COURSE GUIDE

PHY406
OPTICS III

Course Team DR. A. B. Adeloye & Dr. M. A. Chendo
(Course Developers)- University of Lagos
DR. M. Ohakwere-Eze (Course Reviewer)-NOUN

NOUN

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
University Village
Plot 91, Cadastral Zone
Nnamdi Azikiwe Expressway
Jabi, Abuja

Lagos Office
14/16 Ahmadu Bello Way
Victoria Island, Lagos

e-mail: centralinfo@nou.edu.ng
URL: www.nou.edu.ng

Reviewed and printed 2023, 2024

ISBN: 978-978-058-965-3

Course Introduction

In the preceding courses in Optics (I and II), you have learnt that light is an electromagnetic wave. It exhibits polarisation, interference, diffraction, etc. As you have seen, these phenomena are well understood in terms of the wave theory of light. This block, as the title indicates, essentially deals with lasers - a source of coherent light - and their applications, particularly in the areas of photography and optical communication. Lasers owe its invention to the quantum theory according to which, light energy consists of minute packets or quanta. The invention of lasers and related developments has once again brought the field of optics to the forefront of basic research and technological applications. Widespread use of light from laser is because of its high degree of coherence, high directionality, unprecedented brightness, etc. Without fear of exaggeration, we may say that the present period in optics may be called the Laser Age. In this block, we intend to give you a flavour of the basic physical principles involved in the design and operation of lasers and also about some of its important applications.

Laser is a coherent source of light. But what is coherence? You learnt about coherent source of light in OPTICS II in connection with Young's double-slit interference experiment. It was emphasized there that for obtaining observable interference fringe pattern, the light from the slits must be coherent. In Unit 1 of this course, you will learn about the concept of coherence of waves. If the phases of two waves have a definite phase relationship, they are said to be coherent. This phase relationship between waves, which can be in time or space, gives rise to temporal coherence and spatial coherence. You will learn that temporal coherence of electromagnetic waves manifests as monochromaticity and the visibility of the interference fringe pattern indicates the extent of spatial coherence between the interfering waves.

In Unit 2, you will learn the working principle of lasers. In particular, we have discussed the concept of stimulated emission of radiation and the prerequisites for obtaining laser light. Though the first laser used ruby (solid) as active medium, now lasers which employ liquids and gases as the active medium are available. You will learn about different type of lasers. Coherence (Monochromaticity), high directionality and brightness are some of the properties of lasers which are responsible for their so many and so varied applications. In this unit, you will also learn about some of these applications.

Holography is a technique of three-dimensional photography. This technique was invented by Dennis Gabor much before the invention of lasers. However, the full potential of this technique could be realised only after this invention. In Unit 3, you will learn the details of this novel technique. Some of the applications of holography have also been discussed.

The use of lasers has revolutionised communication technology. The monochromaticity of laser light makes it an efficient carrier of information. Communication - transmission of speech, data, etc. - at optical frequency is much faster and more reliable compared to radio and microwave communication. However, optical communication suffers from the drawback that signals get attenuated by dust particles, rains, etc. Thus, for efficient terrestrial optical communication, optical fibres are used. How is light transmitted through optical fibres? What are the characteristics of such fibre? These and other related questions form the subject matter of Unit 4.

We wish you success.

MAIN COURSE

CONTENTS

MODULE 1

UNIT 1 COHERENCE

Unit Structure

1.1 Introduction

In Unit 1 of Optics II, you studied the Young's double-slit interference experiment. We emphasized that for observing interference fringe pattern, the light from two sources must be coherent. By coherence, we mean that the light waves from two slits have a constant phase relationship. Can you recall how this condition of coherence is achieved? If you are unable to do so, you should refer back to relevant pages. Now the question arises why coherence is a prerequisite for observing interference? You will learn about coherence in detail now.

In Sec. 1.2, we elaborate the concept of coherence as applied to waves in general. Further, the most elementary definition of coherence says that the phases of the coherent waves have a predictable relationship at different points and at different times in space. This space and time predictability of the phase relationship of waves gives rise to two types of coherence, namely, spatial coherence and temporal coherence. The concept of temporal coherence, which refers to the phase relationship at different times at a point, has been discussed in Sec. 1.3. You will also learn about the correlation between the width of a spectral line and temporal coherence. In Sec. 1.4, we have discussed spatial coherence which relates to the coherence of two waves travelling side by side. The relationship between the visibility of fringe pattern with spatial coherence is also discussed in detail.

1.2 Objectives

After going through this unit, you should be able to

- explain the concept of coherence
- distinguish temporal coherence from spatial coherence
- relate temporal coherence with the width of spectral lines
- relate spatial coherence with the visibility of fringe pattern, and «
- solve numerical problems based on coherence.

1.3 What Is Coherence?

If you are asked what is coherence, you may say that it is the condition necessary to produce observable interference of light. And if you are asked what is interference, you may say it is connected with the interaction of waves that are coherent. Well, nothing definite follows from such circular arguments! In fact, coherence is a property of light whereas interference is the effect of interaction of light waves. The crucial consideration in interference phenomenon is the relative phase of waves arriving at a given point from two or more sources. That is, in order to observe interference fringes, there must exist a definite phase relationship between the light waves from two sources. Hence, we may say that the necessity of having coherent sources for observing interference fringes essentially implies that the waves from the two sources must have a constant and predictable phase relationship. It is the absence of a definite phase relationship between light waves from ordinary sources that we do not obtain any observable interference fringe pattern.

Now, you may ask: Why is there no definite phase relationship between light waves from two ordinary light sources? Well, the basic mechanism of emission of light involves atoms radiating electromagnetic waves in the form of photons. Each atom radiates for a small time (of the order of 10^{-9} s). Meanwhile, other atoms begin to radiate. The phases of these emitted electromagnetic waves are, therefore, random; if there are two such sources, there can be no definite phase relationship between the light waves emitted from them.

In general, sources, and the waves they emit, are said to be coherent if they

(i) have equal frequencies,
(ii) maintain a phase difference that is constant in time.

If either of these properties is lacking, the sources are incoherent and the waves do not produce any observable interference.

Let us pause for a while and ask ourselves: Why it is a prerequisite for observing interference fringe pattern? To answer this question, let us consider the origin of the bright and dark fringes in the Young's experiment (Fig. 1.1). Let E_1 and E_2 be the electric fields associated with the light waves emanating from slits S_1 and S_2. These waves superpose and the combined electric field at any point on the screen is given by,

$$E = E_1 + E_2 \quad (1.1)$$

You may recall that in the interference pattern, we observe the intensity of light, not the electric field. Since the average intensity of light is proportional to the time-averaged value of the associated electric field, we have

$$I \propto \langle E^2 \rangle \quad (1.2)$$

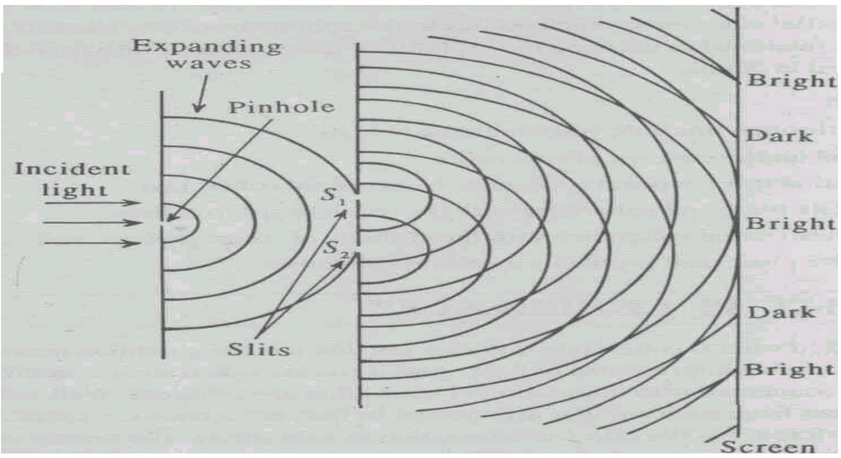

Fig. 1.1 Young's interference experiment

Thus, we have, from Eq. (1.1) and (1.2),

$$\begin{aligned} I &= \langle E_1^2 \rangle + \langle E_2^2 \rangle + 2\langle E_1 E_2 \rangle \\ &= I_1 + I_2 + 2\langle E_1 E_2 \rangle \end{aligned} \quad (1.3)$$

Eq. (1.3) shows that the resultant intensity on the screen is the sum of intensities I_1 and I_2 (due to individual slit sources) and an interference term $2\langle E_1 E_2 \rangle$. The interference term is crucial because it determines whether the resultant intensity is a uniform illumination or a fringe pattern on the screen. The contribution of the interference term to the resultant intensity depends primarily on the phase relationship between the light waves emanating from the two slits.

