1 and r_2 , separation 'L' and reflectivities R_1 and R_2.On multiple reflections, fraction of the cavity radiation spills over the edges of the mirrors. Hence the design of a resonator plays an important role in achieving the total confinement of the radiation to the cavity to obtain useful output continuously.

Resonator Theory

The multiple reflections of the beam can be treated simply as an continuous assessment of beam's position and direction angle θ using suitable matrix for translation.

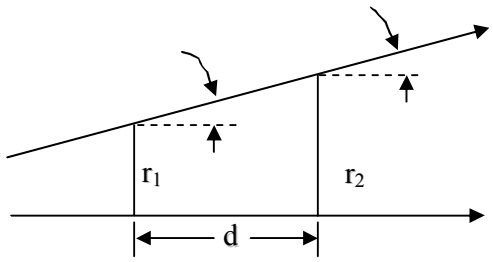


Fig. 2.4.3: Ray tracing

Consider a path of the paraxial ray at two points separated by 'd' when it is at distances r_1 and r_2 from the axis making an angle θ_1 and θ_2 . With the given data at point1 i.e., r_1 and θ_1, the ray's position at point 2 can be expressed as

$$\left.\begin{aligned} r_2 &= r_1 + d\theta_1 \\ \theta_1 &= \theta_2 \end{aligned}\right\} \quad \text{---------- (1)}$$

in the matrix form

$$\begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} \text{---------- (2)}$$

here the matrix of translation over a distance 'd' is $\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$ -------(3)

in general it is represented as $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ matrix

in this case A=1, B=d, C=0 and D=1 Using suitable matrix for each process coming in the way of beam like reflection, refraction, a similar procedure can be adopted to estimate the changes in r_1 and θ_1 as

$$\begin{bmatrix} r_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} = \lambda \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} \text{---------- (4)}$$

after N passes through the cavity

$$\begin{bmatrix} r^N \\ \theta^N \end{bmatrix} = \lambda^N \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} = e^{\pm N\phi} \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} \text{---------- (5)}$$

for large N the beam would clearly diverge leading to a unstable cavity

situation. For convergence of trajectory, the solution is

$$\begin{bmatrix} r^N \\ \theta^N \end{bmatrix} = \lambda^N \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} = e^{-iN\phi} \begin{bmatrix} r_1 \\ \theta_1 \end{bmatrix} \text{ ---------- (6)}$$

since $|e^{-iN\phi}| \leq 1$. the beam would always confine to the region of axis of

the resonator.

The requirement for stability in the case of spherical mirrors of

radius R=2f

$$0<d<2R \text{ ---------(7)}$$

For two mirrors of unequal curvature ($f_1 \neq f_2$)

$$0<\alpha_1\alpha_2<1 \text{ ---------- (8)}$$

$$\left.\begin{aligned} \alpha_1 &= 1-\frac{d}{2f_1} = \left(1-\frac{d}{R_1}\right) = g_1 \\ \alpha_2 &= 1-\frac{d}{2f_2} = \left(1-\frac{d}{R_2}\right) = g_2 \end{aligned}\right\}$$

----------(9)

For stability , the requirement is

$$0 < \left(1-\frac{d}{R_1}\right)\left(1-\frac{d}{R_2}\right) < 1 \text{ ---------- (10)}$$

or

$$0 < g_1 g_2 < 1 \text{ ---------- (11)}$$

This is expressed in stability diagram in XY-plane as shown in the figure.

Stability diagram for two mirrors of curvature R_1 and R_2

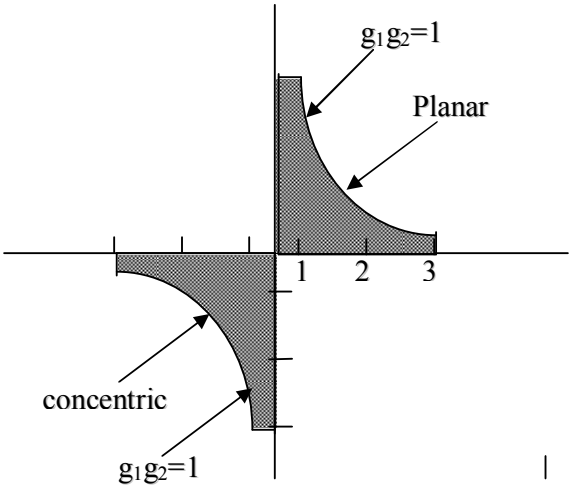


Fig 2.4.4: Stability diagram

In the shaded regions g_1g_2 <1 and the respective cavity is stable, Planar, confocal and concentric combinations are on the boarder line, hence resonator alignment is very critical.

The stable configurations are efficient in the sense that reasonable large fraction of resonator volume is utilized in amplification process. In certain laser systems unstable resonator configurations will also acts as useful designs:

1. They are short resonators and super radiant type.
2. They are suitable for adjustable output coupling simply by changing the spacing of the mirrors.
3. This can be used with an active medium having very high signal gain only.

Different configurations of optical resonators are shown in fig.

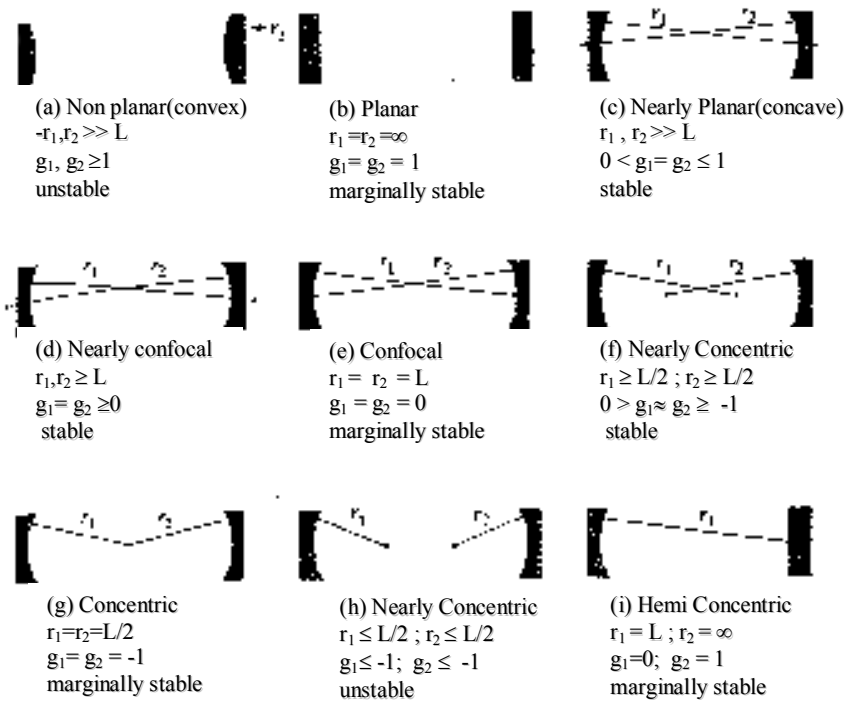


Fig: 2.4.5. Laser cavity mirror configurations. Stability for each of these configurations is indicated

2.4.5. Summary :

The situation in the lasing medium before light amplification has been discussed . the condition has been achieved for a threshold value.(β_{th}). The efficiency in lasing action is discussed in various four level systems. Condition for stability of an optical resonator has been obtained.

2.4.6. Key terminology:

Round trip power gain - threshold condition - efficiency of four level systems -Optical resonator- stability criteria of a resonator- planer- confocal and concentric systems.

2.4.7. Self assessment questions:

1.Discuss the behavior of the active medium at threshold condition.

2.Discuss the efficiency of various four level laser systems.

3.Describe the role played by an optical resonator in lasing

2.4.8. Reference books:

1. Introduction to lasers and their applications D.C. Oshea, W.R.Callen and W.T. Rhodes

2.Laser fundamentals – William T. Silfvast.

Unit II
Lesson - 5

THE LASER OUTPUT

Objective: - To present the discussion on various mechanisms responsible for the total line width in the laser transition.

Structure:

2.5.1 Introduction:

The basic requirements like an active medium enables us to obtain laser output with characteristic qualities, like monochromaticity, directionality, spatial and temporal coherence and high intensity. However, the output consisting of various frequencies called modes on either sides of the centre of transition frequency ν_o gives a definite shape with line ($\Delta\nu$) width and maximum intensity at ν_o.

2.5.2 Absorption and emission line shape broadening

Absorption and emission will not take place at a single frequency but over a range of frequencies spanning the atomic line width. The coefficients of absorption and small signal gain α

and β are really the functions of frequency α(ν) and β(ν). Hence the expression used to represent α is now represented in more general way as

$\alpha(\nu) = \beta_{ij} (N_i - N_j) g(\nu) \frac{h\nu_{ij}}{4\pi c}$ where g(ν) is called the line shape function contains all the frequency dependence of α. g(ν) is equal to unity at ν_0 resonance frequency . Similar dependence of β on ν

$\beta(\nu) = \beta_{ij} (N_j - N_i) \underline{h\nu_{ij}} g(\nu)$

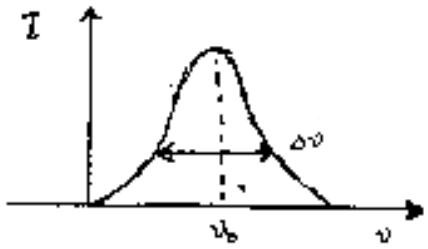


Fig:2.5.1. Line profile

Various line broadening mechanisms account for frequency dependence of 'α' and 'β' resulting finite line widths associated with absorption and emission transitions. The line width of an emission line is the frequency difference between the points halfway down the line shape curve. This is called as full width at half-maximum power (FWHM).

The line width is expressed in Hz, nm and angstroms. Variety of mechanisms are contributing to this line width factor.

The most important broadening mechanisms are

1) Life time broadening or Natural broadening
2) Collision broadening
3) Doppler broadening

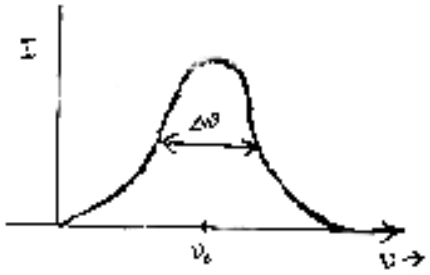


Fig: 2.5.2. Line width

2.5.3. **Line width due to radiative decay**:

Excited atoms radiate at characteristic frequency $\omega_o = \Delta E / \hbar$ or $\nu_o = \Delta E / h$ where ΔE is the separation of energy levels involved in the transition. As is seen from the equation, the emission of radiation is not at single frequency ν_o but spreads on either sides of central frequency (ν_o) called the line width $\Delta\nu$ or $\Delta\omega$. This is associated with decay time of radiating levels as well as other emission line broadening mechanisms.

Classical emission line width:

The decaying electric field of electromagnetic wave is not infinitely long since it has a starting point $t = 0$ and decays exponentially with time constant $T_o = 2 / \nu_o$

$$E_{(t)} = |E_{(t)}|^2 = E_0^2 (e^{-(ro/2)t})^2 = Io\, e^{\frac{-\gamma_0}{2} t} \quad \text{if } t \geq 0$$

$$= 0 \quad \text{if } t < 0$$

The frequency components of this wave can be obtained by taking Fourier transform as

$$E(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} E(t)\ e^{i\omega t}\ dt$$

$$= -\frac{E_0}{\sqrt{2\pi}} \left(\frac{1}{i[(\omega - \omega_0) + i\gamma_0 / 2]} \right)$$

The intensity distribution per unit frequency $I(\omega) \propto |E(\omega)|^2$

$$I(\omega) = I_0 \frac{\gamma_0 / 2\prod}{\left[(\omega - \omega_0)^2 + \frac{\gamma_0^2}{4}\right]}$$

After normalization

Total intensity of emission integrated $I_0 = \int_0^\infty I(\omega) d\omega$

over frequency width of emission line

The line shape is known as Lorentzian distribution and is symmetrical with respect to ω_0

The full width at half maximum (FWHM) $\Delta\omega^{FWHM}$

$$\Delta\omega = 2(\omega - \omega_0) = \gamma_0 = 1 / \tau_0 = 2\prod\Delta\nu$$

This is classical line width

Graph of I(ω) versus ω from classical analysis of a decreasing and radiating electron.

Graph

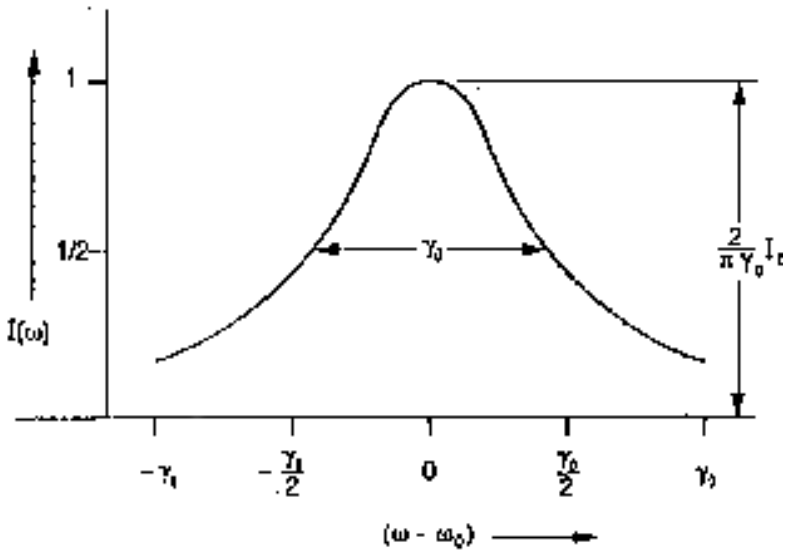


Fig: 2.5.3. Graph of the intensity I(ω) Vs ω from a classical analysis of a decaying & radiating electron as it makes a transition from one energy level to a lower lying level.

For an assembly of N atoms radiating on the same transition with random phases, increased intensity produces identical line shape and width. Such a situation leads to Lorentzian line shape function and is referred as homogeneous broadening.

2.5.4.**Natural emission line width**: Classical treatment has given the relation between line width $\Delta\omega$, the decay rate γ_o and decay time τ_0 . However, this is not accurate description of emission line width associated with the transitions among quantized energy levels.

The famous uncertainty principle will give as an accurate description of line width referred as natural line width ($\Delta\nu_N$).

$$\Delta E.\Delta t \approx \hbar = h/2\pi \quad \text{---------- (1)}$$

The uncertainty of the energy ΔE_u and radiative life time $\Delta\tau_u$ of upper level.

$$\Delta E_u = \frac{\hbar}{\tau_u} = \hbar\gamma_u = \hbar\sum_i A_{ui} \quad \text{---------- (2)}$$

Similarly

$$\Delta E_l = \frac{\hbar}{\tau_l} = \hbar\gamma_l = \hbar\sum_j A_{lj} \quad \text{---------- (3)}$$

As a result, an effective energy width for two levels is

$$\Delta E_\tau = \Delta E_u + \Delta E_l = \hbar(\gamma_u + \gamma_l) + \hbar\left(\sum_i A_{ui} + \sum_j A_{lj}\right) \quad \text{---------- (4)}$$

$$\Delta E\,\tau = \hbar\,\Delta\omega_{ul} = \frac{h}{2\pi}.2\pi\,\Delta\nu_{ul}$$

$$= h\,\Delta\nu_{ul} \text{ ---------- (5)}$$

Equating RHS expressions of (5)&(4)

$$\hbar\left(\sum_i A_{ui} + \sum_j A_{lj}\right) = h\,\Delta\nu_{ul} \text{ -------- (6)}$$

From equation(6) the natural emission line width $\Delta\nu_{ul}^{N}$

$$\Delta\nu_{ul}^{N} = \frac{\left(\sum_i A_{ui} + \sum_j A_{lj}\right)}{2\pi} \text{ -------- (7)}$$

Total decay rate $\gamma_{ul}^{\tau} = \gamma_u + \gamma_l$ ---------- (8)

$$\gamma_{ul}^{\tau} = 2\pi\,\Delta\nu_{ul}^{N} = \left(\sum_i A_{ui} + \sum_j A_{lj}\right) \text{ ---------- (9)}$$

Taking into an account this total decay rate γ_{ul}^{τ} describing minimum emission line width, the line shape function

$$I(\omega) = I_0 \frac{\gamma_{ul}^{\tau} / 2\pi}{(\omega - \omega_0)^2 + (\gamma_{ul}^{\tau})^2 / 4} \text{ ---------- (10)}$$

If lower level is ground state and excited state is metastable Then $\gamma_l = 0$

Then the Intensity Profile I(ν) is

$$I(\nu) = I_0 \frac{\gamma_{ul}^{\tau} / 4\pi^2}{(\gamma - \gamma_0)^2 + (\gamma_{ul}^{\tau} / 4\pi)^2} \text{ ---------- (11)}$$

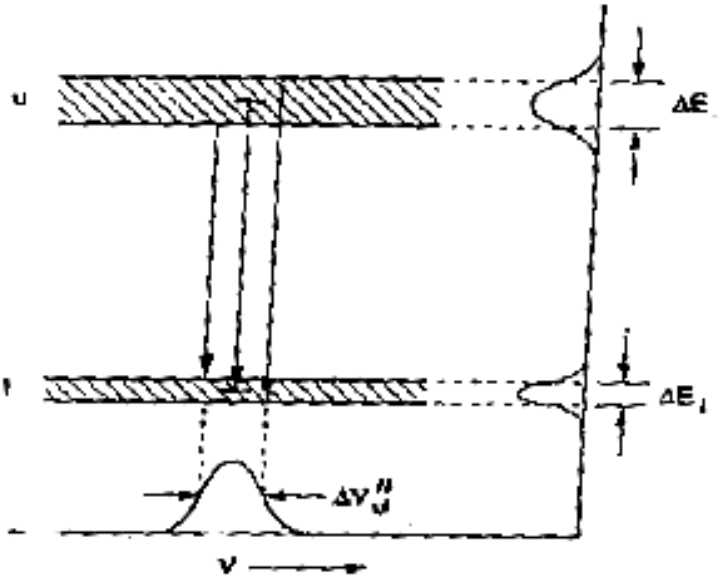


Fig:2.5.4. Quantum mechanical description of the natural line width of emission resulting from a radiating transition between two levels, showing the

2.5.5.**Collisional decay**:-

In a cluster of atoms even at moderate temperatures, the excited populations will collide with surrounding atoms. These collisions can cause the excited atoms to make a transition to lower level without radiating and decreases decay time of the excited energy level. The energy thus lost is given to colliding atom as per the laws of energy and momentum. In some instances, collision will not produce decay of excited level but instead will intercept the phase of radiation atom. This effect will produce broadening of emission.

The decay rate of an excited state 'u' is referred as symbol γ_u of units 1/sec.In general the decay of an excited state population

$$N_u = N_u^0 \; e^{-\Upsilon u \, t}$$

Where $\Upsilon_u = \Upsilon_u^{rad} + \Upsilon_u^{coll}$

and $\Upsilon_u^{rad} = \Sigma A_{ui}$ represents radiative decay rate

and $\Upsilon_u^{coll} = 1 / T_1^u$ represents collisional decay rate

The decay time $\tau_u = 1/\Upsilon_u = 1/\Upsilon_u^{rad} + 1/\Upsilon_u^{coll}$

Collisional decay occurs in all types of media due to interactions of closely located surrounding material with excited atoms. In solids these are referred as phonon collisions.

2.5.6 **Collision broadening** :

Interactions like collisions among atoms, electrons, phonons decreases decay rate of the upper level population, produce significant broadening of line width. In general this type of collisional broadening is divided into two categories 1) Broadening due to non-radiative decay

2) Broadening due to dephasing collisions

Broadening due to non-radiative decay –

An increase in the decay rate of population in the level 'u' due to collisional interactions with surrounding atoms produces increased broadening called 'T_1 broadening'. If the levels participating in a transition are of intermediate excited levels, the increased decay rate is the sum of decay rate associated with each level.

$$\Upsilon_u = \frac{1}{\tau_u} = \sum_i A_{ui} + 1/T_1^u$$

Decay of lower level 'l' with collisional decay T_1^l

$$\Upsilon_l = \frac{1}{\tau_l} = \sum A_{lj} + 1/T_1^l$$

The increased emission line width $\Delta\nu_{ul}$

$$\Delta\nu ul = \Delta\nu_{ul}^N + \frac{1}{2\pi}\left[\frac{1}{T_1^u} + \frac{1}{T_1^l}\right]$$

Where $\Delta\nu_{ul}^{N}$ natural line width due to spontaneous emission.

If T_1 becomes shorter than radiative decay time for a given level, collisional decay rate $1/T_1$ begins to dominate. This reduces the population of the upper laser level and also increases the emission line width.

2.5.7. Broadening due to dephasing collisions :-

When an atom radiating, the collision will produce a sudden phase change in the wave train thus interrupts the phase of radiating atoms at a rate r_2.This broadening is referred as T_2 broadening. The term T_2 represents the average time that occurs between phase-interrupting collisions.

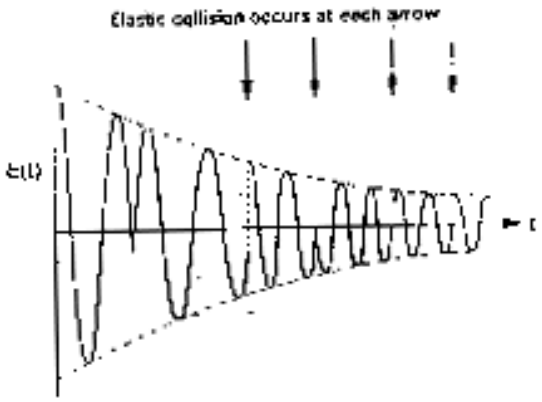


Fig: 2.5.5. Dephasing Collisions

T_2–broadening due to collision dephasing in time domain in which the time duration of each interrupted sinusoidal segment varies but has some average time T_2. Each of these segments represents a sine wave of same frequency.

$$\gamma_2^{u} = \frac{1}{T_{2^u}}$$

and for lower level $\gamma_2^{l} = \frac{1}{T_{2^l}}$

Typically $T_2^{u} = T_2^{l}$ and hence $\gamma_2^{u} + \gamma_2^{l} = \frac{2}{T_2}$

The total broadening by way of collisions for total homogenous emission line width.

$$\Delta\nu_{ul} = \frac{\gamma^{ul\,T}}{2\pi} = \frac{1}{2\pi}\left[\left(\sum_i A_{ui} + \sum_j A_{uj}\right) + \frac{1}{T_1{}^{u}} + \frac{1}{T_1{}^{l}} + \frac{2}{T_2}\right]$$

Where T_2 is short enough, $2/T_2$ factor contributes significantly to the line width. This is major broadening mechanism for many solid state laser transitions. Here γ_2 is a rate that is strictly associated with phase interruption but not associated with a factor associated with decay rate of level.

2.5.8 Doppler broadening :-

Doppler effect associated with sound waves can be experienced by the observer as a change in pitch of the horn of a passing train. While the train approaches, the sound is heard at high frequency and then changes to lower frequency as the train passes. The same frequency shift is characteristic of the emission of light waves from moving atoms in random direction.

Consider light of frequency ν_0 and velocity 'C' being emitted over a time interval Δt from an atom moving towards the observer with a velocity 'ϑ'. At the end of the period Δt, the waves will cover a distance $(C-\vartheta)\,\Delta t$ in the direction of observer. In the absence of the movement of the source, the waves travel a distance $C\Delta t$. Hence the wavelength is little bit compressed to satisfy the relation $\lambda\nu = C$. The observed frequency ν is having shift from the original frequency ν_0 as

$$\nu = \left(\frac{c\Delta t}{(c-\vartheta)\Delta t}\right)\nu_0 \cong \nu_0\left[1 + \frac{\vartheta}{c} + \left(\frac{\vartheta}{c}\right)^2 + \ldots\ldots\ldots\ldots\right] \text{---------- } 1$$

For the source moving away from the observer

$$\nu = \left(\frac{c\Delta t}{(c+\vartheta)\Delta t}\right)\nu_0 \cong \nu_0\left[1 - \frac{\vartheta}{c} + \left(\frac{\vartheta}{c}\right)^2 + \ldots\ldots\ldots\ldots\right] \text{--------- } 2$$

For non relativistic velocities high powers of $\left(\frac{\vartheta}{c}\right)^2$ are neglected.

$$\nu = \nu_0\left[1+\frac{\vartheta}{c}\right]$$ moving towards the observer ----------3

$$\nu = \nu_0\left[1-\frac{\vartheta}{c}\right]$$ moving away from the observer ----------4

The atoms moves at random directions with a range of velocities related to the average temperature T

$$\vartheta = \sqrt{\frac{8kT}{\pi M}}$$ ----------5

Where K=Boltzmann constant.

M=Mass of atom

Considering the component of velocity in x-direction as ϑ_x the frequency shift seen by the observer would be

$$\nu = \nu_0\left[1+\frac{\vartheta_x}{c}\right]$$ ----------6

When the radiating atoms are in thermal equilibrium, they have velocity distribution and probability P(ϑ_x) of the atoms having velocity between ϑ_x and (ϑ_x+dx) is given by Maxwellian Probability distribution function

$$P(\vartheta_x) = \left(\frac{M}{2\Pi kT}\right)^{\frac{1}{2}} \exp\left\{-\frac{M}{2kT}\vartheta_x^{\,2}\right\} d\vartheta_x$$ ----------7

Which has a Gausssian shape of the form e^{-ax^2} where 'a' is positive constant.

Probability of finding the atom with in the velocity range

(ϑ_x + dx, ϑ_y + dy, ϑ_z + dz) is given by

$$P(\vartheta_x, \vartheta_y, \vartheta_z) = \left(\frac{M}{2\Pi kT}\right)^{\frac{3}{2}} \exp\left\{-\frac{M}{2kT}\left\{\vartheta_x^{\,2}+\vartheta_y^{\,2}+\vartheta_z^{\,2}\right\}\right\} d\vartheta_x d\vartheta_y d\vartheta_z$$ ---------8

Any atom can be located some where by normalizing factor

$$\int_{-\alpha}^{+\alpha}\iiint P(\vartheta_x\vartheta_y\vartheta_z)d\vartheta_x d\vartheta_y d\vartheta_z = 1 \qquad \text{----------(9)}$$

Since the Doppler shift is based only on the component of velocity moving either towards or away from the observer, the G(ν) dν ,the probability of the transition frequency between ν and ν+dν is that equal to the probability of atoms having velocity ν_x and

$\nu_x + d\nu_x$.

The solution of definite integral leads to

$$G(\nu) = \frac{c}{\nu_0}\left(\frac{M}{2\Pi kT}\right)^{1/2} \exp\left\{-\left(\frac{M}{2kT}\right)\left(\frac{c^2}{\nu_O^{\,2}}\right)\left(\nu - \nu_o^{\,2}\right)\right\} \text{----------(10)}$$

The Intensity of emission line as a function of frequency is

$$I(\nu) = \left(\frac{M}{2\Pi kT}\right)^{1/2}\left[\frac{c}{\nu_o}\right] I_o \exp\left\{-\left(\frac{M}{2kT}\right)\left(\frac{c^2}{\nu_o^{\,2}}\right)(\nu - \nu_o)^2\right\} \text{----------(11)}$$

which is normalized as $\int_0^a I_{(\nu)}\, d\nu = I_0$ ----------(12)

I_o = Total emission intensity from the transition.

The eqn (11) is the inhomogenous (Gaussian) shape function which dominates the homogenous (Lorentzian) line shape function, if Doppler broadening is larger than homogenous broadening

The Doppler width $\Delta\upsilon^{D} = 2\nu_o\sqrt{\frac{2(\ln 2)kT}{Mc^2}} = \left(7.16x10^{-7}\right)\nu_o\sqrt{\frac{T}{M_N}}$ ----------(13)

Eqn (11) can be interms of $\Delta\nu^D$ as

$$I(\nu) = \frac{2(in2)^{1/2}}{\Pi^{1/2}\Delta\nu^D} I_o \exp\left\{-\left[\frac{4(\ln 2)(\nu - \nu_o)^2}{\left(\Delta\nu^D\right)^2}\right]\right\} \qquad \text{----------(14)}$$

The shape of this emission line consists of sum of Lorentzian function of atoms traveling in different directions radiating different frequencies.

These are all superimposed to produce the total Gaussian emission

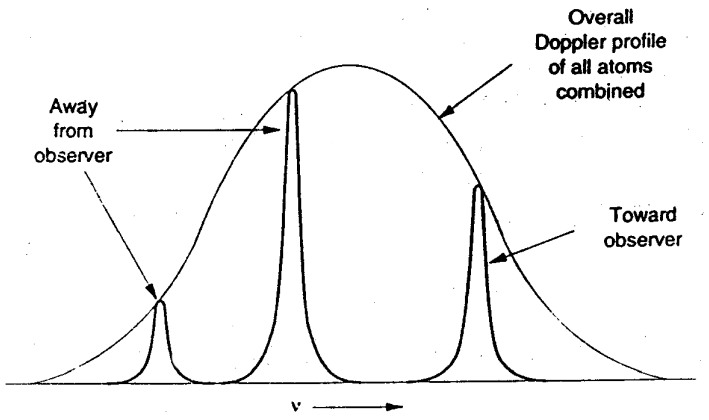


Fig:2.5.6. Natural emission linewidth (Lorentzian profile) of individual atoms traveling in different directions

Doppler verses Natural Broadening

Laser species	λ(nm)	$\Delta\nu_N$(Hz)	$\Delta\nu_D$(Hz)
He-Ne	632.8	$1.4\text{x}10^7$	$1.5\text{x}10^9$
Argon ion	488.0	$4.5\text{x}10^8$	$2.7\text{x}10^9$
Copper	510.5	$3.6\text{x}10^5$	$2.3\text{x}10^9$

The natural line width ($\Delta\nu_N$) is generally much smaller than Doppler line width $\Delta\nu_D$ Doppler broadening is the dominant broadening in most of the gas laser transitions. The Doppler width $\Delta\nu_D$ is proportional to the frequency ν or $\frac{1}{\lambda}$ as well as to the square root of the ratio of the gas temperature to the atomic weight $\left(\sqrt{\frac{T}{M}}\right)$.

2.5.9. Summary of Lesson:-

The emission and absorption mechanisms for an output from a laser are described. It is shown that the nature of line width of a laser transition is due three types mechanisms associated with transition environment. They are

1) Natural line broadening mechanism
2) Collision line broadening mechanism
3) Doppler line broadening mechanism

2.5.10. Key terminology :-Line width – Radiative decay – Natural, Emission line width - Collision decay, Collision broadening - Doppler broadening

2.5.11. Self assessment questions:-

1.Discuss the mechanisms responsible for total line width in a laser transition.
2.Compare the line broadening contributions from each of the responsible mechanism.
3.Discuss the limitation line width due to radiating decay.

2.5.12. Reference Books :-

1) Introduction to lasers and their applications D.C. Oshea, W.R.Callen and W.T. Rhodes
2) Laser Fundamentals – William T.Silfvast

Unit-II
Lesson-6

Different Lasers

Objective: To present a detailed description about the working of the ruby, He-Ne, CO_2 and semiconductor lasers.

Structure:

2.6.1 **Ruby laser**:-

The first laser was constructed in 1960 by T.H.Maiman at Hughes Aircraft Corporation research laboratories. It was operated on pulse basis employing a crystal of pink ruby as the active medium. The term ‘ doped insulator laser’ as used to describe a laser whose active medium is a regular array of atomic crystal with impurity ions intentionally introduced into crystal at the time

of its growth through a process called doping. These are rugged, simple to maintain and capable of generating peak powers.

Active medium: - This dopant- insulator laser consists of an impurity or dopant in a crystalline insulator. The crystal atoms act as a host crystal lattice. The dopants are considered as a 'frozen gas' of heavy ions randomly distributed through out the crystal. The crystalline field partially removes the degeneracy and is important in determining the absorption & emission characteristics of dopant. The dopant Cr^{3+} ion in a free state, has 28 degenerate quantum states of chromium ions in crystal lattice of Al_2O_3, strongly absorb the blue & green bands of the visible spectrum. The chromium ion dopant in Al_2O_3 (sapphire) gives a beautiful deep red coloration to ruby. This absorption is due to the splitting of number of energy states of ion due to its position in crystal lattice. Most of the levels fall into these two bands shown as 2 & 3 (blue absorption, green absorption bands) in the figure. Transitions from 1 ground level to these levels in groups 2 &3 correspond to the observed strong absorption.

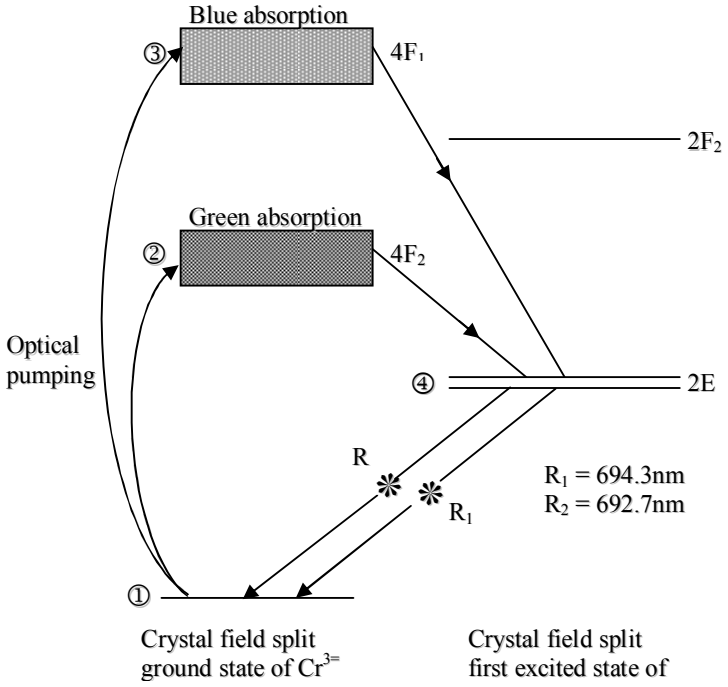


Fig: 2.6.1. Ruby Laser-Energy level diagram

The first excited state of the ion (Cr^{3+}) also splits into number of energy states shown as 4 just below the green absorption band (2). After absorption, the ions undergo non-radiative transitions to lower metastable state 4.These are the upper laser transition levels in ruby responsible for the emission of 694.3nm and 692.7nm called as R_1 & R_2 lines. Stimulated emission at 694.3nm usually dominates. As Ruby operates on three level system, vigorous pumping is necessary to attain threshold population inversion between ground and metastable state.

Pumping:-

The excitation of ions to electronic levels 2 & 3 is obtained by using xenon flash lamp as a pumping source. Ruby rod was placed at the center of helical flash tube and energy from a bank of capacitors was given to flash the tube. The polished and silvered end surfaces of ruby rod act as a resonator mirrors. A schematic diagram of the doped insulator laser system is shown as follows.

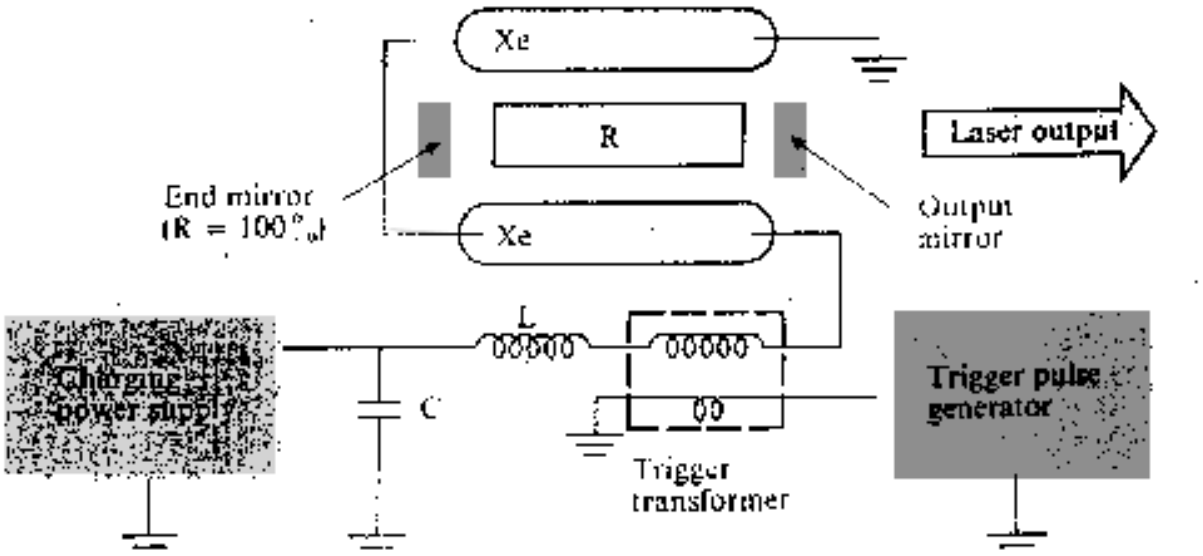


Fig: 2.6.2 : Flash lamp – pumped laser

The flash lamps (Xe) are connected in series with an energy storage bank consisting of more capacitors & with an inductor 'L' to limit the peak current through lamps. The flash lamps are fired applying a trigger pulse to the primary transformer. High voltage pulse of 20 KV ionizes the gas inside the flash lamp tube. R_1 (694.3nm) line predominates over R_2 (692.7nm). Ruby laser originally designed as a pulsed laser. However, improved designs of the system enabled to operate on CW mode also.

2.6.2 He- Ne Laser:-

In 1960 December, Ali Javan from Bell Telephone Laboratories constructed first Cw gas laser using a mixture of helium and neon gas as an active medium. Electrons in the discharge accelerated by the electric field between pair of electrodes collides with atoms of active medium induce transitions to higher energy states and population inversion is created.

A non ionized atom of helium neon mixture in the ratio 10:1 is used as a lasing medium. The energy levels of neon are directly involved in the large transitions while helium gas provides an efficient excitation mechanism to neon atoms. D.C discharge created by placing a high voltage across a gas filled space acts as a source of excitation to various higher energy levels in both species.

Helium atoms are excited by electron impact to low- lying metastable energy states. These metastable states $2\,^3S$ and $2\,^1S$ have almost same energy as that of 4S and 5S of neon atoms. By virtue of resonant collision between helium and neon atoms, the energy of helium levels is transferred to neon species. Thereby the population of neon atomic states of 4s and 5s are enriched to reach the required population inversion. Thus 4s and 5s states are pumped by the metastable states of helium atoms, while 4p and 3p states are depleted because of their short life times. The population inversions between s and p states results amplification by stimulated emission. The lower state populations are depleted by non- radiative transitions to ground state.

There are more number of laser transitions in the He-Ne laser than that are shown in the diagram since each energy state of neon splits into several sub levels. Thus about 130- plus stimulated emission lines are observed in neon. Though the both 633nm and 3.39μm transitions start from the same upper energy level states (5s), 3.39μm infrared transitions has much higher gain to deplete the 5s level eliminating the visible transition at 633nm. However the laser mirrors are designed to be highly reflective at 633nm but highly transmissive at 3.39 μm to stop the infrared transition to reach the necessary threshold gain.

The essential elements of He-Ne laser are the discharge tube containing the He-Ne mixture at the ratio 10:1, power supply and resonator mirrors. A large resistor called ballast resistor is used to limit the current protecting the power supply and stabilizing the operation of the tube.

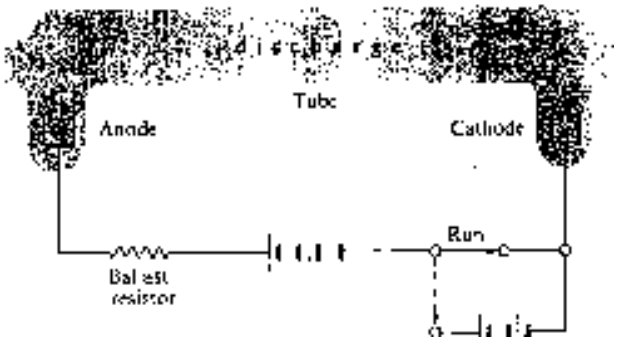


Fig 2.6.3. Simplified electrical circuit for a gas laser. A larger voltage is needed to start the discharge than to maintain it, so a high – voltage is applied to the gas when the laser is turned on. The ballast resistor serves to limit the current once the discharge is initiated

The discharge tube is sealed with Brewster angled windows to give polarized output and to reduce loss due to transmission. The external mirrors M_1 (100%) and M_2 (95%) are used in stable optical resonator configuration. The external mirror configuration is used

1) For the advantage of inserting frequency selective and light switching devices can be inserted into the cavity.
2) To change the dielectric mirrors for different frequency ranges and coherence requirements.

Most of the He-Ne lasers operate to give output powers lie 0.5 to 5.0 mw range and with life times of 50,000 hours.

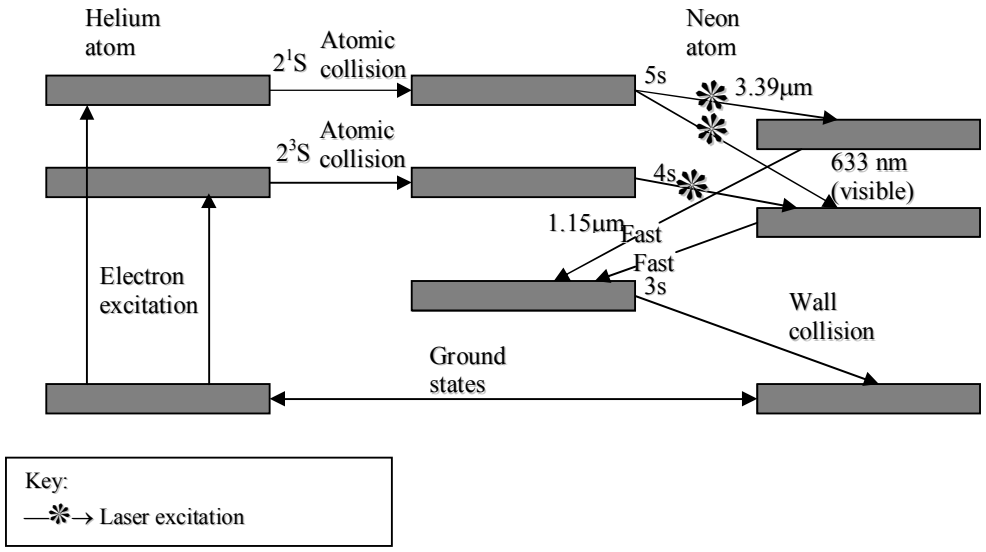


Fig: 2.6.4. Energy-level diagram of the helium-neon laser system

2.6.3.CO_2 Laser

Among molecular lasers, CO_2 laser unquestionably ranks first in both high power and high efficiency operating at infrared wavelengths 10.6 μm and 9.6μm. CO_2 is a linear triatomic molecule composed of 2 oxygen atoms and a carbon atom between them, undergo three different types of vibrational oscillations called vibrational modes.

The energy of oscillation of a molecule in any mode can have only discrete values in integer multiples of some fundamental value. The energy state of the molecule is represented by three numbers (i, j, k). Each number represents amount of energy or number of energy quanta associated with that mode. In addition to vibrational states rotational states are also associated with each vibrational state. The separations between vibration- rotation states are usually much smaller than the separation of electronic states. The vibrational- rotational transitions are in near infrared, for most of the molecular lasers.

The various low lying vibrational energy levels of CO_2 molecule corresponding to ground state are responsible for laser transitions. CO_2 molecule vibrates in

1. Symmetric stretching mode (ν_1) at 1388cm^{-1}.
2. Bending mode (ν_2) at 667cm^{-1}
3. Asymmetric stretching mode (ν_3) at 2349cm^{-1}

the energy level diagram for CO_2 is shown in fig.

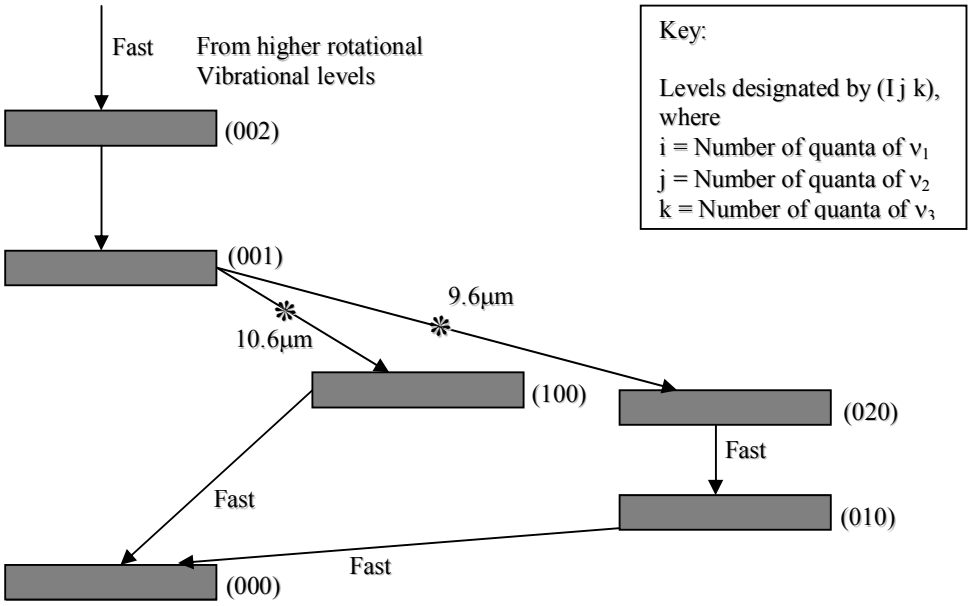


Fig.2.6.5: Energy-level diagram for the CO_2 laser

The operational efficiency of CO_2 laser transition is improved by adding nitrogen and helium to active medium. The energy of the nitrogen molecular vibrational quanta is transferred to CO_2 (001) state by resonant collision and increases the population in 001 state so as to form population inversion between (001) and other lower (100) and (020) states of CO_2. Helium will speed up the depopulation of the lower(100) transition level to produce large population inversion between the upper (001) and lower (100) levels of CO_2. The two laser transitions between (001) to (100) and (001) to (020) produce the output at 10.6μm and 9.6μm respectively. CO_2 lasers are capable of producing tremendous amount of output power at 10.6 μm due to its high 30% efficiency. Gigawatts of peak powers are produced in short nanosec. duration pulses. Because of its high power output efficiency, the CO_2 laser is of great practical importance. The output power is proportional to the active length of the laser medium.

The CO_2 vibrational rotational energy states are shown as bands with their characteristic mode quantum numbers. Laser transitions are shown due to transitions between (001) to (100) gives the infrared at 10.6 μm and the transitions from(001) to (020) gives the infrared 9.6μm output. The addition of N_2 gas to CO_2 increases the efficiency of the operation. The collision transfer of energy of N_2 accumulated in metastable ν=1 level to (001) mode of CO_2 occurs due to approximate energy equivalence of these vibrational levels. The CO_2 lasers are also much more efficient because of the involvement of transition between vibrational rotational levels of the lowest electronic level.

Though the active medium is cheap (CO_2+N_2), the important special components like the resonator mirrors and Brewster angled windows of germanium, cadmium sulphide or sodium chloride, potassium bromide transparent in infrared at 10.6μm ; are costly and very difficult to maintain them due to hygroscopic nature. A diffraction grating mounted an piezo- electric transducer is used to displace of high reflectivity mirror to tune the laser output over 20 distinct lines within either of two major bands at 9.6 μm and10.6 μm. A little H_2O is added to the mixture to convert the CO formed during operation back to CO_2

High power CO_2 lasers are used in industry for welding, hole drilling, cutting, etc…in industrial workshops.

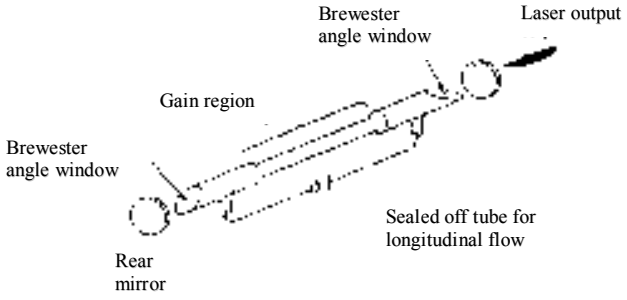


Fig 2.6.6: Longitudinal discharge CO_2 laser

2.6.4.TEA- CO_2 LASER:-

In view of high output power, a new technique has been developed to increase the output by transverse excitation of CO_2 species at atmospheric pressure or more. TEA is an acronym for Transverse Excitation at Atmospheric pressure for greater output. This technique requires about 12kv per cm is necessary to initiate and maintain the discharge. Thus the discharge is arranged to take place at a number of points in a direction transverse to laser cavity. With this arrangement, Gigawatts of peak power can be obtained. Each cathode pin is connected through ballast resistor to avoid negative resistance characteristics of gas. M_1&M_2 are resonator mirrors at stable configuration.