Let us first consider the case when the light waves are in phase at one instance and are out of phase at another instance. In such a situation, the product $E_1 E_2$ will be positive at one instance and negative at the other. As a result, the time average of $E_1 E_2$ will be zero, i.e.

$$< E_1 E_2 > = 0$$

Waves having this kind of phase relationship (varying with time) are said to be **incoherent** and the resultant intensity will be

$$I = I_1 + I_2 \qquad (1.4)$$

Thus, when light waves from two incoherent sources interfere, the resultant intensity will be the sum of individual intensities and the screen will be uniformly illuminated. To give you a simple example, when the headlights of a car illuminate the same area, their combined intensity is simply the sum of two separate intensities. The headlights are incoherent sources and there is no contribution of the interference term.

Now, what will happen if the light waves from two slits have a definite phase relationship i.e. a phase relationship which is constant in time. The source of light emitting such waves is coherent. When light sources are coherent, the resultant intensity is not simply the sum of individual intensities. It is so because in that situation, the interference term in equation (1.3) is non-zero. Let us see what is the form of the interference term when two coherent light waves superpose. There are two cases:

(a) When $E_1 = E_2$, that is, the two waves have the same amplitude, frequency and phase. Thus,

$$I_1 = < E_1^2 > = < E_2^2 > = I_2$$

and

$$2 < E_1 E_2 > = 2 < E_1^2 > = 2I_1$$

The resultant intensity,

$$I = I_1 + I_2 + 2 < E_1 E_2 >$$

$$= I_1 + I_1 + 2I_1$$

$$= 4I_1$$

Thus, at the points on the screen where two interfering waves are in phase, the resultant intensity is four times that due to an individual source. These points will, therefore, appear bright on the screen.

(b) $E_1 = -E_2$, that is, the two waves have same amplitude and frequency but their phases differ by 180° which remains constant in time. In that case, the two waves are completely out of phase and the resultant wave amplitude and intensity will be zero.

$$E = E_1 + E_2 = 0$$

$$\Rightarrow \quad I = 0$$

The points on the screen where the interfering light waves satisfy the above condition will have zero intensity and hence they will appear dark.

Thus, the constant phase relationship between superposing light waves i.e. coherence, is a necessary condition for obtaining interference fringe pattern. When the phase relationship is not constant, the points where superposing light waves arrive in phase at one instant may receive light waves which are completely out of phase at another instance. This results in uniform illumination of the screen and no interference fringe pattern can be observed.

In the above discussion, you have studied the necessity of having coherent sources for observing interference fringe pattern. As mentioned earlier, coherence, which is essentially a correlation phenomenon between two waves, can be with respect to time and/or space. Thus, for expediency, we distinguish two types of coherence: Temporal Coherence and Spatial Coherence. Temporal coherence, or the longitudinal spatial coherence (often called monochromaticity) applies to waves travelling along the same path. It refers to the constancy and predictability of phase relationship as a function of time. Spatial coherence, or transverse spatial coherence refers to the phase relationship between waves travelling side by side, at a certain distance from one another. The further apart the two waves are, the less likely they are to be in phase, and the less coherent the light will be. You will study these two types of coherences in the following sections.

1.4 Temporal Coherence

While studying interference and diffraction of light in the previous two blocks of this course, we assumed that electromagnetic waves remained perfectly sinusoidal for all time. This kind of electromagnetic waves are, however, practically impossible to obtain from ordinary light sources. Why is it so? It is because light emitted from an ordinary source consists of finite size wave trains. Each wave train is sinusoidal in itself and has a characteristic frequency (or wavelength) and phase. However, the collection of wave trains is not sinusoidal. Thus, light waves coming from an ordinary source cannot have one single frequency (monochromatic). Instead, it has a range of frequencies; that is, it has a frequency bandwidth. For these reasons, the so-called monochromatic light, such as from gas discharge tube, is more appropriately called quasi- monochromatic.

This aspect of light (i.e. monochromaticity) refers to its temporal coherence. Temporal coherence can be identified qualitatively as the interval of time during which the phase of the wave motion changes in a predictable manner as it passes through a fixed point in space. And in wave motion corresponding to light from ordinary sources, a predictable phase relationship can be observed only within the average length of the wave trains on time scale.

To elaborate the concept of temporal coherence, let us consider a typical time variation of the amplitude of an electromagnetic wave as shown in Fig. 1.2.

You may notice from the figure that the electric field at time t and $t + \Delta t$ will have a definite phase relationship if $\Delta t << \tau_c$ and will not have any phase relationship if $\Delta t >> \tau_c$ where τ_c represents the average duration of the wave trains. The time τ_c is known as coherence time of the radiation and the wave is said to be coherent for time τ_c. And the path length corresponding to τ_c, given as $L_c = c\tau_c$ is called the **coherence length** of the radiation, where c is the velocity of light.

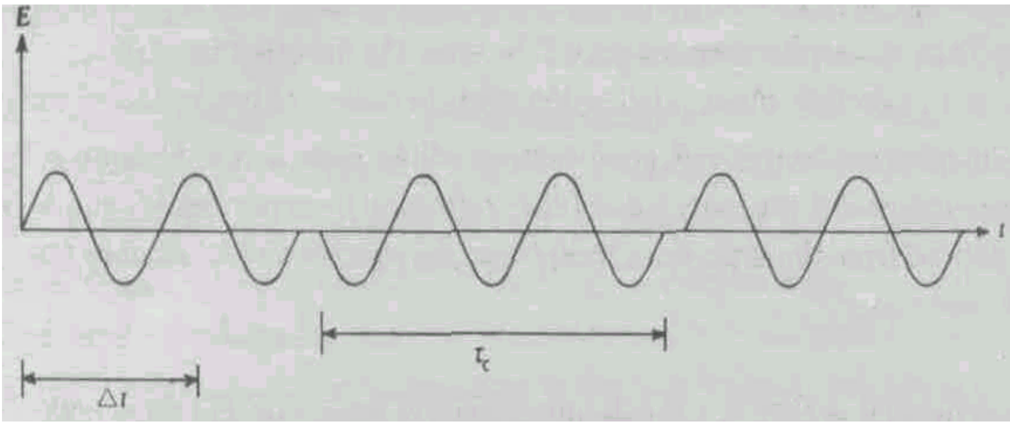

Fig. 1.2 Typical variation or the amplitude of an electromagnetic wave with time. Three typical wave trains have been shown. The coherence time τ_c is the average duration of the wave trains.

In order to study the time-coherence of the radiation, let us re-consider Michelson's interferometer experiment. For completeness, we have reproduced the experimental arrangement in Fig. 1.3. A nearly monochromatic light source is used in the investigation.

For the source (*S*) we may use a neon lamp in front of which we place a filter (*F*) so that radiation corresponding to λ = 6328 Å is allowed to fall on the beam-splitter *G*. Glass plate *G'* is the compensating plate. You may recall from Unit 3 of Optics II, if the eye is in the position as shown in the figure, circular fringes are observed due to the interference of the beams reflected from mirrors M_1 and M_2.

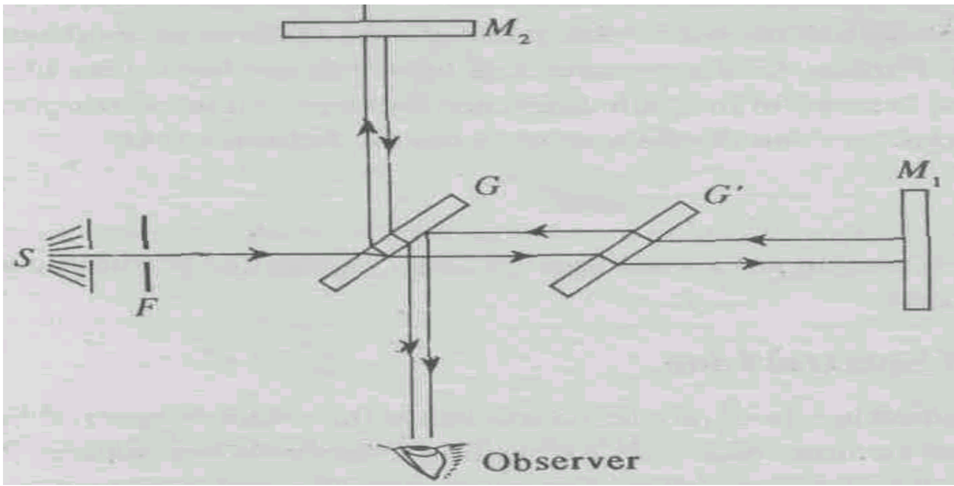

Fig. 1.3 Light paths for Michelson Interferometer

You may also recall that for obtaining these circular fringes, the mirrors should be at right angle to each other and the path difference ($GM_2 - GM_1$) should be small. If mirror M_2 is moved away from the beam splitter *G*, the visibility, and hence the contrast of the interference fringes will become poorer, and, eventually, the fringe pattern will disappear. Why does it happen? Does disappearance of interference fringes has to do something with temporal coherence of light waves from neon lamp? Yes, it is so. The disappearance of the fringes is due to the following phenomenon. When mirror M_2 is moved through a distance d, an additional path $2d$ is introduced for the beam which gets reflected by M_2. As a result, the beam reflected from M_2 interferes with the one reflected from M_1 which had originated $(2d/c)$ s, where c is the velocity of light, earlier from the light source. Clearly, if this time delay $(2d/c)$ is greater than the coherence time (τ_c) of the radiation from the source, the waves reaching the eye after reflection from mirrors M_1 and M_2 will not have any definite phase relationship. In other words, the waves reflected from mirrors M_1 and M_2 are incoherent. Thus, no interference fringes will be seen. On the other hand, if $(2d/c) << \tau_c$, a definite phase relationship exists between the two reflected waves and hence interference fringes with good contrast will be seen. It is so because in this case, we are superposing two wave trains (after reflection from mirrors M_1 and M_2) which are derived from the same wave train (from the source) and hence they are temporally coherent.

For the neon light (λ = 6328 Å), the disappearance of fringes occurs when the path difference between the reflected waves from mirrors M_1 and M_2, is about a few cm. This path difference, $L_c = c\tau_c$ is known as coherence length. Hence for the neon line, $\tau_c \sim 10^{-10}$ s. For commercially available lasers, the coherence length exceeds a few kilometres. Thus, if light beam from a laser be used in the above experiment, we can observe interference fringes for *d* as long as a few kilometres (provided, of course, we have such a big laboratory!).

In short, if the two paths, GM_1 and GM_2 in Fig. 1.3 are equal in length, the fringes have maximum contrast, hence a maximum temporal coherence. If they are not of equal length then the contrast is less. Hence temporal coherence is less. Temporal coherence is, therefore, inversely proportional to the magnitude of the path difference and directly proportional to the length of the wave train. The wave trains are of finite length; each containing only a limited number of waves. The length of a wavetrain is, therefore, the product of the number of waves, *N*, contained in a wave train and of its wave length λ, so $L_c = \lambda N$. Since visibility or the contrast of the interference fringes is directly proportional to the length of the wave train, it can also be taken as proportional to the product of *N* and λ. Further, for a given source of light, you can have some idea about its temporal coherence in terms of the path difference between two interfering waves of the Michelson interferometer. You should now work out the following SAQ.

SAQ 1

If light of 660 nm wavelength has a wavetrain 20λ long, what is its (a) coherence length and (b) coherence time.

1.5 Width of Spectral Line

You might have studied in school physics course about the origin of spectral lines. You may recall that when an atom undergoes a transition from an excited state to the ground state, it emits electromagnetic radiation. The energy (and hence frequency) of the radiation is equal to the difference in energies of the excited and the ground state. Each substance has a unique set of energy state to which its atoms can be excited. Each substance, therefore, has a characteristic set of energy values (and hence frequencies) for the emitted radiations. This set of frequency values constitutes the spectrum of the substance.

Due to one of the fundamental principles of quantum mechanics, namely, the uncertainty principle, the lines in the spectrum are not sharp i.e. corresponding to each spectral line, there is a continuous distribution of frequency in a narrow frequency interval. This narrow frequency or wavelength interval is known as the width of the spectral line. For example, for Cd red line, the width of this interval is about 0.007 Å.

You may now be interested in knowing what determines the width of the spectral lines? Is the width of spectral lines related to temporal coherence? Yes, temporal coherence of the source of light is intimately related to the width of its spectral lines. To see how, let us again consider the interference fringes obtained by Michelson interferometer. You may recall from Unit 3 of Optics II, that Michelson's Interferometer can be used for the measurement of two closely spaced wavelengths. Let us consider a sodium lamp source which emits predominantly two closely spaced wavelengths, λ_1 = 5896 Å and λ_2 = 5890 Å. Now, you may recall from Unit 3 of Optics II that near d = 0, the fringe patterns corresponding to both the wavelengths will overlap. If the mirror is moved away from the plate G by a distance d, Fig. 1.3, the maxima corresponding to the wavelength λ_1 will not, in general, occur at the same angle as for λ_2. It is so because the spacing between the fringes for λ_1 and λ_2 will be different. Indeed, if the distance d is such that the bright fringe corresponding to λ_1 coincides with the dark fringe corresponding to λ_2, we have

$2d = m\lambda_1$ (bright fringe) (1.4a)

and

$2d = \left(m + \frac{1}{2}\right)\lambda_2$ (dark fringe) (1.4b)

and the fringe system will disappear. The condition for disappearance of fringe pattern can, therefore, be expressed as (see box remark below).

> If 'd" is the distance through which one of the mirrors has been moved, the effective path difference will be *2d*. And the condition for bright fringe is
>
> $2d = m\lambda$
>
> and the condition for dark fringe is
>
> $2d = (m + 1/2)\lambda$
>
> where m is an integer.
>
> $2d = m\lambda_1 \Rightarrow m = 2d/\lambda_1$
>
> and
>
> $2d = (m + 1/2)\lambda_2 \Rightarrow (2d/\lambda_2) - 1/2$
>
> $\therefore \frac{2d}{\lambda_2} - \frac{2d}{\lambda_1} = \frac{1}{2}$
>
> $\frac{2d}{\lambda_2} - \frac{2d}{\lambda_1} = \frac{1}{2}$
>
> $\Rightarrow \quad 2d = \frac{\lambda_1\lambda_2}{2(\lambda_1 - \lambda_2)} = \frac{\lambda^2}{2(\lambda_1 - \lambda_2)}$ (1.5)
>
> Since $\lambda_1 \approx \lambda_2$

Now, if we assume that the light beam consists of all wavelengths lying between λ and $\lambda + \Delta\lambda$, instead of two discrete values λ_1 and λ_2, fringes will not be observed if

$$2d \approx \frac{\lambda^2}{\Delta\lambda} \quad (1.6)$$

To arrive at equation (1.6) you should solve the following SAQ.

SAQ 2
Starting from Eq. (1.5) which gives the path difference (2*d*), in terms of two distinct wavelengths λ_1 and λ_2, for which fringes will disappear, derive Eq. (1.6) which is for all wavelengths lying between λ and $\lambda + \Delta\lambda$.

Now, can you see the basic reason why the fringe pattern disappears? Is it somehow related to the non-monochromaticity of the light beam? Yes, it is so. If fact, the moment we consider that the light beam consists of all wavelengths lying between λ and $\lambda + \Delta\lambda$, we are essentially considering the interference pattern produced by non-monochromatic light beam. You may notice from equation (1.6) that as the spread in the wavelength ($\Delta\lambda$) becomes small (more and more monochromatic), the path difference (*2d*) for disappearance of fringes becomes large. And as mentioned earlier, the larger the value of path difference for which fringe pattern does not disappear, the more temporally coherent the light beam is. In other words, monochromaticity or the sinusoidal nature of light beam is strongly related to its temporal coherence. The temporal coherence of the beam is, therefore, directly associated with the width of the spectral line. Since no fringe pattern is observed if the path difference, 2*d*, exceeds the coherence length, L_c, we may assume that the beam consists of all the wavelengths lying between λ and $\lambda + \Delta\lambda$ with

$$\Delta\lambda = \frac{\lambda^2}{L_c} \qquad (1.7)$$

This gives the relation between coherence length and spread in the wavelength of a light beam. Further, since $\nu = c/\lambda$, the spread in frequency $\Delta\nu$ is

$$\Delta\nu \approx \frac{c}{\lambda^2}\Delta\lambda$$

$$= c / L_c$$

And, the coherence time is defined as, $\tau_c = L_c / c$. Therefore, we have

$$\Delta f \sim 1/\tau_c \qquad (1.8)$$

Thus the frequency spread of a spectral line is of the order of the inverse of the coherence time.

In this section, we discussed about temporal (or longitudinal spatial) coherence which relates the predictability or constancy of the phase relationship between two waves arriving at the same point after traversing different optical paths. In other words, we talked about the constancy of phases of waves travelling along the same line. Light beam was considered as a series of wave trains. As per requirement of temporal coherence, if these wave trains are to produce observable interference fringe pattern, they must (a) have the same frequency and (b) overlap at the point of observation (i.e. path difference should be less than the coherence length). Now, what about the phase relationship between two waves travelling side by side at a certain distance from each other? Well, the constancy of the phase relationship of such waves relates to another type of coherence called spatial (or transverse spatial) coherence. This is the subject matter of the next section.

1.6 Spatial Coherence

In unit 2 of Optics II, you studied Young's double-slit experiment for obtaining interference fringe pattern. You may recall that one of the prerequisites for observing the interference pattern was that the source of light should be a point source. Can you say why this condition was imposed? What will happen if, instead of a point source, an extended source of light is used? These are some of the issues which relate to the spatial coherence about which we will study now.