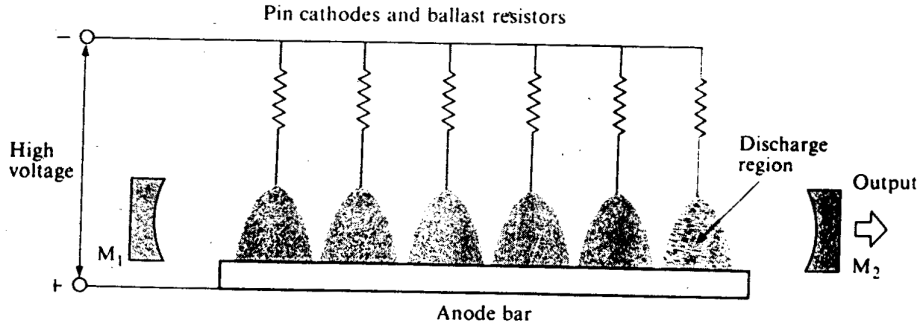


Fig 2.6.7. TEA LASER: - The discharge occurs perpendicular to the laser

2.6.5.The gas dynamic laser –CO_2:-

A population inversion is obtained through the application of thermodynamic principles rather than by standard discharge tube techniques. The CO_2 gas mixture at high pressure is made to flow through a small aperture in a transverse direction to laser axis as shown in the fig. Excitation occurs as a result of heat input into the gas to populate the upper laser level. The rapidly flowing gas is then allowed to expand supersonically through an expansion nozzle into the low pressure region using high speed pumps. The sudden expansion causes the gas to super cool and provide population inversion between upper level and lower level due to rapid relaxation. As mentioned earlier, N_2&H_2O vapour will enhance the efficiency of laser operation. Hundreds of

kilowatts of power has been obtained of 4ms duration. The only disadvantage of this set up is its bulk and rocket like roar that accompanies the gas expansion. The used gas mixture is again re-circulated through high speed pumping system.

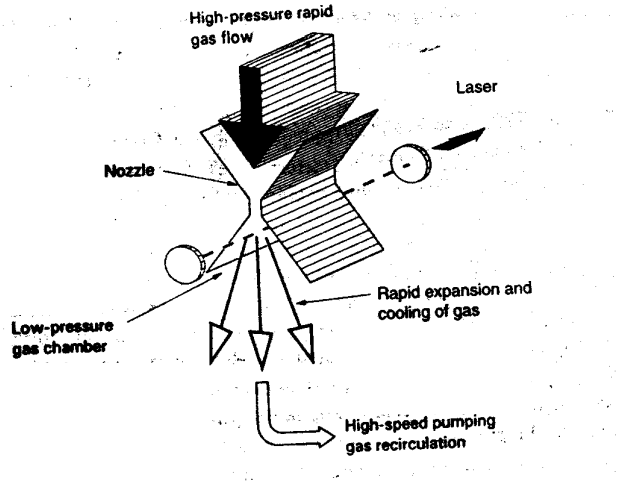


Fig 2.6.8 : CO_2 dynamic laser

2.6.6.Semiconductor laser:-

Materials that have electrical conductivity lying in between the conductors and insulators are called semiconductors. Solid materials are formed by the arrangement of atoms very closely. They are so close that discrete atomic energy levels are perturbed to result energy bands. The energy levels are very close as if they are continuous.

In metals the outermost electrons act as free electrons and constitute a flow of current when electric field is applied. In insulators no such free electrons exist and do not conduct electricity.

In semiconductors, the availability of free electrons depends on temperature and conductivity increases with the increase of temperature. This is in contrast with conductors.

The most popular semiconductors are germanium and silicon in which bonds are formed by sharing adjacent electrons. When the temperature of the solid increases electrons tend to break

up from the band and become free. This leads to form a vacant sites called holes. These holes behave as electrons but with positive charge. When voltage is applied across the material, electrons tend to flow towards positive terminal and holes tend to flow towards negative terminal leading to flow of current.

The energy bands of a semiconductor are separated by energy gaps in which no energy levels exist called forbidden gap. The typical band gap energies are 0.76ev for germanium and 1.1ev for silicon. The forbidden gaps for insulators are considerably larger, for example the diamond has 6ev energy gap. At absolute zero, the electrons occupy the lowest energy levels in valency band separated by a large energy gap over which empty conduction band exists as shown in the figure.

In increasing temperature from absolute zero, some top electrons from valence band gain enough energy to make a transition to the conduction band. The electron thus transferred create a hole in valence band.The excited electronand hole will take part in the conduction process. The excited electrons will occupy lower levels in conduction band and the holes occupy top levels in the valence band.

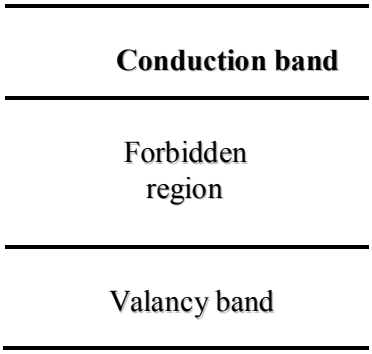


Fig: 2.6.9. Band structure in a semiconductor

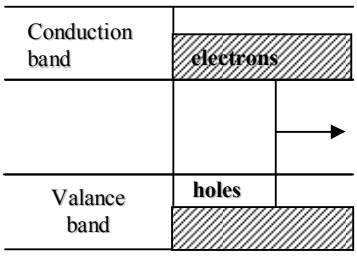


Fig : 2.6.10: An electron and hole may combine and in the process emit a photon

A photon having energy slightly greater than forbidden gap energy produced spontaneously, will stimulate large number of downward transitions of electrons form the conduction to valence band to produce lasing action in the presence of feedback mechanism . Population inversion is easily produced through the use of p-n junctions.

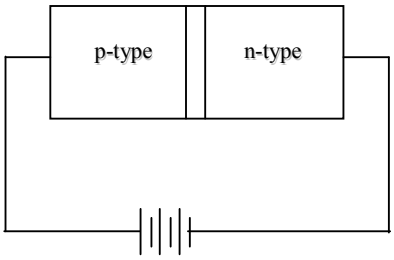


Fig:2.6.11. A p-n junction

Doping with an element containing five outermost electrons (arsenic) to an intrinsic semiconductor, an n-type extrinsic semiconductor is produced. Thus a donor can donate excess electrons to conduction band.

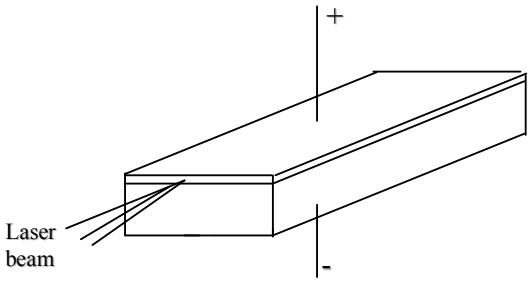


Fig: 2.6.12 A p-n junction laser: the laser emission is confined to a very narrow region around the junction.

Similarly by adding an element with three outermost electrons (gallium/indium), p-type extrinsic semiconductor is produced. The impurity which has fewer valence electrons then the host is called an 'acceptor'.

A number of semiconductors of either form can be produced by adding suitable impurity or a dopant to the material. Tellurium added to gallium arsenide makes it a n-type while zinc makes it p-type semiconductor.

If a junction between p-type and n-type semiconductors is formed, an electric field is created across the junction by the flow of electrons from n-type and holes from p-type semiconductors towards the junction. Connection of positive terminal of a D.C. source to p-type and negative terminal to n-type , produces a flow of electrons form n-region and holes form p-region into the junction. This recombination of electrons and holes produce heat in the case of silicon and germanium which are of no use for lasing action. In gallium arsenide, most of the energy emitted appears as light and it is for this reason GaAs is used as semiconductor laser. The wave length depends upon activation energy i.e., required to free valence electron. This is 1.4 ev. for gallium arsenide to give an emission at 9000 A^0.

A resonant cavity is formed by cleaving the junction ends. At high current density sufficient excitation is provided to have more number of electrons in conduction band than number of holes in valence band to provide population inversion.
The laser emission occurs between two bands of 1μm thickness. As the stimulated emission occurs between two bands of energies, the emission is not as monochromatic as that of a radiation from gas lasers.

The first useful semiconductor laser composed of GaAs materials emitting at 0.8μm to 0.9μm. It is very efficient due to direct conversion of electrical current into light energy and extremely small in size. The output beam can be modulated by modulating diode current. These lasers can be operated for CW operation only at low temperature (77^0k liquid nitrogen temp)

pulsed operation can be obtained at room temperature. However, the improved designs like heterojunction techniques enabled to produce CW operation at IR to UV for a variety of applications (0.5μ.m. to 2.2 μ.m).

2.6.7.Homojunction lasers:-

Lasers of this type consist of a single junction of n and p- doped materials. Due to large amount of heat dissipation, the gain tapering occurs hence this type can be effectively operated at low temperatures. This has historic importance rather than practical utility.

2.6.8.Hetero junction lasers:-

Hetero junction lasers consist of several layers of various materials (semi conduction materials doped and un-doped and metallic layers for conduction of current). A single layer in the center of these layers-the active layer where in gain is produced i.e., a direct band gap material is an efficient radiator, while the adjacent layers are of indirect band gap materials. The materials most often used are either Ga As/Al_xGa_{1-x} As or In_{1-x} Ga_x As_y P_{1-y}/ InP are of III –V group semiconductor alloys. (x and y indicate the concentration of impurity material). Lasers of II-VI compounds such as ZnSe have been produced to operate at wavelengths ranging from 0.460 to 0.530 μ.m. at cryogenic temperatures.

2.6.9.Condition for laser action:-

From Fermi-Dirac statistics the probability of occupation f (E) of any energy state E is given

$$f(E) = \frac{1}{1+\exp[(E-F_o)/KT]}$$

where F_0 is knows as Fermi level for the system

when $T \rightarrow 0\ ^{o}k$, $f(E) = 1$ if $E < F_o$

$= 0$ if $E > F_o$

Thus F_o Fermi level represents the boundary between fully occupied and completely empty levels at $T=0^0k$.

The electrons raised to conduction band occupy the lowest level in that band. The electrons in the valence band also dropped to unoccupied lower levels leaving the top of valence band for 'holes' to occupy. The process that increases population in CB increases f(E) by raising F above equilibrium value F_o. Similarily increase in hole concentration lowers F below F_0. Thus there is population inversion between conduction and valence bands suitable to lasing action.

Condition-

Occupation probability "$f_c (K_f)$" of any state in conduction band is expressed in terms of different Fermi level F_c as

$$f_c(k_f) = \frac{1}{1+\exp[(E-F_c)/KT]}$$

"F_c" is called 'quasi- Fermi level' of electrons in C.B.

"k_j" is the wave vector of the state concerned.

This level separates fully occupied and empty levels in C.B.

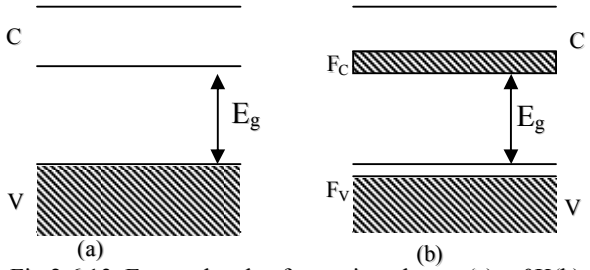


Fig 2.6.13. Energy levels of a semiconductor (a) t=0K(b) electrons excited to the conduction band

Similarly occupational probability of valence band

$$f_v(k_i) = \frac{1}{1+\exp[(E-F_v)/KT]}$$

At equilibrium $F_c=F_v=F_o$

when light beam incident on semiconductor, the number of quanta absorbed 'N_a' per unit time will be proportional to

1. probability per unit time of direct transition from V.B to C.B – B_{vc}
2. density of incident radiation $\rho(w)$
3. the probability of the concerned state in the valence band being occupied $f_v(k_i)$
4. the probability that upper state in CB is empty $[1-f_c(k_j)]$

$$N_a = A\, B_{vc}\, f_v(kj)[1-f_c(k_j)]\, \rho\,(\omega)$$

Number of quanta 'N_e'emitted per limit time by stimulated emission is

$$N_e = A\, B_{cv}\, f_c\,(k_j)\,[1-f_v\,(k_i)]\,\rho\,(\omega)$$

where "A" is a constant of proportionality

For amplification to occur $N_e>N_a$, assuming $B_{vc}= B_{cv}$

substituting and on simplification.

$$F_c(k_j) - F_v(k_i) > E_c(k_j) - E_v(k_i) = \hbar\omega$$

The requirements to be fulfilled for using semiconductor as a laser material are:

1. Transition probability for radiative transition across the gap must be high and must exceed the probability for non radiative transfer of energy to lattice.
2. Excess population should be maintained across the two levels.

Population inversion is obtained by using semiconductor in the form of p-n junction diode heavily doped with donors and acceptors.

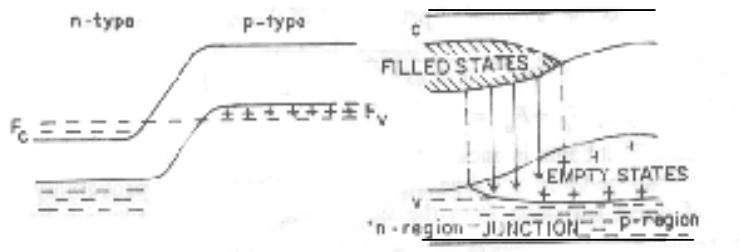


Fig 2.6.14: Energy levels of a p-n junction

In n-type electrons occupied up to F_c in C.B. and in p-type holes are added down to F_v in V.B. The energy difference between p and n-regions is the built in voltage, the Fermi levels lie in the same horizontal level due to potential barrier formed by electrons flow from n-region to p-region. When junction is forward biased, electrons flow to p-side and holes to n-side and overlap in part of the junction region or 'depletion region'. In the depletion region electrons and holes appear in high concentration. The population inversion is formed for lasing action in this region, when the current flow through diode exceeds certain threshold value. Thus the injection of electrons and holes into junction region from opposite sides, population inversion is created between filled levels in C.B. and empty levels at the top of V.B. The recombination of electrons and holes generate coherent radiation when current flow is above threshold value i.e. a stage when gain exceeds the absorption or losses the laser action is confined to a very thin planar junction region.

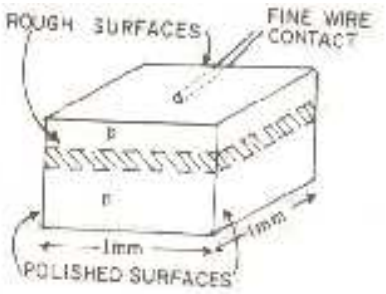


Fig: 2.6.15: injection laser

2.6.10.Injection laser:-

The first laser involving transition between energy bands was reported by Hall et al. in 1962 by direct conversion of electrical energy into coherent emission from GaAs p-n junction.

Diodes used were cubes with edges 0.4 mm long with junction lying in horizontal plane through guide. The front and back faces were polished parallel to each other and perpendicular to the plane of the junction to form an optical cavity. These lasers are called injection lasers since laser action was created by charge carriers injected into semiconductor diode.

The active region consists of layer thickness of micro meter emits a beam of 40 micro meter with a divergence of order 5^0 to 15^0.

2.6.11.Injection laser threshold current:-

Let 'A' be the area and thickness be 'd' of the diode.

N_2 be the number density of C.B and

$N_1 = 0$ of V.B. the electrons decay time τ_1 the current required to maintain upper state population

$$I = \frac{N_2 Ade}{}$$

τ_1

The required threshold current density is $J = I/A = N_2de/\tau_1 \approx 35$ to 100 KA/cm^2

At T> 77 k, "J" increases rapidly due to decrease of probability product resulting decrease in gain. Hence it is not possible to operate above certain temperature "T_c". CW -output powers up to a few watts have been obtained with GaAs laser at 77k.

InAs and InSb lasers have been designed by varying concentrations the output can be tuned in the range 5 to 25 μm with a fine tuning by varying current input.

Advantages:-

Compact, simple and efficient due to direct conversion of current to optical energy it can be linked directly to fiber communication.

Disadvantages:-

It is very difficult to control mode pattern of the output spectral purity, and monochromaticity are much poorer than gas lasers.

2.6.12. Summary:

The construction and working details of ruby, He-Ne, CO_2 and semiconductor lasers are described. The merits and demerits of each class of lasers are mentioned.

2.6.13. Key Terminology-

* Ruby rod formation with doped Cr^{3+} ions.
* He-Ne mixture, resonance transfer of energy
* CO_2 modes pf vibration, Nitrogen, Helium and H_2o in CO_2,
*TEA laser, Gas dynamic laser.
*Semiconductor laser, electron-hole recombination home junction and hetro-junction lasers.

2.6.14. Self assessment question:

1. Describe the construction and working of ruby laser.
2. Explain the construction and working of He-Ne laser
3. Discuss the various types of CO_2-lasers.
4. How lasing is produced in a homojunction laser.
5. Explain the working of injection laser with advantages

2.6.15. Reference Books:-

1. Lasers theory and applications K.Thyagarajan, and A.K.Ghatak.
2. Introduction to lasers and their applications D.C. Oshea, W.R.Callen and W.T. Rhodes
3. Laser and non-linear optics – B.B.Laud.
4. Laser Fundamentals – William T.Silfvast

Unit III

LESSON-1

HOLOGRAPY-BASIC PRINCIPLES

Objective: To present the rudiments of holography like hologram construction and reconstruction and basic required conditions.

Structure

3.1.1 Introduction: - The conventional method of recording optical images is the photographic method in which the intensity variations are recorded on a photograph on two dimensions. The third dimension i.e. the depth of the scene is not at all recorded. Fundamentally a new concept of recording optical images with full information (three dimensional aspects) is known as Holography. The word holography originates from the Greek word ”holos”, meaning the whole. Holography means ‘complete recording’. This technique involves the complete recording of the amplitude of the scattered object wave and phase components using the interference technique. In 1947, Danish Gabor proposed and demonstrated this idea using mercury arc lamp as a source of light. However, till the advent of laser this idea has not been used by scientific community for want of proper source. Leith and Upatnicks have successfully produced laser holograms using off-axis method of recording technique. Thus laser by virtue of its large coherence length has become an indispensable tool in the development of various holographic techniques of paramount importance. This can be understood as given below.

Let us first consider some very general properties of waves and their interference. We know that an atom consists of positively charged nucleus with electrons revolving around it in their respective orbits. A collision with an incoming electron or atom results into the transfer of energy to the atom which causes an electron in the atom to be shifted to a higher level. In this state, the atom possesses more energy than in the ground state and is unstable. After about 10^{-8} seconds, the atom spontaneously returns to its ground state emitting the excess energy in the form of radiation. When an assembly of excited atoms lose their excitation energy by emitting at

random moments randomly phased wave-packets, interference of these waves resulty and what is observed is some mean amplitude of resultant emission.

Let us consider the disturbance produced by the simultaneous action of a number of oscillators. Let ψ_1, ψ_2 ….. be the disturbances produced by individual oscillators. The resultant disturbances can be found using the principle of superposition-a physical hypothesis which can states that for light waves, the disturbance at a point due to the passage of a number of waves is equal to the algebraic sum of the disturbances produced by individual waves. Since the calculations based on this principle can give satisfactory explanation of the observed effects, the principle can be considered as having been confirmed.
Therefore

$$\psi = \psi_1 + \psi_2 + ...(5.1) \qquad \ldots\ldots 3.1.1.$$

Consider an extended source of light consisting of idealized two-level atoms capable of emitting at frequency w. of the total light received at a point Q, (Fig. 3.1.1.)some will come from an atom A_1, some fromA_2, and so on, each acting independently. The waves are randomly phased and we may write for a typical wave

$$\psi_k = a_k \exp\left(i\left(\omega t + \delta_k\right)\right) \_\_\_\_\_(5.2) \quad \ldots\ldots 3.1.2.$$

Where a_k is the amplitude of the wave and δ_K is the phase angle.

Fig. 3.1.1. Light received at a point Q from an extended source.

Classically the energy of the wave is proportional to a_k^2. writing Eq(3.1.2) can be expressed as

$$\psi_k = a_k \exp\left(i\left(X + \delta_k\right)\right) = a_k \left\{\cos\left(X + \delta_k\right) + i\sin\left(X + \delta_k\right)\right\} \_\_(5.3) \quad \ldots..3.1.3.$$

By the principle of superposition, the resultant disturbance is given by

$$\psi = \sum_k \psi_k = \sum_k \cos\left(X + \delta_k\right) + i\sum_k \sin\left(X + \delta_k\right)$$

$$\psi = A\cos\Delta + iA\sin\Delta = A\exp\left(i\Delta\right) \_\_\_\_(5.4) \qquad \ldots\ldots 3.1.4.$$

$$A\cos\Delta = \sum_k a_k \cos\left(X + \delta_k\right)$$

Where we have to put,

$$A\cos\Delta = \sum_k a_k \cos\left(X + \delta_k\right)$$

$$A\sin\Delta = \sum_k a_k \sin\left(X + \delta_k\right)$$

And, hence, the energy of the resultant wave is proportional to A^2 given by

$$A^2 = \left\{\sum_k a_k \cos\left(X+\delta_k\right)\right\}^2 + \left\{\sum_k a_k \sin\left(X+\delta_k\right)\right\}^2$$

$$= \sum_k a_k{}^2 \left\{\cos^2\left(X+\delta_k\right) + \sin^2\left(X+\delta_k\right)\right\}$$

$$+ \sum_{\substack{k \\ k\neq j}} \sum_j a_k a_j \left\{\cos\left(X+\delta_k\right)\cos\left(X+\delta_j\right) + \sin\left(X+\delta_k\right)\sin\left(X+\delta_j\right)\right\}$$

$$= \sum_k a_k{}^2 + \sum_{\substack{k \\ k\neq j}} \sum_j a_k a_j \cos\left(\delta_k - \delta_j\right) \quad \text{_____(5.5) ……..3.1.5.}$$

Since the phase differences vary in a random way, the average value of the summation of the cross products in the second term will be zero. Because for every possible positive value of any term there will be equally probable negative value.

Therefore $\psi = \sum_k \psi_k = \sum_k a_k{}^2 = na^2$ _____(5.6) ……3.1.6.

Where n is the number of excited atoms and a^2 is the mean square amplitude of disturbances. We, thus, see the intensity of illumination –which is proportional to the energy of the waves-is proportional to the number of excited atoms.

Suppose now that by introducing some sort of device we can make the atoms to emit waves which have the same phase, say in which case eq(3.1..4) takes the form

$$\psi = a_1\left\{\cos(X+\delta) + i\cos(X+\delta)\right\}$$

$$+ a_2\left\{\cos(X+\delta) + i\cos(X+\delta)\right\} + ...$$

$$= \cos(X+\delta)\sum_k a_k \qquad \text{3.1.7.}$$

$$= A'\left(\cos\Delta + i\sin\Delta\right) \quad \text{_____(5.7)}$$

$$A'\cos\Delta = \cos(X+\delta)\sum_k a_k$$

That is ,

$$A'\sin\Delta = \sin(X+\delta)\sum_k a_k$$

$E \propto A'^2 = \left(\sum_k a_k{}^2\right) = \left(na^-\right)^2 = n^2a^{-2}$ the intensity is proportional to the square of the number of excited atoms and hence, is much higher than the intensity of the resultant wave produced by randomly phased waves. It is thus obvious that in order to increase the intensity, the atoms must be made to emit waves that are in phase. This leads us to the question of coherence-a property closely linked with the functioning of a laser.

In the plane of photographic emulsion, let the amplitude distribution is A(x,y) exp [i ϕ(x,y)] where A (x,y) and ϕ (x,y) are real functions of x and y (x,y plane). The recorded pattern is proportional to $|A(x, y)\, e^{i\phi(x, y)}|^2 = A^2(x, y)$ in which phase information ϕ(x, y)

is missing. But holographic record contains both the distributions of amplitude and phase. Gabor has shown that the information of both amplitude and phase can be recorded by using the interferometric principle. The scattered wave from the object is made to interfere with a reference beam obtained from the same coherent source. The record of the resulting characteristic interference pattern is called a hologram. When this microscopic interference pattern (hologram) is viewed it will not resemble the object in any aspect as is seen in an ordinary photographic negative or positive. However, the object can be observed with the three dimensional information through a procedure known as wavefront reconstruction.

3.1.2.Principle of Holography:-

Holography is a two step process consisting of

1) Hologram recording
2) Reconstruction of the image.

3.1.2.1. Hologram recording The experimental set-up is shown in fig 3.1.2.

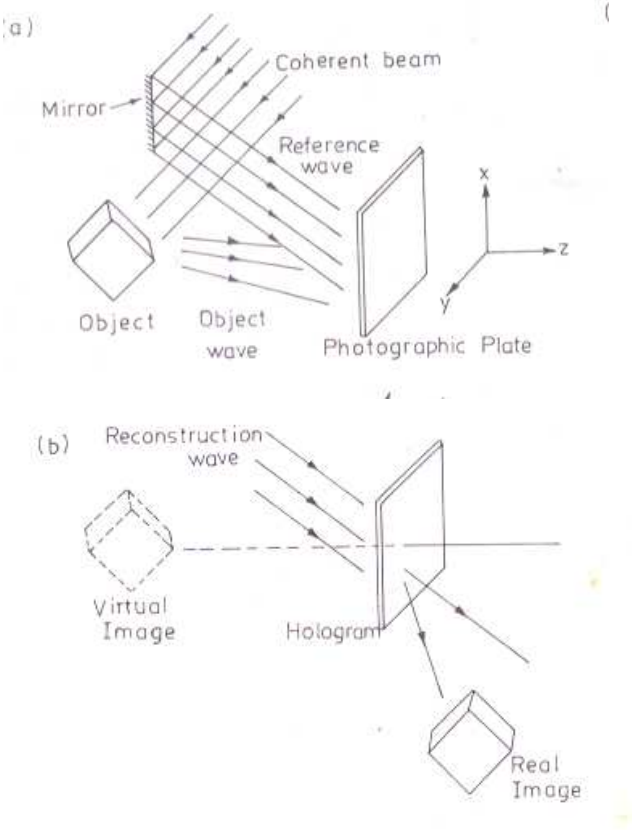


Fig. 3.1.2.

A collimated coherent beam is made into two parts by a beam splitter. One part of the coherent beam is allowed to fall on a mirror and the remaining part is made to fall on the object. The light scattered by the object and the beam reflected by the mirror are made to incident on a photographic plate with an angle θ in between them. The resulting interface pattern of the above

two coherent beams recorded on the photographic plate is called the hologram. Assuming the plane of the photographic plate as x-y plane, and O(x, y) and R(x, y) represent the fields due to object wave and reference waves on photographic plate plane.
The resultant field on the plate is given by

$$U(x, y) = O(x,y) + R(x,y)$$

As the photographic plate responds only to intensity variation, the intensity

$$I(x, y) = |U(x, y)|^2 = |O(x, y)|^2 + |R(x, y)|^2 + O(x,y)\,R^*(x,y) + O^*(x, y)\,R(x, y)$$

Considering time averaging, the resultant pattern recorded on photographic plate after suitable development; represents the hologram of the object.

3.1.2.2 Reconstruction of the image

For reconstruction of an image, the hologram is illuminated by the same reference wave at same angle. The amplitude of the transmitted wave depends on the exposure given on the photographic plate. The exposure is the product of intensity falling on the plate and exposure time. Confining to the linearity of Transmittance- Exposure curve, the transmittance is linearly related to the exposure I

3.1.3. Exposure and transmittance of photgraphic film.

$T = I$

As the hologram is exposed to reference wave R(x,y), then the field emerging from the hologram is

$\phi_{(x, y)} = T_{(xy)}\, R_{(xy)} = [\ |O_{xy}|^2 + |R_{(xy)}|2\ ]\, R_{(xy)} + O_{xy}\ |R_{(xy)}|^2 + O^*_{xy}\, R_{(xy)}\, R_{(xy)}$

Assuming that the reference wave is plane wave with its propagation vector lying in x-z plane, then

$R(x,y) = Ro \exp(-ikx \sin\theta)$

Where $k\sin\theta$ and $k\cos\theta$ represent x and z components of K.

$\phi = [|\ O_{xy}|^2 + R_o^2\ ]\, R_o\, e^{-ikx\sin\theta} + O_{xy}\, R_o^2 + R_o^2\, O^*_{xy}\, e^{-2ikx\sin\theta}$

The first term on RHS corresponds to a wave propagating in the direction of reference wave with an amplitude distortion.
The second term is proportional to the object wave O_{xy} only.

The third term is proportional to O^*_{xy} which is the conjugate of the object wave. This wave in general produces a real image lying on the opposite side of the hologram as shown in fig (b); Additional phase term not only tilts the wave but also introduces distortion in the image.

3.1.3. Basic theory of hologram

To understand the various characteristics of hologram it is necessary to discuss the theory on which holography is based.

Let us first consider the case of a small object, but most of the light falls undisturbed on a photographic plate. The light scattered or diffracted by the object also falls on the plate where it interferes with the direct beam-the reference beam. To find the intensity at a point O on the plate, we may write the field arriving at O as $E=E_r+E_o$ -----(12.1) 3.1.8.

Fig 3.1.4. Hologram of a point object.

Where E_r is the field due to the reference beam and E_o is the field scattered from the object. The scattered field E_o is not simple, both amplitude and phase vary greatly with position. The reflected

wave-fronts are spherical and concentric around the point of origin. We represent the field of such a wave-front by

$$E_o = (A_o / r_o) \exp(i(kr_o - \omega t)) \text{------(12.2)} \quad 3.1.9.$$

And the field E_r by the plane wave $E_r = A_r \exp(i(kz_o - \omega t))$------(12.3) 3.1.10.

Where r_o =PO and z_o is the distance from P to the plate.

Fig. 3.1.5. Reconstruction of the image

The intensity at O is

$$I = (E_r + E_o)^2$$

$$= |A_r|^2 + (A_o^2 / r_o^2) + (A_o A_r / r_o)\exp(ik(r_o - z_o)) + (A_o A_r / r_o)\exp(ik(z_o - r_o)) \text{.....12.4)} 3.1.11$$

choosing the constants K and ϕ suitably we can combine the last two terms from the above relation and write it as

$$I = A_r^2 + (A_o^2 / r_o^2) + K[\cos(k(r_o - z_o) + \phi) / r_o] \text{--------(12.5)} 3.1.12.$$

Because of the cosine term, the total intensity I as a function of r_o shows a series of maxima and minima. The interference of the plane wave E_r with the spherical wave E_o,thus produces a set of circular interference fringes on the plate which, when developed, forms the hologram. If we assume that the plate response is proportional to the intensity I, the power transmission of the plate, T^2, is given by $T^2 = 1 - \alpha I$---(12.6)3.1.13.

Or $T \cong 1 - (1/2)\alpha I$ ---------(12.7)3.1.14.

Where α is a constant.

Let us now see what happens when this hologram is illuminated by the reference beam. The field of the transmitted wave is

$$E = TE_r = (1 - \alpha/2I)A_r \exp(i(kz_o - \omega t))$$

$$= \{1-\alpha/2\ A_r^2 - \alpha/2(A_o^2/r_o^2)\}A_r \exp(i(kz_o-\omega t)) - \alpha/2(A_oA_r/r_o)\exp(Ik(r_o-z_o))A_r\exp(i(kz_o-\omega t)) - \alpha/2(A_oA_r/r_o)\exp(ik(z_o-r_o)\ A_r\exp(i(kz_o-\omega t))$$

$$= \{1-\alpha/2\ A_r^2 - \alpha/2(A_o^2/r_o^2)\}A_r \exp(i(kz_o-\omega t)) -(\alpha A_oA_r^2/2r_o)\exp(i(kr_o-\omega t)) - (\alpha A_oA_r^2/2r_o)\exp(i(2kz_o-kr_o-\omega t))\text{----------(12.8) 3.1.15.}$$

Where we have used (12.4).3.1.11.

The first term of (12.8)3.1.15. represents the same plane wave beyond the plate as the one incident on it, except for the attenuation corresponding to the average blackening of the plate. The second term represents a diverging spherical wave surface identical with the wave surface emitted by the object except for a constant factor. This wave surface when projected back seems to emanate form an apparent object located at the place where the original object was located. This is the virtual image of the object. The third term represents also a spherical wave surface which is a replica of the original wave but has a conjugate or reverse curvature. It converges at a point P^1producing a real image at this point which can be photographed without a lens. The hologram thus produces a virtual image P and a real image P^1.

The general theory of holography is too cumbersome to pursue further. We can however, generalize the treatment given above for a point object to an object of finite size. As before the intensity at the point O is

$$I = (E_r+E_o)^2 = E_r^2+E_o^2+E_oE_r^*+E_o^*E_r. \text{ …….9 3.16.}$$

When the hologram is illuminated by the reference beam, the field on the other side of the plate is

$$E = TE_r = \{1-\alpha/2E_r^2 - \alpha/2E_o^2\}E_r-\alpha/2E_oE_r^2-\alpha/2E_o^*E_r^2\text{----------(12.10)3.1.17}$$

As before the first term gives the attenuated reference wave the second term produces the virtual image and the third term produces a real image of the object at a position which is the mirror image of the virtual image with respect to the plate.

Equation (12.10)3.1.17. shows how holography allows us to make a complete record of the wave coming from the object. In the absence of the reference beam the blackening of the plate would be proportional to E_o^2 i.e. only the modulus of E_o would be recorded. This means that only the amplitude information would be recorded and the phase information is thereby lost. Because of the presence of the reference wave, the field recorded on the plate is proportional to E_o as can be seen from the second term of (12.10)3.1.17., i.e. both amplitude and the phase are recorded thus making the complete reconstruction of the object possible.

Gas lasers operating in the continuous wave (cw) mode, are often used for holography because their coherence is high. However, their emitting power being low, time of exposure has to be large. And, hence, moving objects can not be holographed using gas lasers. Pulsed solid state lasers make it possible to cut down the exposure to about 10^{-3} sec and, hence, can be used to holograph moving objects and to record the development of a process in time.

As the phenomenon involves interference, the radiation should strictly obey certain coherence conditions. The microscopic interference pattern should be stable during recording time. The maximum path difference must be less than the coherence length of the beam. The reconstruction wave must be sufficiently spatially coherent to cover the entire hologram for better resolution. Moving objects cannot be holographed using CW gas lasers. Pulsed solid state lasers are used to holograph the moving objects with $\approx 10^{-3}$ sec exposure

3.1.5. **The hologram Stability, Coherence and other requirements:-**

A non planar reference wave interferes with an arbitrary object wave on the surface of photographic emulsion plate to form a hologram. In very small segments, the curved wave fronts are approximated as planar segments making an angles θ_0 and θ_r with photographic plate which change from region to region. However, interference pattern produce a random grating with a spatial frequency 'f_g'

$$f_g = \left| \frac{Sin\theta_0 - Sin\theta_r}{\lambda} \right|$$

In the process of reconstruction also, the reference wave makes an angle θ_r with hologram to produce the exact object wave in that small particular portion. The entire object wave is reconstructed with proper resolution by the entire hologram.

For an exact re-production of object wave, certain conditions have to be satisfied:
It is desirable to arrange the incidence of reference wave and object wave from two different directions so that the eye can only view the object wave on reconstruction. The maximum path difference between object wave and reference wave should be less than the coherence length of the beam, from the position of beam splitter. A geometric drawing will help to have equal path lengths for both object wave and reference waves. The experimental setup should be arranged on a rigid and vibration free environment so that the interference pattern is stable during recording time. The intensity of the reference wave should be constant and uniform across the holographic plate for proper reconstruction of the object wave. The optical components and the object should be rigidly fixed to a metal table top by magnetic clutches. The air through which beams are passing should remain relatively still to avoid variation of refractive index of air. Variations in refractive index of the medium are responsible for washing out of the hologram. The hologram recording facilities are often typically rigid and supported by pneumatic cylinders to avoid even the small room vibrations caused by neighboring road traffic.

3.1.5. Distinguishing characteristics:

HOLOGRAPH	PHOTOGRAPH
1) It is quite non-intelligible . Since it is an interference pattern of microscopic nature.	1) Every detail is seen clearly
2) The virtual image formed on reconstruction is completely three dimensional form. On tilting the position of head one can notice the objects behind the foreground. New things can be seen on changing the angle of vision	2) Same two dimensional scene can be seen in any angle of vision.
3) Each part of the hologram receives light from all parts of the object. Therefore contains information about all the geometrical characteristics of the entire object. Consequently division of a hologram into separate pieces does not erase a specific portion of the image. Every piece can be used for reconstruction to get entire object image with limited resolution.	3) Destruction of a photograph results an irreparable loss of information corresponding to a part of the object
4) It is reliable storage of information/data and can be shared/used by different persons at the same time	4) The stored information is incomplete and can not be used simultaneously in pieces.
5) It has got enormous information storage capacity since different scenes can be holographed on the same plate by changing the position of the plate through a small angle. Retrieval of the information can be made through orientation of the plate w.r.t incident reference wave. A single hologram of 10 x 10 cms can contain one volume of information.	5) Super position of several scenes on the same photographic plate is an useless act.
6) The hologram itself a negative and the virtual image it produces is a positive. A hologram can be copied by contact printing and can be used for reconstruction purpose.	6) Positives are obtained from negatives only.
7) High resolution photographic emulsions having spatial frequency resolution 3000 lines/mm are required	7) Low resolution films are sufficient
8) Specialized recording equipment is needed	8) Ordinary camera equipment is sufficient.

3.1.6. Holographic Recording materials:-

The photosensitive material coated on optically plane glass plate should be capable of recording interference pattern in microscopic scale, since the spatial frequency of the sinusoidal interference pattern is approximately a few thousand lines per mm ≈ (2000 lines/mm). Eastman kodak type 649 F and Agfa Gavaert type 8E70 photographic emulsions are suitable for large exposures to give good transmittance values. High resolution photographic plates and films should be used for recording holograms, where as ordinary photographic film intended to record intensity variation need not be of high resolution type.

In later development, photoresist, thermoplastics and photopolymers are also used as hologram recording materials. Photoresists commonly used in the manufacture of integrated electronic circuits, are photosensitive materials that form a surface relief pattern on exposing and processing. When the exposure is made from holographic interference pattern, the resultant relief hologram can be used as a master for pressing soft plastic replica phase holograms for mass-production. The principal disadvantage of photoresist materials is of low sensitivity. Thermoplastics and photopolymers are often used in real time holographic applications. The materials are quite inexpensive and of low sensitivity. In data storage applications, photo chromatics and electrooptic crystals are also in use.

3.1.7. Summary of lesson: - Holographic construction and reconstruction procedures have been described. The distinct qualities of a hologram and an ordinary photograph have been presented. The properties of various holographic recording materials have been described.

3.1.8. Key terminology: - Object wave- Reference wave – Hologram- holographic recording materials - Coherence length – Vibrationless experimental setup.

3.1.9. Self assessment questions

1. Distinguish between a hologram and a photograph.
2. Describe the construction of a hologram and reconstruction of the optical image.
3. What are the important conditions to take care while recording a hologram?
4. What are the different materials used to record a hologram and explain their characteristics?
5. Explain the analysis of various images with properties obtained during reconstruction.

3.1.10. Reference and Text books

1. Introduction to lasers and their applications by D.C.Oshea, W.R. Callen and W.T. Rhodes, Addison-Wesley publishing Co., 1978.
2. Laser & Non linear optics by B.B.Laud, New Age International Pvt. Ltd., Publishers, 2001.
3. Contemporary optics by A.K Ghatak & K. Thyagarajan, Plenum Press, 1978.

Unit III

Lesson – 2

TYPES OF HOLOGRAMS

Objective: - To describe various types of holograms qualitatively.

Structure

3.2.1. INTRODUCTION TO TYPES OF HOLOGRAMS:-

The valid assumptions made for the discussion on most of the holograms are:

1. Hologram modifies only the amplitude of the transmitted light in the reconstruction process and does not effect the phase of light.
2. The diffraction effects leading to reconstructed image can be regarded as taking place in a very thin layer essentially within a thin plane:

3.2.2. Absorption hologram:-

The first distinction between different holograms is made on the way in which light is diffracted by the hologram. In absorption holograms, the amplitude of the incident light is modified in the reconstruction process through the absorption by dark areas in the developed emulsion.

3-2-3. Phase hologram:-

In another class of hologram called phase hologram in which optical phase of the incident light is modified in the reconstruction without modifying the amplitude of the light. A phase hologram can be obtained by bleaching, the conventional absorption hologram. Bleaching is the replacement of silver in the developed emulsion with some transparent salt of silver (silver bromide). This salt will produce local variations in the refractive index of emulsion. These will produce a local variation in the refractive index of emulsion. These will introduce corrugations in the reference wave so that a replica of original object wave is produced in the reconstruction process. The advantage of phase hologram is high diffraction efficiency. The reconstructed image from a phase hologram is generally much brighter than the image obtained from absorption holograms.

The second distinction between holograms is much on the basis of thickness on volume of holographic emulsion. Based on the angle between the reference wave and object wave at the emulsion plane, the holograms are classified as:

1) in-line hologram
2) off-axis reference wave holograms
3) Bragg-effect hologram
4) White-light holograms

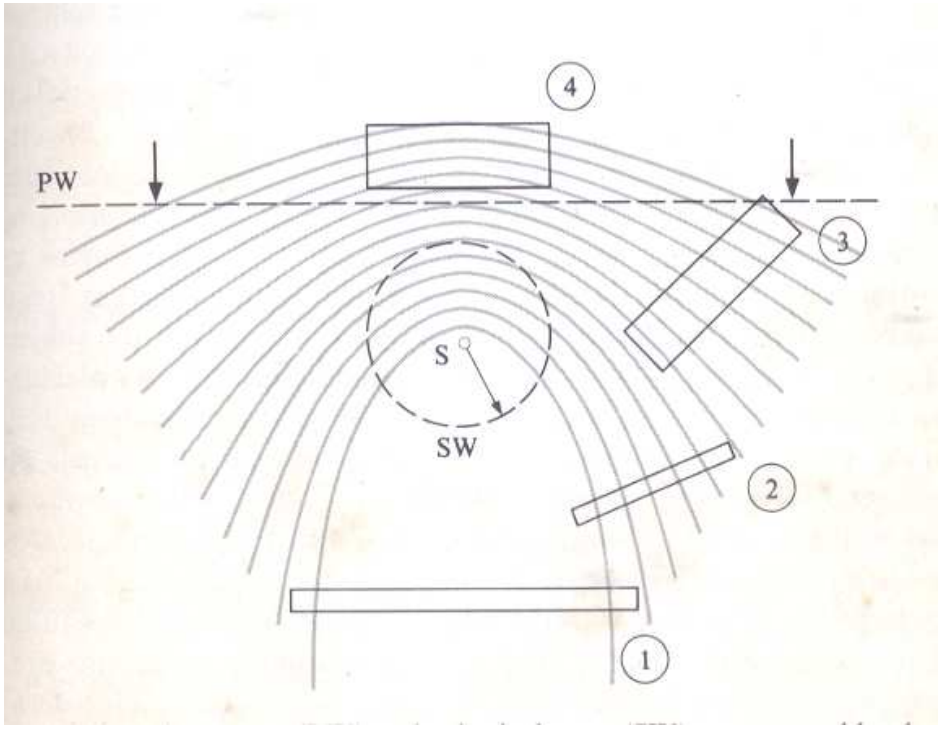


Fig. 3.2.1.: Interfering plane (PW) and spherical wave (SW), represented by dashed lines, produce different grating characteristics in the holographic emulsion. Light lines show loci of interference maxima, where film exposure would be greatest. Thick and thin holographic emulsions are shown in four locations: 1. On-axis, 2. Off-axis reference wave, 3. Bragg-effect, and 4. White-light.

As shown in figure 3.2.1., the variation of the angle between reference wave and object wave can be made by positioning the holographic plate at different angular locations with reference to the above two wavefronts.

Plane waves (PW) interfere with spherical waves (SW) from source and interference maxima are shown. Holographic plates placed at different locations, intercepts the interfering waves.

3.2.4 **In-line hologram:-**

The hologram recorded by placing the holographic plate at position 1 as recorded by Gabor in 1948, is called in –line hologram. The reference wave and object waves almost incident on the plate at same angle. Hence, in the reconstruction both virtual and real images are not separated to give clarity. On the other hand both will superpose one over the other and gives a very clumsy view due to psuedoscopic nature of the real image.

3.2.5 **Off-axis hologram:-**

The technique was developed by Leith and upatnicks in 1962. Holographic plate located at position (2) at which both the waves incident at different angles. On reconstruction, the real and virtual images are clearly separated for an observation of three dimensional effects in virtual image. For small angular offset and for emulsions of not thick, the interference fringes will have large separation compared to thickness of emulsion.

3.2.6 **Bragg-effect hologram:-**

As the angle of the reference wave increases around 90^0 in position (3), the fringe spacing of interference fringes becomes quite small. For an emulsion 12-15 microns thickness, Bragg effect hologram can be reconstructed only if the reference wave is incident at original recording angle.

3-2-7 **White-light holograms:-**

These are known as Lippman-Bragg, reflection, or Denisyuk holograms invented in 1962. Holographic plate located at position
(4) records the interference fringes parallel to the emulsion. Thick emulsions in which several dozen fringes are recorded in its volume, can be used not only for reconstruction but also as an interference filter, for filtering out all but narrow band of colors. Holograms recorded in this position (4) are referred as white-light holograms because they can be used for reconstruction with

white light. The images reconstructed with white light are exceptionally bright and are called Rainbow holograms.

3-2-8 **Multiplex hologram:-**

This is another type of white light hologram, which can produce motion as well as depth. The construction involves a motion picture camera working in relatively circular motion about the object. The motion picture film is developed and a multiple hologram is made by imaging single frames of film in sequence along with the reference wave onto a holographic films.

The developed and bleached hologram is bent into a cylinder and illuminated by a small line source along the cylinder axis. The successive frames recorded by camera in motion on the holographic film corresponding to viewing the object at different angles produce an illusion of depth. The new techniques can enable us to have holographic motion pictures in future.

3-2-9 **Computer generated holograms:-**

The mathematical description of an object an enable the computer to generate the interference fringe pattern similar to that of the pattern of a hologram. These can be plotted with X-Y plotter and photo-reduced to scale. This is exposed to a reference wave for reconstruction and examined from number of perspectives. The disadvantage is the need for great amount of computer processing time. However, they will play an important role in the area of optical data processing and in testing optical elements.

3-2-10 **Fourier transform Holograms:-**

Holograms made with a finite distance from the object are called Fresnel holograms. It is known that the field distribution at the back focal plane of a lens is the Fourier transform of the field distribution at the front focal plane of the lens. Thus the recorded Fourier transform at the back focal plane of the lens will also contain all the information about the object.

Procedure: - In order to record a Fourier transform hologram, the object transparency is placed at the front focal plane of the lens and is illuminated by a plane wave. A photographic plate placed at the back focal plane and is illuminated by a plane wave that forms the reference wave.

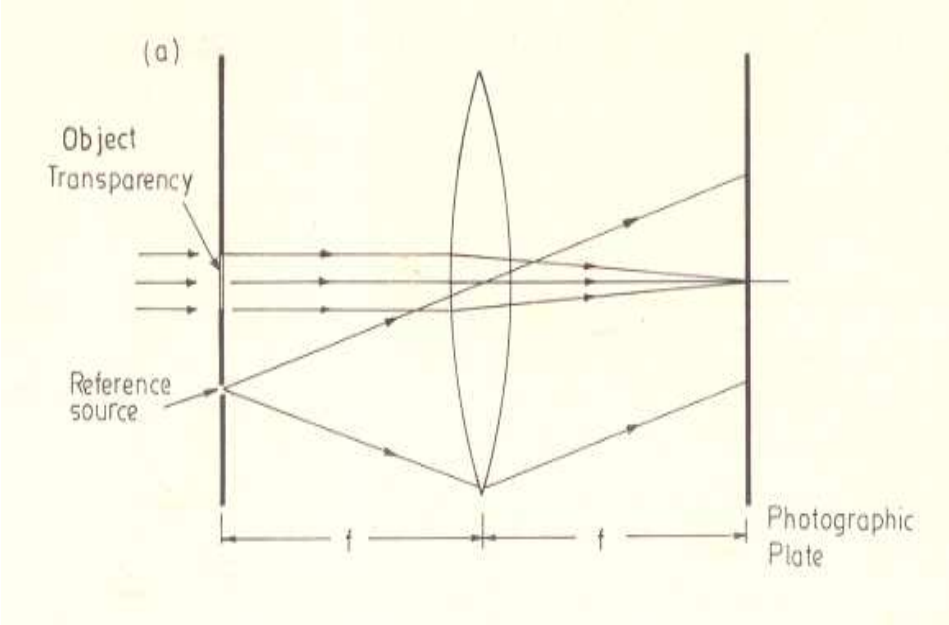


Fig. 3.2.2.: Recording of a Fourier transform hologram

The photographic plate records the resulting interference pattern formed between plane reference wave and the Fourier transform of the object transparency. Recorded pattern is known as Fourier transform hologramThe reconstruction of the object transparency can be made by placing a lens in

front of the hologram and illuminated with a plane wave as shown in figure 3.2.3.