You are aware that in an extended conventional source of light, the radiations emitted from different parts are independent of each other, and in that sense, such sources may be thought of as incoherent. But our interest is not so much in the nature of the source itself as in the quality of the illumination field it produces, for example, in a plane at some distance from the source. Thus, in Young's experiment we are interested in the extent to which there is a constant phase relationship between S_1 and S_2, Fig. 1.4a, so that interference effects can be observed. In other words, we are interested in examining the effect of the finite size of the source, S, on the interference pattern. 12

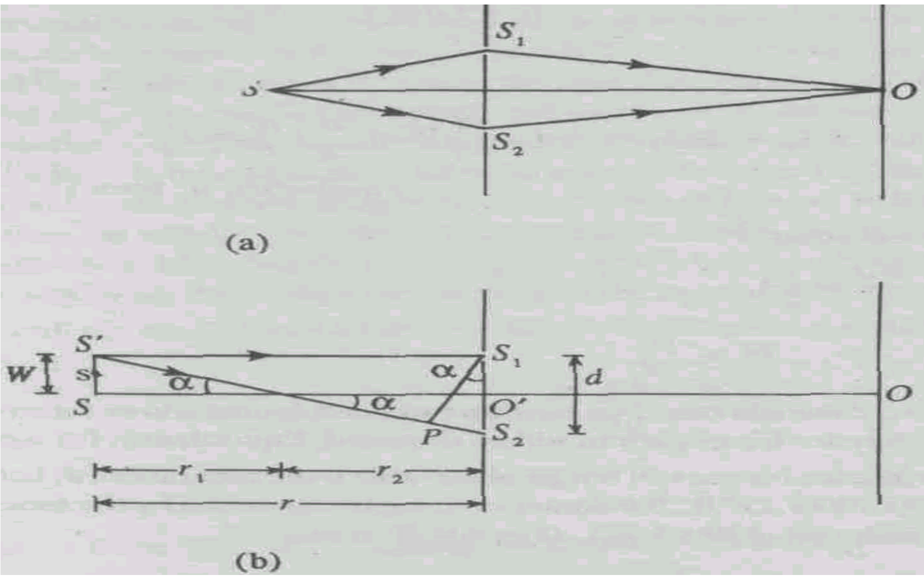

Fig. 1.4 (a) Young's double slit experiment with a point source, S, (b) Young's double silt experiment with an extended source S ' S

In order to understand the effect of an extended source (and hence of spatial coherence) on the interference fringes, let us consider Young's double-slit experiment with an extended source. Fig. 12.4b shows schematically the two slits S_1 and S_2 with an extended light source S' S of width W at a distance r. Light from some point s in the source illuminates the slits, and interference fringes are produced on the screen. If the source consisted of just this single point s (as in an idealised Young's experiment, Fig. 1.4a), the fringes of maximum visibility would have been observed. A real source (such as S ' S in Fig. 1.4b) is, however, of finite size and the fringes produced by illumination from other points of the source are displaced relative to those due to s. Light from the extended source, therefore, produces a spread in fringes with a consequent reduction in the visibility of the fringe pattern.

In order to have some quantitative idea about the spatial coherence, let us assume that the two extreme points of the extended source (Fig. 1.4b), S' and S act as two independent sources. Each source will produce its own interference pattern. Let us assume that $SS_1 = SS_2$ and the point O is such that $S_1O = S_2O$. Clearly the point source S will produce a maximum around O. On the other hand, the intensity at O due to S' will depend on the path length ($S' S_2$ - $S' S_1$). You may recall from Unit 2 of Optics II, that if this path difference

$$S'S_2 - S'S_1 = \lambda/2 \tag{1.9}$$

the minima of interference pattern due to S will fall on the maxima of that due to S'. As a result, there will not be any observable interference pattern. From Fig. 1.4b, we have

$$S'S_2 - S'S_1 = S_2P = \theta d$$

But

$$\theta = \frac{d/2}{r_2} = \frac{W}{r_1}$$

Thus,

$$r = r_1 + r_2 = \frac{1}{\theta}\left(W + \frac{d}{2}\right)$$

or

$$\theta = \frac{W + (d/2)}{r}$$

Therefore,

$$S'S_2 - S'S_1 = \theta d = \left(W + \frac{d}{2}\right)\frac{d}{r} = \frac{Wd}{r}$$

(neglecting d^2 term)

Thus, no fringes will occur if

$$\frac{Wd}{r} = \frac{1}{2}\lambda$$

$$W = \frac{r\lambda}{2d}$$

(1.10)

> For every point on an extended source of extension *r/d*, there is a point at a distance *r/2d* which produces interference fringes separated by half a fringe width. Thus, for sources of such an extension, the visibility of the fringes would be poor.

We may, therefore, conclude that if we have an extended source whose linear dimension is $\sim \lambda r/d$, no interference fringe pattern will be observed. Equivalently, for a given source of width *W*, interference fringes will not be observable if the separation, *d*, between slits S_1 and S_2 is greater than $\lambda r/W$. If θ denotes the angle subtended by the source $(S'S)$ at the point *O'* (midpoint of slits $S_1 S_2$), then $\theta = W/r$. So,

$$d = \frac{\lambda}{\theta}$$

(1.11)

which gives the maximum lateral distance between slits S_1 and S_2 such that the light beam from the extended source may be assumed to have some degree of coherence (i.e., the light waves from an extended source, after passing through slits S_1 and S_2 are able to produce interference fringes). The quantity A/0 is known as **Lateral** (or **transverse**)

Coherence Width and is denoted by l_w. You may note that the coherence width is linear in dimension and is approximately perpendicular to the direction of wave propagation. By contrast, the coherence length, introduced in relation to temporal coherence, is along the direction of wave propagation. For this reason, temporal coherence is sometimes called longitudinal coherence and spatial coherence is sometimes called lateral coherence.

Further, closely related to coherence width is a parameter called coherence area given as

$$a_c = \square\,(l_w/2)^2$$

$$= \square\,(\square/2\square)^2 \qquad (1.12)$$

The waves at any two points within the coherence area are coherent. You may have noticed that Eqs. (1.11) and (1.12) apply to the case in which the extended source is essentially a uniform linear source. If the source is in the form of a uniform circular disc, the lateral coherence width is given as

$$l_w = 1.22\square\,/\square \qquad (1.13)$$

Well, in order to recapitulate what you have studied in this section, how about solving an SAQ!

SAQ3

Suppose we set up Young's experiment with a small circular hole of diameter 0.1 mm in front of a sodium lamp (□ = 589.3nm) source. If the distance from the source to the slits is 1m, how far apart will the slits be when the fringe pattern disappears?

1.7 Angular Diameter of Stars

Now, let us consider an application of the concept of spatial coherence. In the preceding paragraphs, we have seen that the angle subtended by the extended source at the midpoint of the slit separation is related to the lateral coherence width (l_w). Also for a critical value of l_w the interference fringes will disappear. If, instead of an ordinary extended source of light, we consider a terrestrial extended source such as a star, you may like to know: Is it possible to know its angular diameter (i.e. the angle subtended by the star on the slits) by observing the disappearance of fringes? Indeed, it is possible. For measuring the angular diameter of a star, Young's double slit experiment set-up needs modification. Modification in the experimental set-up is necessitated because, for such an arrangement, if we take a typical value of the angular diameter of a star as ~ 10^{-7} radians, the distance *d* between the slits for which fringes disappear will be

$$d = \frac{1.22\lambda}{\theta} = \frac{1.22 \times 5 \times 10^{-7}}{10^{-7}}$$

$$= 6 \text{ m}$$

And for such a large value of *d*, the fringe width will be too small.

To overcome this difficulty, Michelson used an ingenious technique. He achieved an effectively large value of *d* by using two movable mirrors *M* and *M '* as shown in Fig. 1.5. This modified interferometer is known as

Michelson's Stellar Interferometer.

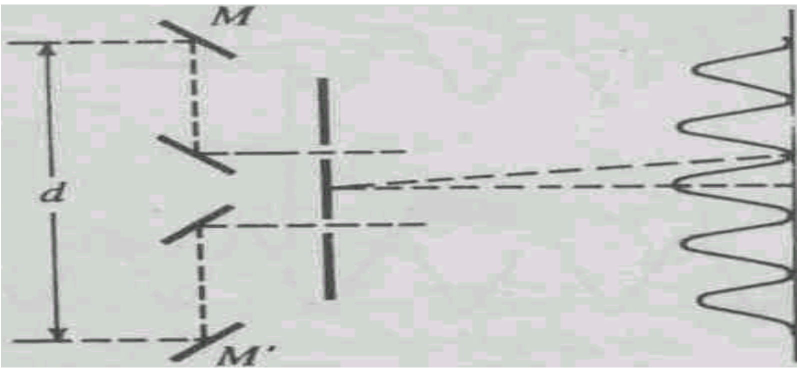

Fig. 1.5 Michelson's stellar interferometer

Since you have studied in detail about the Michelson Stellar interferometer in Unit 3 of Optics II, we would just mention here the results of one typical experiment. In a typical experiment the first disappearance of fringes occurred when the distance between mirrors *M* and *M '* was about 7.3 m which gave the angular diameter, θ as (taking $\lambda \sim 5 \times 10^{-7}$ m)

$$\theta = 1.22\lambda / d$$

$$= \frac{1.22 \times 5 \times 10^{-7}}{7.3}$$

$$= 8.4 \times 10^{-8} \text{ rad}$$

From the known distance of star and the value of its angular diameter, *6,* we can estimate its diameter.

1.8 Visibility of Fringes

Till now, we have been discussing coherence and its importance for observing interference fringes. We have been talking about the disappearance of fringes under different circumstances. For example, in the Young's double slit experiment, interference fringes are seen on a screen with highly spatially coherent light. The fringes are rather distinct; their visibility is high. As the two slits are moved further apart the fringes are more closely spaced and will lose visibility. The degree of visibility, therefore, is the measure of spatial coherence.

Assume that two wave trains of light, each of finite length Δl, overlap to their full extent. Such complete overlap will result in distinct maxima and minima of highest degree of visibility. Even if the wave trains overlap partially, as in Fig. 1.6, interference is possible. However, the degree of visibility of the fringes will diminish depending on the extent of overlap. The question, therefore, is not how much the wave trains must overlap to produce interference; rather, the question is how much visibility we need to see a fringe pattern?

The amount of radiation power incident per unit area is called areana.

The definition of visibility is essentially a matter of comparison. Visibility, V, can be defined as the ratio of the difference between the maximum areana E_{max}, and minimum areana E_{min}, to the sum of the areanas; i.e.

$$V = \frac{E_{max} - E_{min}}{E_{max} + E_{min}} \tag{1.14}$$

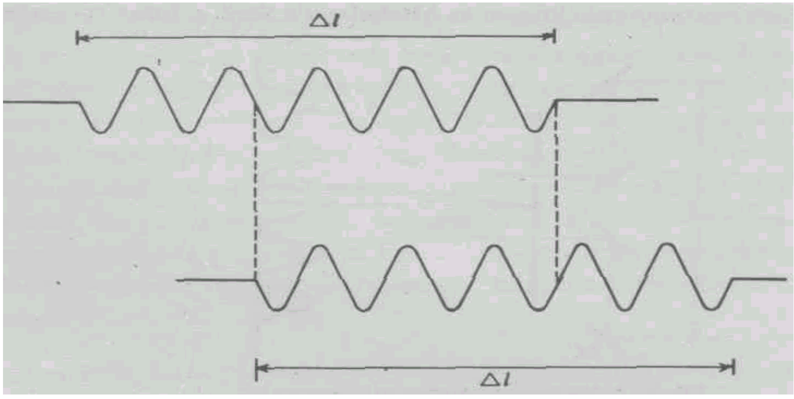

Fig. 1.6 Partial overlap of two wave trains

Let us assume that E_{max} can take any arbitrary value but $E_{min} = 0$. Then visibility, $V = 1$. On the other hand, if $E_{max} = E_{min}$, $V = 0$, fringes cannot be seen. Thus, the visibility may assume any value between 0 and 1. Generally, a visibility of 0.8 is considered high, but a value 0.2 is barely visible.

Now, you may like to know whether visibility is related to coherence? Yes, it is. To see how, let two points on a distant screen be illuminated by two light sources that produce equal areanas E_O. Light waves from each consists of two parts – coherent (*A*) and incoherent (*B*). Areana due to the coherent part, *A* can be expressed as

$$E_A = \gamma E_0$$

where, γ is the degree of coherence, and, the areana due to incoherent part is,

$$E_B = (1 - \gamma)E_O$$

Interference fringes are observed because of part *A*. The coherent part forms fringes whose maxima have intensities. Areana of the maxima is four times as high as the individual contribution. Thus, the maximum areana, $(E_A)_{max}$, is $4\gamma E_O$ and the minimum is zero. Moreover, on this interference fringe pattern, due to the coherent part *A*, a **uniform distribution** due to the incoherent part *B*, is superimposed. The areana of this distribution will be twice as high as the contribution E_B, because it comes from two sources. Hence,

$$(E_B)_{max} = 2E_B = 2(1 - \gamma)E_O = (E_B)_{min}$$

As a result, the areana in the maxima is

$$E_{max} = (E_A)_{max} + (E_B)_{max}$$

$$= 4\gamma E_O + 2(1 - \gamma)E_O$$

$$= 2(1 + \gamma)E_O$$

and the areana in the minima

$$E_{min} = (E_A)_{min} + (E_B)_{min}$$

$$= 0 + 2(1 - \gamma)E_O$$

Therefore, Eq. (1.14) for visibility of the fringes can be written as,

$$V = \frac{2(1+\gamma)E_O - 2(1-\gamma)E_O}{2(1+\gamma)E_O + 2(1-\gamma)E_O}$$

$$= \gamma \text{, the degree of coherence.}$$

Thus, the degree of visibility (or the contrast) of the fringes produced by two light waves is equal to the degree of coherence between them.

The highest visibility and hence the highest degree of coherence will occur when the minimum areana in the expression for *V* is zero. In that case, both the visibility and the degree of coherence are unity. Although conceivable in theory, this is not possible in practice. Complete coherence is merely a theoretical result. However, with the development of laser, about which you would study in the next unit, it is now possible to have light beam of extremely high degree of coherence.

1.9 Summary

Coherence is a property of light. A predictable phase relation exists (when there is coherence) between light waves passing through a point at different times.

Temporal coherence or longitudinal spatial coherence refers to the predictability of the phase of radiation as a function of time. In other words, temporal coherence can be identified as the interval of time during which the phase of the wave changes in a predictable manner as it passes through a fixed point in space. This time interval is known as **coherence time**, τ_c. And the path length corresponding to τ_c given as $L_c = c\tau_c$ is called the **coherence length** of the radiation.

Temporal coherence is related with the width of the spectral lines. The spread in wavelength is given as,

$$\Delta\lambda = \frac{\lambda^2}{L_c}$$

and the corresponding spread in the frequency of the spectral line is $\Delta\nu \approx 1/\tau_c$

Spatial coherence or transverse spatial coherence refers to the correlation between the phases of two light waves travelling side by side. The use of a point source in Young's double slit experiment is essentially to meet the requirement of spatial coherence.

If an extended source of light of width W is used in Young's interference experiment, for observing interference fringe pattern, the following condition must be satisfied

$$W = \lambda / \theta$$

where, λ is the wavelength of the light, and θ is the angle subtended by the extended source on the slits.

The quantity (λ/θ) is known as **lateral** (or **transverse**) **coherence width,** l_w.
For a circular extended source, the coherence width l_w is given as

$$l_w = \frac{1.22\lambda}{\theta}$$

Visibility of an interference pattern is given as

$$V = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}}$$

where, $E_{\max}$ is the maximum areana and $E_{\min}$ is the minimum areana.

In terms of the degree of coherence, *p*, the visibility is given as

$$V = \frac{2(1+\gamma)E_O - 2(1-\gamma)E_O}{2(1+\gamma)E_O + 2(1-\gamma)E_O}$$

where, E_O is the areana produced on the screen by individual light source.

1.10 Terminal Questions

1. The sodium line at λ = 5890 Å, produced in a low-pressure discharge, has spread in wavelength, $\Delta\lambda$ = 0.0194 Å. Calculate (a) the coherence length and (b) line width in hertz.

2. If the visibility in an interference fringe pattern is 50 percent and the maxima receive 15 units of light, how much light does the minima receive?

1.11 Solutions and Answers

SAQs

1. The wavelength of the light, λ = 660 nm and N, the number of waves in the wave train is 20.

(a) So, the coherence length

$L_c = N\lambda$

$= 20 \times 660$ nm

$= 13200$ nm $= 13.2 \times 10^{-12}$ m

(b) Coherence time

$\tau_c = L_c / c$, where c = velocity of light $= 3 \times 10^8$ m/s

$$= \frac{13200 \times 10^{-9}\, m}{3 \times 10^8\, ms^{-1}}$$

$= 4400 \times 10^{-17}$ s

$= 4.4 \times 10^{-14}$ s

2. Eq. (1.5), $2d = \lambda^2 / 2(\lambda_1 - \lambda_2)$ gives the path difference for the disappearance of fringe pattern due to light of wavelengths λ_1 and λ_2. When this expression is to be used for the disappearance of the fringe pattern due to the light beam consisting of all wavelengths lying between λ and $\lambda + \Delta\lambda$, we must divide the interval (width) into two equal parts of $\Delta\lambda / 2$. Thus, the fringe pattern will be produced by wavelength values

$$\lambda_1 = \lambda + (\Delta\lambda / 2)$$

$$\lambda_2 = \lambda$$

With these values, Eq. (1.5) reduces to

$$2d = \frac{\lambda^2}{2((\lambda + \Delta\lambda/2) - \lambda)} = \frac{\lambda^2}{2(\Delta\lambda/2)} = \frac{\lambda^2}{\Delta\lambda}$$

which is Eq. (1.6)

Now, for each wavelength lying between λ and $\lambda + \Delta\lambda / 2$, there will be a corresponding wavelength lying between $\lambda + \Delta\lambda/2$ and $\lambda + \Delta\lambda$, such that the minima of one falls on the maxima of the other. Therefore, the fringe pattern will disappear.