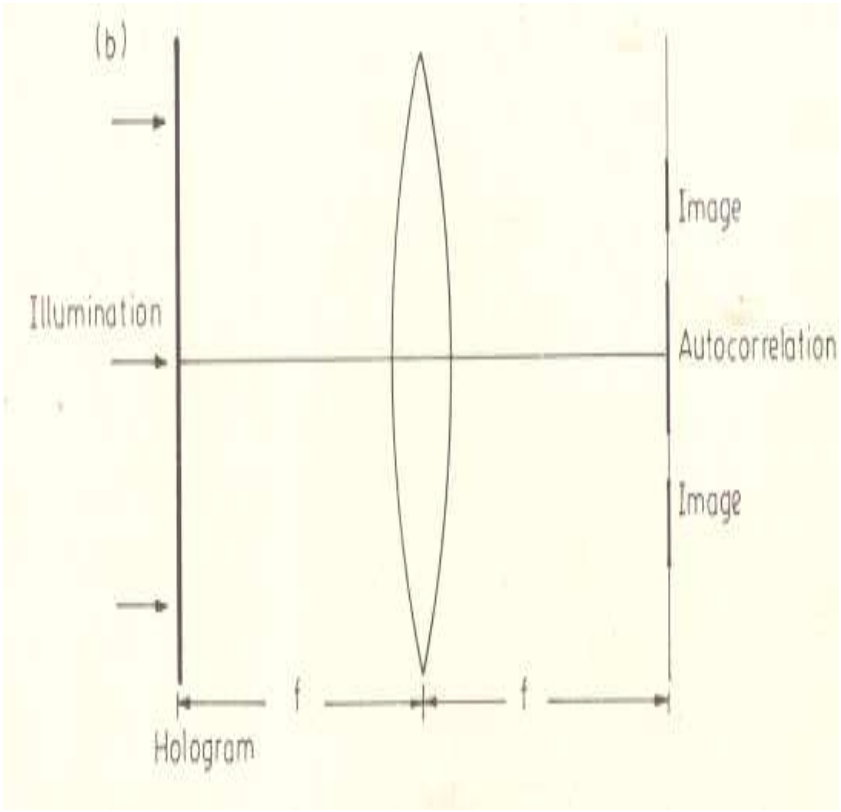


Fig. 3.2.3. Reconstruction of a Fourier transform hologram

The two reconstructed images lie on the back focal plane of the lens. The primary and conjugate images are situated at the back focal plane and selecting proper values of 'a', the primary and conjugate images can be separated. The conditions to be satisfied are $2a \geq 3b$ and $a \approx \theta f$, $\theta \geq 3b/2f$ to obtain two clear images, where 'a' is minimum offset distance of a point reference source and 'b' is the width of transparency.

Fourier transform holograms are widely used in Character recognition problems. Detection of isolated signals from random noise with high resolution.

3-2-11 Volume Holograms:-

In plane holograms the photographic emulsion thickness is negligible compared to fringe spacing of a few microns. When emulsion thickness is in between 5 to 20 μm, the whole volume of emulsion takes part in reconstruction. Such holograms are called as volume holograms.

The interference between two inclined plane waves, the planes of the fringes bisect the propagation directions and forms fringes with separation 'd' according to:

$$2d \sin \theta = \lambda$$

For reconstruction, the fringe system is illuminated by another plane wave. The waves reflected from different planes differ in phase and interfere constructively or destructively depending upon reflective phase differences for maximum intensity the illuminating angle is either θ or $(\Pi - \theta)$. At other angles, the reconstructed image is very weak. At an angle θ, virtual image is reconstructed while at $(\Pi - \theta)$ real image is reconstructed. Due to this directional sensitivity, volume holograms find application in information storage. Various pages of a book can be stored throughout the volume of emulsion. Each page is recorded by changing the orientation of the reference beam. Since the information is stored throughout the volume in a non-localized manner, the hologram is insensitive to dust or scratches etc. Volume holograms find application in recording colored objects by using red, green, and blue regions with same emulsion. When hologram is illuminated with white light, the three colors are reproduced and colored image can be reconstructed.

3-2-12 **Reflection Holograms:-**

In particular, when the object beam and reference beams are traveling in approximately opposite directions as shown in figure. On developing the emulsion the planes of silver atoms are approximately parallel to the emulsion surface with a separation of $(\lambda /2)$ distance. The developed plate behaves like an interference filter i.e., the reflectivity is high only for wavelength with which it was formed. As the hologram becomes highly wavelength selective, the image can be reconstructed even with white light to form a colored image. Such holograms are called reflection holograms since the image can be viewed by reflection

Recording of a Reflection Hologram

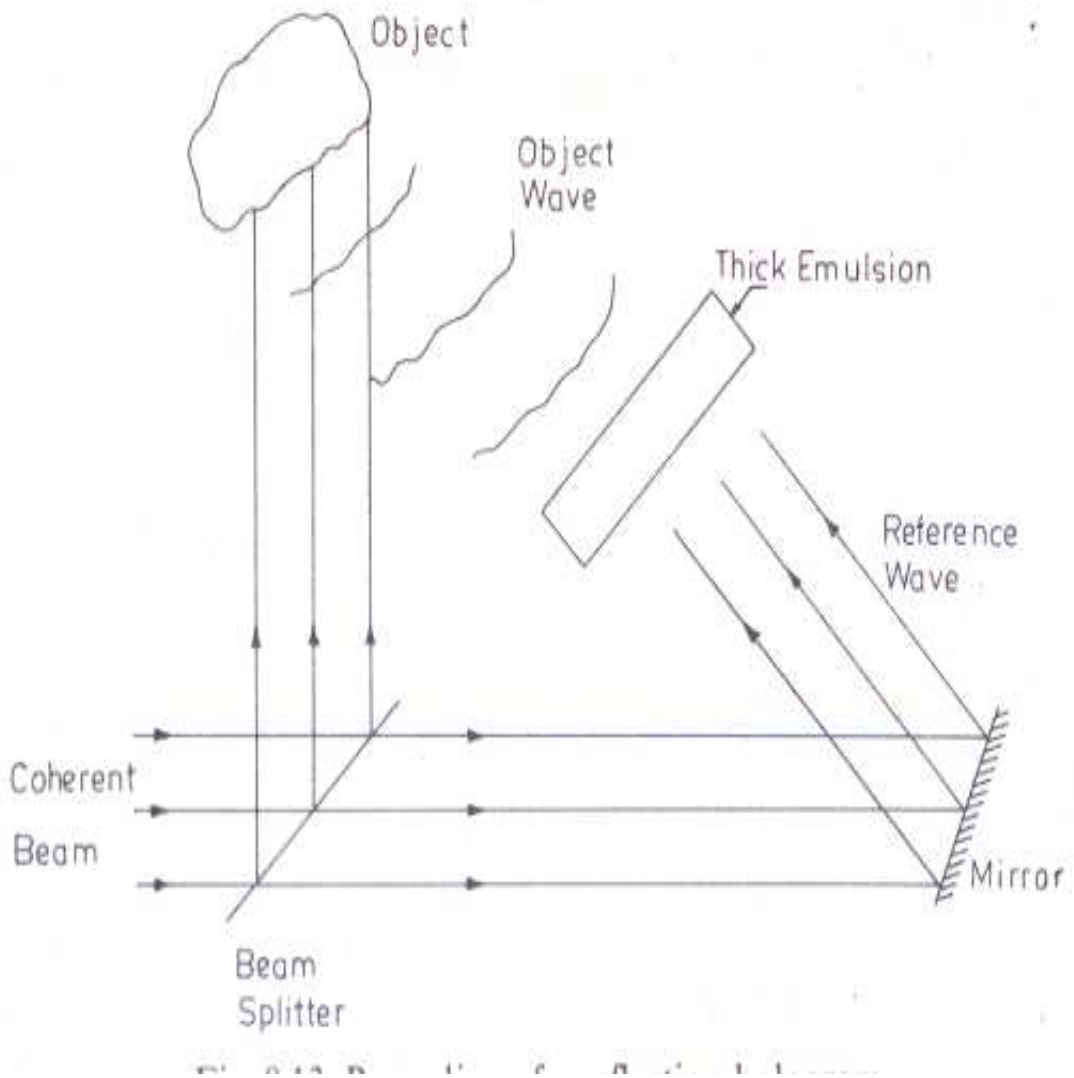


Fig. 3.2.4. Recording of a reflection hologram

3.2.13. Summary:-

Various types of holograms have been described qualitatively with their advantages and disadvantages.

3.2.14. Key Terminology:-

In line hologram – off- axis hologram- white light hologram- Fourier hologram – Volume hologram – reflection hologram.

3.2.15. Self assessment questions:-

1. Describe the advantages of off-axis hologram
2. Discuss the construction of a Fourier transform hologram.
3. Discuss the conditions to record various holograms.

3.2.16. Reference and text books:-

1. Introduction to lasers and their applications by D.C.Oshea, W.R. Callen and W.T. Rhodes, Addison-Wesley publishing Co., 1978.
2. Laser & Non linear optics by B.B.Laud, New Age International Pvt. Ltd., Publishers, 2001.
3. Contemporary optics by A.K Ghatak & K. Thyagarajan, Plenum Press, 1978.

Unit IV
Lesson 3

APPLICATIONS OF HOLOGRAPHY

Objective: - Various applications of holography have been described briefly.

Structure:

3.3.1 Introduction to applications of holography:-

Holography has found application in a number of areas, and new uses are being developed all the time. The novelty and attractiveness of full perspective, three-dimensional displays have inspired the limited use of holography in the advertising business. Of greater technical interest has been the use of holography in the high accuracy measurement of object dimensions, deblurring of photographic images, and the characterization of low-density particulate matter. One especially interesting application has been in microscopy, where a hologram is made of an entire large-volume sample, freezing in one instant a record of many organisms suitable for later inspection at various depths under a microscope.

It is not possible to go into all the applications of holography, because the field is quite extensive and constantly growing. We shall look, however, at three important areas of application: (1) holographic interferometry, (2) holographic optical elements, and (3) holographic optical memories. Each of these applications exhibits a different aspect of holographic principles.

3.3.2. Display of three dimensional images in advertisements

Holography is well known for its capability of projecting third dimensions i.e. depth concept. Advertisers have taken full advantage of using these techniques for displaying the eye catching three-dimensional perspective of their products.

3.3.3. Deblurring of photos:

Two dimensional photos of having blurred nature can be removed with the help of holographic technique. So deblurring of the photos can be done.

3.3.4. Microscopy

The greatest interest in using holography is in the highly accurately measurement of object dimensions and characterization of low-density particulate matter. Very interesting application is in microscopy where a hologram is made of large volume sample, freezing in an instant record of many organisms suitable for later inspection at various depths under microscope.

The magnification associated with the reconstruction (when a radiation of different wavelengths is employed) is proportional to the ratio of wavelengths used for reconstruction to that used during formation. This gives us a method of observing magnified images of objects by increasing the wavelength of the radiation while reconstructing. This was the original idea proposed by Gabor, but because of the lack of sufficiently coherent sources of radiation of shorter wavelengths, this method has still not been put to significant use. At the same time, magnifications of about 100 have been attained with a resolution of several microns. It should be mentioned that a wavelength change during reconstruction always introduces aberrations unless the hologram is also scaled accordingly.

Let S_1 (x_1,y_1,z_1) and $S_2(x_2,y_2,z_2)$ represent two point sources emitting radiation of wavelength λ (fig. 3.3.1.). The plane of the photographic film is represented by z=0. Since z_1 and

Fig. 3.3.1.: Hologram formed by two spherical waves emanating from point sources S_1 and S_2.

z_2 are negative quantities, we introduce

$$\zeta_1 = |Z_1|,\ \zeta_2 = |Z_2|$$

We assume that

$$|X_1|,|Y_1|,|X_2|,|Y_2| \ll \zeta_1,\zeta_2 \quad \text{..... 3.3.1.}$$

The field due to the point source S_1 on the plane of the plate is

$$U_1 = A' e^{-ikr}/r \quad \text{..... 3.3.2.}$$

Where $k = 2\pi/\lambda$ and

$$r_1 = s_1 p = \left[(x_1 - \zeta)^2 + (y_1 - \eta)^2 + z_1^2\right]^{1/2}$$

$$= \zeta_1 \left[1 + \frac{x_1^2 + y_1^2}{\zeta_1^2} + \frac{\zeta^2 + \eta^2}{\zeta_1^2} - \frac{2(x_1\zeta + y_1\eta)}{\zeta_1^2}\right]^{1/2}$$

$$\approx \zeta_1 + \frac{x_1^2 + y_1^2}{2\zeta_1} + \frac{\zeta^2 + \eta^2}{2\zeta_1} - \frac{(x_1\zeta + y_1 n)}{\zeta_1} \quad \text{..... 3.3.3.}$$

The point P (ξ ,η ,0)represents an arbitrary point on the photographic plate. Thus

$$u_1 \approx \frac{A_1}{\zeta_1} \exp\left\{-\frac{ik}{2\zeta_1}\left[\xi^2 + \eta^2 - 2(x_1\xi + y_1\eta)\right]\right\} \quad \text{..... 3.3.4.}$$

Where we have absorbed the constant phase factor

$$\exp\left\{-ik\left[\zeta_1 + \frac{x_1^2 + y_1^2}{2\zeta_1}\right]\right\}$$

in the coefficient A_1 and have put $r_1=\zeta_1$ in the denominator. Equation (3.3.4.) represents the quadratic approximation of a spherical wave.

In an exactly similar manner, if u_2 represents the field at the plane z=0 due to the point source S_2, then

$$u_2 \approx \frac{A_2}{\zeta_2} \exp\left\{-\frac{ik}{2\zeta_2}\left[\xi^2 + \eta^2 - 2x_2\xi - 2y_2\eta\right]\right\} \quad \text{.....3.3.5.}$$

Since the resultant field on the plane z = 0 is u_1+ u_2, the intensity pattern recorded will be

$$I = |u_1 + u_2|^2 = \frac{|A_1|^2}{\zeta_1^2} + \frac{|A_2|^2}{\zeta_2^2} + \frac{A_1 A_2^*}{\zeta_1\zeta_2}\exp^{-i\varphi} + \frac{A_1^* A_2}{\zeta_1\zeta_2}\exp^{i\varphi} \quad \text{.....3.3.6.}$$

Where

$$\phi = \frac{k}{2}\left[(\xi^2 + \eta^2)\left(\frac{1}{\zeta_1} - \frac{1}{\zeta_2}\right) - 2\xi\left(\frac{x_1}{\zeta_1} - \frac{x_2}{\zeta_2}\right) - 2\eta\left(\frac{y_1}{\zeta_1} - \frac{y_2}{\zeta_2}\right)\right] \quad \text{... 3.3.7.}$$

We now illuminate the hologram by a point source placed at the point $S_3\,(x_3,y_3,z_3)$ fig.3.3.2

Fig. 3.3.2. Reconstruction of the hologram

emitting a radiation of wavelength .λ' under the quadratic approximation, the field on the hologram plane is given by

$$u_3 \square \frac{A_3}{\zeta_3}\exp\left\{-\frac{ik'}{2\zeta_3}\left[\xi^2+\eta^2-2x_3\xi-2y_3\eta\right]\right\}\ldots\ldots\ldots \qquad 3.3.8.$$

Where $k' = 2\pi/\lambda'$ and $\zeta_3 = |z_3|$ when this wave illuminates the hologram the emergent wave is given by

$$\phi = u_3 T = u_3 I = \left(\frac{|A_1|^2}{\zeta_1^2}+\frac{|A_2|^2}{\zeta_2^2}\right)\frac{A_3}{\zeta_3}\exp\left[-\frac{ik'}{2\zeta_3}\left(\xi^2+\eta^2-2x_3\xi-2y_3\eta\right)\right]+\frac{A_1A_2^*A_3}{\zeta_1\zeta_2\zeta_3}\exp\left[-\frac{ik'}{2z_p}\left(\xi^2+\eta^2-2\xi x_p-2\eta y_p\right)\right]$$

$$+\frac{A_1^*A_2A_3}{\zeta_1\zeta_2\zeta_3}\exp\left[-\frac{ik'}{2z_c}\left(\xi^2+\eta^2-2\xi x_c-2\eta y_c\right)\right]\ldots\ldots \qquad 3.3.9.$$

Where

$$z_p = \left(\frac{\mu}{\zeta_1}-\frac{\mu}{\zeta_2}+\frac{1}{\zeta_3}\right)^{-1} = \frac{\zeta_1\zeta_2\zeta_3}{\zeta_1\zeta_2+\mu\zeta_2\zeta_3-\mu\zeta_1\zeta_3} \ldots\ldots\ldots\ldots\ldots..3.3.10a$$

$$x_p = z_p\left[\mu\left(\frac{x_1}{\zeta_1}-\frac{x_2}{\zeta_2}\right)+\frac{x_3}{\zeta_3}\right] = \frac{x_3\zeta_1\zeta_2+\mu x_1\zeta_2\zeta_3-\mu x_2\zeta_1\zeta_3}{\zeta_1\zeta_2+\mu\zeta_2\zeta_3-\mu\zeta_1\zeta_3} \ldots.3.10b$$

$$y_p = z_p\left[\mu\left(\frac{y_1}{\zeta_1}-\frac{y_2}{\zeta_2}\right)+\frac{y_3}{\zeta_3}\right] = \frac{y_3\zeta_1\zeta_2+\mu y_1\zeta_2\zeta_3-\mu y_2\zeta_1\zeta_3}{\zeta_1\zeta_2+\mu\zeta_2\zeta_3-\mu\zeta_1\zeta_3} \ldots 3.3.10c$$

$$z_c = \left(\frac{\mu}{\zeta_2}-\frac{\mu}{\zeta_1}+\frac{1}{\zeta_3}\right)^{-1} = \frac{\zeta_1\zeta_2\zeta_3}{\zeta_1\zeta_2+\mu\zeta_1\zeta_3-\mu\zeta_2\zeta_3} \ldots\ldots\ldots\ldots 3.3.10d$$

$$x_c = z_c\left[\mu\left(\frac{x_2}{\zeta_2}-\frac{x_1}{\zeta_1}\right)+\frac{x_3}{\zeta_3}\right] = \frac{x_3\zeta_1\zeta_2+\mu x_2\zeta_1\zeta_3-\mu x_1\zeta_2\zeta_3}{\zeta_1\zeta_2+\mu\zeta_1\zeta_3-\mu\zeta_2\zeta_3} \ldots3.3.10e$$

$$y_c = z_c\left[\mu\left(\frac{y_2}{\zeta_2}-\frac{y_1}{\zeta_1}\right)+\frac{y_3}{\zeta_3}\right] = \frac{y_3\zeta_1\zeta_2+\mu y_2\zeta_1\zeta_3-\mu y_1\zeta_2\zeta_3}{\zeta_1\zeta_2+\mu\zeta_1\zeta_3-\mu\zeta_2\zeta_3} \ldots3.3.10f$$

and $\mu = \lambda'/\lambda$

If we compare the first term on the right -hand side of eq.3.3.9 with Eq.3.3.4. , we see that it represents a spherical wave emanating from the point (x_3,y_3,z_3). Thus the first term is the reconstruction wave itself (with amplitude modification). The second term represents a spherical wave, but the sign of z_p will determine whether the spherical wave is diverging or converging. For $z_p>0$ it will be a diverging spherical wave emanating from the point (x_p,y_p,-zp) . Similarly, if $z_p<0$,it will be a converging corresponds to a diverging spherical wave; hence a virtual image and a negative z_p corresponds to a converging spherical wave and therefore a real image. The last term on the right-hand side of Eq. 3.3.9. can be interpreted similarly it should be mentioned that in the quadratic approximation i.e., when Eq. 3.3.5.is valid there is no distortion in the two images. Further more, since the phase dependence of the second term in Eq. 3.3.9. has the same form as the object wave, it is said to give rise to the primary image; the last term, which has a phase dependence conjugate to that of the object wave gives rise to the conjugate image; hence the subscripts p and c .

The magnification in the primary image is dx_p/dx_1, where dx_p represents the change in x_p when x_1 is changed by dx_1. Thus,

$$M_p = \frac{dx_p}{dx_1} = \frac{dy_p}{dy_1} = \mu \frac{\zeta_2\zeta_3}{\zeta_1\zeta_2 + \mu\zeta_2\zeta_3 - \mu\zeta_1\zeta_3} \quad \ldots\ldots\ldots\ldots \quad \ldots 3.3.11$$

$$M_c = \frac{dx_c}{dx_1} = \frac{dy_c}{dy_1} = -\mu \frac{\zeta_2\zeta_3}{\zeta_1\zeta_2 + \mu\zeta_1\zeta_3 - \mu\zeta_2\zeta_3} \quad \ldots\ldots\ldots\ldots \quad 3.3.12$$

Where M_p and M_e represent the magnification in the primary and conjugate images, respectively, and use has been made of Eq. 3.3.10.

We will now consider some special cases:

(1) When the source corresponding to the reconstruction wave is placed at the same point as the source corresponding to the reference wave and it emits radiation of the same wavelength as the reference source, then we have

$x_3 = x_2$, $y_3 = y_2$, $z_3 = z_2$, $\mu = 1$ ……. 3.3.13.

Thus Eqs. 3.3.10a ….3.3.10f. simplify to

$Z_p = \zeta_1$, $x_p = x_1$, $y_p = y_1$

$$z_c = \frac{\zeta_3\zeta_1}{2\zeta_1 - \zeta_3}, \; x_c = \frac{2x_3\zeta_1 - x_1\zeta_3}{2\zeta_1 - \zeta_3}, \; y_c = \frac{2y_3\zeta_1 - y_1\zeta_3}{2\zeta_1 - \zeta_3} \quad \ldots\ldots\ldots \quad 3.3.14.$$

Thus, the second term on the right-hand side of Eq. 3.3.9represents a virtual image of the object point. The last term of Eq. 3.3.9.represents a real or a virtual image depending on whether z_c is negative of positive, ie. whether ζ_3 is less than a greater than $2\zeta_1$.

The magnification corresponding to the virtual image is unity. However the conjugate image does not, in general, corresponding to unit magnification.

(2) For $\zeta_1 = \zeta_2$, the expression for the magnification [see Eq. 3.3.11. becomes

$$M_p = \mu \zeta_3 / \zeta_1 \quad \text{..... 3.3.15.}$$

which shows that by increasing the wavelength of the reconstruction wave, we can increase the magnification . Further, if ζ_1 is also equal to ζ_3 then the magnification is simply μ. For such a case $z_p = z_c = \zeta_3$ and both primary and conjugate images are virtual and are formed in the original object plane. For such a configuration the terms quadratic in ξ and η in the expression for vanish. This corresponds to the lensless Fourier transform hologram configuration.

(3) Let us consider the case when

$x_1 = y_1 = x_2 = y_2 = x_3 = y_3 = 0$…..…3.3.16.

i.e., the sources for the reference and reconstruction waves are situated on the z axis and the object point also lies on the axis. For such a case Eq. 3.3.10d tells us that $x_p = y_p = x_c = y_c = 0$. Also, if we assume that the object is close to the source, then we may write

$$\zeta_2 = \zeta_1 + \varepsilon \quad \text{..... 3.3.17}$$

Where ε is assumed to be small. If $\zeta_3 = \zeta_2$ and μ =1, Eqs. 3.3.10a and 3.3.10d. give

$$z_p = \zeta_1 = \zeta_2 - \varepsilon, \; z_c \; \zeta_2 + \varepsilon \quad \text{..... 3.3.18}$$

Thus the two images are virtual and situated symmetrically about the reference source.

(4). When the reference and reconstruction waves are plane waves propagating in the +z direction, then in addition to Eq. 3.3.16.we have $\zeta_2 = \zeta_3 = \infty$, and for μ =1,get $z_p = \zeta_1$,

$$z_c = -\zeta_1 \quad \text{.....3.3.19}$$

Thus one of the images is virtual and the other real, and the images are situated symmetrically about the hologram plane.

3.3.5. Holographic interferometry:

Holographic interferometry is a technique for measuring small displacements. In one form or another, it can be used to observe and analyze, either in real time or after some delay,

1. the microscopic flexure of a loaded support beam
2. to observe the strain in a fractured bolt,
3. to record the shock wave from a bullet
4. to observe the vibrational modes of a kettle-drumhead or
5. to detect hidden flaws in an aircraft tire.

The basic principles can be illustrated with the example of

1. single-exposure,
2. Double exposure,
3. Time-average holographic interferometry as applied to stress analysis
4. Speckel interferometry.

3.3.5.1. Single exposure of hologrpahic interferometry:

In this process, a conventional hologram is first made of the object to be tested. The hologram is developed in the normal manner and then returned to the exact position it held during exposure. The object and the hologram are now illuminated just as they were during exposure, so that the reconstructed holographic image falls exactly on the object itself. If the object is now stressed slightly by heating or loading with weights, for example, fringes appear on it. These fringes result from the interference between the light waves from the now-distorted object and the holographically reconstructed waves from the "former" object. Where the distortion is small, the fringes are few in number; where it is great, many fringes appear. A direct qualitative indication of object stress is thus immediately available. An example of such a system of fringes appears in Fig.3.3.3, where the object, a metal bar that is solidly attached at the bottom to a massive base, is

Fig. 3.3.3. :Illustration of single - exposure holographic interferometry. Interference of the actual object wave and reconstructed object wave shows the amount of bending of the metal bar with a slight twist about vertical.

stressed at the top. In the figure, the re-illuminated bar and its holographic image are viewed simultaneously. Note that the rate of change of the beam displacement is greater near the top, the fringes being more closely packed there. Although in general the problem is quite difficult, with special geometries it is possible for one to interpret the fringe pattern to obtain a direct, quantitative measurement of object deformation (and, indirectly, of object stress). for example.

3.3.5.2. Double exposure or real time holographic interferometry:

A variation of real-time holographic interferometry, known as double-exposure holographic interferometry, has found widespread application in several areas of non-destructive testing. Nondestructive testing of tires was, in fact, one of the first large-scale industrial applications proposed for holography. The basic approach is similar to the real-time technique just discussed. In the present case, however, the original and the distorted objects are both recorded holographically in a double-exposure hologram. When the hologram is re-illuminated with the reference wave, both images are viewed simultaneously, and the interference pattern representing the differences in the two recorded light distributions appears localized on the object.

If the object has changed shape during the two exposures, even microscopically, the change will manifest itself as a system of closely-spaced fringes in the reconstructed image.

A common application is illustrated in Fig.3.3.4., which shows the double-exposure hologram of a portion of the interior of an aircraft tire. In preparation for the hologram recording, the tire is

Fig. 3.3.4.:Double-exposure holographic interferogram of an inside portion of an aircraft tire.

placed in a chamber. After the first exposure, the chamber is pumped down to a partial vacuum, The second exposure is then made with the vacuum applied. If there is a local separation between the tread and the outer plies of the tire or between the various plies, both potentially dangerous flaws, air entrapped in the separation will cause a minute local expansion, a microscopic bulging, in the vicinity of the flaw. This bulging is readily discerned in the reconstructed hologram as a series of closely spaced fringes. Commercial holographic testing systems developed for this purpose are able to test one tire every several minutes.

Similarly hologram of a flight of a bullet can be seen in double exposure hologram as shown below.

Fig3.3.5.: Double exposure hologram of a bullet in a flight

3.3.5.3. Time average holographic interferometry:

Closely related to single- and double-exposure holography is a technique known as time-average holographic interferometry. This technique allows the spatial characteristics of low-amplitude vibrations of an object-for example a drumhead, speaker cone, or metal diaphram to be mapped out with great accuracy. Figure 3.3.6.shows a vibrational mode of a guitar. The fringes represent contour lines of equal-amplitude vibrations of the guitar surface; the brightest contours occur at the nodes, or stationary points, of the vibrating surface. This is to be expected, because light reflected from these points always interferes with the reference wave at the hologram in the same way, yielding a recorded fringe pattern of high contrast and, hence, high reconstruction amplitude. For other regions of the vibrating surface, however, the corresponding interference pattern is moving and, for long exposures, produces a diffraction pattern of reduced or even zero reconstruction efficiency. The relationship between the pattern obtained and the vibrational amplitude at a particular location is not a simple one and is not derived here. It is possible, however, to use such time-average interferograms quantitatively.

Fig. 3.3.6.: Time-average holographic image of a guitar.

Similarly surface vibrating amplitudes of a can top and vibrating disk can be understood. (.Fig.3.3.7.and Fig. 3.3.8.).

Fig, 3.3.7.: Hologram of a can top vibrating in (a). the lowest resonance frequency and (b) the second resonance. The rings are loci of equal amplitude.

Fig.3.3.8.: Hologram obtained by stroboscopically illuminating a vibrating disk at the two extreme positions of vibrations

The main advantage of the holographic interferometry technique lies in the fact that the surfaces of the vibrating objects need not be made optically flat and complex vibrating objets can also be studied.

3.3.5.4. .Speckle Interferometry :-

It is also a nondestructive testing technique using laser speckle in time average speckle interferometry. The vibrating surface is illuminated with laser beam and the resulting speckle is made to interfere with a reference wave on the photographic film. As the object vibrates, the regions of the image corresponding to vibration nodes stand out as patterns of stationary, high contrast speckle. In other regions, the speckle move and became blurred due to movement of surface. On development, the result is a photographic of the object that shows clearly with microscopic precision the regions that are stationary.

Speckle interferometry can be used with great versatility to measure object vibration, translation, strain and rotation in stress analysis on models of engineering importance. Using sandwich holography confidential engineering data can be stored as suggested by Abramson *et.al*.

3.3.6. HOLOGRAPHIC OPTICAL ELEMENTS

A hologram takes an incident light wave and, by means of diffraction, generated a different wave. In this sense the hologram can be viewed as a general kind of optical element with the capability of mapping one light distribution into another. The similarity between holograms and conventional optical elements has already been noted: A sine grating, the simplest of holographic

records, changes the direction of a beam of light much as a prism does, and the sinusoidal zone plate behaves simultaneously like a positive lens and a negative lens.

The diffraction process depends upon wavelength as given by the equation

$$f_g = \sin \theta / \lambda$$

Where f_g = Spatial frequency of the grating,

λ = Wavelength of light

$$f_g = 1/d$$

Spatial frequency is the number of periods of the grating per unit distance i.e. reciprocal of grating period 'd'.

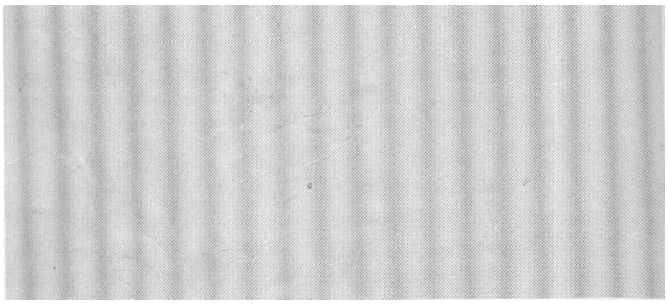


Fig. 3.3.9. :Holographic Sine grating

The sinusoidal zone plate

The sinusoidal zone plat, often referred to as a Gabor zone plate, bears a resemblance to the Fresnel zone plate. The Fresnel zone plate is either transparent or opaque, however, whereas the Gabor zone plate exhiits continuous variations in light transmittance values.

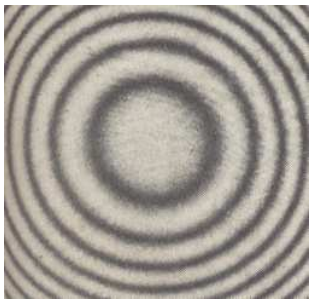


Fig. 3.3.10.: Holographic Zone plate

This similarity between holographic elements and conventional optical elements can be exploited in a number of applications. Because of the strong wavelength dependency of the diffraction process, holographic optical elements are highly dispersive; that is, they behave differently at different wavelengths and must be used with monochromatic light or several of these elements must be combined in such a way as to eliminate wavelength-dependent effects. They can be thin, lightweight, and compact; they can be easily fabricated and replicated; they can be very large in size. In addition, photographic emulsions can be coated onto arbitrary transparent surfaces for the fabrication of special application, conformal optical elements. For example, one might manufacture holographic lenses that could be rolled up in tubes for storage or shipping. One of the earliest and still most successful uses of a holographic optical element has been in optical spectrometers, where a high –resolution, holographically recorded grating replaces the considerably more expensive conventional grating. As holographic recording processes have been improved, so has the manufacture of these holographic grating, to the point where holographic spectrometers now claim a substantial fraction of the spectrometer market. Similarly, holographic lenses (off –axis sinusoidal zone plates) have been used in imaging systems.

Just as glasses of different refractive indexes have been combined in the manufacture of lenses that perform equally well over a wide range of wavelengths, so it is possible to combine multiple holographic zone plates and gratings in a system suitable for achromatic imaging over most of the visible spectrum. Although the individual component elements must be very precisely positioned for them to function properly, achromatic holographic imaging systems constructed in this manner exhibit the attractive features of small mass and reproducibility. To replicate such a system, it is only necessary to duplicate several holograms, where as with conventional systems, glass elements must be ground and polished and tested.

Holographic optical elements are often used in optical data processing systems. A particularly simple example is a coherent optical system (i.e ., employing laser light) for recognizing alphanumeric characters . Let us consider in detail a subsystem for recognizing a specific letter – say the letter K. First, a hologram is made of a prototype letter K with type style and orientation appropriate to the recognition problem, as illustrated in fig 3.3.11a. The reference wave is chosen to be a converging spherical wave. If the hologram is re-illuminated with light from the letter K, a spherical wave is produced that converges to a bright spot in the output plane of the optical system (fig 3.3.11b). There its presence can be detected with a photo detector. The light waves from other alpha numeric characters will also be diffracted by the hologram, but not into the suitably converging wave that is produced by waves from the correct letter (fig 3.3.11c), By detecting the presence or absence of a focused spot of light in the output plane of the system, one can determine whether a particular character is present in the input plane . Additional holographic optical elements can be arranged to allow an entire page of printed material to serve
as input to the system . The locations of focused spots of light in the output plane then signify the locations of the corresponding “matched”* character in the input plane.

Fig 3.3.11.: Holographic character recognition. The holographic filter, recorded as in (a), Maps waves from a "matched" character into converging spherical wave, which are detected as points of light in the output plane (b), Waves from an incorrect character are also diffracted by the filter, but not into suitably converging waves (c).

3.3.7. HOLOGRAPHIC OPTICAL MEMORIES:

The use of photo sensitive materials in large – scale memories to replace magnetic tape or disk bulk storage systems on computers has intrigued scientists and engineers for some time. Optical memories are limited ultimately in their information storage capacity by the wavelength of light used to "read" or " write" the information. Assuming an argon laser operating at 488 nm is the light source, optical memories are theoretically capable of storing in excess of 10^8 bits of information for square centimeter of photosensitive surface, or nearly 10^{13} bits in a 1-cm^3 volume. These numbers exceed the storage density limitations of magnetic systems by several orders of magnitude, and have made optical storage of digital information an area of extensive research. Much of the research conducted has centered on the difficult problem of preparing a photo sensitive material with high resolution that can be written onto, read from, and erased with the fidelity of the recording process preserved over long periods of time.

The introduction of holographic techniques to the optical data storage area has led to optical memory systems for digital data that have several very attractive features. First, As noted earlier, holograms record information redundantly. Thus one bit of information, which in a conventional photographic recording process might be recorded as a transparent spot in a dark background, is in the holographic case recorded as a sinusoidal grating. The information for that one bit is spread out over the entire hologram. With conventional photographic recordings, a peak of dirt on the film could entirely block out the transparent spot corresponding. to the bit . In the holographic case, a single speck irt has no significant effect on the recording. To recover the information recorded holographically i.e., to "read out" the stored bit pattern the holographic recording is placed in front of a positive lens and illuminated by collimated laser light. Each

sinusoid generates a diffracted light beam, which is focused by the lens to a unique spot, there to be detected by a photodiode. The second advantage of the holographic recording of the bit pattern now becomes apparent, since the exact position of the hologram in the playback setup is seen to be unimportant .The hologram can be moved either horizontally or vertically in the input plane, and the focused spots remain stationary. It is only necessary that the hologram be reasonably well illuminated. Binary information stored in sine-grating format thus exhibits simultaneously an immunity to " burst" errors (the speck of dirt) and an insensitivity to improper positioning in the readout system. These two characteristics have made holographic data storage attractive for moderately high – density storage of digital information when records are accessed frequently. Microfilm records of reports and journal articles are presently being made, for example, with cataloging information recorded in holographic form as well as in conventional form. Easy automatic readout of these data in a position – insensitive system is thus assured.

Thus, Holographic techniques are used for optical data storage with the following advantages:

1) Holograms can record information abundantly
2) Holograms are not affected by dust etc. Where as photographic plates totally miss the information.
3) Insensitive to improper positioning in a readout system.

3.3.8. Imaging through Aberrating media:

The principle of holography can also be employed for imaging objects through the media that aberrate any wave passing through them. Thus, for example, the optical components (or intervening medium) may introduced aberrations in the wave. Let us consider the case when both the objet wave and the reference wave pass through the same region of the aberrating medium as shown in Fig. 3.3.12. The photographic plate is situated just behind the aberrating medium. In the

Fig 3.3.12 : A hologram recording in the presence of an aberrating medium.

presence of the aberrating medium the waves get modified. Since both waves the reconstru ction may be effected by the arrangement shown in fig.3.3.13., have been assumed to pass through the same region of aberrating medium the effect on both is same. The fig.3.3.13a corresponds to conventional photography through the aberrating medium is unintelligible.

Fig.3.3.13.: Holographic imaging through an aberrating medium

3.3.9. Summary: -

Holographic interferometric techniques used in non-destructive testing and in stress analysis are described. The advantages in Holographic Optical elements and optical memories are mentioned.

3.3.10. Key terminology:-

Real time holographic interferometry- Double exposure holographic technique- Time average holographic interferometry- Holographic optical elements–Holographic memories.

3.3.11. Self assessment questions:-

1) Describe the application of holographic technique in stress analysis.
2) Describe the principle of holographic elements
3) Describe a holographic technique used for vibration analysis

3.3.12. Reference and Text Books:-

1. Introduction to lasers and their applications by D.C.Oshea, W.R. Callen and W.T. Rhodes, Addison-Wesley publishing Co., 1978.
2. Laser & Non linear optics by B.B.Laud, New Age International Pvt. Ltd., Publishers, 2001.
3. Contemporary optics by A.K Ghatak & K. Thyagarajan, Plenum Press, 1978.

1)

Unit IV

Lesson 1

OVER VIEW OF OPTICAL FIBER COMMUNICATIONS

Objective: To learn about the different forms of communication systems, in particular optical fiber communication system and its advantages and applications.

Structure:

4.1. Introduction:-

Ever since ancient times one of principle interests of human beings has been to device communications systems for sending massages from one place to another place. The fundamental elements of any such communications system are shown in figure 4.1.1.

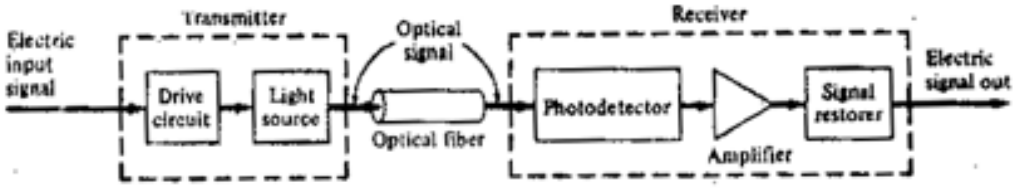


Fig. 4.1.1.: Fundamental elements of a communication system.

These elements include at one end an information source, which inputs a massage to a transmitter. The transmitter couples the massage on to a transmission channel in the form of a signal, which matches the transfer properties of channel. The channel is the medium bridging the distant between the transmitter and the receiver. This can be either a guided the transmission such as a wire or wave guide or it can be an un guided atmospheric or space channel as the signal traverses the channel it may be progressively both attenuated and distorted and increasing distance. For example electric power is lost through heat generation as an electric signal flows along a wire and optical power is attenuated through scattering and absorption by molecules in an atmospheric channel. The function of the receiver to extract the weakened and distorted signal from the

channel, amplify it and restore it to its original form before passing it on to the massage to the destination

4.1.1 FORMS OF THE COMMUNICATION SYSTEMS:-

Many form of the communications systems have appeared over the years. The principle motivation behind each new one were either to improve the transmission fidelity to increase the data rate so that more information could be sent, or to increase the transmission distance between relay stations. Prior to the 19 th century all communication systems were of a very low data rate type and basically involved only optical acoustical means such as signal lamps or horns one of the earliest known optical transmission links for example was the use of a free signal by the Greeks in the eighth century B.C for sending alarms calls for help or announcements of certain events . Only one type of signal was used its meaning being pre arranged between the sender and the receiver in the fourth century B.C the transmission distance was extended to the use of relay stations and by approximately 150 B.C these signals were encoded in relation to the alphabet so that any massage could then be sent improvements of these systems were not actively pursued because of the technology limitations for example the speed of the communication link was limited since the human eye was used as a receiver line of sight transmission paths were required and atmospheric effects such as fog and rain made the transmission path unreliable. Thus it generally termed out to be faster and more efficient to send massages by a courier over the road network .

The discovery of the telegraph by Samuel F.B. Morse in 1838 ushered in a new epoch in communication the era of electrical communications the first commercial telegraph service using wire cables was implemented in 1844 and further installations increased steadily through out the world in the following years the use of wire cables for information expounded with the installation of the first telephone exchange in new HEAVEN Connecticut in 1878 . Wire cable was the only medium electrical communication until the discovery of the long wave length electro magnetic radiation by Heinrich hertz in 1887 .The first implementation of this was the radio demonstration by Guglielmo Marconi in 1895.

In the ensuing years an increasingly larger portion of the electro magnetic spectrum was utilized for conveying information from one place to another the reason for this is that in electrical systems data is usually transferred over the communication channel by super imposing the information signal on to a sinusoidally varying electro magnetic wave which is known as the carrier at the destination information is removed from the carrier wave and processed as desired. since the amount of information that can be transmitted is directly related to the frequency range over which the carrier wave operates increasing the carrier frequency theoretically increases the available transmission band width and consequently provides a larger information capacity .Thus the trend in electrical communication system development was to employ progressively higher frequencies, which offered corresponding increases in band width and hence an increase information capacity. This activity led in turn to the birth of television radar and microwave links.

The portion of the electro magnetic spectrum that is used for electrical communications

is showed in figure 4.1.2. The transmission media used in the spectrum include millimeter and micro wave guides metallic wires and radio waves among the communication systems using these media are the familiar telephone A.M and F.M radio, television, C.B, radar and satellite links all of which have become a part of our every day lives the frequencies of these application range from about 300 Hz in the audio band to about 90 GHZ for the millimeter band.

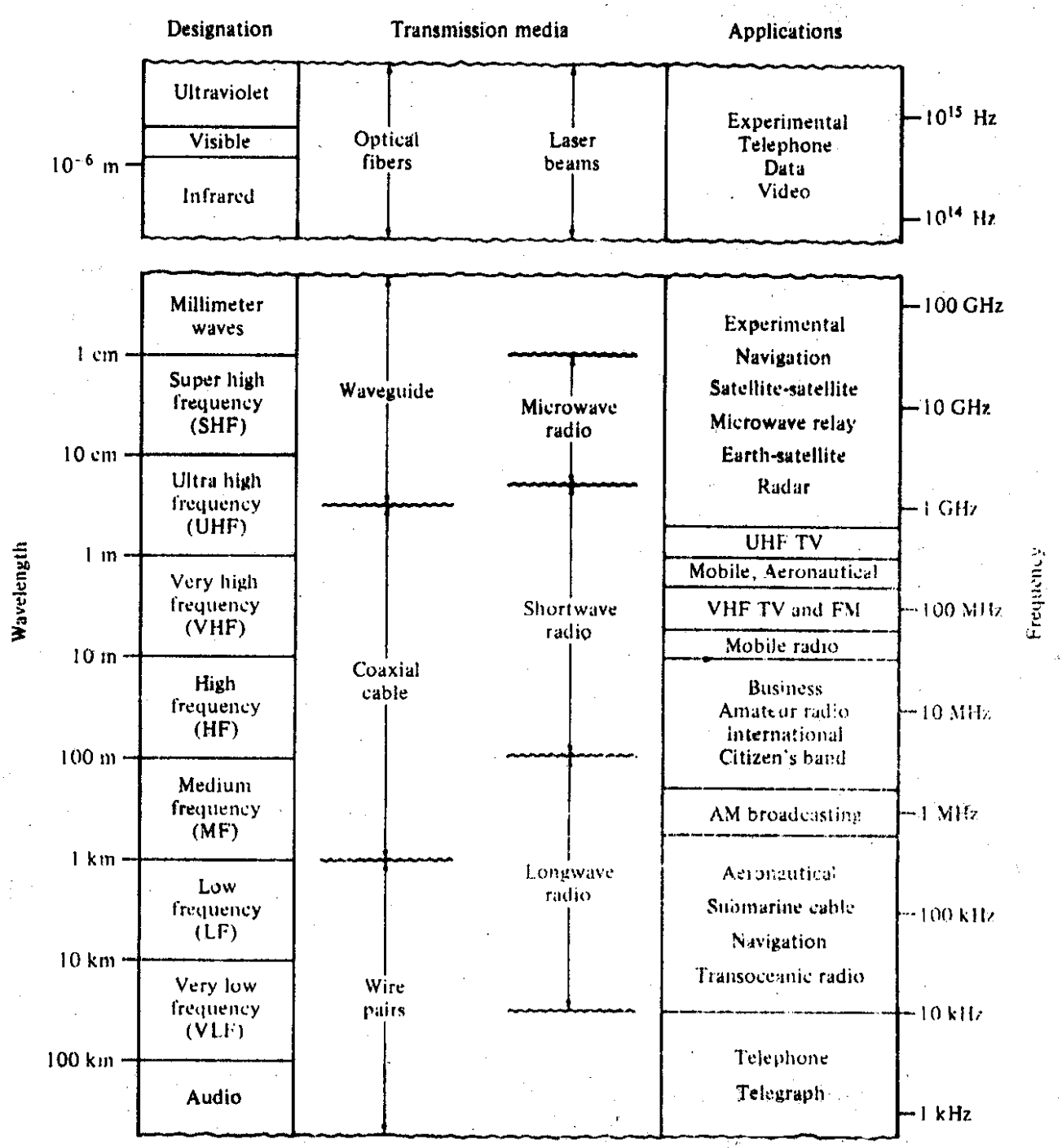


Fig 4.1.2 Examples of communication systems applications in the electromagnetic spectrum.

Another important portion of the electromagnetic spectrum encompasses the optical region shown in fig 4.1.2. in this region it is customary to specify the band of interest in terms of wave lengths instead of frequency as is used in the radio region. The optical spectrum ranges from about 50nm to approximately 100mm , with the visible spectrum being to the 400 –to-700 nm region. Similar to the radio frequency spectrum to transmission media can be used an atmospheric channel or a guided wave channel.

A great interest in communication at these optical frequencies was created in 1960 with the advent of the laser, which made available a coherent optical source. Since optical frequencies are on the order of 5×10^{14} Hz, the laser has a theoretical information capacity exceeding that of microwave systems by a factor of 10^5, which is approximately equal to 10 million TV channels.

With the potential of such wide band transmission capabilities in mind, a number of experiments using atmospheric optical channels were carried out in the early 1960s. These experiments showed the feasibility of modulating a coherent optical carrier wave at very high frequencies. However the high installation expense that would be required, the tremendous costs that would be incurred to develop all the necessary components, and the limitations imposed on the atmospheric channel by rain, fog, snow, and dust make such extremely high speed systems economically unattractive in view of present demands for communication channel capacity. Nevertheless, numerous developments of free –space optical channel systems operating at baseband frequencies are in progress for earth-to-space communications.

Concurrent with the work on atmospheric optical channels are the investigations of optical fibers, since they can provide a much more reliable and versatile optical channel than the atmosphere. Initially, the extremely large losses of more than 1000 dB/km observed in the best optical fibers made them appear impractical . This changed in 1966 when Kao and Hockman and A . Werts almost simultaneously speculated that these high losses were a result of impurities in the fiber material, and that the losses could be reduced to the point where optical waveguides would be a viable medium. This was realized in 1970 when Kapron ,Keck and Maurer of the corning glass works fabricated a fiber having a 20-dB/km attenuation .A whole new era of optical fiber communications was thus launched .

The ensuing development of optical fiber transmission systems grew from the combination of semiconductor technology, which provided the necessary light sources and photo detectors, and optical waveguide technology upon which the optical fiber is based. The result was a transmission link that had certain inherent advantages over conventional copper systems in telecommunications applications. For example, optical fiber have lower transmission losses and wider band widths as compared to copper wires. This means that with optical fiber cable systems more data can be sent over longer distances, there by decreasing the number of wires and reducing the number of repeaters needed over these distances .In addition, the low weight and the small hair-sized dimensions of fibers offer a distinct advantage over heavy, bulky wire cables in crowed underground city ducts. This is also of importance in aircraft where small light –weight cables are

advantageous and intactical military applications where large amounts of cable must be unreeled and retrieved rapidly. An especially important feature of optical fibers relates to their dielectric nature. This provides optical waveguides with immunity to electromagnetic interference, such as inductive pickup from signal carrying wires and lightning, and freedom from electromagnetic pulse (EMP) effects, the latter being of particular interest to military applications. Furthermore there is no need to worry about ground loops, fiber –to-fiber cross talk is very low, and a high degree of data security is afforded since the optical signal is well confined within the waveguide (with any emanations being absorbed by an opaque jacketing around the fiber). Of additional importance is the advantage that silica is the principal material of which optical fibers are made . This material is abundant and inexpensive since the main source of silica is sand.

Early recognition of all these advantages in the early 1970s created a flurry of activity in all areas related to fiber optic transmission systems, this resulted in significant technological advances in optical sources, fibers, photo detectors, and fiber cable connectors. By 1980 this activity had led to the development and worldwide installation of practical and economically feasible optical fiber communication systems that carry live telephone, cable TV, and other types of telecommunications traffic. These installations all operate as base band systems in which the data are sent by simply turning the transmitter on and off. Despite their apparent simplicity such systems have already offered very good solutions to some vexing problems in conventional applications.

4.1.2 ELEMENTS OF AN OPTICAL FIBER TRANSMISSION LINK

An optical fiber transmission link comprises the elements shown in fig. 4.1.3. The key sections are a transmitter consisting of a light source and its associated drive circuitry, a cable offering mechanical and environmental protection to the optical fibers contained inside, and a receiver consisting of a photo detector plus amplification and signal – restoring circuitry. The cabled optical fiber is one of the most important elements in an optical fiber link. In addition to protecting the glass fibers during installation and service, the cable may contain copper wires for powering repeaters, which are needed for periodically amplifying and reshaping the signal when

the link spans long distances. The cable generally contains several cylindrical hair –thin glass fibers, each of which is an independent communication channel.

Fig. 4.1.3. Major elements of optical fiber transmission link.

Analogous to copper cables, the installation of optical fiber cables can be either aerial, in ducts, undersea, or buried directly in the ground. As a result of installation and / or manufacturing limitations, individual cable lengths will range from several hundred meters to kilometers for long – distance applications. The actual length of a single cable section is determined by practical consideration s such as reel size and cable weight. The shorter lengths tend to be used when the cables are pulled through ducts. Longer table lengths are used in aerial or direct burial applications. The complete long distance transmission line is formed by splicing together with the individual cable sections (Fig. 4.1.4.).

Fig. 4.1.4. Optical fiber cables can be installed on poles, in ducts, and undersea, or they canbe buried directly in the ground.

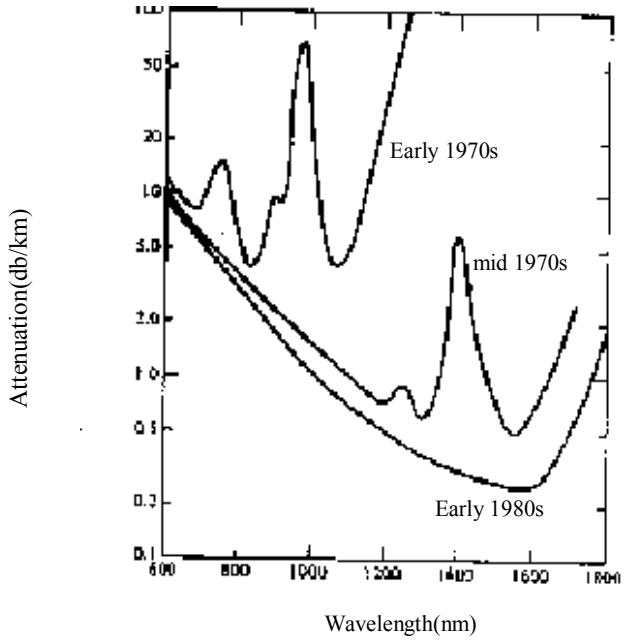


Fig 4.1.5: Optical fiber attenuation as a function of wavelength. Early fiber links were operated exclusively in 800-900nm range where there was a local attenuation minimum. Achievement of lower attenuations created an interest in longer wavelength operation.

One of the principle optical fiber characteristics is its attenuation as a function of wavelength as shown in fig 4.1.5. Early technology made exclusive use of the 800 to 900 nm wavelength band, since in this region the fibers made at that time exhibited a local minimum in the attenuation curve and optical sources and photo detectors operating at these wavelengths were available by reducing the concentrations of hydroxyl ions and metallic ion impurities in the fiber metal, fiber manufacturers were eventually able to fabricate optical waveguides with very low losses in the 1100 to 1600 nm region. This spectral bandwidth is usually referred to as the long wavelength region increased interest thus developed at the 1300 nm wavelength since this is the region of minimum signal distortion in pure silica fibers. Bell laboratories for example begin the first U.S. Field trail of a 1300 nm led based telephone system in Sacramento California on June 26, 1980, and others U.S, European and Japanese telecommunications companies are actively experimenting with advanced long wavelength trail installation.

Research has also commenced on new types of materials for use in the 3 to 5 μm wavelength band the initial interest centered on the polycrystalline metals halides such as zinc chloride (zncl2), thallium, bromide (TlBr) and thallium bromoiodide (known as K.R.S –5). Theoretical predictions expect these fibers we have attenuations of less than 0.01 dB/km . The successful fabrication of fibers having these of losses can have a profound effect on lone distance communication.

Once the cable installed a light source which is dimensionally compatible with the fiber core is used to launch optical power in to fiber. Semi conductor light – emitting diodes (LEDs) and laser diodes are suitable transmitter sources for this purpose since their light output can be modulated rapidly by simply varying the bias current. The electric input signals to the transmitter can be either of an analog or of a digital form. The transmitter circuitry converts these electric signals to an optical signal by varying the current flow through the light source. An optical source is a square –law device , which means that a linear variation in drive current results in a corresponding linear change in the optical output power . In the 800 to 900 nm region the light sources are generally alloys of GaAlAs. At the longer wavelengths (1100 to 1600 nm) , an In GaAsP alloy is the principal optical source material.

After an optical signal has been launched in to the fiber, it will become progressively attenuated and distorted with increase in distance because of scattering, absorption, and dispersion mechanisms in the wave-guide. At the receiver the attenuated and distorted modulated optical power emerging from the fiber end will be detected by a photo diode. Analogous to the light source, the photo detector is also a square-law devise since converts the received optical power directly into an electric current output (photo current). Semi-conductor pin and avalanche photodiodes (APDs) are the two principal photo detectors used in a fiber optic link. Both devise types exhibit high efficiency and response speed. For applications in which a low power optical signal is received, an avalanche photodiode is normally used since it has a greater sensitivity owing to an inherent internal gain mechanism (avalanche effect). Silicon photo-detectors are used in the 800 to 900nm region. A variety of optical detectors are potentially available at the longer wavelengths. The prime material candidate in the 1100 to 1600nm region is an InGaAs alloy. We address these photo-detectors in chapter6.

The design of the receiver is inherently more complex than that of the transmitter since it has to both amplify and reshape the degraded signal received by the photo detector. The principal figure of merit for a receiver is the minimum optical power necessary at the desired data rate to attain either a given error probability for digital systems or a specified signal-to-noise ratio for an analog system. The ability of a receiver to achieve a certain performance level depends on the photo detector type, the effects of noise in the system, and the characteristics of the successive amplification stages in the receiver.
wavelength region increased interest thus developed at the 1300 nm wavelength since this is the region of minimum signal distortion in pure silica fibers. Bell laboratories for example begin the first U.S. Field trail of a 1300 nm led based telephone system in Sacramento California on June 26, 1980, and others U.S, European and Japanese telecommunications companies are actively experimenting with advanced long wavelength trail installation.

Research has also commenced on new types of materials for use in the 3 to 5 μm wavelength band the initial interest centered on the polycrystalline metals halides such as zinc chloride ($ZnCl_2$), thallium, bromide (TlBr) and thallium bromoiodide (known as K.R.S –5). Theoretical predictions expect these fibers we have attenuations of less than 0.01 dB/km . The successful fabrication of fibers having these of losses can have a profound effect on lone distance communication.

Once the cable installed a light source which is dimensionally compatible with the fiber core is used to launch optical power in to fiber. Semi conductor light – emitting diodes (LEDs) and laser diodes are suitable transmitter sources for this purpose since their light output can be modulated rapidly by simply varying the bias current. The electric input signals to the transmitter can be either of an analog or of a digital form. The transmitter circuitry converts these electric signals to an optical signal by varying the current flow through the light source. An optical source is a square –law device , which means that a linear variation in drive current results in a corresponding linear change in the optical output power . In the 800 to 900 nm region the light sources are generally alloys of GaAlAs. At the longer wavelengths (1100 to 1600 nm) , an In GaAsP alloy is the principal optical source material.

After an optical signal has been launched in to the fiber, it will become progressively attenuated and distorted with increase in distance because of scattering, absorption, and dispersion mechanisms in the wave-guide. At the receiver the attenuated and distorted modulated optical power emerging from the fiber end will be detected by a photo diode. Analogous to the light source, the photo detector is also a square-law devise since converts the received optical power directly into an electric current output (photo current). Semi-conductor pin and avalanche photodiodes (APDs) are the two principal photo detectors used in a fiber optic link. Both devise types exhibit high efficiency and response speed. For applications in which a low power optical signal is received, an avalanche photodiode is normally used since it has a greater sensitivity owing to an inherent internal gain mechanism (avalanche effect). Silicon photo-detectors are used in the 800 to 900nm region. A variety of optical detectors are potentially available at the longer wavelengths. The prime material candidate in the 1100 to 1600nm region is an InGaAs alloy. We address these photo-detectors in chapter6.

The design of the receiver is inherently more complex than that of the transmitter since it has to both amplify and reshape the degraded signal received by the photo detector. The principal figure of merit for a receiver is the minimum optical power necessary at the desired data rate to attain either a given error probability for digital systems or a specified signal-to-noise ratio for an analog system. As we shall in chapter 7, the ability of a receiver to achieve a certain performance level depends on the photo detector type, the effects of noise in the system, and the characteristics of the successive amplification stages in the receiver. Services to the home or office such as pay or educational TV library and information retrieval video word processing and electronic mail, banking and shopping. This has led to a great interest in applying fiber optic technology in the subscriber loop form the local switching point to the home or office. In contrast with twisted wire pairs, optical fibers offer low attenuation large information capacity immunity loop to lighting and electro magnetic interference freedom from cross talk between fibers and a virtual independent from use of these features fiber optic subscriber loop plants could thus we designed to meet both present and feature broad band multi service demands.

The wide spread interest in such broad band multi services is evidenced by a number of field trails that are under way are planned in various parts of the world. Some of the major activities include the Highly Interactive Optical Visual Information System(Hi-OVIS) field trail in Japan,. the Yorkville(Bell Canada) and Elie rural systems in Canada, the ambitious Biarritz project in France sponsored by the French governments, and the multiservice system for the business applications developed by GTE .

4.1.4. APPLICATIONS OF OPTICAL FIBERS:

Optical fiber systems in sociological evolution;

The physical and mental needs of the people in an information society are both materialistic and service-oriented. The materialistic aspects refer to the basic production of food and the conversion of raw materials into general merchandise. The service aspects require a new industry to improve the efficiency of the distribution of material goods, balance of resource allocation, and enrich life with new pursuits. At the basis of this service industry are optical fiber systems forming the communications links.

ENERGY RESOURCES

Until the control of thermal nuclear fusion can be perfected in the twenty-first century, available basic energy resources are limited. Many of the readily available resources such as oil and gas are rapidly being depleted. As a result, energy cost is increasing rapidly. This calls for a careful examination of the mode of energy usage, with a view to fulfill the increasing demand more efficiently and safely. It is a process in which the consumption of oil, natural gas, coal, solar energy, and other resources for domestic and industrial activities is balanced for cost and availability. It is tine to trade heavy energy-expending activities, such as the making of articles for rapid consumption, for the manufacture of more durable goods, it is time to explore alternative primary energy resources while taking precaution against undesirable environmental impacts.

Here the fiber system plays a minor but important role. The characteristic of fiber- of immunity to electromagnetic interference-allows fiber systems to be used with advantages in electrical generation and transformer stations to improve control and communications functions.

TRANSPORTATION

The physical movement of people and goods is an essential part of life, and different transportation requirements are met by different means. For mass transit and bulk goods movement, there are buses, trains, ships, and large trucks. For individual journeys, there are automobiles and bicycles, for rapid transit, there are airplanes. Together, these methods of transportation keep the society functioning.

The volume of traffic has been expanding rapidly as population and trade increase. The total expenditure of energy for transportation has reached a level, which is sufficiently large, compared with the level of energy reserve, to warrant attention. More efficient use of the transportation system is important.

The role of fiber system in transportation is indirect. In mass transit systems the operation can be made much more efficient with the provision of reliable means of control. This usually involves locating the position and speed of the vehicles on the railway. Along the high-speed electrified railway line, such a need is most obvious. Fiber control system can be installed advantage along the electrified track where copper wire systems would encounter serious electro magnetic interference problems.

COMMMUNICATIONS

The introduction of optical fiber systems will revolutionize the communications network. The low-transmission loss and the large-bandwidth capability of the fiber systems allow signals to be transmitted for establishing communications contacts over large distances with few or no provisions of intermediate amplification. Also, more information can be transmitted over a shorter time than with the use of alternative systems. This means that the optical fiber communications network can provide more services a lower cost.

The telephone network, based on copper wires as the transmission lines, has provided us with a very valuable communications network. The instant audio contact through telephone has increased business efficiency, reduced the need to travel and provided many services which otherwise would not be available.

The communications network, based on optical fiber as the transmission lines, has several orders of magnitude more information-carrying capacity per unit time that does the telephone network and can provide a host of services, including video services, which require a bandwidth much greater than that used in the telephone service. Such a communications network allows video, audio and data transmission and in association with computers, it literally allows our visual

senses and certain of our brain functions to be extended. The power of such a network would create a completely new sociological environment-one, which befits the information age.

The evolution toward such a communications network is likely to be relatively slow and deliberate. The existing network represents a huge monetary investment and cannot be written Off by overnight. Neither can this investment of an even larger sum of money to create the new network be done instantaneously. There will be a gradual evolution.

The advantages of a fiber system can be exploited most readily in the existing network in a number of important areas. In metropolitan areas copper cables are installed mainly in underground conduits, providing connection between busy exchange offices. Fiber cables with equal traffic-carrying capacity are much smaller in size and can provide interchange office connection without intermediate repeaters in most cases. The lifetime system cost-which includes installation, maintenance, and hardware costs-favours fiber systems, especially where traffic density and growth are high. Intercity routes with high traffic densities can also be served better and more economically by fiber systems. There are the areas where fiber systems are being introduced into the existing network. The network between a satellite ground receiving station and a distribution hub provides another opportunity for early fiber system entry. Trunking of television signals over distances of several kilometers and upward can be most attractive on fiber. The increased usage of fiber systems will result in reduction of their cost, when the economy of scale takes effect. Fiber systems will then be preferred on the basis of cost alone, even in areas where their advantages of large bandwidth and low loss are not needed. Early introduction of fiber cable into the subscriber distribution network will then be possible, thus lying the foundation of the wide-bandwidth information network of the future.

Information services for the home and business premises already being used include document transmission (facsimile); telex; 1200- to 9600-band data for computerized banking; airline, hotel, and other reservations; dial –a-message on telephones for weather, local events, and other information; and stored message from special customer services. Some prospective services being tested include package switching networks for improved data services and computer access, video text for stored information such as restaurant guides, train timetables, programmed teaching, or, for more specific customers, stock market information, inventory control, and soon, and for improved security and energy management, remote alarm and meter monitors. Services envisaged for the future usually involve the use of increased bandwidth to provide faster information transfer and the use of live video information in an interactive manner. While most of services can be provided without the need of a broadband transmission medium, the increased use of information would result in broadband transmission in an increasingly larger area of the communications network. Eventually with the widespread use of broadband services, the capability of broadband distribution to individual subscribers would be required.

Fiber cost is expected to decrease substantially to a level where its use in the distribution network becomes non-cost-prohibitive .Its use is excepted to be very widespread, and it will lay the foundation for the present network to evolve into the future broadband network.

HEALTH SERVISES

Health care delivery requires individual consultation by physicians and massive specialist hospitals with sophisticated equipment. Communications both within large hospitals and to and from these establishments is vital for operational efficiency and effectiveness .If a physician can all for the relevant record of a patient, vital statistics, rapid analyses, and a glossary of information aids at the hospital or at the home of a patient, the health care delivery would be vastly more effective. Fiber systems can provide the communications links capable of handling visual instruments and fast computer –controlled data equipment without suffering from electromagnetic interferences which are often associated with high-powered hospital equipment, such as x-ray machines .The visual observation aids are particularly important. Remote monitoring of patients and recalling of specialized symptomatic information and surgical procedures in visual from can bring the massive resources of a specialist hospital to the assistance of a physician in a remote village.

In a role rather different from that of an information transmission line, an optical fiber as a conductor of light and visual images is exploited in medical instruments for illumination and observation of inaccessible areas. As transmission along the fibers is better controlled, fibers capable of transmitting images in real-time holographic mode can be envisaged. Single-fiber systems can be designed to deliver a precise amount of optical power for pathological analysis and as a surgical tool or as new means of curing certain diseases such as the control and destruction of tumours.

COMMUNITY SERVICES

The wired –city concept has been proposed in which a community is intimately wired together in such a manner that a host of community-related facilities such as local news, shopping guide, neighborhood festivities, community library, rescue squad, and a computer bank are available to the community as a whole.

Experimental system has shown that the value of a wired city is associated with the perception of the people. It is a sociological experiment with far-reaching implications, but the results so far are too influenced by the main social trends to be really meaningful. Even in this context, wired-city development can be seen to promote a new community spirit and a different way of utilizing services. Fiber systems can and will be the leading contender as the principal transmission media for wired cities. The provision of integrated services over a wide-bandwidth, low-loss transmission system is a desirable solution.

MILITARY DEFENSE

The special features of an optical fiber are small size, lightweight, strength, flexibility, wide temperature range, and interference freedom, in addition to wide bandwidth and loss. These features are key to improving the strategic and tactical capabilities of the military forces. More powerful communications networks can be created and installed under various conditions. In the strategic base communications applications, the more compact cable enables the cables to be

easily transported and permits a variety of permanent and mobile configurations to be implemented easily. The remote connection to radar sites from the signal processing station can be rapidly deployed and the low loss allow longer spacing between the radar site and the station, thus allowing a greater safety for the operating personnel. The electromagnetic interference freedom characteristics can be used with great advantage on ships, airplanes, and armored vehicles, where many data are being processed under electrically noisy environment the interference freedom also allows the system to preserve a high degree of privacy. This is utilized in systems where sensitive data are to be transmitted. The high strength and the flexibility allow fiber systems to be envisaged for tactical applications. Wire guided weapons can be made to cover a longer range and with more precise guidance or even with visual target search capabilities. Fibers can also improve the capabilities of towed surveillance vehicles by improving the information and mechanical performance of the towing cable. The propagation characteristics of the fiber can be utilized to indicate strain and temperature change. This can be considered for sensor applications with improved sensitivity. Acoustic and magnetic field sensors and a fiber gyroscope have been identified as possibilities.

BUSINESS DEVELOPMENT

Akin to the wired-city concept is the office-of-the-future concept, where business functions of information retrieval, distribution, dissemination, and analysis are performed by equipment available within the office of the future. A general manager can call for financial and operational data with the touch of a button or a simple voice command. A document can be typed into a word processing machine to allow for easy editing. Subsequent production of personalizing versions can be automatically routed internally thus enabling staff in other parts of the office to access the document on their terminal equipment in a display or hard-copy form and externally, for remote distribution. Inventory control, numerical control of machines and payroll preparation are just a few of the many automated services which an office of the future may have audiovisual conferencing and interactive graphics are other possible features. The technical realization of such a system is not too different from the wired city, but since the system may be confined to a single building, and since the community involved has a more homogeneous and identifiable requirement, the system requirements can be more readily met.

ENTERTAINMENT

The pursuit of happiness often starts with a satisfying job, followed by spending the affordable surplus earnings on entertainment. We are lured to a world of service industry aimed at entertainment. The spectator sports, the stage, the movies, the radio and television (TV) programs, as well as toys and books, are just a handful of more pertinent examples. Perhaps TV has emerged as the most influential of all entertainment media. Backed by revenue from advertising, television broadcasters have captured the attention of millions of people each day and entertain them with a great variety of visual programs. It has dramatically altered the proportion of active participants in entertainment to passive spectators.

Recently, cable television [also referred to as community antenna television (CATV)] through the use of geo-synchronous communication satellites, is changing

the TV industry. Cable TV provides many channels in each house, giving viewers a greater degree of freedom of choice of programs. The commercialization of video tape recorders and videodisks is another entry into the TV industry that is destined to reshape the mode of operation further.

The entertainment market many service vendors who compete for the surplus earnings of the consumers by persuading them to become their customers, promising them entertainment which is worth much more than the payment. For example, Cable TV charges a flat monthly fee for the right to have more programs to choose from, While pay TV charges an extra fee for the right to see mainly movies with no commercials. The charge is low in comparison to the to the cost of going to a movie at a Cinema. The video disk merchants are appealing to the consumer's pride of ownership and individual choice, while video recorder merchants are appealing to the consumer's desire to view scheduled programs at the most convenient time. In the meantime, the cinema, the stage and books continue to flourish but cater to a more select group.

The possibility of providing broadband information delivery to each home with an optical fiber network and the further possibility of delivering video services via that network on a switched basis are important in shaping the future of the entertainment world. The switched TV network can provide programs of wide audience appeal to cater to the majority while serving programs of limited appeal to more selected audiences by suitably arranging are venue base compatible with a narrow business basis. Such a scheme resembles a video library and provides the ultimate in individual choice.

EDUCATION

With the technology available, the education process could be made more effective. For instance, the availability of pocket tape recorders at a certain university prompted some students to skip classes and ask friends to place their recorders at the class to record the lecture. the continued for some time, and the popularity of such a practice grew until one day a few students walked into the classroom and found it full of recorders but with the professor absent. His recorder was there with the "play" button down. This story has more than one moral. One of the morals prompts the question, "Is a prerecorded lecture better or worse?"

In fact, with computers and videotape, a prepared teaching tape instead of a live lecture together with the use of computer –aided instruction can be most effective. The role of the professor then becomes that of the mentor who leads live sessions of debates and criticisms and opens the minds of the experienced.

Already educational TV and program learning on computers are being used in schools and universities. The successful open university in the United Kingdom (at Waltin, Buckinghamshire, England) demonstrated convincingly the power of TV as an educational tool. Universities with large student bodies in several campuses found closed - circuit television (CCTV)to be an effective means of marking courses available. In fact, prepared lecture series are appearing as educational commodities. In due course authoritative texts should be made accessible as references at information banks.

The provision of broadband communications systems in educational institutions is a definite necessity. The role of fiber optics for this application is very clear.

4.1.5. Summary: Development of communication systems, applications in the electromagnetic spectrum, advantages of optical fibers due to small size, large bandwidth, light weight, less repeater stations between long distance communications, elimination of cross-talk, applications of optical fiber communications in different field have been given.

4.1.6. Keywords: Optical fiber, communications, electromagnetic spectrum, bandwidth, applications of optical fibers

4.1.7. Self-assessment questions

1. Give an account of communication systems applications in the electromagnetic spectrum
2. What are the advantages of fiber optic communications
3. Discuss the applications of optical fibers.

4.1.8. Text and Reference Books

1. "Optical fiber communications" by G. Keiser, McGraw-Hill International Edition, 2000, Third edition, and also see first edition.
2. "Optical fiber systems: Technology, design, and applications" by Charles K. Kao, McGraw -Hill, 1986.

Unit IV

Lesson 2

FUNDAMENTALS OF LIGHT AND OPTICAL FIBERS

Objective: To explanation of nature of light, optical fiber structures, mode theory of waveguides.

Structure:

4.2.1. THE NATURE OF LIGHT

The concepts concerning the nature of light has undergone several variations during the history of physics. Until the early seventeenth century it was generally believed that light consisted of a stream of minute particles that were emitted by luminous sources. These particles were pictured as traveling in straight lines, and it was assumed that they could penetrate transparent materials but were reflected from opaque ones. This theory adequately described certain large-scale optical effects such as reflection and refraction, but failed to explain finer-scale phenomena such as interference and diffraction.

The correct explanation of diffraction was given by Fresnel in 1815. Fresnel showed that the approximately rectilinear propagation character of light could be interpreted on the assumption that light is a wave motion, and that the diffraction fringes could thus be accounted for in detail. Later the work of Maxwell in 1864 theorized that light waves must be electromagnetic in nature. Furthermore observation of polarization effects indicated that the light waves are transverse (that is, the wave motion is perpendicular to the direction in which the wave travels). In this wave or physical optics viewpoint the electromagnetic waves radiated by a small optical source can be represented by a train of spherical wave fronts with the source at the center as shown in fig. 4.2.1. A wave front is defined as the locus of all points in the wave train, which have the same phase. Generally one draws wave fronts passing either through the maxima or the minima of the

wave such as the peak or trough of a sine wave for example. Thus the wave fronts are separated by one wavelength.

When the wavelength of the light is much smaller than the object which it encounters the wave fronts appear as straight lines to this object or opening .in this case the light wave can be represented as a plane wave and its direction of travel can be indicated by a light ray which is drawn perpendicular to the phase front . Thus large scale optical effects such as reflection and refraction can be analyzed by the simple geometrical process of ray tracing. This view of optics is referred to as ray or geometrical optics. The concept of light rays is very useful because the rays show the direction of energy flow in the light beam.

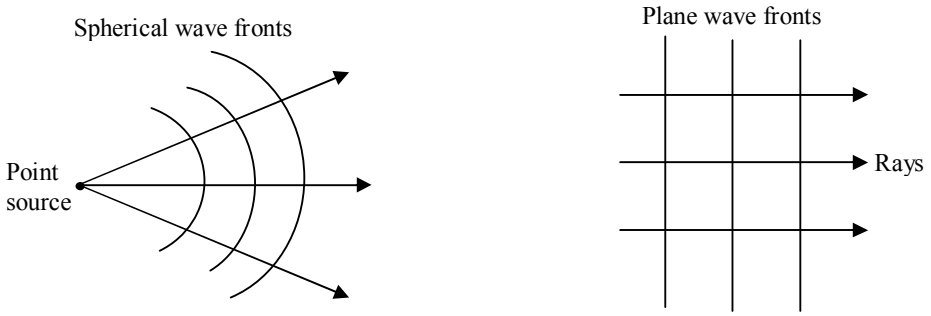


Fig. 4.2.1. Representation of spherical and plane wave fronts and their associated rays

A train of plane or linearly polarized waves traveling in a direction **k** can be represented in general form.

$$A(x,t) = e_1 A_0 \exp[j(\omega t - k.x)] \ldots\ldots\ldots (4.2.1)$$

with $\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ representing a general position vector and $\mathbf{k} = \mathbf{k}_x\mathbf{e}_x + \mathbf{k}_y\mathbf{e}_y + \mathbf{k}_z\mathbf{e}_z$ representing the wave propagation vector.

Here A_o the maximum amplitude of the wave vector, $\omega = 2\pi\nu$, where ν is the frequency of the light, the magnitude of the wave vector **k** is $k = 2\pi/\lambda$, which is known as the wave propagation constant with λ being the wavelength of the light and **e**, is a unit vector lying parallel to an axis designated by j.

Note that the components of the actual electromagnetic field represented by Eq(4.2.1) are obtained by taking the real part of this equation.

For example and if $\mathbf{k} = k\mathbf{e}_z$ and if **A** denotes the electric field **E** with the coordinate axes chosen such that $\mathbf{e}_y = \mathbf{e}_x$ then the real measurable electric field varies harmonically in the x direction and is given by

$$E_x(z,t) = \mathrm{Re}(E) = e_x E_{ox} \cos(\omega t - kz) \ldots\ldots (4.2.2)$$

Which represents a plane wave traveling in the z direction. The reason for using the exponential form is more easily handled mathematically than equivalent exponential form is that it is more easily expression given in terms of sine and cosine. In addition the rationale for using harmonic function is that any waveform can be expressed in terms of sinusoidal waves using Fourier technique

The electric and magnetic field distribution in a train of plane electromagnetic waves at a given instant in time are shown in fig 4.2.2. The waves are moving in the direction indicated by the vector **k**. Based on Maxwell's equations it is easily shown that **E** and **H** both perpendicular to the direction of propagation. Such a wave is called a transverse wave. Furthermore **E** and **H** are mutually perpendicular so that **E**, **H**, and **k** form a set of orthogonal vectors.

The plane wave example given by eq4.2-2 has its electric field vector always pointing in the $\mathbf{e}_x$ direction. Such a wave is linearly polarized with polarization vector $\mathbf{e}_x$. A general state of polarization is described by considering another linearly polarized wave, which is independent of the first wave and orthogonal to it.

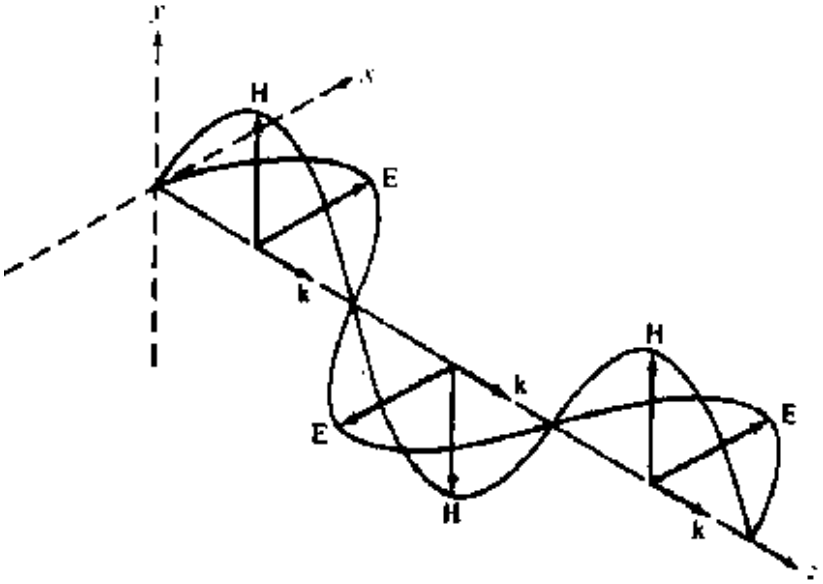


fig 4.2.2: Field distributions in a train of plane electromagnetic waves at a given instant in time

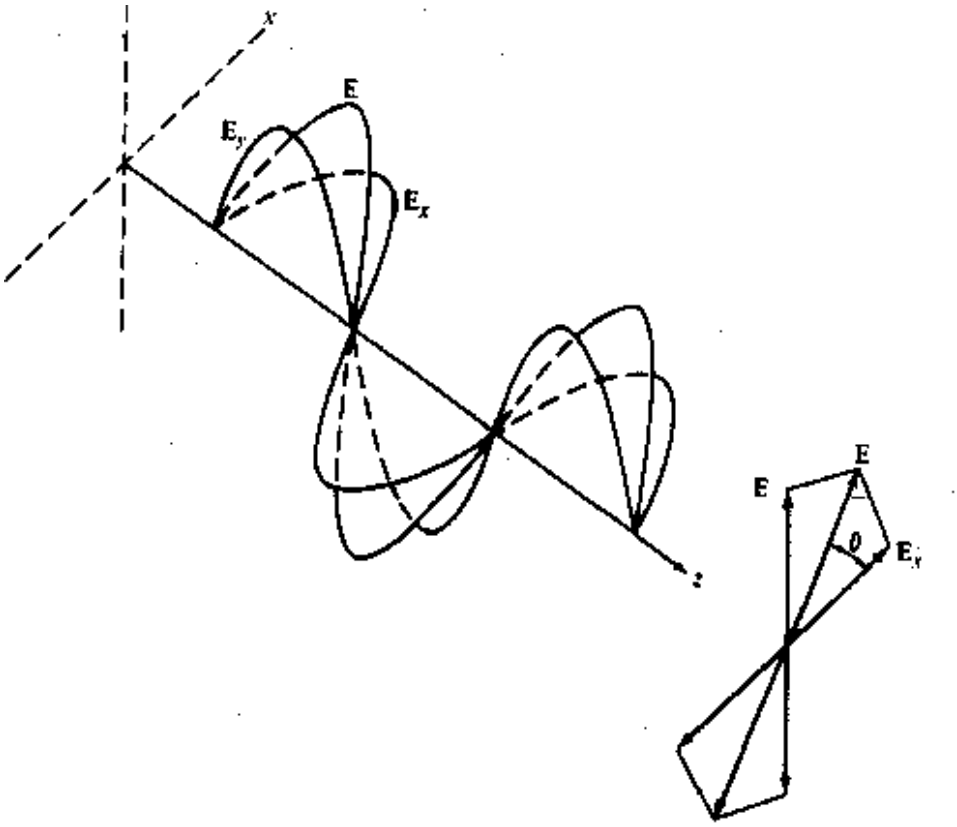


fig4. 2.3: Addition of two linearly polarized waves having zero relative phase between them.

Let this wave be $E_y(z,t) = e_y E_{oy} \cos(\omega t - kz + \delta)$ .....(4.2.3)

Where δ is the relative phase difference between the waves. The resultant wave is then simply
$E(z,t) = E_x(z,t) + E_y(z,t)$.....(4.2.4)

If δ is zero or an integer multiple of 2π, the wave are in phase. Equation 4.2.4 is then also a linearly polarized wave with a polarization vector making an angle $\theta = \arctan \frac{E_o}{E_{ox}}$

with respect to $\mathbf{e}_x$ and having a magnitude
$E = (E_{ox}{}^2 + E_{oy}{}^2)^{1/2}$

This case is shown schematically in fig 4.2.3. Conversely, just as any two orthogonal plane waves can be combined in to a linearly polarized wave an arbitrary linearly polarized wave can be resolved in to two independent orthogonal plane waves that are in phase.

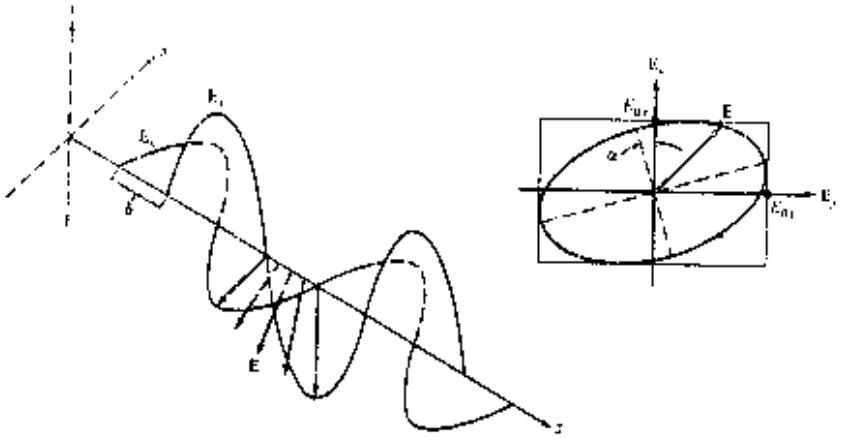


Fig 4.2.4: Elliptically polarized light resulting from the addition of two linearly polarized waves of unequal amplitude having a non-zero phase difference between them.

For general values of δ the wave given by eq 4.2.4 is elliptically polarized. The resultant field vector **E** will both rotate and change its magnitude as a function of the angular frequency was shown in fig 4.2.4 the endpoint of **E** will trace out an ellipse at a given point in space with the axis by angle α given by

$$\tan 2\alpha = \frac{2E_{ox}E_{oy}\cos\delta}{E_{ox}^{\ 2} - E_{oy}^{\ 2}} \text{ .....(4.2.5)}$$

where $\mathbf{E}_{ox} = \mathbf{E}_{oy}$ the resultant wave is circularly polarized.

The wave theory of light adequately accounts for all phenomena involving the transmissions of light. However in dealing with the interaction of light and matter such as occurs in dispersion and in the emission and absorption of light neither the particle theory nor the wave theory of light is appropriate. Instead we must turn to quantum theory, which indicates that optical radiation has particle as well as wave properties. The particle nature arises from the observation that light energy is always emitted or absorbed in discrete in units called quanta or photons

In all experiments used to show the existence of photons the photons the energy is found to depend only on the frequency ν. This frequency must be measured by observing a wave property of light

The relationship between the energy E and the frequency ν of a photon is given by

$$E = h\upsilon \text{ .....(4.2.6)}$$

where h = 6.625× 10^{-34} J-s is Planck's constant. When light is incident is on an atom a photon can transfer its energy to an electron within this atom thereby exciting it to a higher energy level. In this process either all or none of the photon energy is the imparted to the electron . The energy absorbed by the electron must be exactly equal to that required to excite the electron to a higher energy level conversely an electron in an exited state can drop to a lower state separated from state it by an energy hν by emitting a photon of exactly this energy.

4.2.2. Basic Optical Laws and definitions

The concepts of reflection and refraction can be interpreted most easily by considering the behavior of light rays associated with plane waves traveling in a dielectric material. When a light ray associated with encountered a boundary separating two different media part of the ray encounters two different media part of the ray is reflected back into the first medium and the reminder is bent as it enters the second material. This is shown in fig 4.2.5. $n_2 < n_1$ where the bending or refraction of the light ray at the interface is a result of the difference in the speed of light in two materials having difference in the speed of light in interface indices.

The frequency of the wave is $\nu = c / \lambda$ ……. (4.2.7)

The relationship at the interface is known as Snell's law and is given by

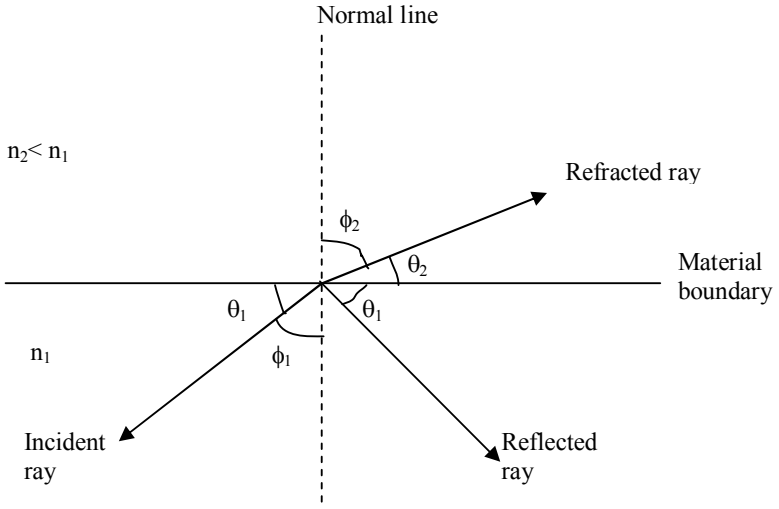
$$n_1 \sin\phi_1 = n_2 \sin\phi_2 \text{ ……(4.2.8)}$$

or equivalently as

$$n_1 \cos\theta_1 = n_2 \cos\theta_2 \text{ ……..(4.2.9)}$$

Where the angles are defined in Fig.4.2.5.

According to the law of reflection the angle θ, at which the incident ray strikes the interface is exactly equal to the angle the reflected ray makes with the same interface. In addition, the incident ray, the normal to the interface, and the reflected ray all lie in the same plane, which is perpendicular to the interface plane between the two materials. When light traveling in a certain medium is reflected off an optically denser material (one with a higher refractive index), the process is referred to as external reflection. Conversely, the reflection of light off of less optically



` Fig 4.2.5. Refraction and reflection of a light ray at a material boundary

dense material (such as light traveling in glass being reflected at a glass-to-air interface) is called internal reflection.

As the angle of incidence θ_1 in an optically denser material (higher refractive index) becomes smaller, the refracted angle θ_2 approaches zero. Beyond this point no refraction is possible and the light rays become totally internally reflected. The conditions required for total internal reflection can be determined by using Snell's law [Eq. (4.2.9)]. Consider Fig 4.2.6, which shows a glass surface in air. A light ray gets bent toward the glass surface as it leaves the glass in accordance with Snell's law. If the angle of incidence θ_1 is decreased, a point will eventually be reached where the light ray in air is parallel to the glass surface. This point is known as the critical angle of incidence. When the incident angle θ_1 is less than the critical angle, the condition for total internal reflection is satisfied; that is, the light is totally reflected back into the glass with no light escaping from the glass surface. (This is an idealized situation. In practice there is always some tunneling of optical energy through the interface. This can be explained in terms of the electromagnetic wave theory of light which is presented in Sec.4.2.4.)

As an example consider the glass-air interface shown in Fig4.2.6. When the light ray in air is parallel to the glass surface then $\theta_2=0$ so that $\cos\theta_2 = 1$. The critical angle in the glass is thus

$$\theta_c = \arccos\frac{n_2}{n_1} \quad \text{.....(4.2.10)}$$

Using $n_1=1.50$ for glass and $n_2=1.00$ for air, is about 48^0. Any light in the glass incident on the interface at an angle θ_1 less than 48^0 is totally reflected back into the glass.

In addition, when light is totally internally reflected, a phase change δ occurs in the reflected wave. This phase change depends on the angle $\theta_1 < \theta_C$ according to the relationships

$$\tan\frac{\delta_N}{2} = \frac{\sqrt{n^2\cos^2\theta_1 - 1}}{n\sin\theta_1} \quad \text{.....(4.2.11a)}$$

$$\tan\frac{\delta_p}{2} = \frac{\sqrt{n^2\cos^2\theta_1 - 1}}{\sin\theta_1} \quad \text{.....(4.2.11b)}$$

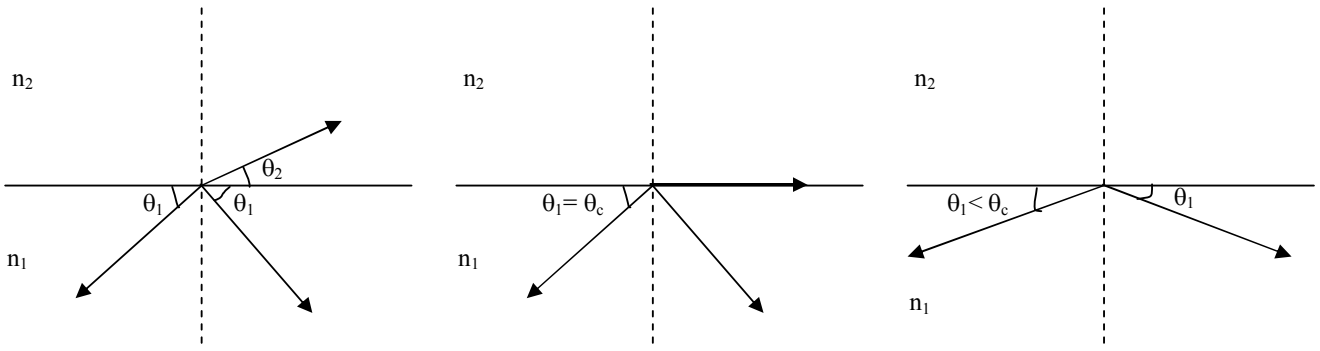


Fig 4.2.6:Representation of the critical angle and total internal reflection at a glass-air interface

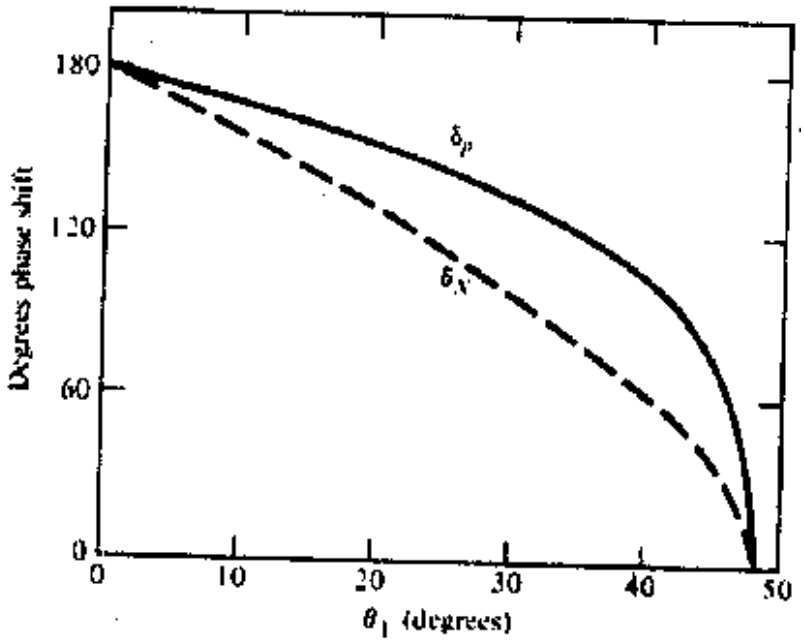


fig 4.2.7. Phase shifts occurring from the reflection of wave components normal (δ_N) and parallel (δ_p) to the reflecting surface

Here δn and δ_P are the phase shifts of the wave components normal and parallel to the reflecting surface, respectively, and $n= n_1/n_2$. A derivation of Eqs.(4.2.11a) and (4.2.11b) can be found in Klein. These phase shifts are shown in Fig. 4.2.7 for a glass-air interface (n=1.5 and $\theta_C = 48^o$). The values range from zero immediately at the critical angle to π when $\theta_1 = 0^o$.
These basic optical principles will now be used to illustrate how optical power is transmitted along a fiber.

4.2.3 OPTICAL FIBER MODES AND CONFIGURATIONS

Before going into details on optical fiber characteristics in Sec.4.2.3.3, we first present a brief overview of the underlying concepts of optical fiber modes and optical fiber configurations.

4.2.3.1 Fiber Types:

An optical fiber is a dielectric waveguide that operates at optical frequencies. This fiber waveguide is normally cylindrical in form. It confines electromagnetic energy in the form of light to within its surfaces and guides the light in a direction parallel to its axis. The transmission properties of an optical waveguide are dictated by its structural characteristics, which have a major effect in determining how an optical signal is affected as it propagates along the fiber. The structure basically establishes the information-carrying capacity of the fiber and also influences the response of the waveguide to environmental perturbations.

The propagation of light along a waveguide can be described in terms of a set of guided electromagnetic waves called the modes of the waveguide. These guided modes are referred to as the bound or trapped modes of the waveguide. Each guided mode is a pattern of electric and magnetic field lines that is repeated along the fiber at intervals equal to the wavelength. Only a certain discrete number of modes are capable of propagating along the guide. As will be seen in Sec.4.2.4, these modes are those electromagnetic waves that satisfy the homogeneous wave equation in the fiber and the boundary condition at the waveguide surfaces.

Although many different configurations of the optical waveguide have been discussed in the literature, the most widely accepted structure is the single solid dielectric cylinder of radius a and index of refraction n_1 shown in Fig. 4.2.8. This cylinder is known as the core of the fiber. The core is surrounded by a solid dielectric cladding having refractive index n_2 that is less than n_1. Although, in principle, a cladding is not necessary for light to propagate along the core of the fiber, it serves several purposes. The cladding reduces scattering loss resulting from dielectric discontinuities at the core surface, it adds mechanical strength to fiber, and it protects the core from absorbing surface contaminants with which it could come in contact.

In low-and medium-loss fibers the core material is generally glass which is surrounded by either a glass or a plastic cladding. Higher-loss plastic core fibers with plastic claddings are also widely in use. In addition, most fibers are encapsulated in an elastic, abrasion-resistant plastic material. This material adds further strength to the fiber and mechanically isolates or buffers the fibers from small geometrical irregularities, distortions, or roughnesses of adjacent surfaces. These perturbations could otherwise cause scattering losses induced by random microsopic bends that can arise when the fibers are incorporated into cables or supported by other structures.

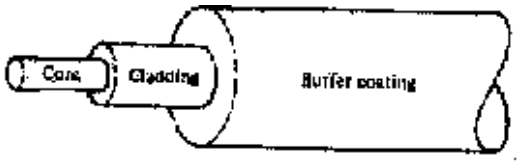


Fig 4.2.8. Schematic of a single fiber structure. A circular solid core of refractive index n_1 is surrounded by a cladding having a refractive index $n_2<n_1$. An elastic plastic buffer encapsulates the fiber.

Variations in the material composition of the core give rise to the two commonly used fiber types shown in Fig.4.2.9. In the first case the refractive index of the core is uniform throughout and undergoes an abrupt change (or step) at the cladding boundary. This is called a step-index fiber. In the second case the core refractive index is made to vary as a function of the radial distance from the center of the fiber. This type is a graded-index fiber.