3. Width (or the diameter) of the source

$$W = 0.1 \text{ mm} = 1 \times 10^{-4} \text{ m}$$

And distance between the source and slits

$$r = 1 \text{ m}$$

Hence the angle subtended by the source on slits

$$\theta = \frac{W}{r} = \frac{1 \times 10^{-4} m}{1 m} = 10^{-4} \text{ rad}$$

Wavelength of the light

$$\lambda = 589.3 \times 10^{-9} \text{ m}$$

The lateral coherence width for a circular extended source

$$l_w = \frac{1.22\lambda}{\theta} = \frac{1.22 \times 589.3 \times 10^{-9} m}{10^{-4} rad}$$

$$= 0.72 \text{ cm}$$

Thus, if the separation between the slits is more than 0.72 cm, the fringe pattern will disappear.

TQs

1. λ = 5890 Å 5890×10^{-10} m

$\Delta\lambda$ = 0.0194 Å = 0.0194×10^{-10} m

(a) From equation (1.7), we have

$$\Delta\lambda = \frac{\lambda^2}{L_c}, \text{ where } \quad L_c = \text{coherence length}$$

$$\therefore \quad L_c = \frac{\lambda^2}{\Delta\lambda} = \frac{(5890 \times 10^{-10})^2 m^2}{0.0194 \times 10^{-10} m}$$

$= 0.18$ m

(b) The spread in frequency $\Delta\nu$ (line width in hertz) and coherence time τ_c is related as (equation (1.8))

$$\Delta\nu = \frac{1}{\tau_c} = \frac{1}{L_c / c} = \frac{c}{L_c}$$

where c = velocity of light = 3×10^8 m/s

$$\therefore \quad \Delta\nu = \frac{3 \times 10^8 \, m/s}{0.18 \, m}$$

$$= 1.6 \times 10^9 \text{ Hz}$$

2. The visibility of an interference fringe pattern is given as

$$V = \frac{E_{max} - E_{min}}{E_{max} + E_{min}}$$

where E_{max} is the maximum areana, i.e., the amount of radiation power contained in the maxima of the fringe pattern; and E_{min} is the minimum areana.

From the problem, we have

$$V = 50 \text{ percent} = \frac{1}{2}, E_{max} = 15 \text{ units}, \; E_{min} = ?$$

So, from above equation for visibility, we have

$$\frac{1}{2} = \frac{15 - E_{min}}{15 + E_{min}}$$

$$\therefore \quad (15 + E_{min}) = 2(15 - E_{min})$$

$$\therefore \quad E_{min} = 5 \text{ units}$$

Hence, 5 units of light will be received in the minima of the fringe pattern.

UNIT 2 PHYSICS OF LASERS

Structure

2.1 Introduction

In the previous unit, you learnt about coherence and coherent sources of light. It was explained there, why conventional thermal sources of light emit radiation which have very low degree of coherence. However, phenomenon like interference, which requires coherent light sources, can indeed be observed with conventional light sources. The quest for obtaining a light source with high degree of coherence led to the invention of lasers. As you know, a useful indicator of the degree of coherence is the coherence length. For ordinary light, the coherence length is of the order 10^{-2} m, whereas the coherence length for a laser light can be as long as 10^{5} m! So, you may appreciate the difference in the degree of coherence between an ordinary light and the laser light. In the present unit, we will discuss this source of highly coherent light beam – the LASER.

The name laser is an acronym for **Light Amplification by Stimulated Emission of Radiation**. You must realise that the key words here are amplification and stimulated emission. The existence of stimulated emission of radiation, when radiation interacts with matter, was predicted by Einstein in 1916. His theoretical prediction was realised by C. H. Townes and co-workers in 1954 when they developed microwave amplification by stimulated emission of radiation (maser). The principle of maser was adapted for light in visible range by A. Schawlow and C. H. Townes in 1958 but the first laser device was developed by T. H. Maiman in 1960. Once the laser was invented, it has found applications in such diverse fields as basic research, industry, medicine, space, photography, communication, defence, etc.

In Sec. 2.2, you will learn about the quantum mechanical description of the emission Physics of Lasers and absorption of light. In particular, you will learn about spontaneous emission and stimulated emission of radiation. In Sec. 2.3, the physical principles involved in the operation of lasers viz., excitation (or pumping), the need of an active medium and the feedback mechanism have been explained. Since the invention of laser by Maiman using small ruby rod as active medium, Lasers have come a long way. Presently, lasers are built using solid or liquid or gas as active media. Apart from these, now semi-conductor based lasers are finding wide applications. These different types of lasers have been briefly discussed in Sec. 2.4. The applications of lasers are so many and so varied that their detailed account will take us too far. In Sec. 2.5, we have, however, briefly discussed applications of lasers in industry, medicine, communication and basic research. In the next unit, you will study about holography, which would not have been possible without laser light. And in Unit 4, you will study about optical fibres – a medium of transporting light – which is a very active area of research and development for long distance optical communication purposes.

2.2 Objectives

After going through this unit, you should be able to:

- explain the concept of stimulated emission of radiation and differentiate it from spontaneous emission
- describe the need and methods of pumping
- list the characteristics of the active medium for lasers
- describe different types of lasers, and
- describe the important applications of lasers.

2.3 Light Emission and Absorption

As you are aware, most of the man-made sources of light are the solids and gases heated to high temperatures. For example, in case of incandescent bulb, the tungsten filament is heated, and in case of mercury tube light, the gas is heated. The energy of the heating source is absorbed by the atoms or molecules of the solid or the gas, which, in turn, emit light. The basic mechanism of the origin of light from within gas molecules, liquids and solids is similar in many respect to that from an individual atom. And the process of emission and absorption of light from atoms can be understood in terms of Bohr's atomic model. Though you might have studied Bohr's model in a previous course, we briefly discuss it here for the sake of completeness.

2.3.1 Quantum Theory: A Brief Outline

According to Bohr's theory, the energy of an atom or a molecule can take on only definite (discrete) values. These are known as the energy levels of the atom. The transition of an atom from one energy level to another energy level occurs in quantum jump. This was one of the basic assumptions of Bohr's theory. On the basis of this presumption, Bohr postulated that light is not emitted by an electron when it is revolving in one of its allowed orbits (and hence has a fixed value of energy). Light emission takes place when the atom makes a transition from an excited state (of energy E_i) to a state of lower energy E_f. The frequency of the emitted radiation is given by

$$h\nu = E_i - E_f$$

where E_i is the energy of the initial orbit, E_f is the energy of the final orbit, ν the frequency of the emitted light and *h* is the Planck's constant. The quantized orbits of the electron and the energy level diagram of the simplest atom – the hydrogen atom – are shown in Fig. 2.1.

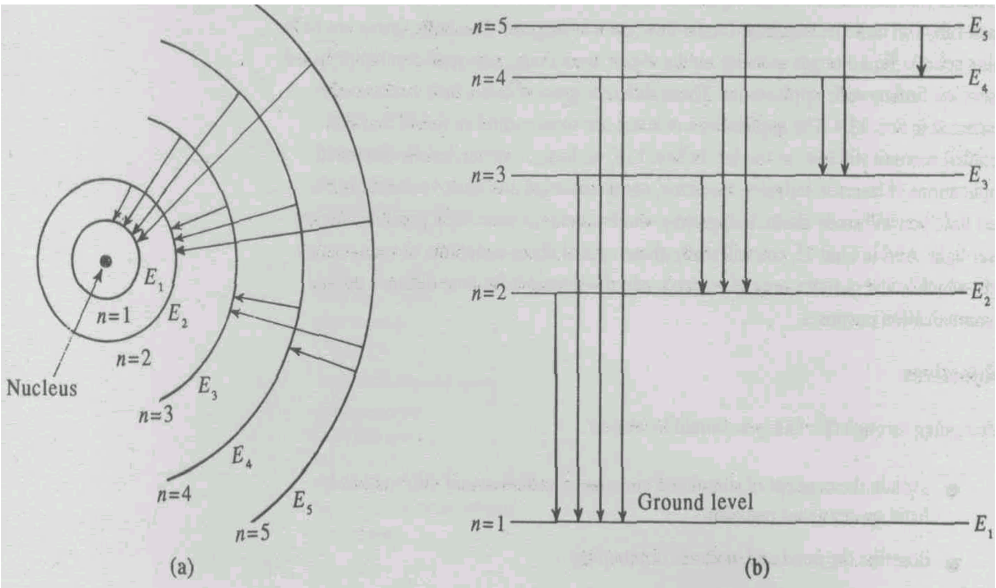

Fig. 2.1 (a) Bohr circular orbits for the revolving electron of hydrogen atom, showing transitions, giving rise to the emitted light waves of different frequencies; (b) Energy level diagram for the hydrogen atom