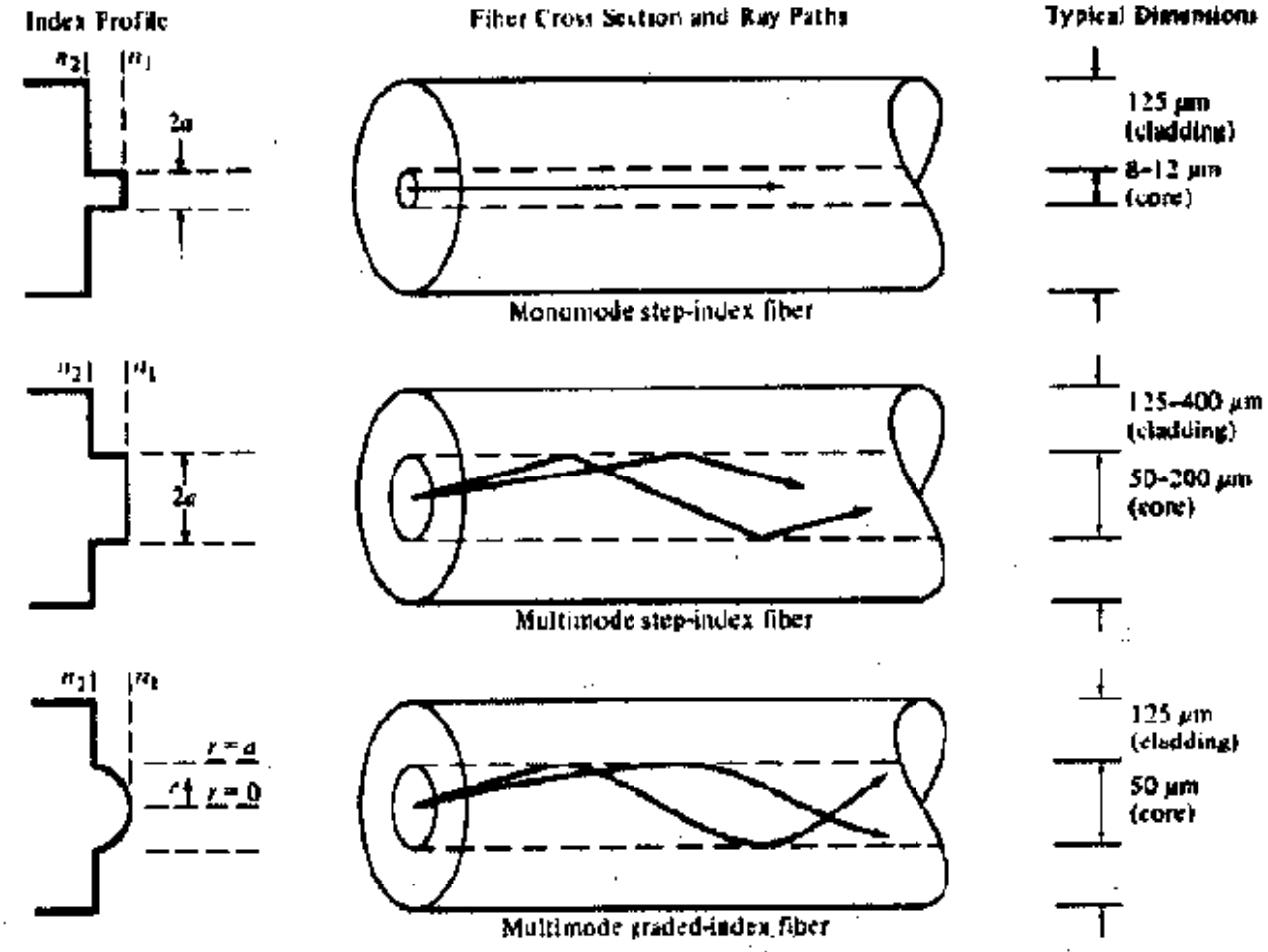


Fig 4.2.9 comparison of single mode and multi mode step index and graded index optical fibers.

Both the step-and the graded-index fibers can be further divided into single-mode and multimode classes. As the name implies, a single-mode fiber sustains only one mode of propagation, whereas multimode fibers contain many hundreds of modes. A few typical sizes of single-and multimode fibers are given in Fig4..2-9 to provide an idea of the dimensional scale. Multimode fibers offer several advantages compared to single-mode fibers. The larger core radii of multimode fibers make it easier to launch optical power into the fiber and facilitate the connecting together of similar fibers. Another advantage is that light can be launched into a multimode fiber using a light-emitting-diode (LED) source, whereas single-mode fibers must be excited with laser diodes. Although LEDs have less optical output power than laser diodes, they are easier to make, are less expensive, require less complex circuitry, and have longer lifetimes than laser diodes, thus making them more desirable in many applications.

A disadvantage of multimode fibers is that they suffer from intermodal dispersion. Briefly, intermodal dispersion can be described as follows. When an optical pulse is launched into a fiber, the optical power in the pulse is distributed over all (or most) of the modes of the fiber.

Each of the modes that can propagate in a multimode fiber travels at a slightly different velocity. This means that the modes in a given optical pulse arrive at the fiber end at slightly different times, thus causing the pulse to spread out in time as it travels along the fiber. This effect, which is known as intermodal dispersion, can be reduced by using a graded-index profile in the fiber core. This allows graded-index fibers to have much larger bandwidths (data rate transmission capabilities) than step-index fibers. Even higher bandwidths are possible in single-mode fibers where intermodal dispersion effects are not present.

4.2.3.2 Rays and Modes

As we mentioned in Sec.4.2.3.1, the electromagnetic light field that is guided along an optical fiber can be represented by a superposition of bound or trapped modes. Each of these guided modes is composed of a s et of simple electromagnetic field configurations which form a standing-wave pattern in the transverse direction (that is, transverse to the waveguide axis). For mono-chromatic light fields of radian frequency w, a mode traveling in the positive z direction (that is, along the fiber axis) has a time and z dependence given by

$$\mathrm{Exp}[j(\omega t-\beta z)]$$

The factor β is the z component of the wave propagation constant $k=2\pi/\lambda$ and is the main parameter of interest in describing fiber modes. For guiding modes β can only assume certain discrete values, which are determined from the requirement that the mode field must satisfy Maxwell's equations and the electric and magnetic field boundary conditions at the core-cladding interface. This is described in detail in Sec.4.2.4.

Another method for theoretically studying the propagation characteristics of light in an optical fiber is the geometrical optics or ray-tracing approach. This method provides a good approxima-tion to the light acceptance and guiding properties of optical fibers when the ratio of the fiber radius to the wavelength is large. This is known as the small wavelength limit. Although the ray approach is strictly valid only in the zero wavelength limit, it is still relatively accurate and extremely valuable for nonzero wavelengths when the number of guided modes is large, that is, for multimode fibers. The advantage of the ray approach is that, compared to the exact electromagnetic wave (modal) analysis, it gives a more direct physical interpretation of the light propagation characteristics in an optical fiber.

Since the concept of a light ray is very different from that of a mode, let us qualitatively what the relationship is between them. A guided mode traveling in the z direction (along the fiber axis) can be decomposed into a family of superimposed plane waves that collectively form a standing-wave pattern in the direction transverse to the fiber axis. Since with any plane wave we can associate a light ray that is perpendicular to the phase front of the wave, the family of plane waves corresponding to a particular mode forms a set of rays called a ray congruence. Each ray of this particular set travels in the fiber at the same angle relative to the fiber axis. We note here that, since only certain number M of discrete guided modes exist in a fiber, the possible angles of the ray congruences corresponding to these modes are also limited to the same number M. Although a simple ray picture appears to allow rays at any angle less than the critical angle to propagate in a fiber, the allowable quantized propagation angles result when the phase condition for standing waves is introduced into the ray picture. This is discussed further in Sec.4.2.3.5.

Despite the usefulness of the approximate geometrical optics method, a number of limitations and discrepancies exist between it and the exact modal analysis. An important case is the analysis of single-mode o r few-mode fibers which must be dealt with by using electromagnetic theory. Problems involving coherence or interference phenomenon must also be solved with an electromagnetic approach. In addition, a modal analysis is necessary when a knowledge of the field distribution of individual modes is required.
This arises, for example when analyzing the excitation of an individual mode or when analyzing the coupling of power between modes at waveguide imperfections. Another discrepancy between the ray optics approach and the modal analysis occurs when an optical fiber is uniformly bent with a constant radius of curvature. Wave optics correctly predicts that every mode of the curved fiber experiences some radiation loss. Ray optics, on the other hand, erroneously predicts that some ray congruences can undergo total internal reflection at the curve and, consequently, can remain guided without loss.

4.2.3.3 STEP-INDEX FIBER STRUCTURE:

We begin our discussion of light propagation in an optical waveguide by considering the step-index fiber. In practical step-index fibers the core of radius a has a refractive index n_1 which is typically equal to 1.48. This is surrounded by a cladding of slightly lower index n_2, where

$$n_2 = n_1(1-\Delta) \quad \ldots\ldots\ldots\ldots (4.2.12)$$

The parameter Δ is called the core-cladding index difference or simply the index difference. Values of n_2 are chosen such that Δ is nominally 0.01. Since the core refractive index is larger than the cladding index, electromagnetic energy at optical frequencies is made to propagate along the fiber waveguide through internal reflection at the core-cladding interface.

4.2-3-4 RAY OPTICS REPRESENTATION:

Since the core size of multimode fibers is much larger than the wavelength of the light we are interested in (which is approximately 1 μm), an intuitive picture of the propagation mechanism in an ideal multimode step-index optical waveguide is most easily seen by a simple ray (geometrical) optics representation. For simplicity in this analysis we shall consider only a particular ray belonging to a ray congruence which represents a fiber mode. The two types of rays that can propagate in a fiber are meridional rays and skew rays. Meridional rays are confined to the meridian planes of the fiber, which are the planes that contain the axis of symmetry of the fiber (the core axis). Since a given meridional ray lies in a single plane, its path is easy to track as it travels along the fiber. Meridional rays can be divided into two general classes: bound rays that are trapped in the core and propagate along the fiber axis according to the laws of geometrical optics, and unbound rays that are refracted out of the fiber core.

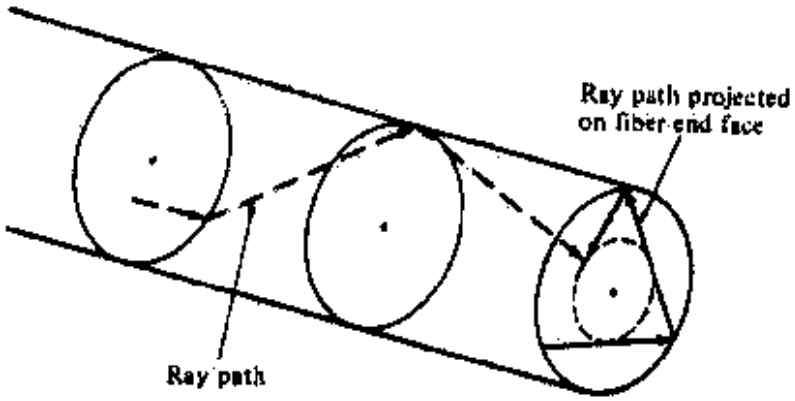


fig 4.2.10. Ray optics representation of skew rays traveling in a step index optical fiber core.

Skew rays are not confined to a single plane but, instead, tend to follow a helical type path along the fiber as shown in Fig.4.2.10. These rays are more difficult to track as they travel along the fiber since they do not lie in a single plane. Although skew rays constitute a major portion of the total number of guided rays their analysis is not necessary to obtain a general picture of rays propagating in a fiber. The examination of meridional rays will suffice for this purpose. However, a detailed inclusion of skew rays will change such expressions as the light acceptance ability of the fiber and power losses of light traveling along a waveguide.

A greater power loss arises when skew rays are included in the analyses, since many of the skew rays that geometric optics predicts are trapped in the fiber are actually leaky rays. These leaky rays are only partially confined to the core of the circular optical fiber and attenuate as the light travels along the optical waveguide. This partial reflection of leaky rays cannot be described by pure ray theory alone. Instead, the analysis of radiation loss arising from these types of rays must be described by mode theory. This is explained further in Sec.4.2.4.

The meridional theory is shown in Fig.4.2.11. for a step-index fiber. The light ray enters the fiber core from a medium of refracative index n at an angle θ_0 with respect to the fiber axis and strikes the core-cladding interface at a normal angle ϕ. If it strikes this interface at such an angle that it is totally internally reflected, the meridional rays follows a zig-zag path along the fiber core passing through the axis of the guide after each reflection.

From Snell's law the minimum angle that supports total internal reflection for the meridional ray is $\sin(\phi_{\min}) = \frac{n_2}{n_1}$ .....(4.2.13)

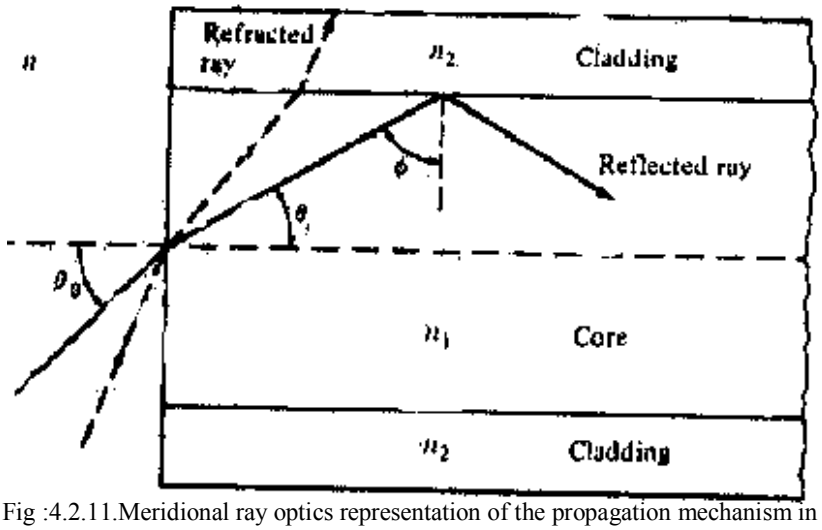


Fig :4.2.11.Meridional ray optics representation of the propagation mechanism in an ideal step-index optical waveguide

Rays striking the core-cladding interface at angles less than ϕ_{min} will refract out of the core and be lost in the cladding. The condition of Eq. (4.2.13) can be related to the maximum entrance angle $\theta_{0,\,max}$ through the relationship

$n \sin\theta_{o,\max} = n_1 \sin\theta_c = (n_1^2 - n_2^2)^{1/2}$ .....(4.2.14)

Where θ_C is the critical angle. Thus those rays having entrance angles θ_0 less than $\theta_{0,\,max}$ will be totally internally reflected at the core-cladding interface.

Equation (2-14) also defines the numerical aperture NA of a step-index fiber for meridional rays.

$NA = n \sin\theta_{o,\max} = (n_1^2 - n_2^2)^{1/2} \approx n_1\sqrt{2\Delta}$ .....(4.2.15)

The approximation on the right-hand side is valid for the typical case where Δ as defined by Eq.(4.2.12) is much less than 1. Since the numerical aperture is related to the maximum acceptance angle, it is commonly used to describe the light acceptance or gathering capability of a fiber and to calculate source-to-fiber optical power coupling efficiencies. The numerical aperture is a dimensionless quantity, which is less than unity, with values normally ranging from 0.14 to 0.50.

4.2.3.5. WAVE REPRESENTATION:

The ray theory appears to allow rays at any angle θ_1 less than the critical angle to propagate along the fiber. However, when the phase of the plane wave associated with the ray is taken into account, it is seen that only rays at certain discrete angles less than or equal to θ_C are capable of propagating along the fiber.

To see this, consider a light ray in the core incident on the reflective surface at an angle θ as shown in Fig.4.2.12. The plane wave associated with this ray is of the form given in Eq.(4.2.1). As the wave travels it undergoes a phase change δ given by

$$\delta = k_1 s = n_1 k s = \frac{n_1 2\pi s}{\lambda} \quad \text{.....(4.2.16)}$$

where k_1 = the propagation constant in the medium of refractive index n
$k = k_1/n_1$ is the free-space propagation constant
s = the distance traveled along the ray by the wave

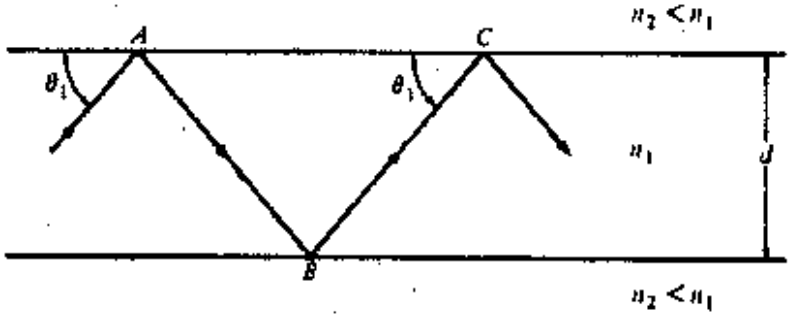


Fig 4.2.12. Light ray propagating along a fiber waveguide. Phase changes occur both as the wave travels through the fiber medium and at the reflection points.

The phase of the wave changes not only as the wave travels but also upon reflection from a dielectric interface, as shown in Sec.4.2.2.

In order for the wave associated with a given ray to propagate along the waveguide shown in Fig.4. 2.12, the phase of the twice reflected wave must be the same as that of the incident wave. That is, the wave must interfere constructively with itself. If this phase condition is not satisfied, the wave would interfere destructively with itself and just die out. Thus the total phase shift that results when the wave traverses the guide twice (from points A to B to C) and gets reflected twice (at points A and B) must be equal to an integer multiple of 2π rad. Using Eqs. (4.2.16) and (4.2.11), we let the phase change that occurs over the distance ABC be $\delta_{AC} = n_1 k \left(2d / \sin\theta_1 \right)$ and the phase changes upon reflection each be (assuming for simplicity that the wave is polarized normal to the reflecting surface)

$$\delta_1 = 2\arctan \frac{(n^2 \cos^2\theta_1 - 1)^{1/2}}{n \sin\theta_1} \quad \text{.....(4.2.17)}$$

Where $n = n_1/n_2$. Then the following condition must be satisfied

$$\frac{2n_1 kd}{\sin\theta_1} + 2\delta_1 = 2\pi M \quad \text{.....(4.2.18)}$$

Where M is an integer that determines the allowed ray angles for waveguiding.

4.2.3 MODE THEORY FOR CIRCULAR WAVEGUIDES:

To attain a more detailed understanding of the optical power propagation mechanism in a fiber, it is necessary to solve Maxwell's equations subject to the cylindrical boundary conditions of the fiber. This has been carried out in extensive detail in a number of works. Since a complete treatment is beyond the scope of this book, only a general outline of the analyses will be given here.

Before we progress with our discussion of mode theory in circular optical fibers, let us first qualitatively examine the appearance of modal fields in the planar dielectric slab waveguide shown in Fig.4. 2.13. This waveguide is composed of a dielectric slab of refractive index n_1 sandwiched between dielectric material of refractive index $n_2 < n_1$.which we shall call the cladding. This represents the simplest form of an optical waveguide and can serve as a model to gain an understanding of wave propagation in optical fibers. In fact, a cross-sectional view of an optical fiber cut along its axis. Figure 4.2.13 shows the field patterns of several of the lower-order modes (which are solutions of Maxwell's equations for the slab waveguide). The order of a mode is equal to the number of field maxima across the guide. The order of the mode is related to the angle that the ray congruence corresponding to this mode makes with the plane of the waveguide (or the axis of a fiber); that is, the steeper the angle, the higher the order of the mode. The plots show that the electric fields of the guided modes are not completely confined to the central dielectric slab (that is, they do not go to zero at the guide-cladding interface), but, instead, they extend partially into the cladding. The fields vary harmonically in the guiding region of refractive index n_1 and decay exponentially outside of this region. For low-order modes the fields are tightly concentrated near the center of the slab (or the axis of an optical fiber) with little penetration into the cladding region. On the other hand, for higher-order modes the fields are distributed more toward the edges of the guide and penetrate further into the cladding region.

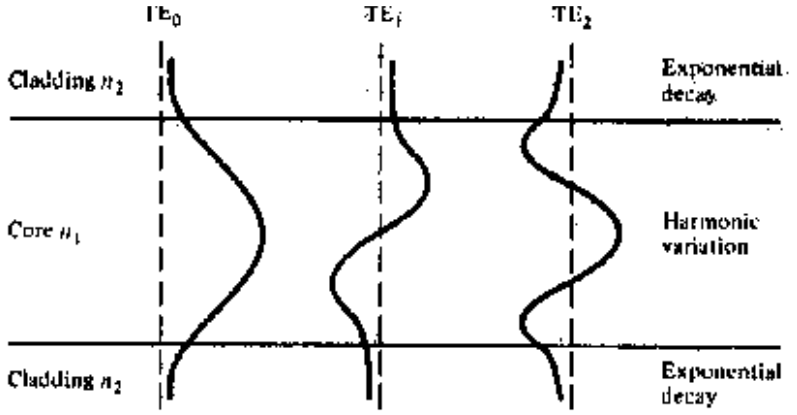


Fig 4.2.13 . Electric field distribution for several of the lower order guided modes in a symmetrical slab waveguide.

Solving Maxwell's equations shows that, in addition to supporting a finite number of guided modes, the optical fiber waveguide has an infinite continuum of radiation modes that are

not trapped in the core and guided by the fiber but are still solutions of the same boundary-value problem. The radiation field basically results from the optical power that is outside the fiber acceptance angle being refracted out of the core. Because of the finite radius of the cladding, some of this radiation gets trapped in the cladding, hereby causing cladding modes to appear. As the core and cladding modes propagate along the fiber, mode coupling occurs between the cladding modes and higher-order core modes. This coupling occurs because the electric fields of the guided core modes are not completely confined to the core but extend partially into the cladding (see Fig.4.2.13) and likewise for the cladding modes. A diffusion of power back and forth between the core and cladding modes this occurs, which generally results in a loss of power from the core modes. In practice, the cladding modes will be suppressed by a lossy coating which covers the fiber or they will scatter out of the fiber after traveling a certain distance because of roughness on the cladding surface.

In addition to bound and refracted modes, a third category of modes called leaky modes is present in optical fibers. These leaky modes are only partially confined to the core region, and attenuate by continuously radiating their power out of the core as they propagate along the fiber. This power radiation out of the waveguide results from a quantum mechanical phenomenon known as the tunnel effect. However, it is essentially based on the upper and lower bounds that the boundary conditions for the solutions of Maxwell's equations impose on the propagation constant β. Mode remains guided as long as β satisfies the condition.

$$n_2 k < \beta < n_1 k$$

Where n_1 and n_2 are the refractive indices of the core and cladding, respectively, and $k=2\pi/\lambda$. The boundary between truly guided modes and leaky modes is defined by the cutoff condition $\beta = n_2 k$. As soon as β becomes smaller than $n_2 k$. Power leaks out of the core into the cladding region. Leaky modes can be carry significant amounts of optical power in short fibers. Most of these modes disappear after a few centimeters, but a few have sufficiently low losses to persist in fiber lengths of a kilometer.

4.2.5. Summary: In this lesson we have examined the structure of optical fibers and have considered two mechanisms that show how light propagates along these fibers. In its simplest form optical fiber is a coaxial cylindrical arrangement of two homogeneous dielectric materials. This fiber consists of a central core of refractive index n_1 surrounded by a cladding region of refractive index n_2 that is less than n_1. This configuration is referred to as a step-index fiber because the cross-sectional refractive index profile has a step function at the interface between the core and clad. Graded index fibers are those in which the refractive index profile varies as a function of the radial coordinate r in the core but is constant in the cladding (often profile represented as a power law). A general picture of light propagation in an optical fiber can be obtained by considering a ray-tracing (or geometrical optics) model in a slab waveguide. Light rays propagate along the slab waveguide by undergoing total internal reflection in accordance with snell's law in the interfaces of these two materials. Further, mode theory approach has been explained.

4.2.6. Keywords:

Light wave, waveguides, total internal reflection, propagation constant, step-index and graded-index optical fibers. Guided modes

4.2.7. Self assessment questions

1. What is total internal reflection and apply it to the optical fibers.
2. Explain the mode theory of circular waveguides. Obtain the wave guide equations.
3. Explain the terms step-index fibers, graded index fibers and modes and configurations in fibers.

4.2.8. Text and reference books

1. Optical fiber communications by G. Keiser, McGraw-Hill International Edition, 2000, Third edition, and also see first edition.
2. Optical fiber communications by Kato, McGraw-Hill, 1986
3. Inhomogeneous optical waveguides by M.S. Sodha and A.K. Ghatak, Plenum Press, 1977

Unit IV
Lesson 3

OPTICAL FIBERS: WAVEGUIDING FUNDAMENTALS

Objective: To discuss about the wave propagation in the step and graded index optical fibers, their numerical apertures and different modes in the respective optical fibers

Structure:

4.3.1 MAXWELL'S EQUATIONS :

To analyze the optical waveguide we need to consider Maxwell equations that give the relationship between the electrical and magnetic fields. Assuming a linear isotopic dielectric material having no currents and free charges these equations take the form .

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad \text{.....(4.3.1a)}$$

$$\nabla \times H = \frac{\partial D}{\partial t} \quad \text{.....(4.3.1b)}$$

$$\nabla . D = 0 \quad \text{..... (4.3.1c)}$$

$$\nabla . B = 0 \quad \text{..... . (4.3.1d)}$$

where $\mathbf{D} = \varepsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$. The parameter ε is the permittivity (or dielectric constant) and μ is the permeability of the medium.
A relationship defining the wave phenomenon of the electro magnetic fields can be derived from maxwell's equations. Taking the curl of equation (4.3.1a) and making use of equation (4.3.1b)

yields

$$\nabla\times(\nabla\times E)=-\mu\frac{\partial}{\partial t}(\nabla\times H)=-\varepsilon\mu\frac{\partial^2 E}{\partial t^2}\ldots\ldots\ldots(4.3.2)$$

Using the vector identity $\nabla\times(\nabla\times E)=\nabla(\nabla.E)-\nabla^2 E$

And using eq (4.31c) (that is,$\nabla.\mathbf{E}=0$), Eq (4.3.2) becomes

$$\nabla^2 E=\in\mu\frac{\partial^2 E}{\partial t^2}\ldots\ldots\ldots..(4.3.3)$$

Similarly, by taking the curl of equation (4.3.1 b) it can be shown that

$$\nabla^2 H=\in\mu\frac{\partial^2 H}{\partial t^2}\ldots\ldots\ldots\ldots\ldots(4.3.4.)$$

Equations (4.3.3) and (4.3.4.) are the standard wave equations.

4.3.2 WAVEGUIDE EQUATIONS:

Consider electromagnetic waves propagating along a cylindrical fiber shown in fig. 4.3.1 For this fiber a cylindrical coordinate system (r,ϕ,z) is defined with the z-axis lying along axis of the waveguide. If the electromagnetic waves are to propagate along the z-axis, they will have functional dependence of the form

$$E=E_o(r,\phi)e^{j(\omega t-\beta z)}\ldots\ldots\ldots\ldots(4.3.5.)$$

$$H=H_o(r,\phi)e^{k(\omega t-\beta z)}\ldots\ldots\ldots\ldots(4.3.6)$$

Which are harmonic in time t and coordinate z. the parameter β is the z-component of the propagation vector and will be determined by boundary conditions on the electromagnetic fields at the core –cladding interface.

When eqs(4.3.5) and (43.6) are substituted into Maxwell's curl equations. We have from eq (4.3.1a),

$$\frac{1}{r}(\frac{\partial E_z}{\partial\phi}+jr\beta E_\phi)=-j\omega\mu H_r\ldots\ldots\ldots\ldots..(4.3.7a)$$

$$j\beta E_r+\frac{\partial E_z}{\partial r}=j\omega\mu H_\phi\ldots\ldots\ldots\ldots\ldots\ldots\ (4.3.7b)$$

$$\frac{1}{r}[\frac{\partial}{\partial r}(rE_\phi)-\frac{\partial E_r}{\partial\phi}]=-j\mu\omega H_z\ldots\ldots\ldots\ (4.3.7c)$$

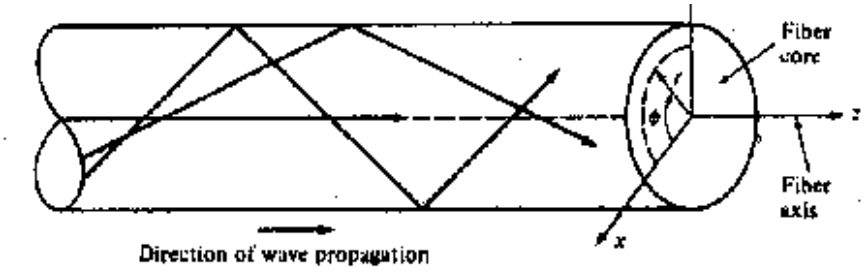


Fig 4.3.1: Cylindrical coordinate system used for analyzing electromagnetic wave propagation in an optical fiber.

and from Eq (4.3.1b)

$$\frac{1}{r}\left(\frac{\partial H_z}{\partial \phi} + jr\beta H_\phi\right) = j \in \omega E_r \quad \text{.....(4.3.8a)}$$

$$j\beta H_r + \frac{\partial H_z}{\partial r} = -j \in \omega E_\phi \quad \text{..... (4.3.8b)}$$

$$\frac{1}{r}\left[\frac{\partial}{\partial r}(rH_\phi) - \frac{\partial H_r}{\partial \phi}\right] = j \in \omega E \quad \text{..... (4.3.8c)}$$

By eliminating variables these equations can be rewritten such that when E_z and H_z are known. The remaining transverse components E_r, E_ϕ, H_r and H_ϕ can be determined. For example E_ϕ or Hr can be eliminated from eq (4.3.7b) and (4.3.8b) so that the components H_ϕ or E_r respectively can be found interms E_z or H_z doing so yields

$$E_r = -\frac{j}{q^2}\left(\beta\frac{\partial E_z}{\partial r} + \frac{\mu\omega}{r}\frac{\partial H_z}{\partial \phi}\right) \quad \text{..... (4.3.9a)}$$

$$E_\phi = -\frac{j}{q^2}\left(\frac{\beta}{r}\frac{\partial E_z}{\partial \phi} - \mu\omega\frac{\partial H_z}{\partial r}\right) \quad \text{..... .(4.3.9b)}$$

$$H_r = -\frac{j}{q^2}\left(\beta\frac{\partial H_z}{\partial r} - \frac{\omega e}{r}\frac{\partial E_z}{\partial \phi}\right) \quad \text{..... (4.3.9c)}$$

$$H_\phi = -\frac{j}{q^2}\left(\frac{\beta}{r}\frac{\partial H_z}{\partial \phi} + \omega \in \frac{\partial E_z}{\partial r}\right) \quad \text{.....(4.3.9d)}$$

where $q^2 = \omega^2\varepsilon\mu - \beta^2 = k^2 - \beta^2$,
substitution of equations (4.3.9c) and (4.3.9d) into equation (4.3.8c) results in the wave equation in cylindrical coordinates.

$$\frac{\partial^2 E_z}{\partial r^2}+\frac{1}{r}\frac{\partial E_z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2}+q^2E_z=0 \ldots\ldots\ldots(4.3.10)$$

substitution of equations (4.3.9a) and (4.3.9b) into equation (4.3.7c)

$$\frac{\partial^2 H_z}{\partial r^2}+\frac{1}{r}\frac{\partial H_z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 H_z}{\partial \phi^2}+q^2H_z=0 \ldots\ldots(4.3.11)$$

It is interesting to note that eqs(4.3.10) and (4.3.11) each contain either only E_z or H_z . this appears to imply that the longitudinal components of **E** and **H** are uncoupled and can be chosen arbitrarily provided that they satisfy eqs(4.3.10) and (4.3.11). However, in general, coupling E_z and H_z is required by the boundary conditions of the electromagnetic field components described in (4.3.4) if the boundary conditions do not lead to components mode solutions can be obtained in which either the modes are called transverse, electric or TE modes and then transverse magnetic are TM modes result. Hybrid modes exist if both E_z and H_z are non-zero. It is the designated as HE or EH modes depending on whether H_z or E_z respectively makes a larger contribution to the transverse field. The fact that hybrid modes are present in optical waveguides make the analysis more complex compared to the simpler case hollow metallic waveguides. There only TE and TM modes are found.

4.3.3 WAVE EQUATIONS FOR STEP INDEX FIBERS:

We now use the above results to find the guided modes in a step index fiber .A standard mathematical procedure for solving eqs such as eq4.3.10 is to use the separation of variables methods, which assumes a solution of the form.

$$E_z = AF_1(r)F_2(\phi)F_3(z)F_4(t) \ldots\ldots\ldots\ldots(4.3.12)$$

As was already assumed the time and z dependent factors are given by

$$F_3(z)F_4(t)=e^{j(\omega t-\beta z)} \ldots\ldots\ldots\ldots\ldots\ldots \quad (4.3.13)$$

Since the wave is sinusoidal in time and propagates in the z direction. In addition, because of the circular symmetry of the waveguide, each field component must not change when the coordinate ϕ is increased by 2π. We thus assume a periodic function of the term.

$$F_2(\phi)=e^{j\upsilon\phi} \ldots\ldots\ldots\ldots\ldots \quad (4.3.14)$$

The constant ν can be positive or negative. But it must be an integer since the fields must be periodic in ϕ with a period of 2π.

Substituting eq 4.3.14 into eq 4.3.12. The wave equation for E_Z eq 4.3.10 becomes

$$\frac{\partial^2 \mathrm{F_1}}{\partial r^2}+\frac{1}{r}\frac{\partial \mathrm{F_1}}{\partial r}+\left(\mathrm{q}^2-\frac{v^2}{r^2}\right)\mathrm{F_1}=0 \ldots\ldots .(4.3.15)$$

which is a well-known differential equation for Bessel functions. An exactly identical equation can be derived for Hz .

For the configuration of the step index fiber, we consider a homogeneous core of refractive index n_1 and radius a, which is surrounded by an infinite cladding of index n_2. The reason for assuming an infinitely thick cladding is that the guided modes in the core have exponentially

decaying fields outside the core, which must have in significant values at the outer boundary of the cladding. In practice optical fibers are designated claddings that are sufficiently thick. So that the guided mode field thus not reach the outer boundary of the cladding .To get an idea of the field patterns the electric field distribution for several of the lower order guided modes in a symmetrical slab waveguide we are shown in fig (4.2.13). The fields vary harmonically in the guiding region of refractive index in n_1 and decay exponentially outside of this region.

Eq (4.3.15) must now be solved for the regions inside and outside .The core for the inside region the solutions for the guided modes must remain finite as $r \rightarrow 0$ whereas on the outside the solution must decay to zero as $r \rightarrow \infty$. Thus for r<a the solutions are Bessel functions of the first kind of order ν, for these functions we use the common designation $J_v(ur)$. Here $u^2=k_1^2-\beta^2$ with $k_1= 2\pi n_1/\lambda$.. The expressions for E_z and H_z inside the core are thus

$$E_z(r<a) = AJ_v(ur)e^{jv\phi}e^{j(\omega t-\beta z)} \quad \text{.....(4.3.16)}$$

$$H_z(r>a) = BJ_v(ur)e^{jv\phi}e^{j(\omega t-\beta z)} \quad \text{..... (4.3.17)}$$

Where A and B are arbitrary constants.

Outside of the core the solutions to Eq(4.3.15) are given by modified Bessel functions of the second kind $K_v(wr)$, where $w^2= \beta^2 - k_2^2$ with $k_2= 2\pi n_2/\lambda$. The expressions for E_z and H_z outside the core are, therefore,

$$E_z(r>a) = CK_v(\omega r)e^{jv\phi}e^{j(\omega t-\beta z)} \quad \text{.....(4.3.18)}$$

$$H_z(r>a) = DK_v(\omega r)e^{jv\phi}e^{j(\omega t-\beta z)} \quad \text{..... .(4.3.19)}$$

With C and D being arbitrary constants.

From the definition of the modified Bessel function, it is seen that $K_v(wr) \rightarrow e^{-ur}$ as $wr \rightarrow \infty$. Since $K_v(wr)$ must go to zero as $r \rightarrow \infty$, it follows that w > 0. This, in turn , implies that $\beta \geq K_2$ which represents a cutoff condition. The cutoff condition is the point at which a mode is no longer bound to the core region. A second condition on β can be deduced from the behaviour of $J_v(ur)$ Inside the core the parameter u must be real for F_1 to be real , from which it follows that $K_1 \geq \beta$. The permissible range of β for bound solutions is ,therefore,

$$n_2k = k_2 \leq \beta \leq k_1 = n_1k \quad \text{..... (4.3.20)}$$

Where $k = \frac{2\pi}{\lambda}$ is the free-space propagation constant.

4.3.4 Modal Equation:

The solutions for β must be determined from the boundary conditions. The boundary conditions require that the tangential components E_ϕ and E_z of **E** inside and outside of the dielectric interface at r=a must be the same and, similarly, for the tangential components H_ϕ and H_z. Consider first the tangential components of **E**. For the z component we have, from Eq(4.3.16) at the inner core-cladding boundary ($E_z = E_{z1}$) and from Eq(4.3.18) at the outside of the boundary ($E_z = E_{z2}$) that,

$$E_{z1} - E_{z2} = AJ_v(ua) - CK_v(\omega a) = 0 \quad \text{..... (4.3.21)}$$

The ϕ component is found from Eq(4.3.9b) .Inside the core the factor q^2 is given by

$$q^2 = u^2 = k_1^2 - \beta^2 \quad \text{.....(4.3.22)}$$

Where $k_1 = \frac{2\pi}{\lambda} = \omega\sqrt{\varepsilon_1 \mu}$, while outside the core

$$\omega^2 = \beta^2 - k_2^2 \quad \text{.....} \qquad (4.3.23)$$

while outside the core $k_2 = {2\pi n_2}/{\lambda} = \omega\sqrt{\varepsilon_2 \mu}$, Substituting Eqs.(4.3.16)and (4.3.17) into Eq.(4.3.9b) to find $E_{\phi 1,}$ and similarly using Eqs.(4.3.18) and (4.3.19) to determine $E_{\phi 2}$ yields at r = a.

$$E_{\phi 1} - E_{\phi 2} = -\frac{j}{u^2}\left[A\frac{j\nu\beta}{a} J_\nu(ua) - B\omega\mu u J_\nu^1(ua) \right] - \frac{j}{w^2}\left[C\frac{j\nu\beta}{a} K_\nu(wa) - D\omega\mu w K_\nu^1(wa) \right] = 0 \ldots$$

.....(4.3.24)

Where the prime indicates differentiation with respect to the argument. Similarly, for the tangential components of **H** it is readily shown that at r = a.

$$H_{z1} - H_{z2} = BJ_\nu(ua) - DK_\nu(wa) = 0 \text{ .....} \qquad (4.3.25)$$

and

$$H_{\phi 1} - H_{\phi 2} = \frac{-j}{u^2}\left[B\frac{j\nu\beta}{a} J_\nu(ua) + A\omega\varepsilon_i u J_\nu^1(ua) \right] - \frac{j}{w^2}\left[D\frac{j\nu\beta}{a} K_\nu(wa) + C\omega\varepsilon_2 w K_\nu^1(wa) \right] = 0 \ldots$$

.....(4.3.26)

Equations (4.3.21),(4.3.24),(4.3.25)and(4.3.26) are a set of four equations with four unknown coefficients, A, B, C, and D. a solution to these equations exists only if the determinant of these coefficients is zero.

$$\begin{vmatrix} J_\nu(ua) & 0 & -K_\nu(wa) & 0 \\ \frac{\beta\nu}{au^2} J_\nu(ua) & \frac{j\omega\mu}{u} J_\nu^1(ua) & \frac{\beta\nu}{aw^2} K_\nu(wa) & \frac{j\omega\mu}{w} K_\nu^1(wa) \\ 0 & J_\nu(ua) & 0 & -K_\nu(wa) \\ -\frac{j\omega\varepsilon_1}{u} J_\nu^1(ua) & \frac{\beta\nu}{au^2} J_\nu(ua) & -\frac{j\omega\varepsilon_2}{w} K_\nu^1(wa) & \frac{\beta\nu}{aw^2} K_\nu(wa) \end{vmatrix} = 0 \text{ .....} \quad \ldots \ (4.3.27)$$

Evaluation of this determinant yields the following eigen value equation for β :

$$(I_\nu + H_\nu)(k_1^2 I_\nu + k_2^2 H_\nu) = \left(\frac{\beta\nu}{a}\right)^2 \left(\frac{1}{u^2} + \frac{1}{w^2}\right)^2 \text{.....(4.3.28)}$$

where

$$I_\nu = \frac{J_\nu^1(ua)}{uJ_\nu(ua)} \text{ and } H_\nu = \frac{K_\nu^1(wa)}{wK_\nu(wa)}$$

Upon solving Eq(4.3.28) for it will be found that only discrete values restricted to the range given by Eq(4.3.20) will be allowed. Although Eq(4.3.28) is a complicated transcendental equation which is generally solved by numerical techniques, its solution for any particular mode will

provide all the characteristics of that mode. We shall now consider this equation for some of the lower-order modes of a step-index waveguide.

4.3.5 MODES IN STEP-INDEX FIBERS:

To help describe the modes we shall first examine the behavior of the J-type .Bessel functions . these are plotted in fig.4.3.2. for the first three orders. The J-type Bessel functions are similar to harmonic functions. Since they exhibit oscillatory behaviour for real k as is the case for sinusoidal functions. Because of the oscillatory behaviour of J_ν
There will be m roots of Eq(4.3.28)for a given ν value. These roots will be designated by $\beta_{\nu m}$ and the corresponding modes are either $TE_{\nu m}$, $TM_{\nu m,}$ $EH_{\nu m}$, $HE_{\nu m}$.

For the dielectric fiber waveguide all modes are hybrid modes except those for which ν=0. when ν=0 the right –hand side of Eq(4.3.28) vanishes and two different eigen value equations result. These are

$$I_0 + H_0 = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.3.29a)$$

or,

$$\frac{J_1(ua)}{uJ_0(ua)} + \frac{K_1(wa)}{wK_0(wa)} = 0 \ldots\ldots\ldots (4.3.29b)$$

which corresponds to $TM_{\nu m,}$ modes (E_z=0), and

$$k_1^2 I_0 + k_2^2 H_0 = 0 \ldots\ldots\ldots\ldots (4.3.30a)$$

or

$$\frac{k_1^2 J_1(ua)}{uJ_0(ua)} + \frac{k_2^2 K_1(wa)}{wK_0(wa)} = 0 \ldots\ldots (4.3.30b)$$

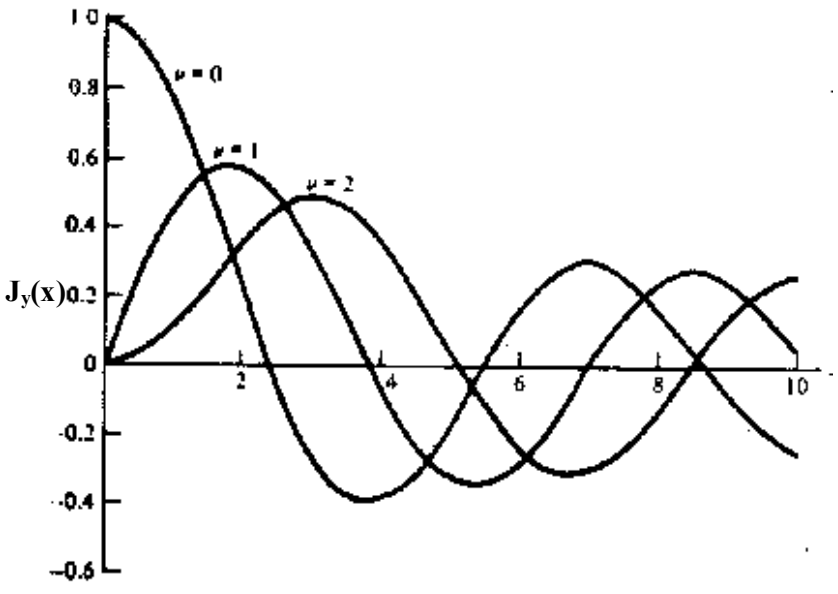


Fig: 4.3.2: Variation of the Bessel function Jx(x) for the first three orders ($_{\nu}$=1,2,3) plotted as a functions of x.

which corresponds to TM_{vm}modes (H_z=0).

When $\nu \neq o$ the situation is more complex and numerical methods are needed to solve Eq(4.3.28) exactly. However, simplified and highly accurate approximations based on the principle that the core and cladding refractive indices are nearly the same have been derived by Synder and Gloge. The condition that n_1-n_2<<1 was referred to by Gloge as giving rise to weakly guided modes.

Let us next examine the cutoff conditions for fiber modes. As was mentioned in relation to Eq(4.3.20), a mode is referred to as being cutoff when it is no longer bound to the core of the fiber so that its field no longer decays on the outside of the core. The cutoffs for the various modes are found by solving Eq(4.3.28) in the limit. This is, in general fairly complex so that only the results, which are listed in Table4.3.1, is given here.

Table4.3.1. Cutoff conditions for some lower order modes

ν	Modes	Cutoff condition
0	TE_{vm} TM_{vm}	$J_\nu(ua) = 0$
1	HE_{vm} EM_{vm}	$J_\nu(ua) = 0$
≥ 2	EH_{vm}	$J_\nu(ua) = 0$
	HE_{vm}	$[(n_1^2/n_2^2)+1]J_{\nu-1}(ua) = J_\nu(ua)\,[ua/\nu-1]$

An important parameter connected with the cutoff condition is the normalized frequency V (also called the V-number or V-parameter) defined by

$$V^2 = (u^2 + w^2)a^2 = \left(\frac{2\pi a}{\lambda}\right)^2 (n_1^2 - n_2^2) \quad \text{.....(4.3.31)}$$

Which is a dimensionless number that determinates how many modes a fiber can support. The number of modes that can exists in a waveguide as a function of V may be conveniently represented interms of a normalized propagation constant b defined by

$$b = \frac{a^2 w^2}{V^2} = \frac{(\beta/k)^2 - n_2^2}{n_1^2 - n_2^2}$$

A plot of b as a function of V is shown in Fig.4.3.3. for a few of the low-order modes. This figure shows that each mode can exist only for values of V that exceed a certain limiting value. The modes are cut off when $\beta/k = n_2$. The HE_{11}mode has no cutoff and ceases to exist only when the core diameter is zero. This is the principle on which the single-mode fiber is based. By appropriately choosing a, n_1, and n_2 so that

$$V=\frac{2\pi a}{\lambda}\left(n_1^2-n_2^2\right)^{1/2}\geq 2.405 \text{ ……..(4.3.32)}$$

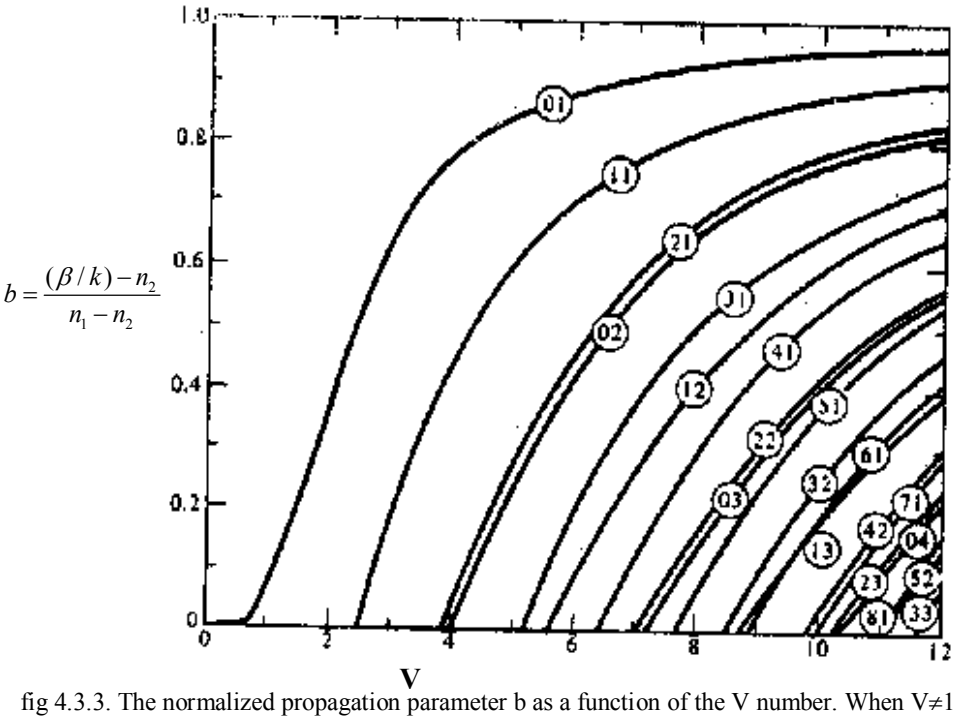


fig 4.3.3. The normalized propagation parameter b as a function of the V number. When V≠1. The curve numbers vm designate $HE_{v+1,m}$ and $EH_{v-1,m}$ modes. For v=1, the curve numbers vm give the HE_{2m}, TE_{0m}, TM_{0m} modes.

Which is the value at which the lowest –order Bessel function J_0 is zero (see Fig.4.3.2), all modes except the HE_{11}mode are cut-off.

Single –mode fibers are constructed by letting the dimension of the core diameter be a few wavelength (usually 8 to 12) and by having small index differences between the core and the cladding. From Eq(4.3.32)with V=2.4 , it can be seen that single –mode propagation is possible for fairly large variations in values of the physical core sizes a and the core-cladding index differences Δ. However, in practical design of single mode fibers, the core –cladding index difference varies between 0.1 and0.2 percent, and the core diameter should be chosen to be just below the cutoff of the first higher-mode; that is, for V slightly less than 2.4. for example , a typical single mode fiber may have a core radius of 3 μm and a numerical aperture of 0.1 at a wavelength of 0.8μ m. From Eqs.(4.2.15) and (4.3-31) this yields V=2.356.

The parameter V can also be related to the number of modes M in a multimode fiber when M is large. An approximate relationship for step-index fibers can be derived from ray theory. A

ray congruence incident on the end of a fiber will be accepted by the fiber if it lies within an angle defined by the numerical aperture as given in Eq.(4.2-15)

$$\mathrm{NA} = \sin\theta = \left(\mathrm{n}_1^2 - n_2^2\right)^{1/2} \ldots\ldots\ldots\ldots\ldots\ldots (4.3.33)$$

For practical numerical apertures sin θ is small so that sin θ ≈ θ. the solid acceptance angle for the fiber is therefore

$$\Omega = \pi\theta^2 = \pi\left(n_1^2 - n_2^2\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.3.34)$$

For electromagnetic radiation of wavelength λ emanating from a laser or a waveguide the number of modes per unit solid angle is given by $2A/\lambda^2$, where A is the area the mode is leaving of entering. The area A in this case is the core cross section. The factor 2 comes from the fact that the plane wave can have two polarization orientations. The total number of modes M entering the fiber is thus given by

$$\mathrm{M} = \frac{2\mathrm{A}}{\lambda^2}\Omega = \frac{2\pi^2 a^2}{\lambda^2}\left(n_1^2 - n_2^2\right) = \frac{\mathrm{V}^2}{2} \ldots (4.3.35)$$

4.3.6. Power Flow in Step-Index Fibers:

A final quantity of interest for step-index fibers is the fractional power flow in the core and cladding for a given mode. As is illustrated in Fig 4.2-13,the electromagnetic field for a given mode does not go to zero at the core – cladding interface, but changes from an oscillating form in the core to an exponential decay in the cladding. Thus the electromagnetic energy of a guided mode is carried partially in the core and partially in the cladding. The further away a mode is from its cutoff frequency the more concentrated its energy is in the core. A s cutoff is approached, energy travels in the cladding. At cutoff the field no longer decays outside the core and the mode now becomes a fully radiating mode.

The relative amounts of power flowing in the core and the cladding can be obtained by integrating the pointing vector in the axial direction.

$$S_2 = \frac{1}{2}\mathrm{Re}\left(\mathrm{E}\times\mathrm{H}^*\right).\mathrm{e}_2 \ldots\ldots\ldots\ldots\ldots \qquad (4.3.36)$$

over the fiber cross-section. Thus the power in the core and cladding, respectively, is given by

$$P_{core} = \frac{1}{2}\int_0^a\int_0^{2\pi} r\left(E_x H_y^* - E_y H_x^*\right) d\phi dr \ldots\ldots \ldots (4.3.37)$$

$$P_{clad} = \frac{1}{2}\int_a^x\int_0^{2\pi} r\left(E_x H_y^* - E_y H_x^*\right) d\phi dr \ldots\ldots \ldots (4.3.38)$$

Where the star denotes the complex conjugate. Gloge has shown that, based on the weakly guided mode approximation which has an accuracy on the order of the index difference Δ between the core and cladding. The relative core and cladding powers for a particular mode ν is given by

$$\frac{P_{core}}{P} = \left(1 - \frac{u^2}{V^2}\right)\left[1 - \frac{J_y^2(ua)}{J_{v+1}(ua)J_{v-1}(ua)}\right] \quad \text{.....(4.3.39)}$$

and $$\frac{P_{clad}}{P} = 1 - \frac{P_{core}}{P} \quad \text{.....(4.3.40)}$$

Where P is the power ν. The relationship between P_{core} and P_{clad} are plotted in Fig 4.3.4. in terms of the fractional powers P_{core}/P and P_{clad}/P. In addition, far from cutoff the average total power in the cladding has been derives for fibers in which many modes can propagate. Because of this large number of modes, those few modes that are appreciably close to cutoff can be ignored to a reasonable approximation. The derivation assumes an incoherent source, such as a tungsten filament lamp or a light–emitting diode, which , in general , excites every fiber mode with the same amount of power. The total average cladding power is thus approximated by

$$\left(\frac{P_{clad}}{P}\right)_{total} = \frac{4}{3}M^{-1/2} \quad \text{.....(4.3.41)}$$

From fig. 4.3.4. and Eq (4.3.41) it can be seen that, since M is proportional to V^2. The power flow in the cladding decreases as V increases.

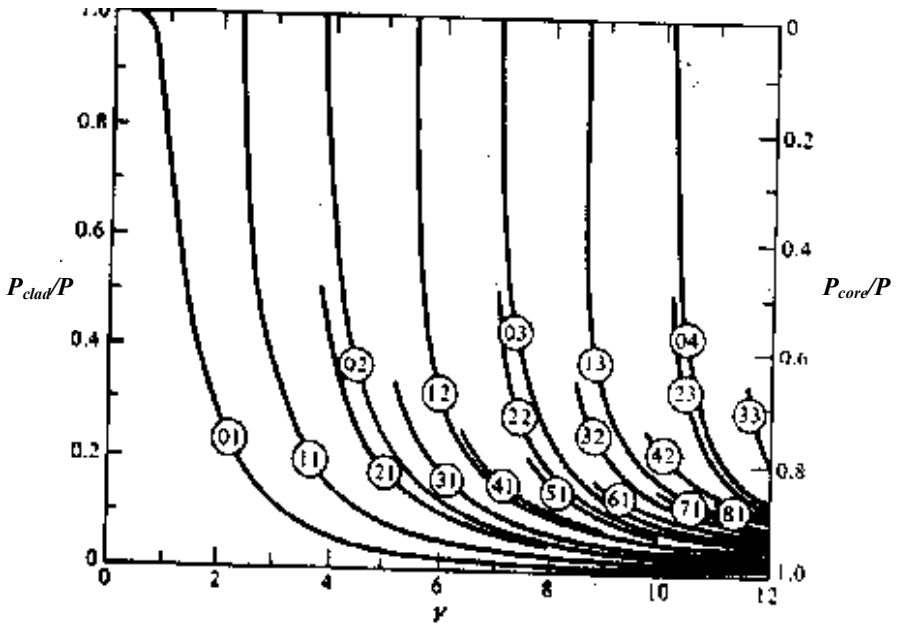


Fig 4.3.4. Fractional power flow in the cladding of a step-index optical fiber as a function of the V number. When V≠1. The curve numbers vm designate $HE_{v+1,m}$ and $EH_{v-1,m}$ modes. For v=1, the curve numbers vm give the HE_{2m}, TE_{0m}, TM_{0m} modes.

As an example, consider a fiber having a core radius of 25 , a core index of 1.48,and Δ=0.01. At an operating wavelength 0.84 the value of V is 39 and there are 760 modes in the

fiber .From Eq(4.3.41) approximately 5 percent of the power propagates in the cladding. If Δ is decreased to, say 0.003, in order to decrease signal dispersion, then 242 modes propagate in the fiber and about 9 percent of the power resides in the cladding. For the case of the single-mode fiber ,considering the HE_{11} mode in fig.4.3.4, it is seen that for V=1 about 70 percent of the propagates in the cladding , whereas for V=2.405, which is where the TE01 mode begins, approximately 84 percent of the power is now within the core.

4.3.7. GRADED-INDEX FIBER STRUCTURE:

In the graded-index fiber design the core refractive index decreases continuously with radial distance r from the center of the fiber but is generally constant in the cladding. The most commonly used construction for the refractive index variation in the core is power law relationship

$$n(r) = \begin{cases} n_1[1-2\Delta(\frac{r}{a})^{\alpha}]^{1/2} & for 0 \le r \le a \\ n_1(1-2\Delta)^{1/2} \approx n_1(1-\Delta) = n_2 & for\ r \ge a \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots (4.3.42)$$

Here r is the radial distance from the fiber axis, a is the core radius, n_1 is the refractive index at the core axis n_2 is the refractive index of the cladding, and the dimensionless parameter α defines the shape of the index profile. The index difference Δ for the graded-index is given by

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \approx \frac{n_1 - n_2}{n_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.3.43)$$

The approximation on the right hand side of this equation reduces the expression for Δ to that of the step-index fiber given by eq(4.2-12). Thus the same symbol is used in both cases. For $\alpha=\infty$, eq(4.3.42) reduces to the step-index profile $n(r)=n_1$.

4.3.7.1. GRADED INDEX NUMERICAL APERTURE (NA):

The determination of the NA for graded – index fibers is more complex than for step-index fibers. In graded –index fibers the NA is a function of position across the core end face. This is in contrast to the step-index fiber where the NA is constant across the core. Geometrical optics considerations show that light incident on the fiber core at position r will propagate as a guided mode only if it is within the local numerical aperture NA(r) at that point. The local numerical aperture is defined as

$$NA(r) = \begin{cases} [n^2(r) - n_2^2]^{1/2} \approx NA(0)\sqrt{1-\left(r/a\right)^{\alpha}} & for\ r \le a \\ 0 \quad for\ \ r > a \end{cases} \quad \ldots\ldots\ldots (4.3.44)$$

Where the axial numerical aperture is defined as

$$NA(0) = \left[n^2(0) - n_2^2 \right]^{1/2} = \left(n_1^2 - n_2^2 \right)^{1/2} \approx n_1 \sqrt{2\Delta} \quad \text{.....(4.3.45)}$$

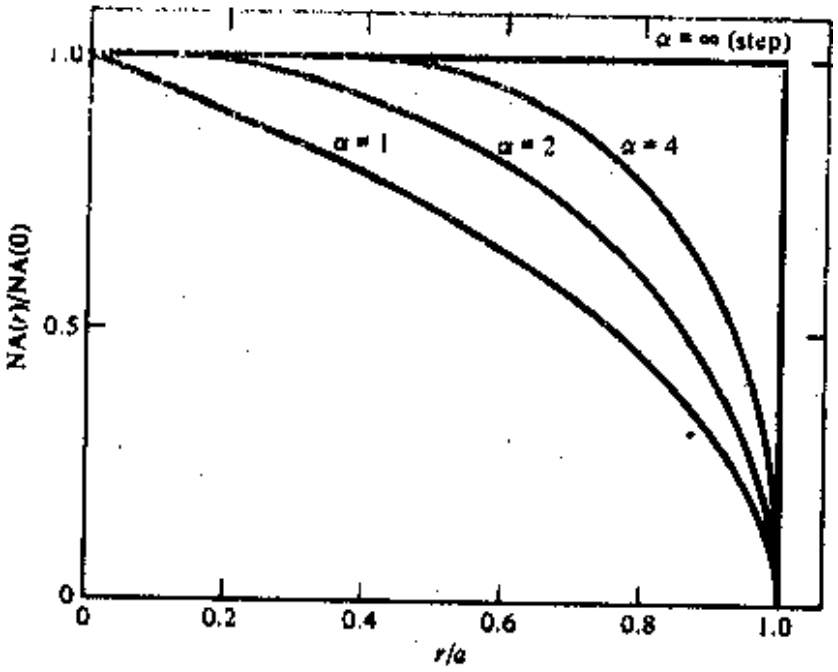


Fig..4. 3.5. A comparison of the numerical aperture for fibers having various α profiles.

4.3.7.2. MODES IN GRADED-INDEX FIBERS

A modal analysis of an optical fiber based on solving Maxwell's equations can only be carried out rigorously if the core refractive index is uniform, that is, for step-index fibers. In other cases, such as for graded-index fibers, approximation methods are needed. The most widely used analysis of modes in a graded-index fiber is an approximation cased on the WKB method (named after Wenzel, Kramers, and Brilliouin) which is the commonly used in Quantum Mechanics. The purpose of the WKB method is to obtain an asymptotic representation for the solution of a differential equation containing a parameter that varies slowly over the desired range of the equation. That parameter in this case in the refractive-index profile n® which varies only slightly over distance on the order of the optical wavelength.

Analogous to the step- index fiber, Eq.(4.3.15) for the radial component of the wave equation must be solved.

$$\frac{d^2F_1}{dr^2} + \frac{1}{r}\frac{dF_1}{dr} + \left[k^2 n^2(r) - \beta^2 - \frac{\nu^2}{r^2} \right] F_1 = 0 \quad \text{.....(4.3.46)}$$

Where n (r) is given by Eq. (4.3.42). The general procedure in the WKB method is to let

$$F_1 = Ae^{jkQ(r)} \text{ ……..(4.3.47)}$$

where the coefficient A is independent of r. Substituting this into Eq(4.2.64) gives

$$jkQ^{11} - \left(kQ^1\right)^2 + \frac{jk}{r}Q^1 + \left[k^2n^2(r) - \beta^2 - \frac{\nu^2}{r^2}\right] \text{ ……..(4.3.48)}$$

where the primes denote differentiation with respect to r. Since n (r) varies slowly over a distance on the order of a wavelength, an expansion of the function Q (r) in powers of λ or, equivalently, in powers of $k^{-1} = \lambda/2\pi$ is expected to converge rapidly. Thus we let

$$Q(r) = Q_0 + \frac{1}{k}Q_1 + \text{.....} \text{ ……(4.3.49)}$$

Where Q_o, Q_1,…….. are certain functions of r. Substituting Eq. (4.3.49) into Eq. (4.3-48) and collecting equal powers of k yield

$$\left\{-k^2\left(Q_0^1\right)^2 + \left[k^2n^2(r) - \beta^2 - \frac{\nu^2}{r^2}\right]\right\} + \left(jkQ_0^{11} - 2kQ_0^1Q_1^1 + \frac{jk}{r}Q_0^1\right) + termsoforder(k^0, k^{-1}, k^{-2}, ...) = 0$$

……..(4.3.50)

A sequence of defining relations for the Q_1 functions are obtained by setting to zero the terms in equal powers of k. Thus, for the first two terms of Eq. (4.3.50),

$$-k^2\left(Q_0^1\right)^2 + \left[k^2n^2(r) - \beta^2 - \frac{\nu^2}{r^2}\right] = 0 \text{ ……(4.3.51)}$$

$$jkQ_0^{11} - 2kQ_0^1Q_1^1 + \frac{jk}{r}Q_0^1 = 0 \text{ …….. .(4.3.52)}$$

Integration of Eq. (4.2-69) yields

$$kQ_0 = \int_{r_1}^{r_2}\left[k^2n^2(r) - \beta^2 - \frac{\nu^2}{r^2}\right]^{1/2} dr \text{ ……..} \qquad (4.3.53)$$

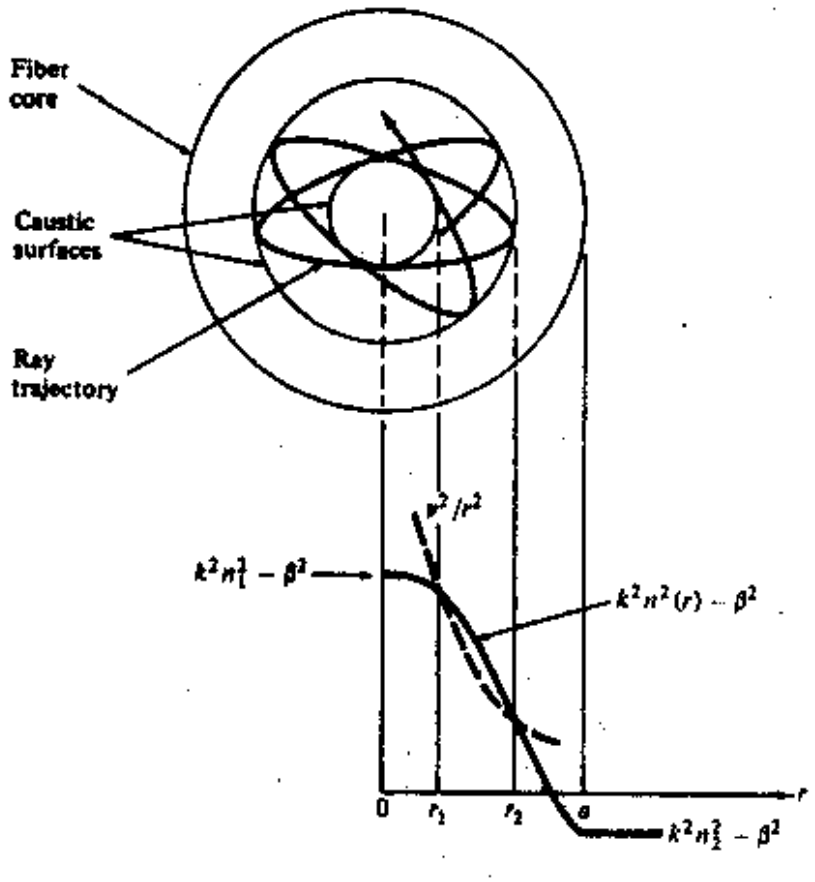


Fig.4.3.6. Cross-sectional projection of a skew ray in a graded index fiber and the graphical representation of its mode solution from the WKB method. The field is oscillatory between the turning points r_1 and r_2 and is evanescent outside of this region.

A mode is bound in the fiber core only if Q_o is real. For Q_o to be real, the radical in the integrand must be greater than zero. In general, for a given mode v, there are two values r_1 and r_2 for which the radical is zero as is indicated by the limits of integration in Eq. (4.3.52). Note that these values of r are functions of v. Guided modes exist for r between these two values. For other values of r the function Q_o is imaginary, which leads to decaying fields.

To help visualize the solutions to Eq. (4.3.52), consider the cross-sectional projection of a skew ray in a graded-index fiber shown in Fig..4.3.6. The path followed by the ray lies completely within the boundaries of two coaxial cylindrical surfaces, known as the caustic surfaces, that have inner and outer radii r_1 and r_2, respectively. The radii r_1 and r_2 are those points at which the radical in the integrand of Eq. (4.3.52.) becomes zero. They are called turning points since the ray turns from increasing to decreasing values of r or vice versa. To evaluate the turning points, consider the functions $k^2n^2(r)-\beta^2$ and ν^2/r^2 plotted in Fig.4.3.6. as solid and dashed curves, respectively. The crossing points of these two curves give the points r_1 and r_2. An oscillating field exists when the solid curve lies above the dashed curve, which indicates bound mode solutions.

Evanescent (nonbound decaying mode) fields occur when the solid curve lies below the dashed curve.

To form a bound mode of the graded-index fiber, each wave associated with the ray congruence corresponding to this mode must interfere constructively with itself in such a way as to form a standing-wave pattern in the radial cross-sectional direction. This requirement imposes the condition that the phase function Q_o between r_1 and r_2 must be a multiple of π (that is, an integer number of half-periods), so that

$$m\pi = \int_{r_1}^{r_2} \left[k^2 n^2(r) - \beta^2 - \frac{\nu^2}{r^2} \right]^{1/2} dr \quad \text{.....(4.3.53)}$$

Where m=0,1,2,……. is the radial mode number that counts the number of half-periods between the turning points. The total number of bound modes m(β) can be found by summing Eq. (4.3.53) over all v from o to v max, where v_{max} is the highest-order bound mode for a given value of β. If v_{max} is a large number, the sum can be replaced by an integral yielding

$$m(\beta) = \frac{4}{\pi} \int_0^{r_{max}} \int_{r_1(\nu)}^{r_2(\nu)} \left[k^2 n^2(r) - \beta^2 - \frac{\nu^2}{r^2} \right]^{1/2} dr d\nu \quad \text{.....(4.3.54)}$$

The factor 4 arises from the fact that each combination (m,v) designates a degenerate group of four modes of different polarization or orientation. If we change the order of integration, the lower limit on r must be r_1=0 in order to count all the modes, and the upper limit on v is found from the condition

$$\left[k^2 n^2(r) - \beta^2 - \frac{\nu_{max}^2}{r^2} \right] = 0 \quad \text{.....(4.3.55)}$$

$$\text{thus } m(\beta) = \frac{4}{\pi} \int_0^{r_2} \int_0^{r_{max}} \left[k^2 n^2(r) - \beta^2 - \frac{\nu^2}{r^2} \right]^{1/2} d\nu dr \quad \text{.....(4.3.56)}$$

Evaluating the integral over v with v_{max} given by Eq. (4.3-55) yields

$$m(\beta) = \int_0^{r_2} [k^2 n^2(r) - \beta^2] r dr \quad \text{.....(4.3.57)}$$

To evaluate this further, we choose the index profile n(r) given by Eq. (4.3.42). The upper limit of integration r_2 is determined from the condition that

$$kn(r) = \beta$$

Combining this condition with Eq. (4.3.42) gives

$$r_2 = a \left[\frac{1}{2\Delta} \left(1 - \frac{\beta^2}{k^2 n_1^2} \right) \right]^{1/a} \quad \text{.....(4.3.58)}$$

Using Eqs. (4.3.42) and (4.3.58), the number of modes is

$$m(\beta) = a^2 k^2 n_1^2 \Delta \frac{\alpha}{\alpha+2}\left(\frac{\left(k^2 n_1^2 - \beta^2\right)}{2\Delta k^2 n_1^2}\right)^{2+\alpha/\alpha} \quad \text{.....(4.3.59)}$$

All bound modes in a fiber must have $\beta \geq kn_2$. If this condition does not hold, the mode is no longer perfectly trapped inside the core and loses power by leakage into the cladding. The maximum number of bound modes M is thus found by letting

$$\beta = kn_2 = kn_1(1-\Delta)$$

Where Eq. (4.3.42) was used for the relationship between n_1 and n_2. Thus

$$M = m(kn_2) = \frac{\alpha}{\alpha+2} a^2 k^2 n_1^2 \Delta \text{ .....(4.3.60)}$$

gives the total number of bound modes in a graded-index fiber having a refractive-index profile given by Eq. (4.3.51).

4.3.8. SUMMARY: Solving Maxwell's equations for a dielectric medium subject to the boundary conditions of the step and graded index optical fibers, the propagation of waves are understood. The boundary conditions at the core-cladding interface lead to a coupling between the longitudinal components of the E and H field. This coupling leads to involve hybrid mode solutions.

4.3.9. Keywords: Maxwell's equations, boundary conditions, Numerical aperture, modes,

4.3.10. Self assessment questions

1. Discuss the wave equations for step index fibers and hence derive mode equations. Describe modes there in.
2. Write notes on power flow in step index fibers
3. Discuss the wave equations for graded index fibers and hence derive mode equations. Describe modes there in.
4. Define numerical aperture in step and graded index fibers. Explain the importance of it.

4.3.11. Reference Books

1. Optical fiber communications by G. Keiser, McGraw-Hill International Edition, 2000, Third edition, and also see first edition.
2. Optical fiber communications by Kato, McGraw-Hill, 1986
3. Inhomogeneous optical waveguides by M.S. Sodha and A.K. Ghatak, Plenum Press, 1977

Unit IV
Lesson 4

SIGNAL ATTENUATION IN OPTICAL FIBERS

Objective: To know about signal attenuation while it is propagating in optical fibers, different attenuation mechanisms operating in the optical fibers.

Structure

4.4. Introduction

In lessons 2 and 3 we showed the structure of optical fibers and examined the concept of how light propagates along a cylindrical dielectric optical waveguide. Here we shall continue the discussion of optical fibers by answering two very important questions:

1. What are the loss or signal attenuation mechanisms in a fiber?
2. Why and to what degree do optical signals get distorted as they propagate along a fiber?

Signal attenuation (also known as fiber loss or signal loss) is one of the most important properties of an optical fiber, because it largely determines the maximum repeater less separation between a transmitter and a receiver. Since repeaters are expensive to fabricate, install, and maintain, the degree of attenuation in a fiber has a large influence on system cost. Of equal importance is signal distortion. The distortion mechanisms in a fiber cause optical signal pulses to broaden as they travel along a fiber. If these pulses travel sufficiently far, they will eventually overlap with neighboring pulses, thereby creating errors in the receiver output. The signal distortion mechanisms thus limit the information carrying capacity of a fiber.

Since these two factors are closely tied to how and of what a fiber is constructed, we shall first discuss fiber materials and construction methods. This will be in the form of a very brief overview that defines the terminology and gives the necessary background concepts. A more detailed treatment of fiber materials and fabrication methods is given in chap.10. Here we shall mainly concentrate on low loss glass fibers that are suitable for long distance information transmission.

4.4.1 Fiber Materials and Fabrication Methods: An Overview

In selecting materials for optical fibers a number of requirements must be satisfied. For example:

1. It must be possible to make long, thin, flexible fibers from the materials.
2. The material must be transparent at a particular optical wavelength in order to the fiber to guide light efficiently.
3. Physically compatible materials having slightly different refractive indices for the core and cladding must be available.

Materials satisfying these requirements are glasses and plastics.

The largest categories of optically transparent glasses from which optical fibers are made are the oxide glasses. Of these the most common is silica SiO_2, which has a refractive index of 1.458 at 850 nm. To produce two similar materials having slightly different indices of refraction for the core and cladding, trace amounts of either fluorine or various oxides (referred to as dopants) such as B_2O_3 ,GeO_2 ,or P_2O_5 are added to the silica . As shown in Figure . 4.4.1. the addition of GeO_2 or P_2O_5 increases the refractive index ,where as doping the silica with fluorine or B_2O_3 decreases it. When referring to a doped silicon glass, notations such as GeO_2-SiO_2 are used, for example, to denote a GeO_2 dopant.

Two basic techniques are used in the fabrication of all-glass optical wave-guides. These are the vapor phase oxidation processes and the direct melt methods. The direct melt method follows traditional glass-making procedures in that optical fibers are made directly from the molten state of purified components of silicate glasses. In the vapor phase oxidation processes, highly pure vapors of metal halides (for example, $SiCl_4$ and $GeCl_4$) react with oxygen to form a white powder of SiO_2 particles. The particles are then collected on the surface of a bulk glass by one of three commonly used processes and are then sintered

(transformed to a homogeneous glass mass by heating without melting) by a variety of techniques to form a clear glass rod or tube (depending on the process).

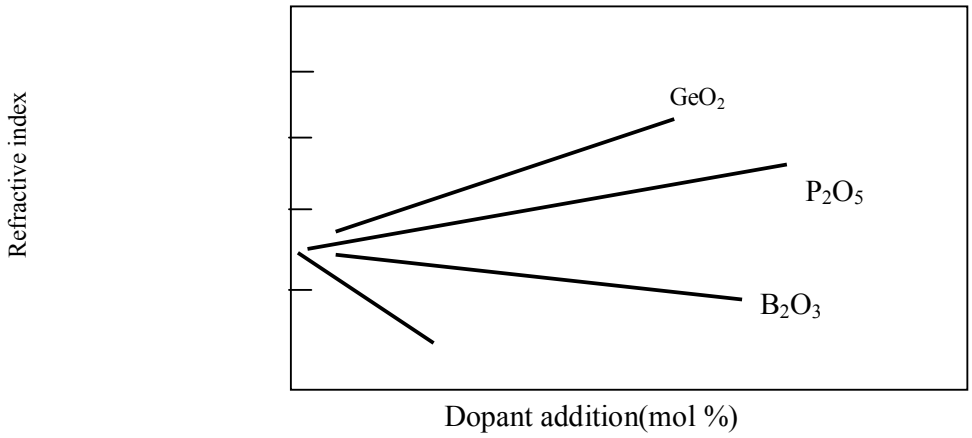


Fig 4.4.1: Variation in refractive index as a function of doping concentration in silica glass.

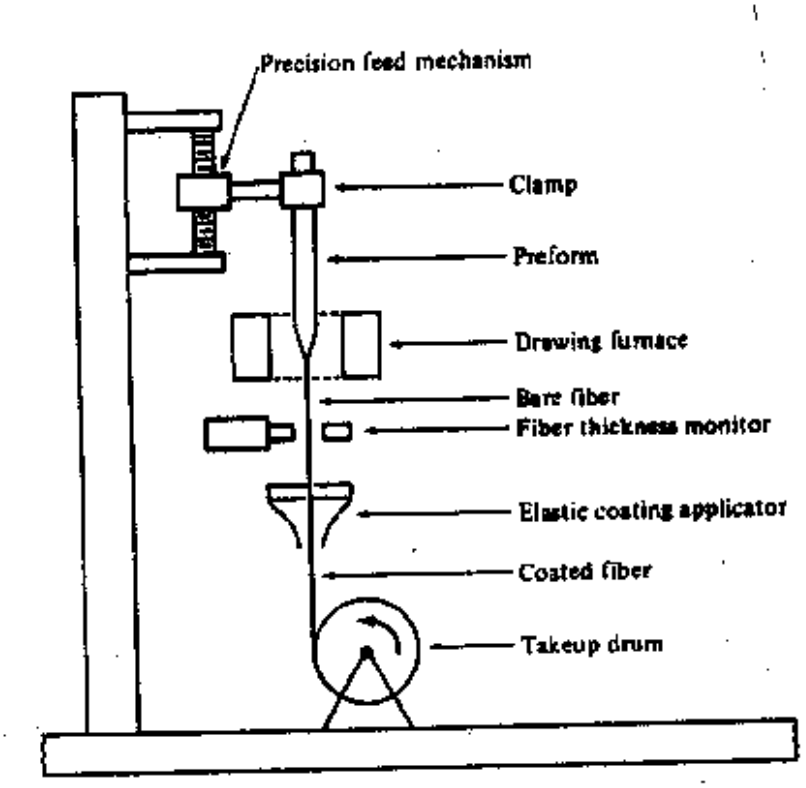


Fig 4.4.2: Schematic of fiber drawing apparatus.

This rod or tube is a called a preform. It is typically around 10mm in diameter and 60 to 90 cm long. Fibers are made from the preform by using the equipment shown in fig 4.4.2. The preform is precision fed into a circular heater called the drawing furnace. Here the perform end is softened to the point where it can be drawn into a very thin filament which becomes the optical fiber.

4.4.2 Attenuation

Attenuation of a light signal as it propagates along a fiber is an important consideration in the design of an optical communication system, since it plays a major role in determination the maximum transmission distance between a transmitter and a receiver. The basic attenuation mechanisms in a fiber are absorption, scattering and radiative losses of the optical energy absroption is related to the fiber material, where as scattering is associated both with the fiber material and with structural imperfections in the optical waveguide. Attenuation owing to radiative effects originates from perturbation (both microscopic and macroscopic) of the fiber geometry.

We shall discuss the units in which fiber losses are measured and then present the physical phenomenon giving rise to attenuation.

4.4.2.1 Attenuation Units

Signal attenuation (or fiber loss) is defined as the ratio of the optical output power P_{out} from a fiber length L to the optical input power P_{in}. This power ratio is a function of wavelength, as is shown by the general attenuation curve in fig 4.4.3.

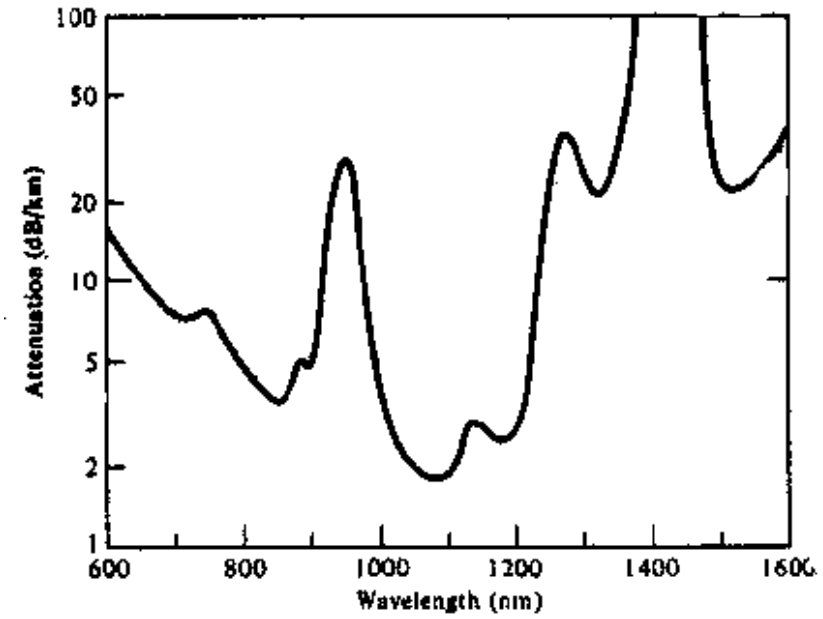


fig 4.4.3: Attenuation Vs wavelength curve of a typical early technology fiber having a high water impurity content.

The symbol α is commonly used to express attenuation in decibels per kilometer

$$\alpha = 10 \log \left[\frac{(P_{in} / P_{out})}{L} \right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4.4.1)$$

As ideal fiber would have no loss so that Pout =Pin. This corresponds to a 0-dB attenuation, which in practice is impossible. An actual low loss fiber may have a 3-dB/km average loss, for example. This means that the optical signal power would decrease by 50% over a 1-km length and would decrease by 75% (a 6-dB loss) over a 2-km length since loss contributions expressed in decibels are additive.

4.4.2.2. Absorption

Absorption is caused by three different mechanisms:

1. Absorption by atomic defects in the glass composition.
2. Extrinsic absorption by impurity atoms in the glass materials.
3. Intrinsic absorption by the basic constituent atoms of the fiber material.

Atomic defects are imperfections of the atomic structure of the fiber material such as missing molecules, high density clusters of atom groups, or oxygen defects in the glass structure. Usually absorption losses arising from these defects are negligible compared to intrinsic and impurity absorption effects. However, they can be significant if the fiber is exposed to intense nuclear radiation levels, as might occur inside a nuclear reactor, during a nuclear explosion, or in the earths Van Allen belts.

The dominant absorption factor in fibers prepaid by the direct melt method is the presence of impurities in the fiber material. Impurity absorption results predominantly from transition metal ions such as iron, chromium, cobalt, and copper, and from OH (water) ions. The transition metal impurities which are present in the starting materials used for direct melt fibers range between 1and 10 parts per billion (ppb) causing losses from 1 to 10 dB/km. The impurity levels in vapor phase deposition processes are usually one to two orders of magnitude lower. Impurity absorption losses occur either from electronic transitions between the energy levels associated with the incompletely filled inner sub-shell of these ions or because of charge transitions from one ion to another. The absorption peaks of the various transition metal impurities tend to be broad, and several peaks may overlap, which further broadens the absorption region.

The presence of OH ion impurities in fibers performs results mainly from the oxyhydrogen flame used for the hydrolysis reaction of the $SiCl_4$, $GeCl_4$, and $POCl_3$ starting materials. Water impurity concentrations of less than a few parts per billion are required if the attenuation is to be less than 20 dB /km. Early optical fibers had high levels of OH ions which resulted in large absorption peaks occurring at 1400, 950, and 725 nm. These are the first, second, and third overtones, respectively, of the fundamental absorption peak of water near 2.7 μm, as shown in fig 4.4.3. Between these absorption peaks there are regions of low attenuation.

The peaks and valleys in the attenuation curve resulted in the assignment of various transmission windows to early optical fibers. Significant progress has been made in reducing

the residual OH content of fibers to less than 1 ppb. For example, the loss curve of a fiber prepared by the VAD method with an OH content of less than 0.8 ppb is shown in fig 4.4.4.

Fig. 4.4.4. Attenuation versus wavelength curve of a VAD fiber with very low OH

Intrinsic absorption is associated with the basic fiber material (for example, pure SiO_2) and is the principle physical factor that defines the transparency window of a material over a specified spectral region. It occurs when the material is in a perfect state with no density variations, impurities, material inhomogeneties, etc. Intrinsic absorption thus sets the fundamental lower limit on absorption for any particular material.

Intrinsic absorption results from electronic absorption bands in the ultraviolet region and from atomic vibration bands in the near infrared region. The electronic absorption bands are associated with the band gaps of the amorphous glass materials. Absorption occurs when a photon interacts with an electron in the valence band and excites it to a higher level, as is

described in sec. 4.2.1. The ultraviolet edge of the electron absorption bands of both amorphous or crystalline materials follow the empirical relationship

$$\alpha_{uv} = Ce^{E/E0} \text{ ---------- -(4.4.2)}$$

Which is known as Urbach's rule. Here C and E_o are empirical constants and E is the photon energy. The magnitude and characteristic exponential decay of the ultraviolet absorption are shown in fig 4.4.5.

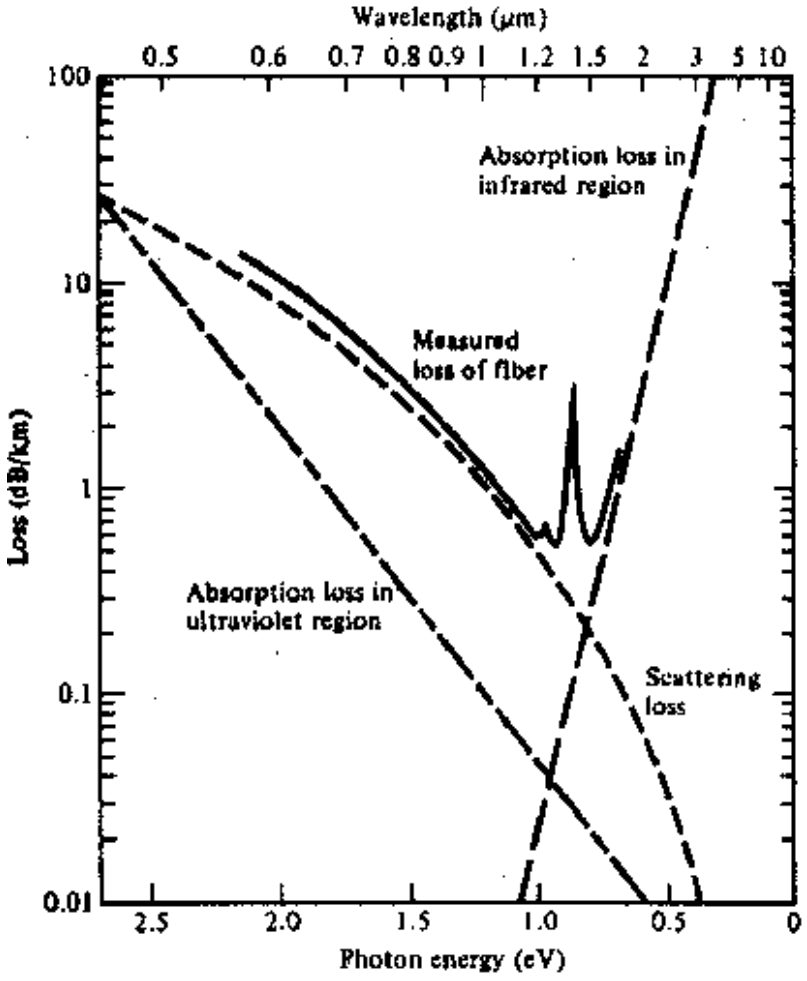


Fig 4.4.5 Optical fiber attenuation characteristics and their limiting mechanisms for a GeO_2-doped low-loss low-OH content fiber.

In the near infrared region above 1.2 μm, the optical wave-guide loss is predominantly determined by the presence of OH ions and the inherent infrared absorption of the constituent material. The inherent infrared absorption is associated with the characteristic vibration frequency of the particular chemical bond between the atoms of which the fiber is composed. An interaction between the vibrating bond and the electromagnetic field of the optical signal results in a transfer of energy from the field to the bond, thereby giving rise to absorption. This absorption is quite strong because of the many bond present in the fiber.

These mechanisms result in a wedge shaped spectral loss characteristic. Within these wedge losses as low as 0.2 dB/km at 1.55 μm in a single mode fiber have been measured. A comparison of the infrared absorption induced by various doping materials in low water content fibers is shown in fig. 4.4.6. This indicates that for operation at longer wavelengths a GeO_2 - doped fiber material is the most desirable. Note that the absorption curve in fig 4.4.5 is for a GeO_2-doped fiber.

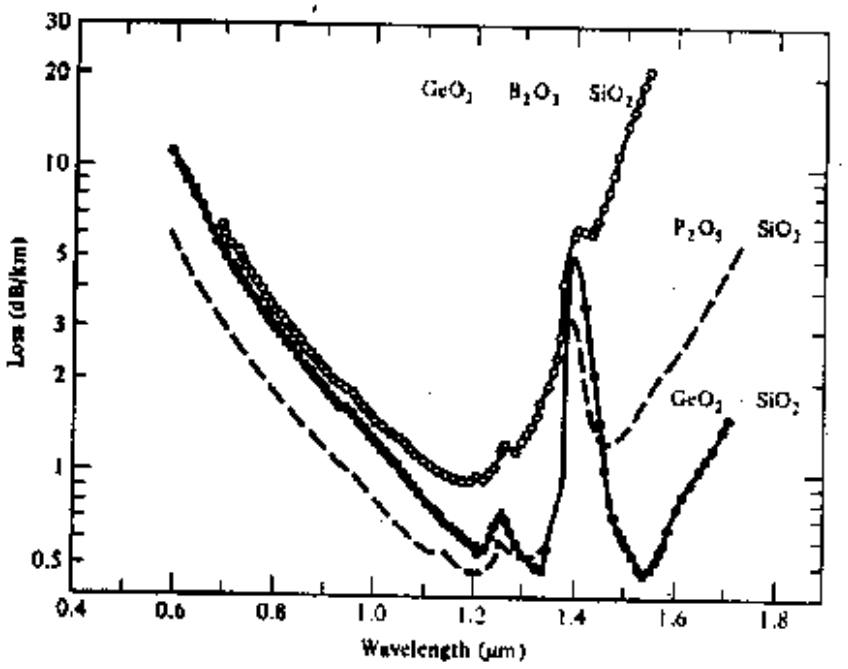


Fig 4.4.6 A comparison of the infrared absorption induced by various doping materials in low-loss fibers.

4.4.2.3 Scattering Losses:

Scattering losses in glass arise from microscopic variations in the material density, from compositional fluctuations, and from structural inhomogeneities or defects occurring during fiber manufacture. Glass is composed of a randomly connected network of molecules. Such a structure naturally contains regions in which the molecular density is either higher or lower than the average density in the glass. In addition since glass is made up of several oxides, such as SiO_2,GeO_2, and P_2O_5, compositional fluctuations can occur. These two effects give rise to refractive index variations, which occur within the glass over distances that are small compared to the wavelength. These index variations cause a Rayleigh- type scattering of the light. Rayleigh scattering in glass is the same phenomenon that scatters light from the sun in the atmosphere, thereby giving rise to a blue sky.

The expressions for scattering induced attenuation are fairly complex owing to the random molecular nature and the various oxide constituents of glass. For single component glass the scattering loss at a wavelength λ resulting from density fluctuations can be approximated by (in base e units)

$$\alpha_{scat} = \left(\frac{8\pi^3}{3\lambda^4}\right)(n^2 - 1)^2 k_B T_f \beta_T \text{ ----------4.4.3}$$

here n is the refractive index, k_B is Boltzmann constant, β_τ is the isothermal compressibility of the material , and the fictive temperature T_f is the temperature at which the density fluctuations are frozen into the glass as it solidifies (after having been drawn into a fiber). Alternatively the relation (in base e units).

$$\alpha_{scat} = \left(\frac{8\pi^3}{3\lambda^4}\right) n^8 p^2 k_B T_f \beta_T \text{ ----------4.4.4}$$

has been derived, where p is the photoelastic coffecient. Note that Eqs. 4.4.3 and 4.4.4 are given in units of nepers (that is, base e units). To change this to decibels for optical power attenuation calculations, multiply these equations by 10 log e = 4.343.

For multi-component glasses the scattering is given by

$$\alpha = \left(\frac{8\pi^3}{3\lambda^4}\right)(\delta n^2)^2 \delta V \text{ ----------4.4.5}$$

where the square of the mean- squared refractive index fluctuation $(\delta n^2)^2$ over a volume of δV

is $$(\delta n^2)^2 = \left(\frac{\partial n}{\partial p}\right)^2 (\delta p)^2 + \sum_{i=1}^{m}\left(\frac{\partial n^2}{\partial C_i}\right)^2 (\delta C_i)^2 \text{ ---------4.4.6}$$

Here δp is the density fluctuation and δC_i is the concentration fluctuation of the ith glass component. The magnitudes of the composition and density flutuations are generally not known and must be determined from experimental scattering data. Once they are known the scattering loss can be calculated.

Structural inhomogeneties and defects created during fiber fabrication can also cause scattering of light out of the fiber. These defects may be in the form of trapped gas bubbles, unreacted starting materials, and crystallized regions in the glass. In these extrinsic effects to the point where scattering resulting from them is negligible compared to the intrinsic Rayleigh scattering.

Since Rayleigh scattering follows a characteristic λ^{-4} dependence it decreases dramatically with increasing wavelength, as shown in fig. 4.4.5. for wavelengths below about 1μm it is the dominant loss mechanism in a fiber and gives the attenuation versus wavelength plots their characteristic downward trend with increasing wavelength. At wavelengths longer than 1μm, infrared absorption effects tends to dominate optical signal attenuation.

4.4.2.4 Radiative Losses

Radiative losses occur whenever an optical fiber undergoes a bend of finite radius of curvature. Fibers can be subject to two types of bends:

(a) bends having radii that are large compared to the fiber diameter, such as occur when a fiber cable turns a corner, and

(b) random microscopic bends of the fiber axis that can arise when the fibers are incorporated into cables.

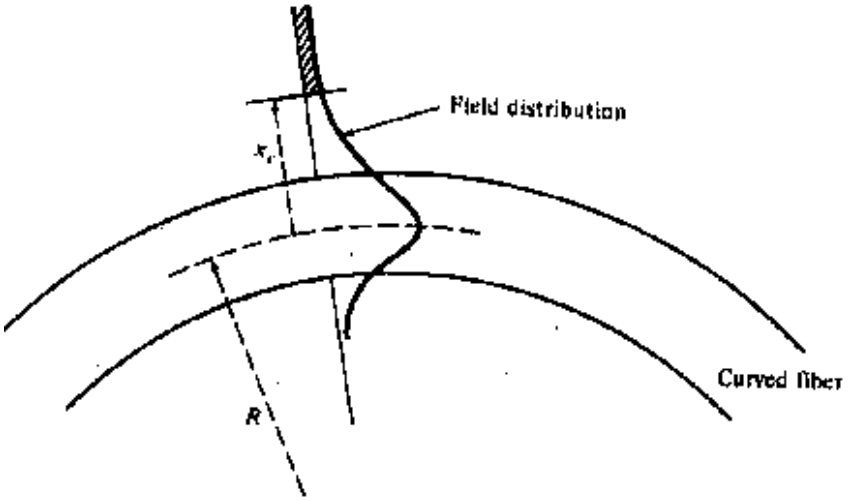


Fig 4.4-7:Sketch of the fundamental mode field in a curved optical wave-guide.

Let us first examine large curvature radiation losses. For slight bends the excess loss is extremely small and is essentially unobservable. As the radius of curvature decreases, the loss increases exponentially until at a certain critical radius the curvature loss becomes observable. If the bend radius is made a bit smaller once this threshold point has been reached, the losses suddenly become extremely large.

Qualitatively these curvature loss effects can be explained by examining the modal electric field distributions shown in fig 4.2.13. Recall that this figure shows that any bound core mode has an evanescent field tail in the cladding which decays exponentially as a function of distance from the core, part of the energy of a propagating mode travels in the fiber cladding. When a fiber is bent, the field tail on the far side of the center of curvature must move faster to keep up with the field in the core, as is shown in fig 4.4.7 for the lowest order fiber mode. At a certain critical distance x_c from the center of the fiber the field tail would have to move faster than the speed of

light to keep up with the core field. Since this is not possible the optical energy in the field tail beyond x_c radiates away.

The amounts of optical radiation from a bent fiber depends on the field strength at x_c and on the radius of curvature R. Since higher order modes are bound less tightly to the fiber core than lower-order modes, the higher-order modes will radiate out of the fiber first. Thus the total number of modes that can be supported by a curved fiber is less than in a straight fiber. Gloge has derived the following expression for the effective number of modes N_{eff} that are guided by a curved fiber of radius a:

$$N_{eff} = N_x \left\{1-\left(\frac{(\alpha+2)}{2\alpha\Delta}\right)\left[\frac{2\alpha}{R}+\left(\frac{3}{2n_2kR}\right)^{2/3}\right]\right\} \text{----------(4.4.7)}$$

Where α defines the graded index profile, Δ is the core cladding index difference, n_2 is the cladding refractive index, $k = 2\pi/\lambda$ is the wave propagation constant, and

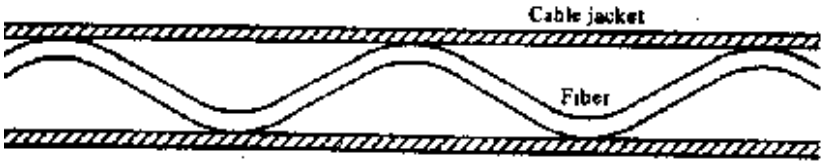


Fig-4.4.8:Microbends shown as repetitive changes in the radius of curvature of the fiber axis

$$N_\alpha = \frac{\alpha}{\alpha+2}(n_1ka)^2\,\Delta \text{----------(4.4.8)}$$

is the total number of modes in a straight fiber. As an example, let us find the radius of curvature R at which the number of modes decreases by 50% in a graded index fiber. For this fiber let $\alpha = 2$, $n_2 = 1.5$, $\Delta = 0.01$, $a = 25$ μm, and let the wavelength of the guided light be 1.3 μm. solving Eq.4.4.7. yields R =1cm.

Another form of radiation loss in optical wave-guides results from mode coupling caused by random microbends of the optical fiber. Microbends are repetitive changes in the radius of curvature of the fiber axis, as is illustrated in the fig. 4.4.8. they are caused either by non-uniformities in the sheathing of fiber or by non-uniform lateral pressures created during the cabling of the fiber. The latter effect is often referred to as cabling or packing losses. An increase

in attenuation results from microbending because the fiber curvature causes repetitive coupling of energy between the guided modes and the leaky or non-guided modes in the fiber.