The quantum mechanical explanation about the origin of light, as discussed above, applies to all the known light sources. To focus our attention on the atomic processes involved in the emission and absorption of light, let us consider only two energy levels of an atom. Let the energy of the lower level be E_1 and that of the upper level be E_2. An atom lying in level E_2 will tend to make a transition to level E_1 so that it occupies a state of lower energy. Such emission process is known as spontaneous emission because it occurs in the absence of any external stimulus. The process of spontaneous emission is shown in Fig. 2.2(a). The photon emitted in spontaneous emission will have the energy ($E_2 - E_1$), while its other characteristics such as momentum, polarisation, will be arbitrary. The light emitted by ordinary sources results due to spontaneous emission. Absorption of light is the converse process of emission. The atom in a lower energy state can absorb a photon of energy $h\nu$ $(= E_2 - E_1)$ and get excited to the upper level E_2. The absorption process is depicted in Fig. 2.2(b).

Now, can you guess what will happen if an atom is in the higher energy level, E_2 and a photon of energy $h\nu$ $(= E_2 - E_1)$ interacts with it? Well, in such a situation, the photon may trigger the atom in the upper level to emit radiation. This emission process is known as stimulated emission. When the atom is already in the higher energy level, the photon, instead of being absorbed, may play the role of a trigger, and induce the transition from E_2 to E_1. As a result, the atom falls into lower energy level and an **additional** photon of energy $h\nu = E_2 - E_1$ is emitted. **In**

this process of stimulated emission, shown in Fig. 2.2(c), both the inducing and the induced photons have the same energy. The light from laser is due to the stimulated emission of radiation.

It is worth mentioning here that of the three processes mentioned above, only the first two, that is, the spontaneous emission and the absorption of light were postulated on the basis of Bohr's theory. It was only when Einstein considered the whole idea of emission and absorption of radiation in terms of thermodynamic equilibrium between matter and radiation that stimulated emission of radiation could be predicted. What were Einstein's theoretical arguments for the prediction? Let us learn these now.

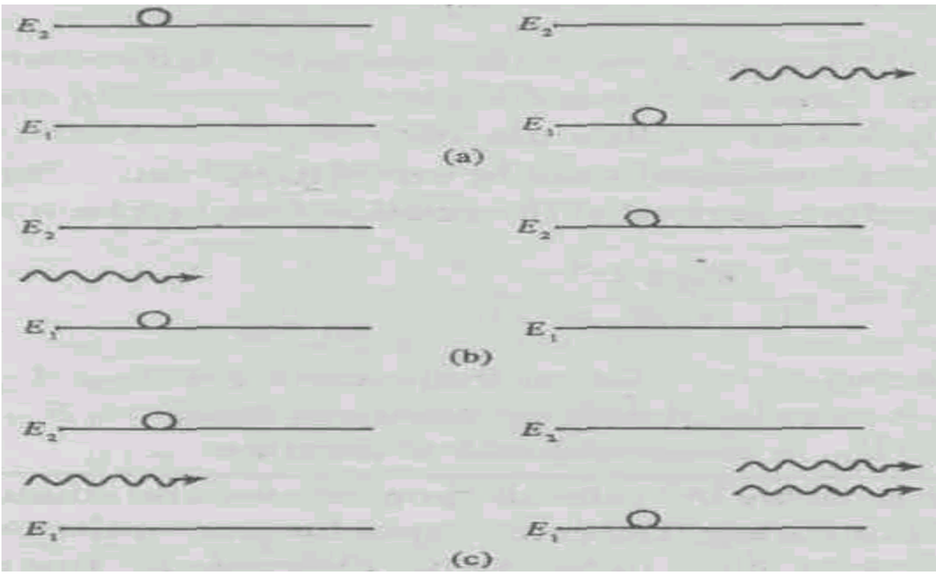

Fig. 2.2 (a) Spontaneous Emission (b) Absorption and (c) Stimulated Emission of light

2.3.2 Stimulated Emission of Radiation: Einstein's Prediction

Stimulated emission, as mentioned above, is the reverse of the process in which electromagnetic radiation or photons are absorbed by the atomic systems. When a photon is absorbed by an atom, the energy of the photon is converted into the internal energy of the atom. The atom is then raised to an excited (higher energy) state and it may radiate this energy spontaneously, emitting a photon and reverting to the ground (or some lower energy) state. However, during the period the atom is in the excited state, it can be stimulated to emit a photon if it interacts with another photon. This stimulating photon should have precisely the energy of the one that would otherwise be emitted spontaneously. Let us took at the theoretical arguments put forward by Einstein for the existence of stimulated emission.

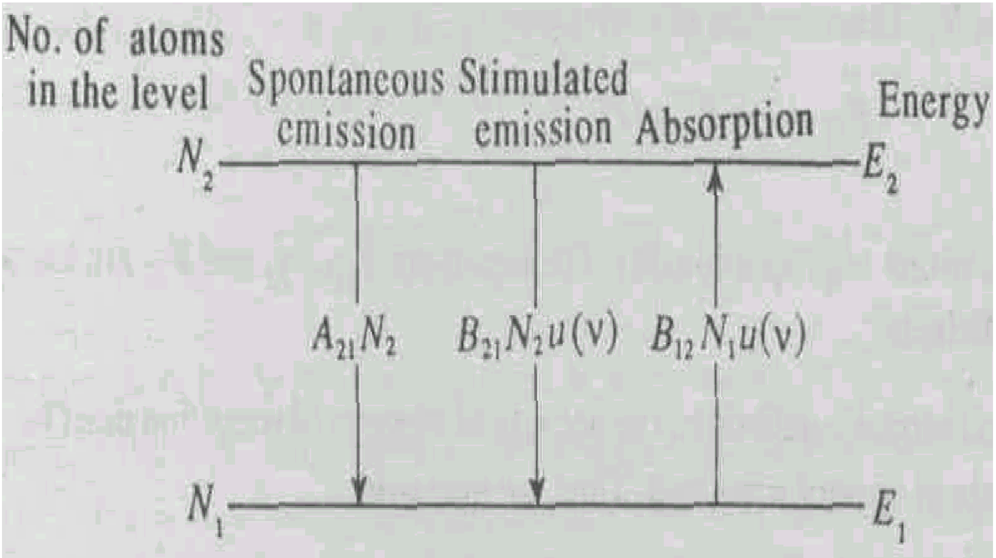

Fig. 2.3 An atomic system of two energy levels showing different emission and absorption processes

Refer to Fig. 2.3 which shows a system of two energy levels E_1 and E_2 with population of atoms N_1, and N_2 respectively. Let $E_1 < E_2$. According to Maxwell-Boltzmann distribution, the ratio of population of atoms in different levels for the system in thermal equilibrium is given

$$\frac{N_2}{N_1} = e^{-(E_2 - E_1)/k_B T}$$

or

$$N_2 = N_1 e^{-h\nu/k_B T}$$

where, k_B is the Boltzmann constant and T is the absolute temperature.

Now what will be the ratio of the population of the energy levels if radiation of energy $h\nu$ is introduced into the system? Einstein proposed that if this system of energy levels and the radiations is to remain in thermal equilibrium, the rate of downward transition (due to spontaneous and stimulated emission) must be equal to the rate of upward transition (due to absorption). He, therefore, arrived at the relation (see box below),

$$\frac{N_2}{N_1} = \frac{B_{12}u(\nu)}{A_{21} + B_{21}u(\nu)} \quad (2.3)$$

where $u(\nu)$ is the energy density of radiation at frequency ν and B_{12}, A_{21}, B_{21} are Einstein's co-efficient. A_{21} is associated with spontaneous emission, B_{21} is associated with stimulated emission and B_{12} is associated with absorption.