One method of minimizing micro-bending losses is by extruding a compressible jacket over the fiber. When external forces are applied to this configuration, the jacket will be deformed but the fiber will tend to stay relatively straight, as shown in fig.4.4.9. for a graded index fiber having a core radius a, outer radius b (excluding the jacket), and an index difference Δ, the micro-bending loss α_M of a jacketed fiber is reduced from that of an unjacketed fiber by a factor

$$F(\alpha_M) = \left\{ \left[1 + \pi\Delta^2 \left(\frac{b}{a} \right)^4 \right] \frac{E_f}{E_j} \right\}^{-2} \quad \text{----------(4.4.9)}$$

Here E_j and E_f are the Young modulii of the jacket and fiber, respectively. The Young modulus of common jacket materials ranges from 20 to 500 Mpa. The Young modulus of fused silica glass is about 65 Gpa.

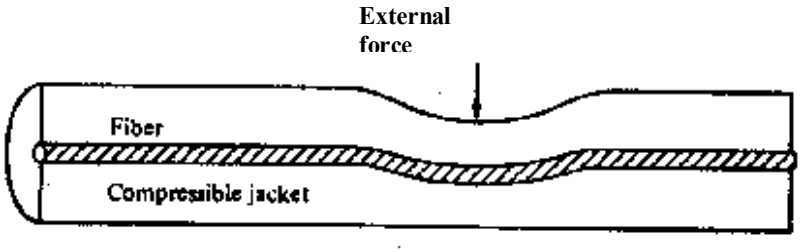


Fig 4.4.9.: A compressible jacket extruded over a fiber minimizes micro-bending resulting from external forces.

4.4.2.5 Core and cladding losses:

Upon measuring the propagation losses in an actual fiber, all the dissipative and scattering losses will be manifested simultaneously. Since the core and cladding have different indices of refraction and, therefore, differ in composition, the core and cladding generally have different attenuation co-efficients denoted α_1 and α_2, respectively. If the influence of modal coupling is ignored, the loss for a mode of order (ν, m) for a step index wave-guide is

$$\alpha_{\upsilon m} = \alpha_1 \frac{P_{core}}{P} + \alpha_2 \frac{P_{clad}}{P} \text{----------(4.4.10)}$$

where the fractional power P_{core} /P and P_{clad} /P are shown in fig 4.2-17 for several low order modes. Using Eq. 2.58, this can be written as

$$\alpha_{\upsilon m} = \alpha_1 + (\alpha_2 - \alpha_1)\frac{P_{clad}}{P} \text{ ----------(4.4.11)}$$

The total loss of the wave-guide can be found by summing over all modes weighted by the fractional power in that mode.

For the case of a graded index fiber the situation is much more complicated. In this case, both the attenuation coefficients and the modal power tend to be functions of the radial coordinate. At a distance r from the core axis the loss is

$$\alpha(r) = \alpha_1 + (\alpha_2 - \alpha_1)\frac{n^2(0) - n^2(r)}{n^2(0) - n_2^2} \text{ ----------(4.4.12)}$$

where α_1 and α_1 are the axial and cladding attenuation coefficient, respectively. And the ns are defined by Eq. 2.60. The loss encountered by a given mode is then

$$\alpha_{gl} = \frac{\int_0^\infty \alpha(r)\, p(r)\, r dr}{\int_0^\infty P(r)\, r dr} \text{ ----------(4.4.13)}$$

Where p(r) is the power density of that mode at r. the complexity of the multimode wave-guide has prevented an experimental correlation with a model. However, it has generally been observed that the loss increase mode number.

4.4.3 Summary: In this lesson the definition of attenuation and its units, the responsible operating mechanism like scattering, absorption, radiative losses and core-cladding losses is discussed.

4.4.4 Keywords: Signal losses, Attenuation, dB/km, Scattering, Absorption losses, Radiative losses, Core-cladding losses

4.4.5 Self assessment questions:

1. What is signal loss. Define attenuation constant. What are the units. Explain its importance.
2. Explain briefly scattering, absorption and radiative losses in optical fibers.
3. Discuss about the core-cladding losses in optical fibers.

4.4.6 Reference Books

1. Optical fiber communications by G. Keiser, McGraw-Hill International Edition, 2000, International student edition 1980
2. Optical fiber communications by Kato, McGraw-Hill, 1986

Unit IV
Lesson 5

SIGNAL DISTORTION IN OPTICAL WAVE-GUIDES

Objective: To understand about the signal distortion in optical fibers while it is propagating. The signal distortion is due to intrinsic and extrinsic properties of the fiber materials.

Structure

4.5. Introduction

An optical signal becomes increasingly distorted as it travels along a fiber. This distortion is a consequence of intra-modal dispersion and inter-modal delay effects. These distortion effects can be explained by examining the behavior of the group velocities of the guided modes, where the group velocities is the speed at which energy in a particular mode travels along the fiber.

Intra-modal dispersion is pulse spreading that occurs within a single mode. It is a result of the group velocity being a function of the wave-length λ and is, therefore, often referred to as chromatic dispersion. Since intra-modal dispersion depends on the wavelength, its effect on signal distortion increases with the spectral width of the optical source. This spectral width is the band of wavelengths over which the source emits light. It is normally characterized by the root-mean-square (rms) spectral width σ_λ for light emitting diodes (LEDs) the rms spectral width is approximately 5% of a central wavelength. For example, if the peak emission wavelength of an LED source is 850 nm, a typical source spectral width would be 40 nm; that is, the source emits

most of its optical power in the 830 to 870 nm wavelength band. Laser diode optical sources have much narrower spectral widths, typical values being 1 to 2 nm.

The two main causes of intra-modal dispersion are:

1. Material dispersion, which arises from the variation of the refractive index of the core material as a function of wavelength. This causes a wavelength dependence of the group velocity of any given mode.
2. Wave-guide dispersion, which occurs because the modal propagation constant β is a function of a/λ (the optical fiber dimension relative to the wavelength λ, where a is the core radius).

The other factor giving rise to pulse spreading is inter-modal delay which is a result each mode having a different value of the group velocity at a single frequency.

Of these three, wave-guide dispersion usually can be ignored in multimode fibers. However, this effect can be significant in single mode fibers. The full effects of these three distortion mechanisms are seldom observed in practice since they tend to be mitigated by other factors, such as non-ideal index profiles, optical power launching conditions (different amounts of optical power launched into the various modes), non-uniform mode attenuation, mode mixing in the fiber and in splices, and by statistical variations in these effects along the fiber. In this section we shall first discuss the general effects of signal distortion and then examine the various dispersion mechanisms.

4.5.1 Information capacity determination:

A result of the dispersion-induced signal distortion is that a light pulse will broaden as it travels along the fiber. As shown in fig 4.5.1 this pulse broadening will eventually cause a pulse to overlap with neighboring pulses. After a certain amount of overlap has occurred, adjacent pulses can no longer be individually distinguished at the receiver and errors will occur. Thus the dispersive properties determine the limit of the information capacity of the fiber a measure of the information capacity of an optical wave-guide is usually specified by the band-width distance product in MHz. Km. For a step index fiber the various distortion effects tend to limit the bandwidth distance product to about 20 MHz. Km. In graded index fibers the radial refractive index profile can be carefully selected so that pulse broadening is minimized at a specific operating wavelength. This had led to band width distance products as high as 2.5 GHz. Km.

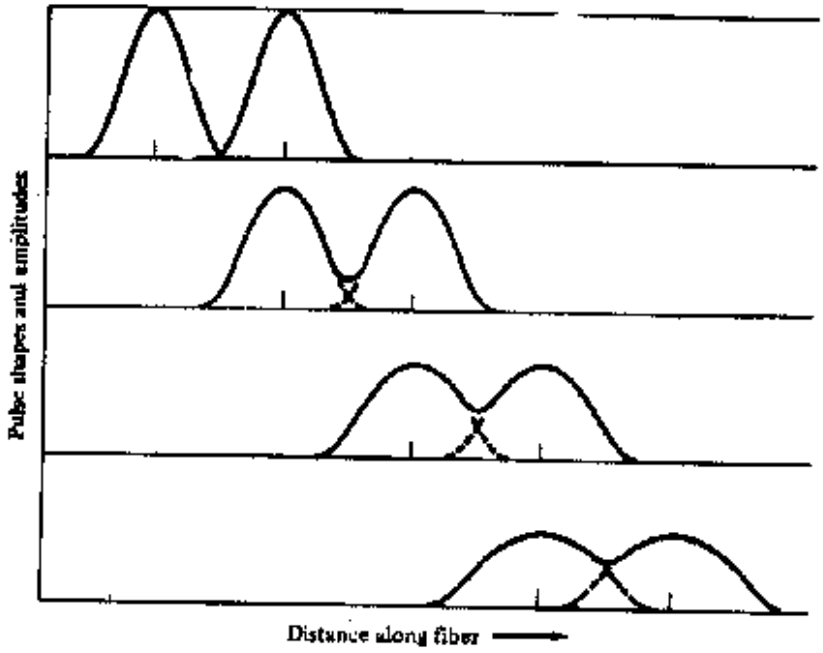


Fig. 4.5.1 Broadening and attenuation of two adjacent pulses as they along a fiber.

Single mode fibers can have capacities well in excess of this. A comparison of the information capacity of an 800 MHz. Km. Optical fiber with the capacities of typical coaxial cables used for UHF and VHF transmission is shown in fig 4.5.2. The curves are shown in terms of signal attenuation versus data rate. The fiber has a 6-dB/km low frequency attenuation. The flatness of the attenuation curve for this fiber extends up-to the micro wave spectrum.

The information carrying capacity can be determined by examining the deformation of short light pulses propagating along the fiber. The following discussion on signal distortion is thus carried out primarily from the stand point of pulse broadening. Which is representative of digital transmission.

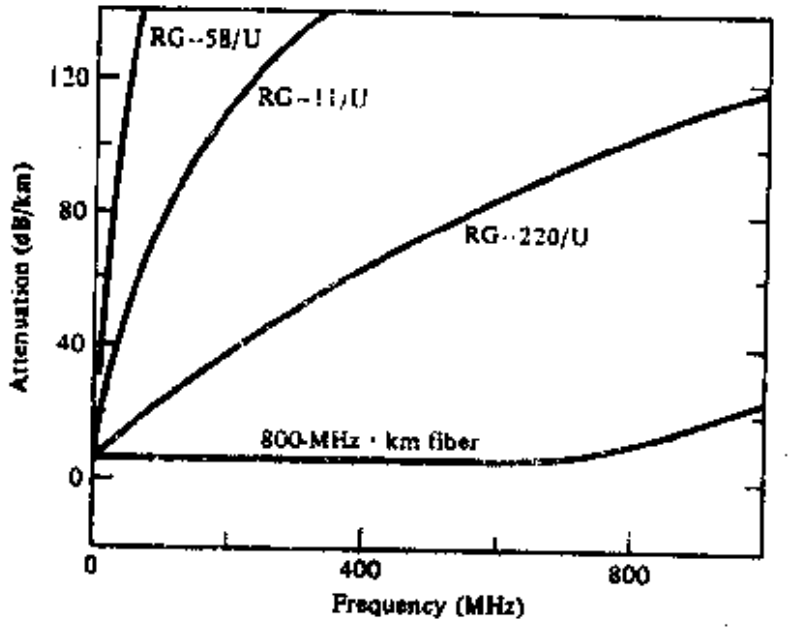


Fig 4.5.2: A comparison of the attenuation as a function of frequency or data rate of various coaxial cables and an 800
MHz. Km optical wave-guide.

4.5.2 Group delay :

Let us examine a signal that modulates in optical source. We shall assume that the modulated optical signal excites all modes equally at the input end of the fiber. Each mode thus carries an equal amount of energy through the fiber. Furthermore, each mode contains all of the spectral components in the wavelength band over which the source emits. The signal may be considered as modulating each of these spectral components in the same way. As the signal propagates along the fiber, each spectral component can be assumed to travel independently and to under go a time delay or group delay per unit length in the direction of propagation given by

$$\frac{t_g}{L} = \frac{1}{V_g} = \frac{1}{C}\frac{d\beta}{dk} = -\frac{\lambda^2}{2\pi c}\frac{d\beta}{d\lambda} \text{ ---------- } \mathbf{4.5.1}$$

Here L is the distance traveled by the pulse, β is the propagation constant along the fiber axis, k =2π/λ, and the group velocity

$$V_g = c\left(\frac{dk}{d\beta}\right)^{-1}$$

is the velocity at which the energy in a pulse travels along a fiber.

Since the group delay depends on the wavelength, each spectral component of any particular mode takes a different amount of time to travel a certain distance. As a result of this difference in time delays, the optical signal pulse spreads out with time as it is transmitted over the fiber. The quantity we are thus interested in is the amount of pulse spreading that arises from the group delay variation.

If the spectral width of the optical source is not too wide the delay difference per unit wavelength along the propagation path is approximately $dt_g/d\lambda$. for spectral components which are $\Delta\lambda$ apart and which lie $\Delta\lambda/2$ above and below a central wavelength λ_0, the total delay difference τ over a distance L is

$$\tau = \frac{dt_g}{d\lambda}\Delta\lambda \text{ ----------4.5.2}$$

if the spectral width $\Delta\lambda$ of an optical source is characterized by its rms value σ_λ, then the pulse spreading can be approximated by

$$\tau_g = \frac{dt_g}{d\lambda}\sigma_\lambda = -\frac{L\sigma_\lambda}{2\pi c}\left(2\lambda\frac{d\beta}{d\lambda} + \lambda^2\frac{d^2\beta}{d\lambda^2}\right) \text{----------4.5.3}$$

The factor $$D = \frac{1}{L}\frac{dt_g}{d\lambda} \text{ ----------4.5.4}$$

is designated as the dispersion. It defines the pulse spread as a function of wavelength and is measured in nanoseconds per kilometer. It is a result of material and waveguide dispersion. In many theoretical treatments of intra-modal dispersion it is assumed for simplicity that material dispersion and wave-guide dispersion can be calculated separately and then added to give the total dispersion of the mode. In reality these two mechanisms are intricately related since the dispersive properties of the refractive index (which gives rise to material dispersion) also affects the wave-guide dispersion. However, an examination of the inter-dependence of material and waveguide dispersion has shown that , unless a very precise value is desired, a good estimate of the total intra-modal dispersion can be obtained by calculating the effect of signal distortion arising from one type of dispersion in the absence of the other, and then adding the results. Material dispersion and waveguide dispersion are, therefore, considered separately.

4.5.3 Material Dispersion :

Material dispersion occurs because the index of refraction varies as a non-linear function of the optical wavelength. This is exemplified in fig 4.5.3. for silica. As a consequence, since the group velocity V_g of a mode is a function of the index of refraction the various spectral components of a given mode will travel at different speeds, depending on the wavelength. Material dispersion is, therefore, an intra-modal dispersion effect, and is of particular importance for single mode waveguides and for LED systems (since an LED has a broader output spectrum than a laser diode).

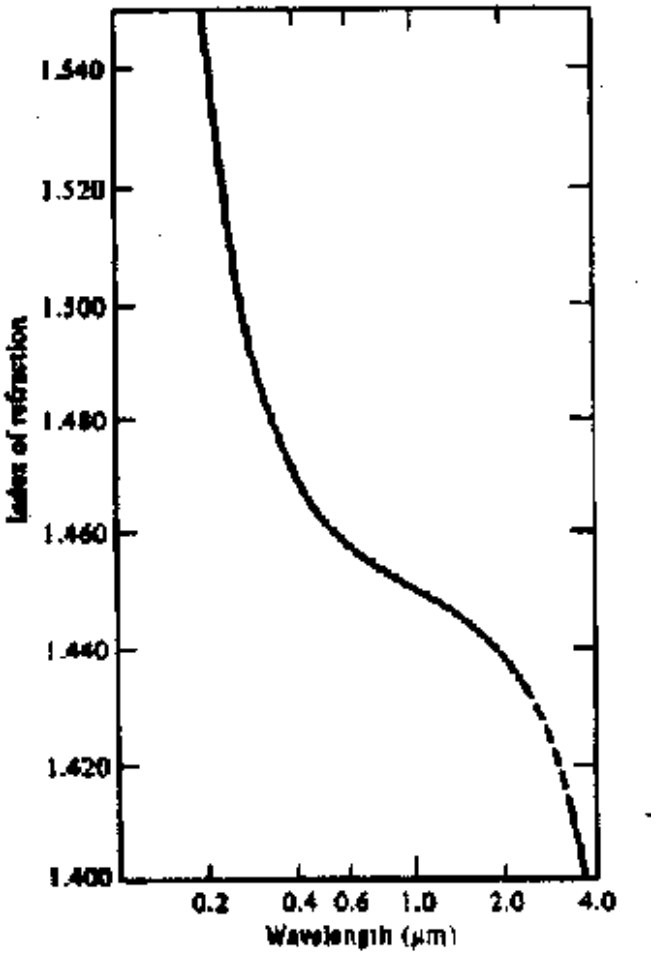


Fig 4.5.3 Variation in the index of refraction as a function of the optical wavelength for silica.

To calculate material induced dispersion, we consider a plane wave propagating in an infinitely extended dielectric medium that has a refractive index n (λ) equal to that of the fiber core. The propagation constant β is thus given by

$$\beta = \frac{2\pi n(\lambda)}{\lambda}$$ ----------4.5.5.

Substituting this expression for β, into Eq. 4.5.1. with $k = 2\pi / \lambda$ yields the group delay t_{mat} resulting from material dispersion

$$t_{mat} = \frac{L}{c}\left(n - \lambda \frac{dn}{d\lambda} \right)$$ ----------4.5.6.

Using Eq. 4.5.3. the pulse spread τ_{mat} for a source of spectral width σ_λ is found by differentiating this group delay with respect to wavelength and multiplying by σ_λ to yield

$$\tau_{mat} = \frac{dt_{mat}}{d\lambda}\sigma_\lambda = -\frac{L}{c}\lambda \frac{d^2 n}{d\lambda^2}\sigma_\lambda$$ --------**4.5.7**..

A plot of Eq. 4.5.7. for unit length L and unit optical source spectral width σ_λ is given in fig 4.5.4. for the silica material shown in fig 4.5.3. from Eq.4.5.7. and fig 4.5.4. it can be seen that material dispersion can be reduced either by choosing sources with narrower spectral output widths (reducing σ_λ) or by operating at longer wavelengths. As an example, consider a typical GaAlAs LED having a spectral width of 40 nm at an 800 nm peak output so that $\sigma_\lambda / \lambda = 5\%$. As can be seen from fig 4.5.4. and Eq.4.5.7. this produces a pulse spread of 4.4ns / km. Note that material dispersion goes to zero at 1.27 μm for pure silica.

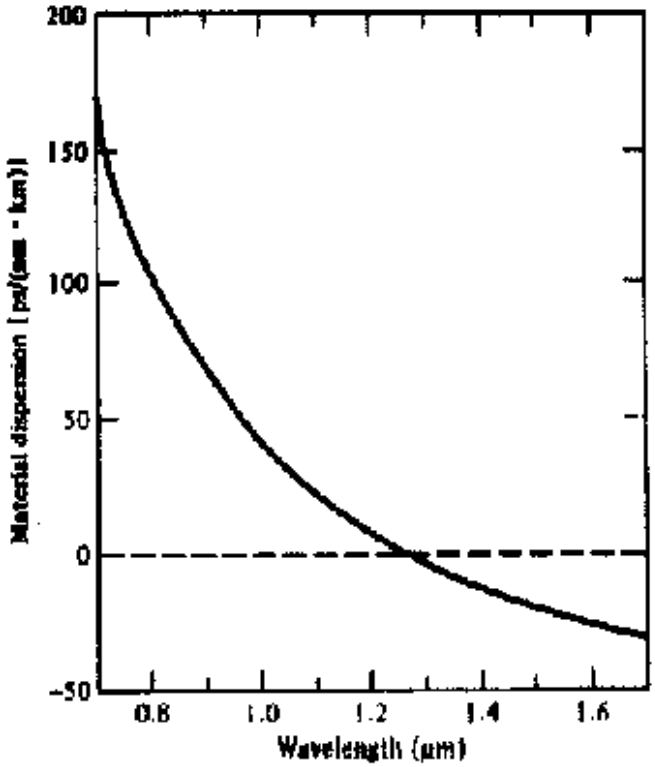


Fig 4.5.4 : material dispersion as a function of optical wavelength for silica.

4.5.4 WAVEGUIDE DISPERSION:

The effect of waveguide dispersion on pulse spreading can be approximately by assuming that the refractive index of the material is independent of wavelength. Let us first consider the group delay, that is, the time required for a mode to travel along a fiber of length L. To make the results independently of fiber configuration, we shall express the group delay in terms of the normalized propagation constant 'b' defined by

$$b = 1 - \left(\frac{ua}{V}\right)^2 = \frac{\beta^2/k^2 - n_2^2}{n_1^2 - n_2^2} \text{ ……..4.5.8}$$

For small values of the index difference $\Delta=(n_1 - n_2)/n_1$, Eq.4.5.8. can be approximated by

$$b = \frac{\beta/k - n_2}{n_1 - n_2} \text{ ……4.5.9.}$$

Solving Eq.4.5.9 for β, we have

$$\beta = n_2 k\left(b\Delta + 1\right) \text{ ……. 4.5.10.}$$

With this expression for β and using the assumption that n_2 is not a function of wavelength, we find that the group delay t_{ug} arising from wave-guide dispersion is

$$t_{wg} = \frac{L}{c}\frac{d\beta}{dk} = \frac{L}{c}\left[n_2 + n_2\Delta\frac{d(kb)}{dk} \right] \text{.....} \mathbf{4.5.11.}$$

The modal propagation constant β is obtained from the eigen-value equation expressed by Eq.2.46, and is generally given in terms of the normalized frequency V defined by eq,4.3.31. we shall therefore use the approximation

$V = ka\,(n_1^2 - n_2^2)^{1/2} \cong kan_2\sqrt{2\Delta}$ which is valid for small values of Δ, to write the group delay in Eq. 4.5.11. in terms of V instead of k, yielding

$$t_{wg} = \frac{L}{c}\left[n_2 + n_2\Delta\frac{d(Vb)}{dV} \right] \text{.....} \mathbf{4.5.12}$$

. The first term in Eq.4.5.12. is a constant and the second term represents the group delay arising from waveguide dispersion. The factor d(Vb)/dV can be expressed as

$$\frac{d(Vb)}{dV} = b\left[1 - \frac{2J_v^2(ua)}{J_{v+1}(ua)J_{v-1}(ua)} \right]$$

waveguide dispersion is generally very small compared where u is defined by Eq.2.404.3.22 and a is the fiber radius. This factor is plotted in fig 4.5.5. as a function of V. the plots show that for a fixed value of V, the group delay is different for every guided mode. When a light pulse is launched into a fiber, it is distributed among many guided modes. These various modes arrive at the fiber end at different times depending on their group delay, so that a pulse spreading results. For multimode fibers the waveto material dispersion and can therefore be neglected.

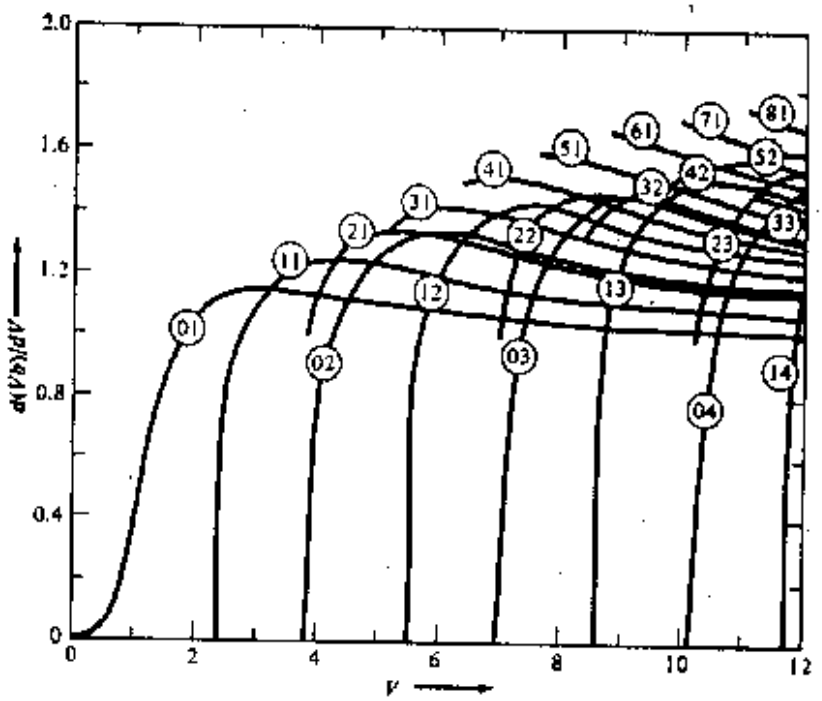


Figure 4.5.5: The group delay arising from wave-guide dispersion as a function of the V number for a step-index fiber. When ν≠1, the curve numbers νm designate the $HE_{\nu+1,\,m}$ and $EH_{\nu-1,m}$ modes. For ν=1, the curve numbers νm give the HE_{2m}, TE_{0m} modes.

For single mode fibers, however, wave-guide dispersion is of importance and can be of the same order of magnitude as a material dispersion. To see this let us compare the two dispersion factors. The pulse spread $\tau_{\omega g}$ occurring over a distribution of wavelengths σ_λ is obtained from the derivative of the group delay with respect to wavelength

$$\tau_{wg} = \sigma_\lambda \frac{dt_{wg}}{d\lambda} = -\frac{V}{\lambda}\sigma_\lambda \frac{dt_{wg}}{dV} = -\frac{n_2 L \Delta \sigma_\lambda}{c\lambda} V \frac{d^2(Vb)}{dV^2} \quad \text{..... 4.5.13.}$$

The factor $Vd^2(Vb)/dV^2$ is plotted as a function of V in fig 4.5.6. for the fundamental mode shown in fig 4.5.5. this factor reaches a maximum at V=1.2 but runs between 0.2 and 0.1 for a practical single-mode operating range of V=2 to 2.4. thus for values of Δ=0.01 and n_2=1.5,

$$\frac{\tau_{wg}}{L} = -\frac{0.003\sigma_\lambda}{c\lambda} \quad \text{..... 4.5.14}$$

.Comparing this with the material dispersion induced pulse spreading from Eq 4.5.7. for λ=900 nm, where $\frac{\tau_{mat}}{L} \approx -\frac{0.02\sigma_\lambda}{c\lambda}$ **.....4.5.15**

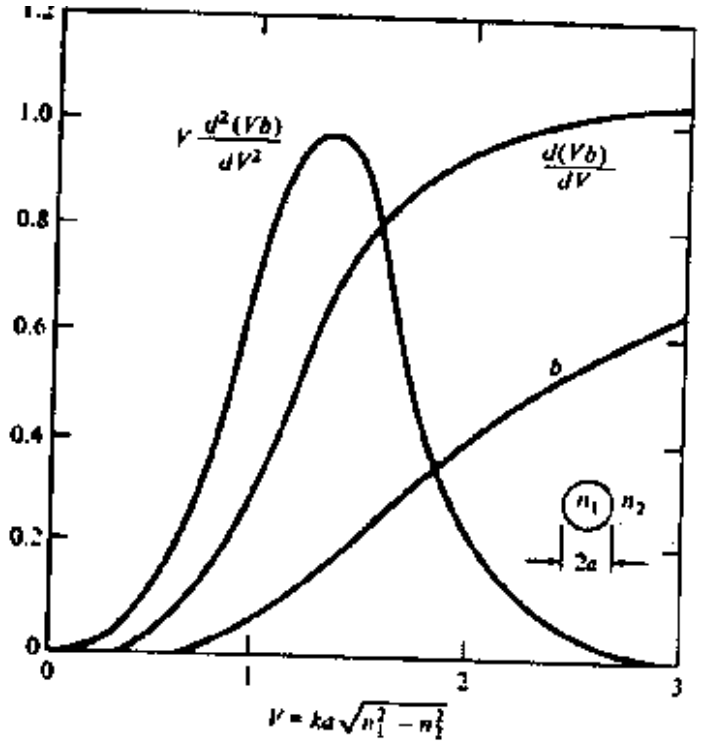


Fig 4.5.6 The waveguide parameter b and its derivatives d(Vb)/dV and $Vd^2(Vb)/dV^2$ plotted as a function of the V number.

It is clear that material dispersion dominates at lower wavelengths. However, at longer wavelengths such as at 1.3 μm, which is the spectral region of extremely low material dispersion in silica waveguide dispersion can become the dominating pulse distorting mechanism. Examples of the magnitudes of material and wave-guide dispersions are given in fig 4.5.7. for a fused silica core single mode fiber having V=2.4. in this figure the approximation that material and wave-guide dispersions are additive was used.

Figure 4.5.7. shows that in single mode fibers the total dispersion can be reduced to zero at a particular wavelength through the mutual cancellation of material and wave-guide dispersions. The particular wavelength at which the total dispersion is reduced to a minimum can be selected anywhere in the 1.3 to 1.7μm spectral range. For GeO_2 doped fibers this is achieved by varying the amount of GeO_2 dopant to obtain different material behavior, and by controlling the wave-guide effects through variations in core diameter and core cladding index difference. For example fibers designed for zero dispersion in the minimum attenuation region of 1.55μm could

have core diameters of approximately 4 to 4.8μm, core dopant concentrations of about 13 mol % GeO_2 , and index differences ranging from 0.55 to 1.8%.

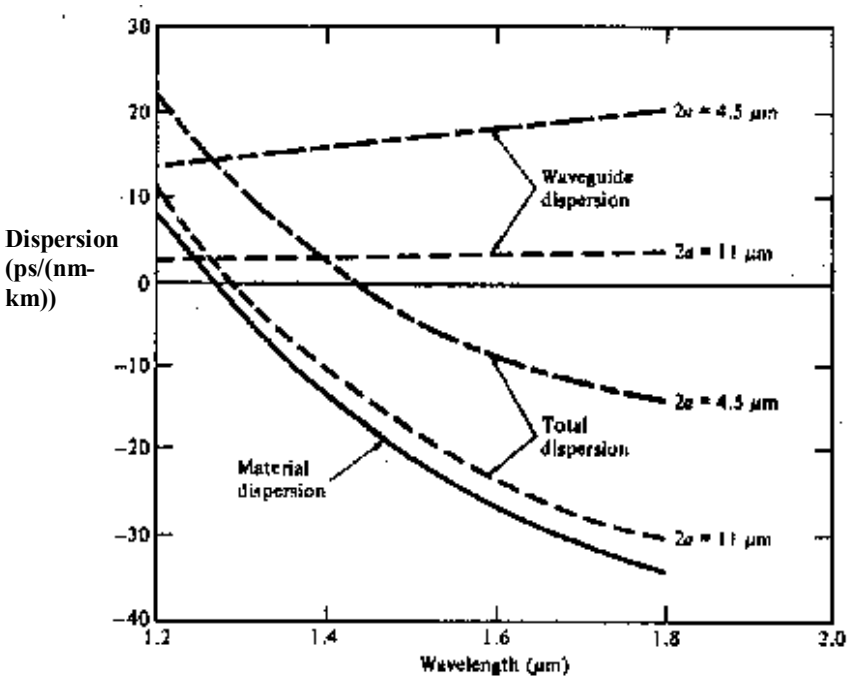


Fig 4.5.7 Examples of the magnitudes of material and waveguide dispersions as a function of optical wavelength for a single mode fused silica-core fiber.

4.5.5 Inter-modal Dispersion:

The final factor giving rise to signal distortion is inter-modal dispersion, which is a result of different values of the group delay for each individual mode at a single frequency. The steeper the angle of propagation of the ray congruence, the higher is the mode number and consequently the slower the axial group velocity. This variation in the group velocities of the different modes results in a group delay spread or inter-modal dispersion. This distortion mechanism is eliminated by single mode operation, but is important in multi-mode fibers. The pulse broadening arising from inter-modal dispersion is the difference in travel time between the longest ray congruence paths (the highest-order mode) and the shortest ray congruence paths (the fundamental mode). This is simply obtained from ray tracing and is given by

$$\tau_{\text{mod}} = T_{\max} - T_{\min} = \frac{n_1 \Delta L}{c} \quad \text{.....4.5.16}$$

For values of n_1 =1.5 and Δ=0.01the modal spread is τ_{mod} =0.015L/c. comparing this with Eqs. 4.5.14. and 4.5.15. with a relative spectral width of σ_λ/λ=4% for a LED, we have $\tau_{\omega g} = 1.2\times10^{-4}$ L/c and $\tau_{mat} = 8\times10^{-4}$ L/c, which shows that τ_{mod} dominates the pulse spreading by about an order of magnitude in step-index fibers. The relative dominance of τ_{mod} is even greater if a laser diode light source, which has a narrower spectral output width than an LED, is used.Pulse broadening in graded index wave-guide**.**

4.5.6. Pulse broadening in graded index wave-guide.

The analysis is more involved owing to the radial variation in n_1. The feature of this grading of the refractive index profile is that it offers multimedia propagation in a relatively large core together with the possibility of very low inter-modal delay distortion. This combination allows the transmission of high data rates over long distances while still maintaining a reasonable degree of light launching and coupling and coupling ease.

Since the index of refraction is lower at the outer edges of the core, light rays will travel faster in this region that in the center of the core where the refractive index in higher. This can be seen from the fundamental relationship $\nu=c/n$, where ν is the speed of light in a medium of refractive index n. Thus the ray congruence characterizing the higher order mode will tend to travel further than the fundamental ray congruence, but at a faster rate. The higher order mode will thereby tend to keep up with the lower order mode, which, in turn, reduces the spread in the modal delay. The root mean square (rms) pulse broadening σ in a graded index fiber can be obtained from the sum

$$\sigma = \left(\sigma^2_{inter\,modal} + \sigma^2_{intra\,modal}\right)^{1/2} \text{.....4.5.17.}$$

where $\sigma_{\text{inter-modal}}$ is the rms pulse width resulting from inter-modal delay distortion and $\sigma_{\text{inter-modal}}$ is the rms pulse width resulting from pulse broadening within each mode. To find the inter-modal delay distortion, we use the relationship connecting inter-modal delay to pulse broadening derived by Personick

$$\sigma_{intermodal} = \left(\left\langle \tau_g^{\,2} \right\rangle - \left\langle \tau_g \right\rangle^2\right)^{1/2} \text{..... 4.5.18.}$$

where the group-delay τ_g of a mode is given by Eq. 4.5.1. and the quantity <A> is defined as the average of the variable A_{vm} over the mode distribution, that is, it is given by

$$\langle A \rangle = \sum_{v,m} \frac{P_{vm} A_{vm}}{M} \quad \text{..... 4.5.19}$$

where P_{vm} is the power contained in the mode of order (ν, m).

The group delay

$$\tau_g = \frac{L}{c} \frac{\partial \beta}{\partial k} \quad \text{..... 4.5.20}$$

is the time it takes energy in a mode having a propagation constant β to travel a distance L. to evaluate τ_g we solve Eq.2.78 4.3.59.for β, which yields

$$\beta = \left[k^2 n_1^2 - 2\left(\left(\frac{\alpha+2}{\alpha} \right) \frac{m}{a^2} \right)^{\alpha/\alpha+2} (n^2{}_1 k^2 \Delta)^{2/\alpha+2} \right]^{1/2} \quad \text{..... 4.5.21.}$$

or, equivalently,

$$\beta = k n_2 \left[1 - 2\Delta (m/M)^{\alpha/\alpha+2} \right]^{1/2}$$

where m is the number of possible guided modes having propagation constants between n,k and β, M is the total number of possible guided modes given by Eq. (2.79)4.3.60. substituting Eq. (4.5.21) into Eq. (4.5.20), keeping in mind that n_1 and also depend on k, we obtain

$$\tau = \frac{L}{c} \frac{k n_1}{\beta} \left[N_1 - \frac{4\Delta}{\alpha+2} \left(\frac{m}{M} \right)^{\alpha/\alpha+2} \left(N_1 + \frac{n_1 k \partial \Delta}{2\Delta \partial k} \right) \right]$$

$$= \frac{L N_1}{c} \frac{k n_1}{\beta} \left[1 - \frac{\Delta}{\alpha+2} \left(\frac{m}{M} \right)^{\alpha/\alpha+2} (4+\varepsilon) \right] \quad \text{..... 4.5.22}$$

Where we have used Eq. (2.79)4.3.60. for M and have defined the quantities

$$N_1 = n_1 + k \frac{\partial n_1}{\partial k} \quad \text{..... 4.5.23a}$$

$$\varepsilon = \frac{2 n_1 k}{N_1 \Delta} \frac{\partial \Delta}{\partial k} \quad \text{.....4.5.23b}$$

As we noted in Eq. (2.38)4.2.20, guided modes only exit for values of β lying between kn_2 and kn_1. Since n_1 differs very little from n_2, that is

$$n_2 = n_1 (1 - \Delta)$$

Where $\Delta<<1$ is the core-cladding index difference, it follows that $\beta \cong n_1 k$. Thus we can use the relationship

$$y = \Delta\left(\frac{m}{M}\right)^{\alpha/\alpha+2} << 1 \quad \textbf{..... 4.5.24}.$$

In order to expand Eq.(4.5.22.) in a power series in y. Using the approximation

$$\frac{kn_1}{\beta} = (1-2y)^{-1/2} \approx 1 + y + \frac{3y^2}{2} \quad \textbf{.....4.5.25}$$

We have that

$$\tau = \frac{N_1 L}{c}\left[1 + \frac{\alpha - 2 - \varepsilon}{\alpha+2}\Delta\left(\frac{m}{M}\right)^{\alpha/\alpha+2} + \frac{3\alpha - 2 - 2\varepsilon}{2(\alpha+2)}\Delta^2\left(\frac{m}{M}\right)^{2\alpha/\alpha+2} + O(\Delta^3)\right] \ldots\ldots \textbf{4.5.26}$$

Equation (4.5.26.) shows that to first order . The group delay difference between modes is zero if

$$\alpha = 2 + \in \ldots\ldots\ldots 4.5.27$$

Since $\in$ is generally small, this indicates that minimum intermodal distortion will result from core refractive-index profiles which are nearly, that is,$\alpha \cong 2$.

If we assume that all modes are equally excited, that is, p_{mn} =p for all modes, and if the number of fiber mode is assumed to be large, then the summation in Eq. (4.5.19) can be replaced by integral. Using these assumptions, eq.(4.5.26) can be sub. Into eq. (4.5.18) to yield

$$\sigma_{intermodal} = \frac{LN_1\Delta}{2c}\frac{\alpha}{\alpha+1}\left(\frac{\alpha+2}{3\alpha+2}\right)^{1/2} \times \left[c_1^{\,2} + \frac{4c_1c_2(\alpha+1)\Delta}{2\alpha+1} + \frac{16\Delta^2 c_2^{\,2}(\alpha+1)^2}{(5\alpha+2)(3\alpha+2)}\right]^{1/2} \quad \textbf{...4.5.28.}$$

Where we have used the abbreviations

$$c_1 = \frac{\alpha - 2 - \varepsilon}{\alpha+2}$$

$$c_2 = \frac{3\alpha - 2 - 2\varepsilon}{2(\alpha+2)} \quad \textbf{.....4.5.2}$$

To find the intramodal pulse broadening, we use the definition

$$\sigma^2_{intra\,modal} = L^2\left(\frac{\sigma_\lambda}{\lambda}\right)^2\left\langle\left(\lambda\frac{d\tau_g}{d\lambda}\right)^2\right\rangle \quad \textbf{.....4.5.30}$$

Where σ_λ is the rms spectral width of the optical source. Eq.(4.5.26) can be used to evaluate the parentheses. If we neglect all terms second and higher order in Δ. We obtain

$$\lambda\frac{d\tau_g}{d\lambda} = -\frac{L}{C}\lambda^2\frac{d^2n_1}{d\lambda^2} + \frac{N_1L\Delta}{c}\frac{\alpha-2-\varepsilon}{\alpha+2}\frac{2\alpha}{\alpha+2}\left(\frac{m}{M}\right)^{\alpha/\alpha+2} \quad \textbf{.....4.5.31}$$

Here we have kept only the largest terms , that is term involving factors such as $d\Delta/d\lambda$ and Δ $dn_1/d\lambda$ are negligibly small. Both term in Eq.(4.5.21) contributes to $\lambda\, d\tau_g/d\lambda$. For large values of α , since $\lambda^2\, d^2n_1/d\lambda^2$ and Δ are the same order of magnitute. However , the second term in Eq.(4.5.21) is small compared to the first term when α is closely to 2 .

To evaluate $\sigma_{inter\text{-}modal}$ we again assume that all modes are equally excited and that the summation in eq.(4.5.19) can be replaced by an integral. Thus substituting eq.(4.5.31) into Eq.(4.5.30) we have

$$\sigma_{intra\,modal} = \frac{L}{c}\frac{\sigma_\lambda}{\lambda}\left[\left(-\lambda^2\frac{d^2n_1}{d\lambda^2}\right)^2 - N_1c_1\Delta\left(2\lambda^2\frac{d^2n_1}{d\lambda^2}\frac{\alpha}{\alpha+1} - N_1\Delta\frac{2\alpha}{3\alpha+2} - \right)\right]^{1/2} \quad .....4.5.32$$

Olshansky and Keck have evaluated σ as a function of α at λ=900 nm for a titania-dopped silica fiber having a numerical aperture of 0.16. This is shown in fig. 4.5.8. here the uncorrected curve assumes $\in$=0 and includes only intermodal dispersion(no material dispersion). The inclusion of the effect of $\in$ shifts the curve to higher values of α. The effect of the spectral width of the optical source on the rms pulse width is clearly demonstrated in fig.4.5.8. The light sources shown of an LED, an injection laser diode, and a distributed – feedback laser having rms spectral widths of15,1and 0.2nm, respectively. The data transmission capacities of the sources are approximately 0.13, 2, and 10 (Gb.km)/s, respectively.

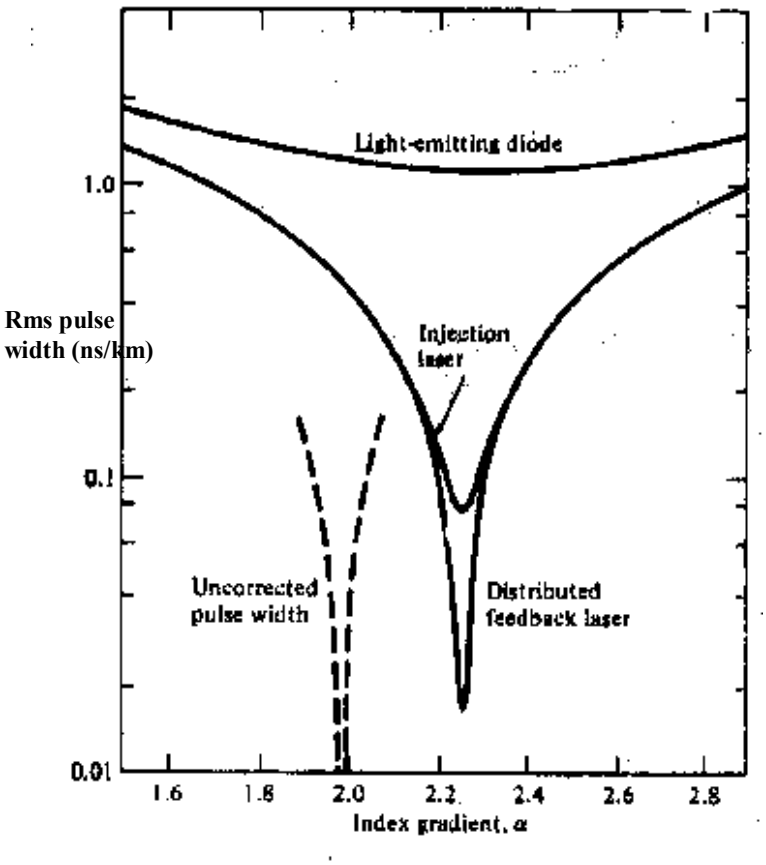


Figure: 4.5.8 calcutated rms pulse spreading in a graded index fiber versus the index parameter alpha at 900 nm . the uncorrected pulse curve is for ε=0 and assumes mode dispersion only . the other curves in to material dispersion for an LED , an injection laser diode , and a distributed feed back laser having spectral widths of 15, 1 , and 0.2 nm respectively .

The value of α which minimizes pulse distortion depends strongly on wavelength. To see this, let us examine the structure of a graded index fiber. A simple model of this structure is to consider the core to be composed of concentric cylindrical layer the refractive index has a different variation with wavelength λ since the glass composition is different in each layer. Consequently , a fiber with a given index profile α will exhibit different pulse spreading according to the source wavelength used. This is generally called profile dispersion. An example of this is given in fig.4.5.9. for a $GeO_2 - SiO_2$ fiber. This shows that the optimum value of α decreases with

increasing wavelength. Suppose one wishes to transmit at 900nm.a fiber having an optimum profile α_{opt} at 900nm should exhibit a sharp bandwidth peak at that wavelength. Fibers with under compensates profiles, characterized by $\alpha > \alpha_{opt}$ (900nm) , tend to have a peak bandwidth at a shorter wavelength. On the other hand, over compensated fibers which have an index profile $\alpha < \alpha_{opt}$ (900nm) become optimal at a longer wavelength.

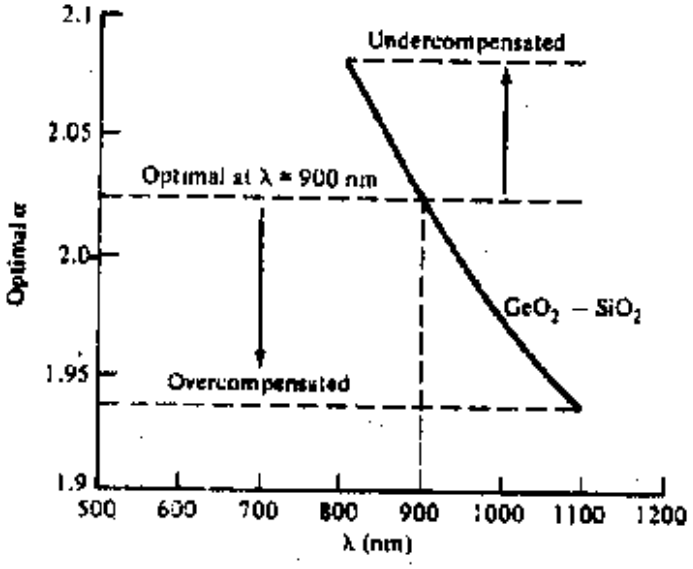


Figure 4.5.9: Profile dispersion effect on the optimum value of α as a function of a wavelength for a $GeO_2 - SiO_2$ graded index fiber.

If the effect of material dispersion is ignored (that is for $dn_1/d\lambda = 0$), an expression for the optimum index profile can be found from the minimum of Eq. 4.5.28 as a function of α. This occur at

$$\alpha_{opt} = 2 + \varepsilon - \Delta \frac{(4+\varepsilon)(3+\varepsilon)}{5+2\varepsilon} \quad \text{..... 4.5.33}$$

If we take $\varepsilon=0$ and $dn_1/d\lambda=0$, then Eq. 4.5.28. reduces to

$$\sigma_{opt} = \frac{n_1 \Delta^2 L}{20\sqrt{3}c} \text{.....454.34.}$$

This can be compared with the dispersion in a step-index fiber by setting $\alpha=\infty$ and $\varepsilon=0$ in Eq.4.5.28, yielding $\sigma_{step} = \frac{n_1 \Delta L}{c} \frac{1}{2\sqrt{3}} \left[1 + 3\Delta + \frac{12\Delta^2}{5}\right]^{1/2} = \frac{n_1 \Delta L}{2\sqrt{3}c}$ .....4.5.35.

Thus under the assumption made in Eqs. 4.5.34 and 4.5.35,

$$\frac{\sigma_{step}}{\sigma_{opt}} = \frac{10}{\Delta} \text{.....4.5.36}$$

Hence since typical values of Δ are 0.01, Eq.4.5.36 indicates that the capacity of a graded index fiber is about three orders of magnitude larger than that of a step index fiber. For Δ=1% the rms pulse spreading in a step index fiber is about 14 ns/km, whereas that for a graded index fiber is calculated to be 0.014 ns/km.

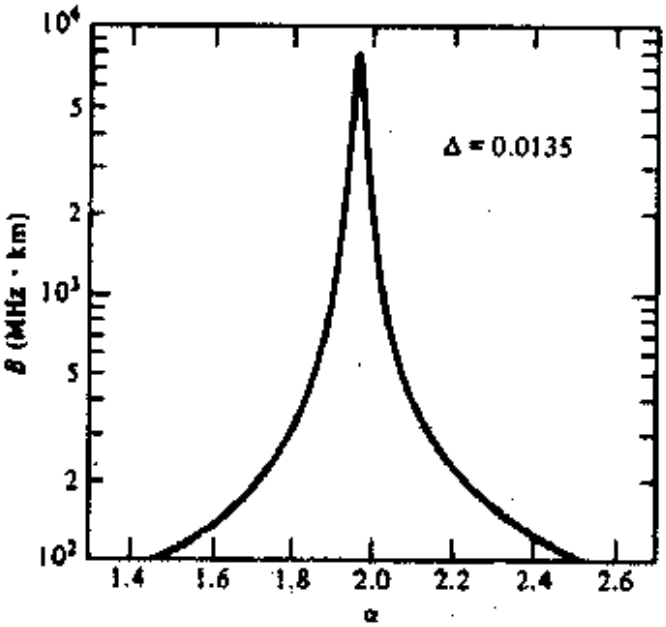


Fig 4.5.10 Variations in bandwidth resulting from slight deviations in the refractive index profile for a graded-index fiber with Δ=0.0135.

In practice these values are greater because of manufacturing difficulties. For example, it has been shown that although theory predicts a band-width of about 8 GHz. Km, in practice very slight deviations of the refractive index profile from its optimum shape, owing to unavoidable manufacturing tolerances, can decrease the fiber band width dramatically. This is illustrated in fig 4.5.10. for a fiber with Δ =0.0135. A change in α of a few percent can decrease the bandwidth by an order of magnitude.

4.5.7. Mode Coupling:

In real systems pulse distortion will increase less rapidly after a certain initial length of fiber because of mode coupling and differential mode loss. In this initial length of fiber, coupling of energy from one mode to another arises because of structural imperfections, fiber diameter and refractive index variations, and cabling induced micro-bends. The mode coupling tends to average out the propagation delays associated with the modes, thereby reducing inter-modal dispersion. Associated with this coupling is an additional loss, which has units of dB/km. The result of this phenomenon is that, after a certain coupling length L_c, the pulse distortion will change from an L dependence to a $(L_cL)^{1/2}$ dependence. The improvement in pulse spreading caused by mode coupling over the distance $Z<L_c$ is related to the excess loss hZ incurred over this distance by the

$$hZ\left(\frac{\sigma_s}{\sigma_o}\right)^2 = C \quad \text{.....4.5.37.}$$

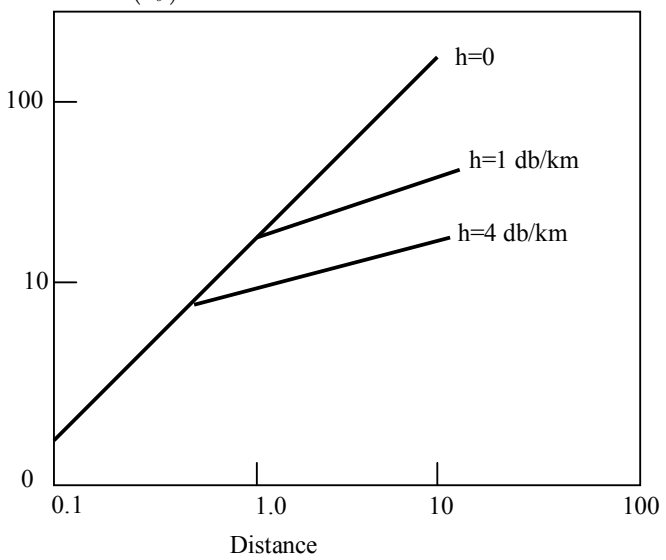


Fig 4.5.11: Mode coupling effects on pulse distortion in long fibers for various coupling losses.

Here C is a constant, σ_c is the pulse width increase in the absence of mode coupling, σ_c is the pulse broadening in the presence of strong mode coupling, and hZ is the excess attenuation resulting from mode coupling. The constant C in Eq, (4.5.37) is independent of all dimensional quantities and refractive indices. It depends only on the fiber profile shape, the mode-coupling strength, and the modal attenuations.

The effect of mode coupling on pulse distortion can be significant for long fibers, as is shown in Fig. 4.5.11. for various coupling losses in a graded-index fiber. The parameters of this fiber are Δ=1 percent, α=4,and C=1.1. The coupling loss h must be determined experimentally, since a calculation would require a detailed knowledge of the mode coupling introduced by the various waveguide perturbations. Measurement of bandwidth as a function of distance have produced values of L_c ranging from about 100 to 550m.

4.5.8. Summary: Optical signals are increasingly distorted as they travel along a fiber. This distortion is a consequence of intramodal and intermodal dispersion effect. Intramodal dispersion is pulse spreading that occurs within a single mode due to two main causes. In multimode fibers, pulse distortion also occurs since each mode travels at a different group velocity. Mode delay is the dominant pulse-distorting mechanism in step-index fibers. Intermodal delay distortion can be made very small by careful tailoring of the core refractive index profile. Material dispersion thus tends to be the dominant pulse-distorting effect in graded index fibers

4.5.9. Keywords: Signal distortion, information capacity, group delay, dispersion - intermodal, intramodal, waveguide , pulse broadening,

4.5.10. Self assessment

1. Explain information capacity and group delay in optical fibers.
2. Explain in detail about the intermadal and intramodal dispersions in the optical fibers.
3. Discuss the pulse broadening mechanism in optical fibers.
4. Write short notes on waveguide dispersion and material dispersion.
5. Write notes on mode coupling.

4.5.11. Reference and text Books

1. Optical fiber communications by G. Keiser, McGraw-Hill International Edition, 2000, Third edition, and also see first edition.
2. Optical fiber communications by Kato, McGraw-Hill, 1986

Unit IV

Lesson 6

OPTICAL FIEBER FABRICATION AND CABLING

Objective: To know the materials used for optical fibers, the different fabrication processes for optical fibers and their mechanical properties along with the cabling of optical fibers.

Structure

4.6.1. FIBER MATERIALS:

The structure and material composition of a fiber also dictate how and what degree optical signals get attenuated and distorted as they propagate along a fiber. To conclude the discussion of optical fibers, we shall show here what materials are used and how fibers are manufactured. We shall analyze their mechanical strengths and illustrate how they can be incorporated into cable structures, which protect the waveguides from the external environ-ment. In selecting material for optical fibers, a number of requirements must be satisfied.
For example:

1. It must be possible for make long, thin, flexible fibers from the material.
2. The material must be transparent at a particular optical wavelength in order for the fiber to guide light efficiently.

3. Physically compatible materials having slightly different refractive indices for the core and clad must be available.

Materials satisfying these requirements are glasses and plastics.

The majority of fibers are made of glass consisting either of silica or a silicate. The variety of available glass fibers ranges from high-loss glass fibers with large cores used. For short-transmission, distances to very transparent, (low-loss) fibers employed in long-haul applications. Plastic fibers are less widely used because of their substantially higher attenuation compared to glass fibers. The main use of plastic fibers is in short distance applications and in abusive environments, where the greater mechanical strength of plastic fibers offer an advantage over the use of glass fibers.

4.6.1.1. GLASS FIBERS:-

Glass is made by fusing mixtures of metal oxides, sulfides, or selenides. The resulting material is a randomly connected molecular network rather than a well defined ordered structure as found in crystalline materials. A consequence of this random order is that glasses do not have well defined melting points. When glass is heated up from room temperature, it remains a hard solid up to several hundred degrees centigrade. As the temperature is increased further, the glass gradually begins to soften until at very high temperatures it becomes a viscous liquid. The expression "melting temperature" is commonly used in glass manufacture. This term refers only to an extended temperature range in which the glass becomes fluid enough to free itself fairly quickly of gas bubbles.

The largest category of optically transparent glasses from which optical fibers are made consists of the oxide glasses. Of these the most common is silica, which has a refractive index of 1.458 at 850nm. To produce two similar materials having slightly different indices of refraction for the core and cladding, either fluorine or various oxides (referred to as dopants) such as B_2O_3, GeO_2 are added to the silica. As shown in fig.4.6.1., the addition of GeO_2 increases the refractive index whereas doping the silica with fluorine or B_2O_3 decreases it. Since the cladding must have a lower index than the core, examples of fiber compositions are:

1. GeO_2-SiO_2 core; SiO_2 cladding
2. P_2O_5-SiO_2 core; SiO_2 cladding
3. SiO_2 core; B_2O_3- SiO_2 cladding
4. GeO_2- B_2O_3- SiO_2 core; B_2O_3- SiO_2 cladding

Here the notation G_eO_2-SiO_2, for examples, denotes a GeO_2-doped silica glass.

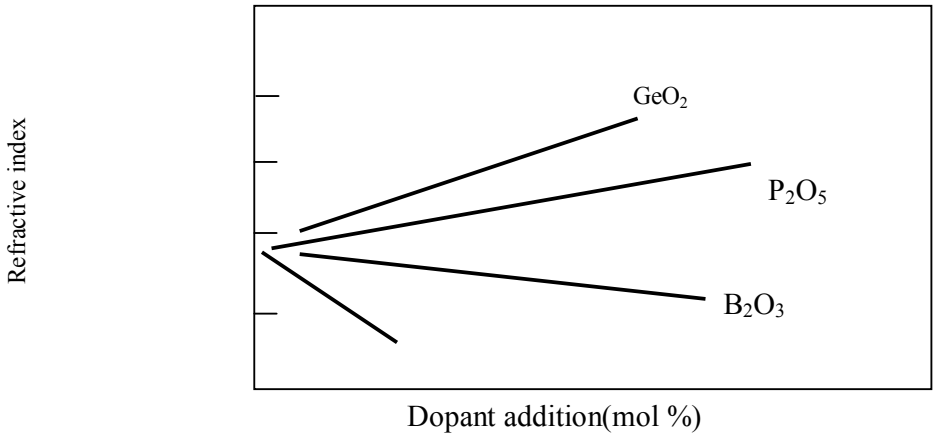


Fig 4.6.1: Variation in refractive index as a function of doping concentration in silica glass.

The principal raw material for silica is sand. Glass composed of pure silica is referred to either as silica glass, fused silica, or vitreous silica. Some of its desirable properties are a resistance to deformation at temperatures as high as 1000 centigrade degrees, a high resistance to breakage from thermal shock because of its low thermal expansion, good chemical durability, and high transparency in both the visible and infrared regions of interest fiber optic communication systems. Its high melting temperature is a disadvantage if the glass is prepared from a molten state. However, this problem is partially avoided when using vapour deposition techniques.

An alternative to the silica glasses is the low-melting silicates. Typical glasses used for optical fibers are the soda lime silicates, the germanosilicates, and various borosilicates. The soda lime silicates, for example, are combinations of silica, an alkaline oxide such as Na_2O, or Li_2O, and a second oxide such as CaO (lime), MgO, ZnO, or BaO. These glasses are relatively easy to melt and fabricate. The raw materials of soda-lime silicates are ultra pure powdered forms of oxides (SiO_2, GeO_2 & B_2O_3) and carbonates (Na_2CO_3, K_2CO_3, $CaCO_3$, and $BaCO_3$).

4.6.1.2. Halide Glass fibers:

In 1975 at the Universite de Rennes discovered fluoride glasses that have extremely low transmission losses at mid-infrared wavelengths (0.2-0.8 μm) with the lowest loss being around 2.55 μm). Fluoride glasses belong to a general family of halide glasses in which the anions are from elements in a heavy metal fluoride glass, which uses ZrF_4 as the major component and glass network former. Several other constituent need to be added to make that has moderate resistance to crystallization. Table 4.6.1. lists the constituents and their molecular percentages of a particular fluoride glass refereed to as ZBLAN (after its elements ZrF4, BaF3, LaF3, AlF3 and NaF). This material forms the core of a glass fiber. To make a lower-refractive index glas, one partially replaces ZrF_4 by HaF_4 to get a ZHBLAN cladding. Although these glasses potentially offer intrinsic minimum losses of 0.01-0.001 dB/km, fabricating long lengths of these fibers is difficult. First, ultra pure materials must be used to reach this low level. Second, fluoride glass is prone to devitrification. Fiber making techniques have to take this into account to avoid the formation of microcrystallinetes, which have a drastic effect on scattering losses.

Table 4.6.1. : Molecular composition of a ZBLAN fluoride glass

Material	Molecular percentage
ZrF4	54
BaF2	20
LaF2	4.5
AlF3	3.5
NaF	18

In 1978 investigations were reported on materials having extremely low light transmission loss in the 2 to 5 mm wavelength region. The principle materials in this mid-infrared range are various halide crystals (for example, TiBrI, TiBr, KCl, CSi, and AgBr) and fluoride glasses composed of mixtures of GdF_3,BaF_2,ZrF_4 and AlF_3. One limitation of these materials was that only short lengths of fibers could be fabricated because conventional glass-drawing techniques can not be used for crystalline materials. However, investigations are now actively continuing in achieving longer lengths by extrusion methods. Another problem is that, since these materials are inherently weaker than oxide glasses, special cabling must be done to strengthen the fiber and to protect it from moisture.

4.6.1.3. Active Glass fibers:

Incorporating rare-earth elements (atomic number 57 – 71) into a normally passive glass gives the resulting material new optical na diamagnetic properties. These new properties allow the material to perform amplification, attenuation, and phase retardation on the light passing through it. Doping can be carried out for both silica and halide glasses. Two commonly used materials for fiber lasers are erbium and neodymium. The ionic concentrations of the rare earth elements are low to avoid clustering effect. By examining the absorption and fluorescence spectra of these materials, one can use an optical source, which emits at an absorption wavelength to excite electrons to higher energy level in the rare earth dopants. When these excited electrons drop to lower energy levels, they emit light in a narrow optical spectrum at the fluorescence wavelengths.

4.6.1.4. Chalgenide Glass fibers:

In addition to allowing the creation of optical amplifiers, the nonlinear properties of glass fibers can be exploited for other applications, such as all-optical switches and fiber lasers. Chalgenide glass is one candidate for these uses because of its high optical nonlinearity and its long interaction length. These glasses contain at least one chalcogen element (S, Se or Te) and typically one other element such as P,I,Cl, Br, Cd, Bas, Si, or Tl for tailoring the thermal , mechanical and optical properties of the glass. Among the various chalgenide glasses As_2S_3 is one fo the most well known materials. Single-mode fibers have been made using As_{40} S_{58} Se_2 and As_2S_3 for core and cladding materials, respectively. Losses in these glasses typically ranged around 1dB/km.

4.6.1.5. Plastic-Clad Glass Fibers

Optical fibers constructed with glass cores and glass cladding are very important for long – distance applications where the very low losses achievable in these fibers are needed for short-distance applications (up to several hundred meters), where higher losses are tolerable, the less expensive plastic clad silica fibers can be used. These fibers are composed of silica cores with the lower refractive index cladding being a polymer (plastic) material. These fibers are often referred to as PCS (plastic-clad silica) fibers.

A common material source for the silica core is selected high purity natural quartz. A common cladding material is a silicone resin having a refractive index of 1.405 at 850 nm. Silicone resin is also frequently used as a protective coating for other types of fibers. Another popular plastic cladding material is perfluoronated ethylene propylene (Teflon FEP). The low refractive index of 1.338 of this material results in fibers with potentially large numerical apertures Plastic claddings are only used for step-index fibers. The core diameters are larger (510 to 600 mm) than the standard 50 mm diameter core of all glass graded index fibers, and the larger difference in the core and cladding indices results in a high numerical aperture. This allows low cost large are a light sources to be used for coupling optical power into these fibers, there by yielding comparatively inexpensive but lower quality systems which are quite satisfactory for many applications.

able 4.6.2. : Sample characteristics of PMMA and PFP polymer optical fibers

Characteristics	PMMA POF	PFP POF
Core diameter	0.4μm	0.125-0.30 mm
Cladding diameter	1.0mm	0.25-0.60 mm
Numerical aperture	0.25	0.20
Attenuation	150bB/km at 650nm	60-80 dB/km at 650-1300 nm
Bandwidth	2.5 Gb/s over 100m	2.5 Gb/s over 300m

All plastic multimode step index fibers are good candidates for fairly short (up to about 100 m) and low –cost links. Although they exhibit considerably greater optical signal attenuations than glass fibers, the toughness and durability of plastic allow these fibers to be handled with out special care. The high refractive index differences that can be achieved between the core and cladding materials yield numerical apertures as high as 0.6 and large acceptance angles of up to 70. In addition, the mechanical flexibility of plastic allows these fibers to have cores, with typical diameters ranging from 110 to 1400 mm. These factors permit the use of inexpensive large area light emitting diodes, which, in conjunction with the less expensive plastic fibers, make an economically attractive system.

Examples of plastic fiber constructions are;

1. A polystyrene core (n_1 =1.60) and a methyl methacrylate cladding (n_2 = 1.49) to give an NA of 0.60
2. A polymethacrylate core (n_1= 1.49) and a cladding made of its copolymer (n_2 =1.40) to give an NA of 0.50

FIBER FABRICATION

Two basic techniques are used in the fabrication of all glass optical waveguides. These are the vapour phase oxidation processes and the direct melt methods. The direct melt method follows traditional glass making procedures in that optical fibers are made directly from the molten state of purified components of silicate glasses. In the vapour phase oxidation process, highly pure vapours of metal halides (e.g., $SiCl_4$ and $GeCl_4$) react with oxygen to form a white powder of SiO_2 particles .The particles are then collected on the surface of a bulk glass by one of three different commonly used processes and are sintered (transformed to a homogeneous glass mass by heating with out melting) by one of a variety of techniques to from a clear glass rod or tube (depending on the process). This rod or tube is called a preform. It is typically around 10 mm in diameter and 60 to 90 cm long. Fibers are made from the preform by using the equipment shown in fig. 4.6.2.

The preform is precision –fed into a circular heater called the drawing furnace. Here the preform end is softened to the point where it can be drawn into a very thin filament, which becomes the optical fiber. The turning speed of the take up drum at the bottom of the draw tower determines how fast the fiber is drawn. This, in turn, will determine the thickness of the fiber, so that a precise rotation rate must be maintained. An optical fiber thickness monitor is used in a feedback loop for this speed regulation. To protect the bare glass fiber from external contaminants such as dust and water vapor, an coating is applied to the fiber immediately after it is drawn.

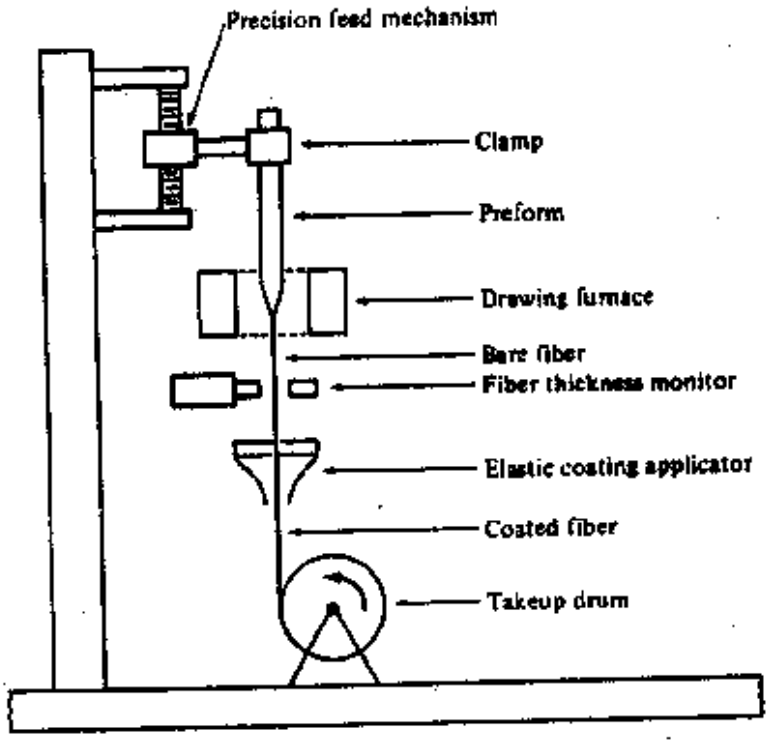


fig 4.5.2. schematic of fiber drawing apparatus

We hall now briefly examine some details of the direct melt and vapour phase oxidation processes.

4.6.2.1. OUTSIDE VAPOUR PHASE OXIDATION

The first fiber to have a loss of less than 20 db/km was made at the Corning glass works by the out side vapour phase oxidation (OVPO) process. This method is illustrated in fig .4.6.3. First, a layer of SiO_2 particles called a soot is deposited from a burner on to a rotating or ceramic mandrel. The glass soot adhere to this bait rod and, layer by layer, a cylindrical, porous glass preform is built up .By properly controlling the constituents of the metal halide vapor stream during the deposition process, the glass compositions and dimensions desired for the core and cladding can be incorporated into the preform . Either step or graded index preforms can thus be made.

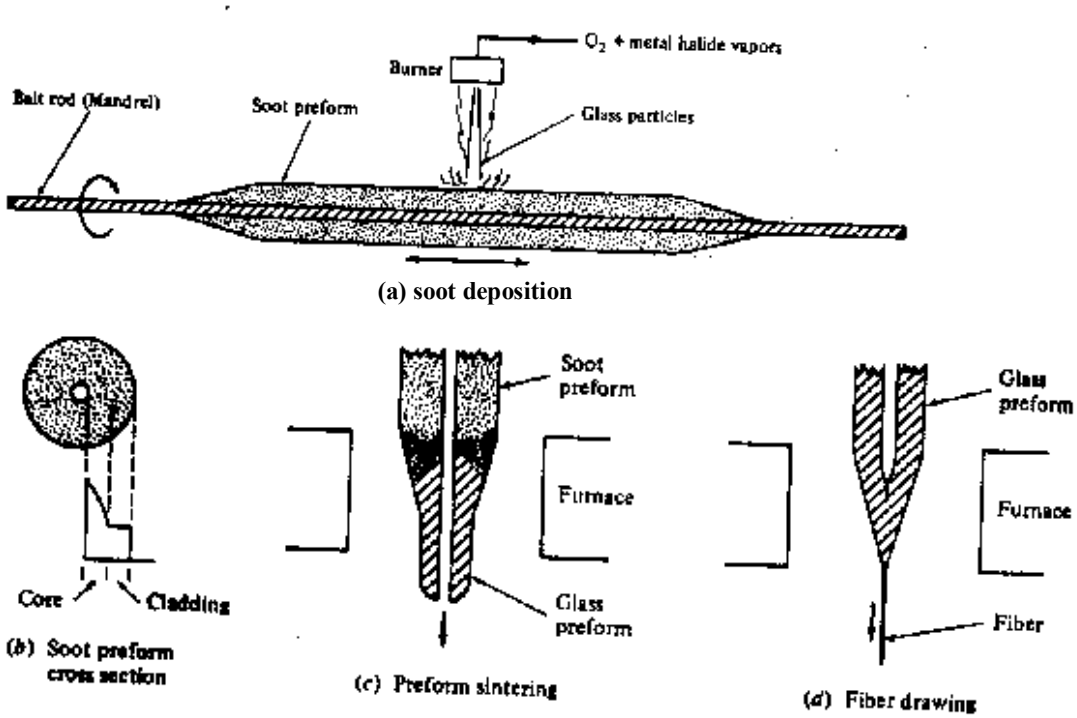


Fig 4.6.3 Basic steps in preparing a perform by the OVPO process

(a) Balt rod rotates and moves back and forth under the burner to produce a uniform deposition of a glass soot particles along the rod .
(b) Profiles can be step or graded index
(c) Following deposition, the soot perform is sintered into a clear glass preform

When the deposition process is completed, thus mandrel is removed and the porous tube is then vitrified in a dry atmosphere at a high temperature (above 1400 ^{o}C) to a clear glass preform. This clear preform is subsequently mounted in a fiber drawing tower and made into a fiber as shown in fig. 4.6.2. The central hole in the tube preform collapses during this drawing process.

4.6.2.2. VAPOUR PHASE AXIAL DEPOSITION (VAD)

The OVPO process described in sec.4.6.2.1. is a lateral deposition method. Another OVPO type process is the vapor phase axial deposition method (VAD), illustrated in fig.10-4. In this method the SiO_2 particles are formed in the same way as described in the OVPO process. As these particles emerge from the torches. They are deposited on to the end surface of a silica glass rod, which acts as a seed. A porous preform is grown in the axial direction by moving the rod upward. The rod is also continuously rotated to maintain cylindrical symmetry of the particle deposition.

As the porous Preform moves upward, it is transformed in to a solid, transparent rod preform by melting (heating in a narrow localized zone) with the carbon ring heater shown in fig. 10-4. The resultant preform can be drawn into a fiber by heating it in another furnace, as shown in fig.4.6.2.

Both step and graded index fibers in either multimode or single mode varieties can be made by the VAD method. The advantage of the VAD method are : (1) the preform has no central hole as occurs in the OVPO process: (2) the preform can be fabricated in continuous lengths which can affect process costs and product yields; and (3) the fact that the deposition chamber and the zone melting ring heater are tightly connected to each other in the same enclosure allows the achievements of a clean environment

.

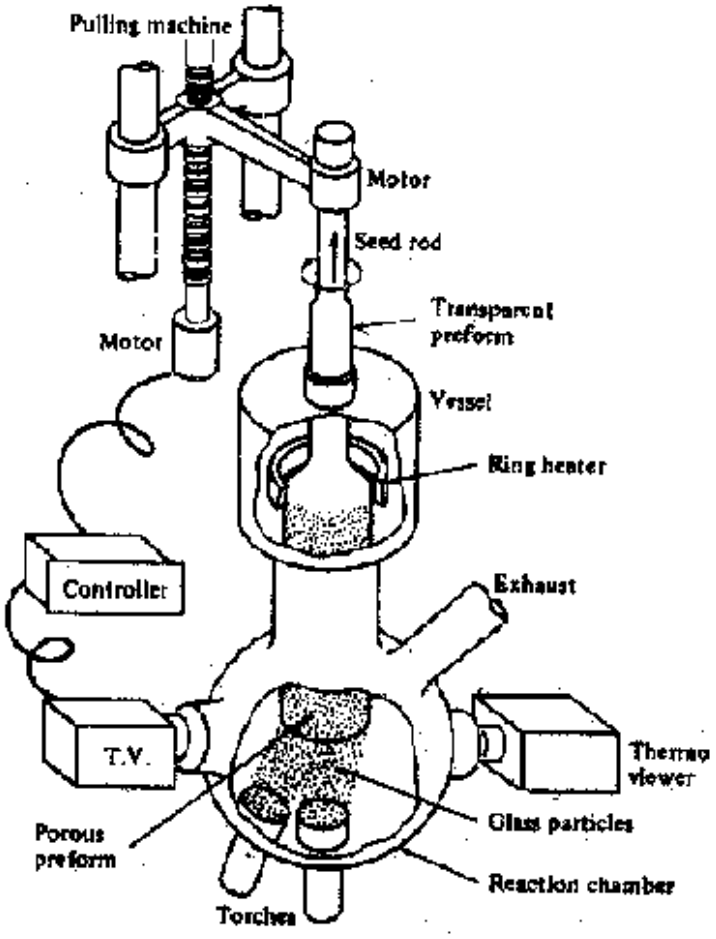


Fig 4.6.4. Apparatus used for the VAD (vapour phase axial deposition) process

.4.6.2.3. MODIFIED CHEMICAL VAPOUR DEPOSITION

The modified chemical vapour deposition (MCVD) process shown in fig. 4.6.5. was pioneered at Bell laboratories and widely adopted else where to produce very low loss graded index fibers. The glass vapor particles arising from the reaction of the constituent metal halide gases and oxygen flow through the inside of a revolving silica tube. As the SiO_2 particles are deposited, they are sintered to a clear glass layer by an oxyhydrogen torch, which travels back and forth along the tube. When the desired thickness of glass has been deposited, the vapour flow is shut off and the tube is heated strongly to cause it to collapse into a solid rod preform. The fiber that is subsequently drawn from this perform rod will have a core that consists of the vapour-deposited material and a cladding that consists of the original silica tube.

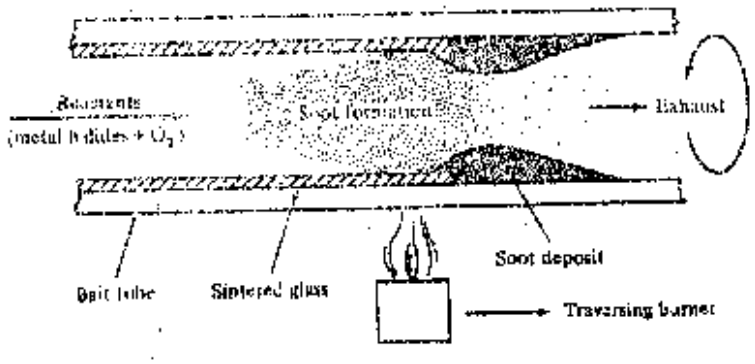


fig. 4.6.5 Schematic view of MCVD (modified chemical vapour deposition) process

4.6.2.4. Plasma-Activiated Chemical Vapour Deposition

Philips Research Scientists invented the Plasma-Activiated Chemical Vapour Deposition (PCVD) process. PCVD method is similar to the MCVD process in that deposition occurs within a silica tube. However, a nonisothermal plasma operating at low pressure initiates the chemical reaction. With the silica tube held at temperatures in the range 1000-1200 ^{o}C to reduce mechanical stresses in the growing glass films, a moving microwave resonator operating at 2.45 GHz generates a plasma inside the tube to activate the chemical reaction. Thus, no sintering is required. When one has deposited the desired glass thickness, the tube is collapsed into a preform just as in the MCVD case.

Figure 4.6.6.: Schematic of PCVD process

4.6.2.5. Double-Crucible Method

Multi component glasses are generally drawn into fibers by using a direct-melt technique. In this method, glass rods for the core and cladding materials are first made separately melting mixtures of purified powders to make the appropriate glass composition. These rods are then used as feed stock for each of two concentric crucibles, as shown in Fig. 4.6.7. The inner crucible contains molten core glass and the outer one contains the cladding glass. The fibers are drawn from the molten state through orifices in the bottom of the two concentric crucibles in a continuous production process.

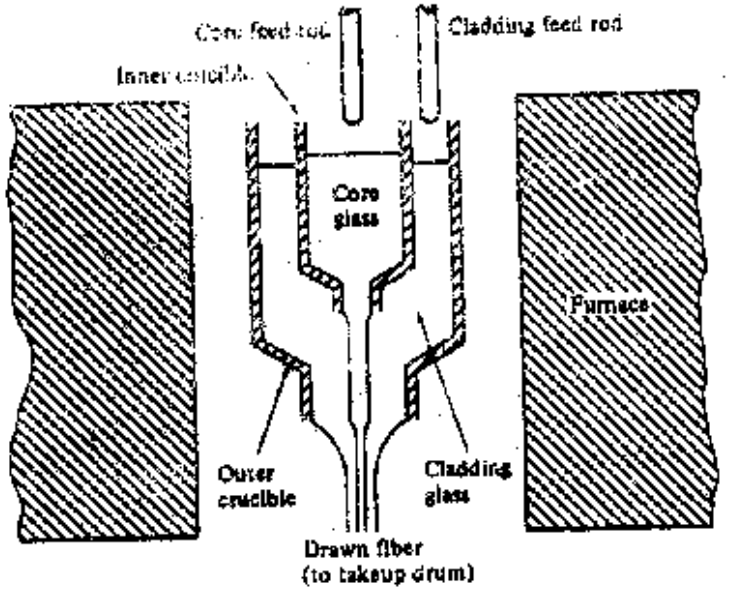


Fig 4.6.7. Double crucible arrangement for drawing fibers from the molten state of glass.

Although this method has the advantage of being a continuous process, careful attention must be paid to avoid contaminants during the melting process. The main sources of contamination arise from the furnace environment and from the crucible. Silica crucibles are usually used in preparing the glass seed rods, whereas the concentric double crucibles used in the drawing furnace are made from platinum.

4.6.3. MECHANICAL PROPERTIES OF FIBERS

In addition to the transmission properties of optical wavedguides, their mechanical characters play a very important role when they are used as the transmission medium in optical communication systems. Fibers must be able to withstand the stresses and strains that occur during the cabling process and the loads induced during the installation and service of the fiber can be either impulsive or gradually varying. Once the cable is in place, the service loads are usually slowly varying ones, which can arise from temperature variations or a general settling of the cable following installation.

Strength and static fatigue are the two basic mechanical characteristics of glass optical fibers. Since the sight and sound of shattering glass are quite familiar, one intuitively suspects that glass is not a very strong material. However, the longitudinal breaking stress of pristine glass fibers is comparable to that of metal wires. The cohesive bond strength of the constituent atoms of a glass fiber governs its theoretical intrinsic strength. Maximum tensile strengths of 14 Gpa (2×10^6 lb/in^2) have been observed in short-gauge-length glass fibers. This is close to the20-Gpa tensile strength of steel wire. The difference between glass and metal is that, under an applied stress, glass will extend elastically up to its breaking strength, whereas metals can be stretched plastically well beyond their true elastic range. Copper wires, for example, can be elongated plastically by more than 20 percent before they fracture. For glass fibers elongations of only about one percent are possible before fracture occurs.

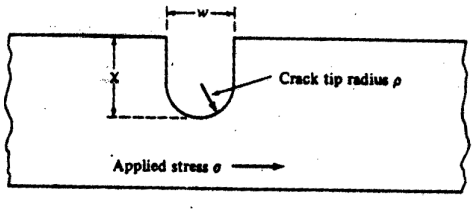


Fig 4.6.8. A hypothetical model of a microcrack in an optical fiber

In practice the existence of stress concentrations at surface flaws or micro-cracks limits the median strength of long glass fibers to the 700-to 3500-Mpa (1 to 5×10^5 lb/in^2) range. The fracture strength of a given length of glass fiber is determined by the size and geometry of the severest flaw (the one that produces the largest stress concentration) in the fiber. A hypothetical, physical flaw model is shown in fig.4.6.8.. This elliptically shaped crack is generally referred to as a Griffith microcrack. It has a width w, a depth χ and a tip radius ρ. The strength of the crack for silica fibers follows the relation.

$$K = Y\chi^{1/2}\sigma \quad \text{.....(4.6.1)}$$

Where the stress intensity factor K is given in terms of the stress σ in megapascals applied to the fiber, the crack depth is given in millimeters, and Y is a dimensionless constant that depends on flaw geometry. For surface flaws, which are the most critical in glass fibers, $Y = \sqrt{\pi}$. From this

equation the maximum crack size allowable for a given applied stress level can be calculated. The maximum values of k depend upon the glass composition but tend to be in the range of 0.6 to 0.9 $MN/m^{3/2}$

Since an optical fiber generally contains many flaws having a random distribution of size, the fracture strength of a fiber must be viewed statistically. If $F(\sigma,L)$ is defined as the cumulative probability that a fiber of length L will fail below a stress level σ then, under the assumption that the flaws are independent and randomly distributed in the fiber and that the fracture will occur at the most severe flaw, we have

$$F(\sigma, L) = 1 - e^{-LN(\sigma)} \quad \text{.....(4.6.2)}$$

Where $N(\sigma)$ is the cumulative number of flaws per unit length with a strength less than σ. A widely used form for $N(\sigma)$ is the empirical expression proposed by weibull

$$N(\sigma) = \frac{1}{L_0}\left(\frac{\sigma}{\sigma_0}\right)^m \quad \text{.....(4.6.3)}$$

Where m, σ_0, L_0 are constants related to the initial inert strength distribution. This leads to the so-called Weibull expression.

$$F(\sigma, L) = 1 - \exp\left[-\left(\frac{\sigma}{\sigma_0}\right)^m \frac{L}{L_0}\right] \quad \text{.....(4.6.4)}$$

A plot of a weibull expression is shown in fig. 4.6.9. for measurements performed on long-fiber samples. These data were obtained by testing to destruction a large number of fiber samples. The fact that a single curve can be drawn through the data indicates that the failures arise from a single type of flaw. Earlier works showed a double-curve weibull distribution with different slopes for short fiber manufacturing process and the other from fundamental flaws occurring in the glass perform and the fiber. By careful environmental control of the fiber-drawing furnace, numerous 1-km lengths of silica fiber having a single failure distribution and a maximum strength of 3500 Mpa have been fabricated.

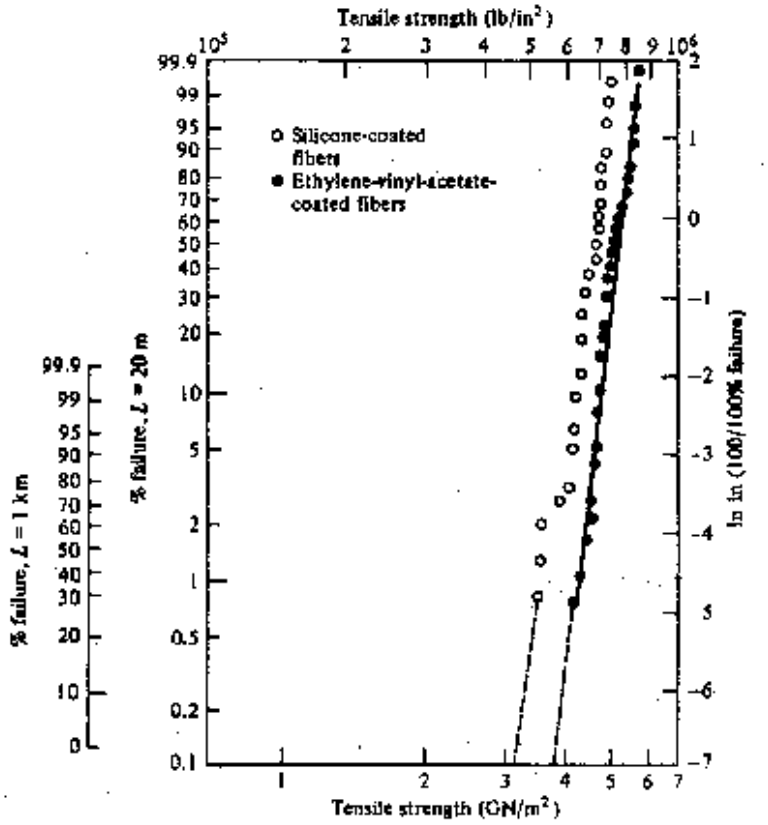


Fig 4.6.9.: A weibull plot showing the cumulative probability that fiber of 20m and 1km lengths will fracture at the indicated applied stresses.

In contrast to strength, which deals with instantaneous failure under an applied load, static fatigue relates to the slow growth of preexisting flaws in the glass fiber under humid conditions and tensile stress. This gradual flaw growth causes the fiber to fail at a lower stress level than that which could be reached under a strength test. A flaw such as the one shown in fig. 4.6.8. propagates through the fiber because of chemical erosion of the fiber material at the flaw tip. The primary cause of this erosion is the presence of water in the environment, which reduces the strength of the bonds in the glass. The speed of the growth reaction is increased when the fiber is put under stress. However based on experimental investigations, it is generally believed (but not yet fully substantiated) that static fatigue does not occur if the stress level is less than approximately 0.20 if the inert strength (in a dry environment, such as a vacuum.). Certain fiber materials are more resistant to static fatigue than others, with fused silica being the most resistant of the glasses in water. In general, coatings, which are applied to the fiber immediately during the manufacturing process, afford a good degree off protection against environmental corrosion.

Another important factor to consider is dynamic fatigue. When an optical cable is being installed in a duct, it experiences repeated stress owing to surging effects. The surging is caused by varying degrees of friction between the optical cable and the duct or guiding tool in a manhole

on a curved route. Varying stress also arise in aerial cables that are set into transverse vibration by the wind. Theoretical and experimental investigations have shown that the time to failure under these conditions is related to the maximum allowable stress by the same lifetime parameters that are found from the cases of static stress and stress that increases at a constant rate.

A high assurance of fiber reliability can be provided by proof testing. In this method, an optical fiber is subjected to a tensile load greater than that expected at any time during the cable manufacturing, installation, and service. Any fibers, which do not meet the proof test, are rejected. Empirical studies of slow crack growth show that the growth rate $d\chi/dt$ is approximately proportional to a power of the stress intensity factor, that is

$$\frac{d\chi}{dt} = AK^b \quad \text{.....(4.6.5)}$$

Here A and b are material constants and the stress intensity factor is given by Eq(4.6.1). For most glasses b ranges between 15 to 50.

If a proof test stress σ_0 is applied for a time t_s, then from Eq(4.6..5) we have

$$B\left(\sigma_i^{b-2} - \sigma_p^{b-2}\right) = \sigma_p^b t_p \quad \text{.....(4.6.6)}$$

Where σ_0 is the initial inert strength and

$$B = \frac{2}{b-2}\left(\frac{K}{Y}\right)^{2-b} \frac{1}{AY^b} \quad \text{.....(4.6.7)}$$

When this fiber is subjected to a static stress σ_0 after proof testing, the time to failure t_s is found from eq. (4.6.5) to be

$$B\left(\sigma_p^{b-2} - \sigma_s^{b-2}\right) = \sigma_s^b t_s \quad \text{.....(4.6.8)}$$

Combining Eqs. (4.6.6) and (4.6.8) yields

$$B\left(\sigma_i^{b-2} - \sigma_s^{b-2}\right) = \sigma_p^b t_p + \sigma_s^t t_s \quad \text{.....(4.6.9)}$$

To find the failure probability F, of a fiber after a time t_s after proof testing, we first define $N(t, \sigma)$ to be the number of flaws per unit length which will fail in a time t under an applied stress σ. Assuming that $N(\sigma_i) >> N(\sigma_s)$, then,

$$N(t_s, \sigma_s) \approx N\left(\sigma_i\right) \quad \text{.....(4.6.10)}$$

Solving Eq. (4.6.9) for σ_I and substituting into Eq. (4.6.3), we have from Eq.(4.6.10),

$$N(t_s, \sigma_s) = \frac{1}{L_0}\left\{\frac{\left[\left(\sigma_p^b t_p + \sigma_s^b t_s\right)/B + \sigma_s^{b-2}\right]^{1/b-2}}{\sigma_0}\right\}^m \quad \text{.....(4.6.11)}$$

The failure number $N(t_p, \sigma_p)$ per unit length during proof testing is found from Eq. (4.6.11) by setting $\sigma_s = \sigma_p$ and letting $t_s = 0$, so that

$$N(t_p,\sigma_p)=\frac{1}{L_0}\left[\frac{\left(\sigma_p^b t_p/B+\sigma_p^{b-2}\right)^{1/b-2}}{\sigma_0}\right]^m \quad \text{.....(4.6.12)}$$

Letting $N(t_x, \sigma_x)= N_x$ the failure probability F_s for a fiber after it has been proof-tested is given by

$$F_s=1-e^{-L(N_s-N_p)} \quad \text{.....(4.6.13)}$$

Substituting Eqs. (4.6.11) and (4.6.12) into Eq. (4.6.13), we have

$$F_s=1-\exp\left(-N_p L\left\{\left[\left(1+\frac{\sigma_s^b t_s}{\sigma_p^b t_p}\right)\frac{1}{1+C}\right]^{m/b-2}-1\right\}\right) \quad \text{.....(4.6.14)}$$

Where $C=B/\sigma_p^2 t_p$ and where we have ignored the term

$$\left(\frac{\sigma_s}{\sigma_p}\right)^b \frac{B}{\sigma_s^2 t_p} \ll 1 \quad \text{.....(4.6.15)}$$

This holds since typical values of the parameters in this term are $\sigma_x/\sigma_p \approx 0.3$ to 0.4, $t_p \approx 10$ s, $b > 15$, $\sigma_p = 350$ MN/m^2, and $B \approx 0.05$ to 0.5 (MN/m^2)^2 s

The expression for F_s given by Eq. (4.6.14) is valid only when the proof stress is unloaded immediately, which is not the case in actual proof testing of optical fibers. When the proof stress is released within a finite duration, the C value should be rewritten as.

$$C=\gamma\frac{B}{\sigma_p^2 t_p} \quad \text{.....(4.6.16)}$$

where γ is a coefficient of slow-crack growth effect arising during the unloading period.

4.6.4. FIBER OPTIC CABLES

In any practical application of optical waveguide technology, the fibers need to be incorporated in some type of cable structure. The cable structure will vary greatly, depending on whether the cable is to be pulled into underground or intra-building ducts, buried directly in the ground, installed on outdoor poles, or submerged under water. Different cable designs are required for each type of application, but certain fundamental cable design principles will apply in every case. The objectives of cable manufacturers have been that the optical fiber cables should be installable with the same equipment, installation techniques, and precautions as those used in conventional wire cables. This requires special cable designs because of the mechanical properties of glass fibers.

One important mechanical property is the maximum allowable axial load on the cable since this factor determines the length of cable that can be reliably installed. In copper cables the

wires themselves are generally the principal load-bearing members of the cable, and elongations of more than 20 percent are possible without fracture. On the other hand, extremely strong optical fibers tend to break at 4 percent elongation, whereas typical good-quality fibers exhibit long-length breaking elongations of about 0.5 to 1.0 percent. Since static fatigue occurs very quickly at stress levels above 40 percent of the permissible elongation and very slowly below 20 percent of the breaking limit, fiber elongations during cable manufacture and installation should be limited to 0.1 to 0.2 percent.

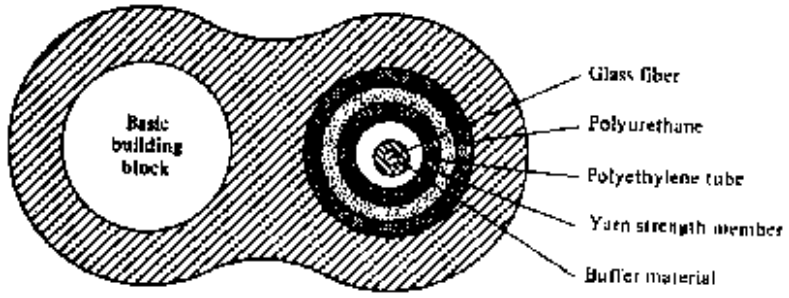


Fig 4.6.10. A hypothetical two fiber cable design. The basic building block on the left is identical to that shown for the right hand fiber.

Steel wire, which has a young's modulus of 2x104Mpa, has been extensively used for reinforcing conventional electric cables and can also be employed for optical fiber cables. For some applications it is desirable to use nonmetallic constructions, either to avoid the effects of electromagnetic induction or to reduce cable weight. In this case plastic strength members and high-tensile-strength organic yarns such as Kelvar (a product of the Dupont Chemical Corporation) are used. With good fabrication practices the optical fibers are isolated from other cable components, they are kept close to the neutral axis of the cable, and room is provided for the fibers to move when the cable is flexed or stretched.

Another factor to consider is fiber brittleness. Since glass fibers do not deform plastically, they have a low tolerance for absorbing energy from impact loads. Hence, the outer sheath of an optical cable must be designed to protect the glass fibers inside from impact forces. In addition, the outer sheath should not crush when subjected to side forces, and it should provide protection from corrosive environmental elements. In underground installations, a heavy-gauge-metal outer sleeve may also be required to protect against potential damage from burrowing rodents, such as gophers.

In designing optical fiber cables, several types of fiber arrangements are possible and a large variety of components could be included in the construction. The simplest designs are one- or two-fiber cables intended for indoor use. In a hypothetical two-fiber design shown in Fig 4.6.10, a fiber is first coated with a buffer material and placed loosely in a tough, oriented polymer tube, such as polyethylene. For strength purposes this tube is surrounded by strands of aramid yarn which, in turn, is encapsulated in a polyurethane jacket. A final outer jacket of polyurethane, polyethylene, or nylon binds the two encapsulated fiber units together. Larger cables can be

created by stranding several basic fiber building blocks (as shown in Fig.4.6.10) around a central strength member. This is illustrated in Fig 10-10 for a six-fiber cable. The fiber units are bound onto the strength member with paper or plastic binding tape, and then surrounded by an outer jacket. If repeaters are required along the route where the cable is to be installed, it may be advantageous to include wires within the cable structure for powering these repeaters. The wires can also be used for fault isolation or as an engineering order wire for voice communications during cable installation.

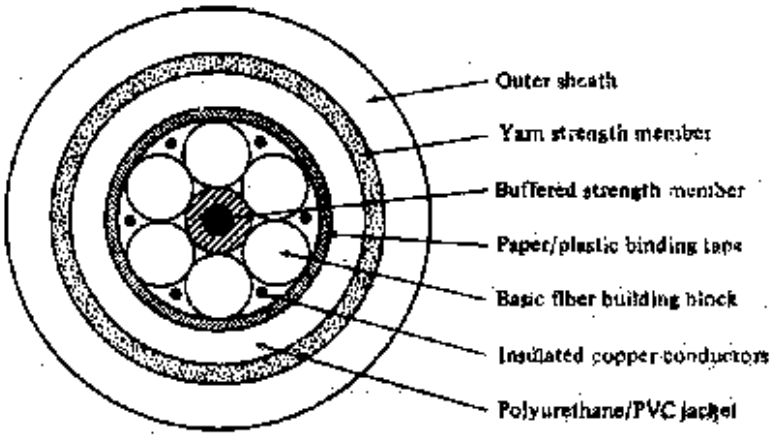


Fig 4.6.11.: A typical six-fiber cable created by stranding six basic fiber blocks around a central member.

4.6.1. **Summary:**

In this lesson, different materials used for optical fibers, different fabrication process for optical fibers are given. The characteristic mechanical properties are also given. Finally, the prepared optical fibers can be used actually. So their cabling is mentioned.

4.6.6. **Keywords:** Fiber materials – fiber fabrication techniques – mechanical properties of optical fibers, cabling

4.6.7. Self- assessment questions

1. Give in detail about the materials used for optical fibers.
2. Explain in detail about the different fabrication techniques for the optical fibers. Mention about the merits.
3. Write the mechanical properties of optical fibers.
4. Explain the different cabling of optical fibers.

4.6.8. Text books

1. “Optical fiber communications” by G. Keisser, McGraw-Hill International Edition, 2000, Third edition, and also see first edition.
2. “Optical fiber systems: Technology, design, and applications” by Charles K. Kao, McGraw -Hill, 1986.