Form Eqs. (2.2) and (2.3), we have

$$\frac{B_{12}u(\nu)}{A_{21}+u(\nu)B_{21}} = e^{-h\nu/k_BT}$$

Or

$$u(\nu) = \frac{A_{21}}{B_{12}} \frac{1}{e^{h\nu/k_BT} - (B_{21}/B_{12})} \quad (2.4)$$

The energy density of black body radiation is given by Planck's radiation law:

$$u(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/k_BT} - 1} \quad (2.5)$$

Equation (2.5) must be same as Eq. (2.4). So we must have

$$B_{21} = B_{12} \quad (2.6)$$

and

$$A_{21}/B_{21} = \frac{8\pi h\nu^3}{c^3} \quad (2.7)$$

These are Einstein's relations. On the basis of Einstein's relations, we can conclude the following:

Following Einstein, let us write down the rates of spontaneous and stimulated emission and the rate of absorption of radiation. The rate of spontaneous emission will be independent of the energy density of the radiation field because for this process to occur, the presence of photon is not required. This emission process will be proportional to the number of atoms, N_2, in the higher energy state. So, we may write the rate of spontaneous emission as

$$P_{21} = N_2 A_{21} \tag{i}$$

where A_{21} is constant of proportionality.

Assume next that the system of atoms is subject to some external radiation field. In that case, as mentioned earlier, one of the two processes, namely, the stimulated emission and absorption, may occur. The probability of their occurrence depends on the energy density of radiation at the particular frequency separating the two levels and the population of states from which transition takes place. Therefore, the rate of stimulated emission will be proportional to the energy density of the radiation and the population of higher energy state, N_2. Thus, the rate of stimulated emission

$$P_{21} = N_2 B_{21} u(\nu) \tag{ii}$$

where B_{21} is another constant of proportionality and $u(\nu)$ is energy density of radiation at frequency ν.

On the other hand, the rate of absorption will depend on $u(\nu)$ and the population of the lower energy state, N_1. Thus, the rate of absorption

$$P_{12} = N_1 B_{12} u(\nu) \tag{iii}$$

where B_{12} is the constant of proportionality. The constants A_{21}, B_{12} and B_{21} are known as **Einstein's coefficients**.

With the system in thermal equilibrium, the net rate of downward transition must be equal to the net rate of upward transition. Thus, we may write

$$N_2 A_{21} + N_2 B_{21} u(\nu) = N_1 B_{12} u(\nu) \tag{iv}$$

Dividing both side by N_1, we get

$$\frac{N_2}{N_1} A_{21} + \frac{N_2}{N_1} B_{21} u(\nu) = B_{12} u(\nu)$$

or

$$\frac{N_2}{N_1} (A_{21} + u(\nu) B_{21}) = B_{12} u(\nu)$$

so that

$$\frac{N_2}{N_1} = \frac{B_{12} u(\nu)}{A_{21} + u(\nu) B_{21}} \tag{v}$$

(a) Eq. (2.6) indicates that the probabilities of absorption and stimulated emission are the same. In other words, when an atomic system is in equilibrium, absorption and emission take place side by side. Normally, $N_2 < N_1$, and absorption dominates stimulated emission. An incident photon is more likely to be absorbed than to cause stimulated emission. But, if we could find a material that could be induced to have a majority of atoms in the higher state than in the lower state, i.e. $N_2 > N_1$, the stimulated emission may dominate absorption. This condition of the atomic system (where $N_2 > N_1$) is known as population inversion. And when the stimulated emission dominates over absorption in the atomic system, it is said to lase.

(b) If we substitute $B_{12} = B_{21}$ in equation (2.4), we get the ratio of the number of spontaneous emission to stimulated emission

$$\frac{A_{21}}{B_{21}u(\nu)} = e^{h\nu / k_B T} - 1 \qquad (2.8)$$

When the system is in thermal equilibrium at temperature T, for $h\nu << k_B T$, Eq. (2.8) suggests that stimulated emission will dominate spontaneous emission. On the other hand, when $h\nu >> k_B T$, spontaneous emission will dominate stimulated emission. Now which of these two processes will dominate for ordinary thermal sources of light? To know that, you should do the following SAQ.

SAQ 1

The absolute temperature, T, for an ordinary source of light is typically of the order of 10^3 K. With the help of Eq. (2.8), show that in such sources, the process of spontaneous emission will dominate over the stimulated emission.

2.3.3 Einstein's Prediction Realised

You now know that when matter and radiation are in thermal equilibrium, besides spontaneous emission and absorption of radiation by matter, there must be a third process, called stimulated emission. This prediction did not attract much attention until 1954, when Townes and co-workers developed a microwave amplifier (MASER) using NH_3. In 1958, Shawlow and Townes showed that the maser principle could be extended into the visible region. In 1960, the prediction was realised by Maiman who built the first laser, using Ruby as an active medium. Maiman found that a suitable active component for a laser could be made from a single crystal of pink ruby: aluminium oxide ($A1_2O_3$), coloured pink by the addition of about 0.5 percent chromium. For any laser action to take place, the condition of population inversion must be met. By population inversion we mean that the number of atoms in the higher energy state is larger than the ground (or some lower energy) state. The energy states of the chromium atom, as shown in Fig. 2.4, are ideal for obtaining population inversion. The chief characteristics of energy levels of a chromium atom is that the levels labelled as E_1 and E_2 have a life time 10^{-8} s, whereas the state marked *M* has a life time 3×10^{-3} s. The energy state *M* with such a long life time (as compared to other excited states) is called a **metastable state**.

When an atom undergoes a non-radiative transition the energy is not released in the form of photons; rather, the energy is transferred via atomic collisions, collision with the crystal lattice, etc.

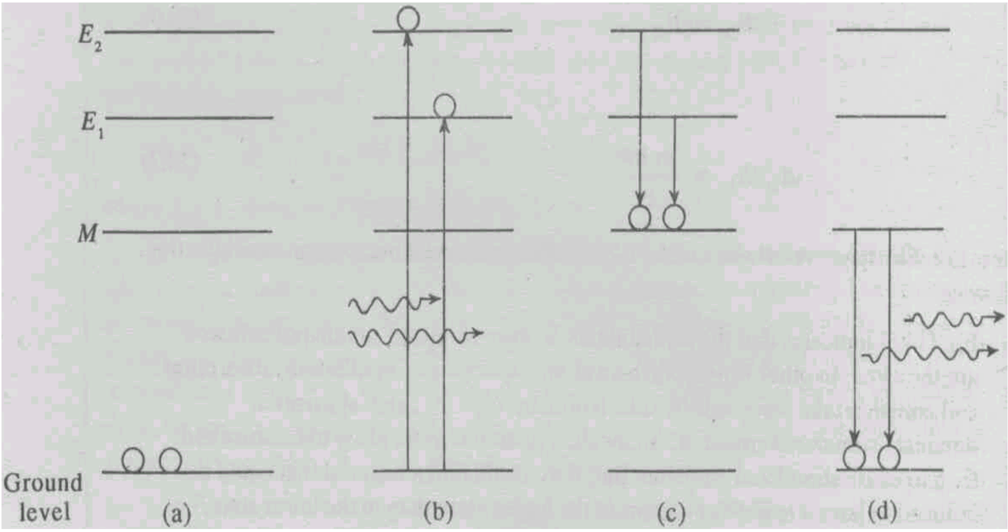

Fig.2.4: Energy levels of chromium atom: (a) atoms in the ground state

(b) on absorbing photons, atoms are excited to one of the two energy levels E_1 and E_2 . (c) atoms give up some of its energy to the crystal lattice and fall to a metastable level, M, (d) When stimulated by photons, the atoms in metastable level emit photon and fall to ground state.

A chromium atom in its ground state can absorb a photon (λ = 6600 Å) and make a transition to the level E_1 it could also absorb a photon of λ = 4000 Å and make a transition to the level E_2 . In either case, it subsequently makes a non-radiative transition, in time 10^{-8} s, to the metastable state *M.* Since the slate *M* has a very long life, the number of atoms in this state keeps on increasing and we may achieve a population inversion between the state *M* and *G* (the ground state). Thus, we may have a larger number of atoms in the level *M* compared to those in the state *G*. Once population inversion is achieved, light amplification can take place.

In the original set up of Maiman, the pink ruby was machined into a rod of length nearly four centimetre and diameter half a centimetre. Its ends were polished optically flat and parallel and were partially silvered. The rod was placed near an electronic flash tube (filled with xenon gas) that provided intense light for pumping chromium atoms to higher energy states. The set up of ruby laser is shown in Fig. 2.5. When the required population inversion was achieved with the help of electronic flash tube, the first few photons released (at random) by atoms dropping to the ground state stimulated a cascade of photons, all having the same frequency.

You now know how a ruby laser, developed by Maiman, works. You will appreciate that production of laser light demands that certain conditions be met beforehand. (We deliberately avoided reference to these in the above paragraphs.) First, is it possible to achieve laser light from any medium? If not, what are the characteristics of the medium which can produce laser light after proper excitation? (The media capable of producing laser light are called active media.) Secondly, how do we achieve population inversion? Further, for sustained laser light, it is necessary to feed some of the output energy back into the active medium. This is known as feedback and is achieved by resonant cavity. What is the nature of this resonant cavity for lasers? These are some of the important aspects of laser operation and design about which you will learn now.

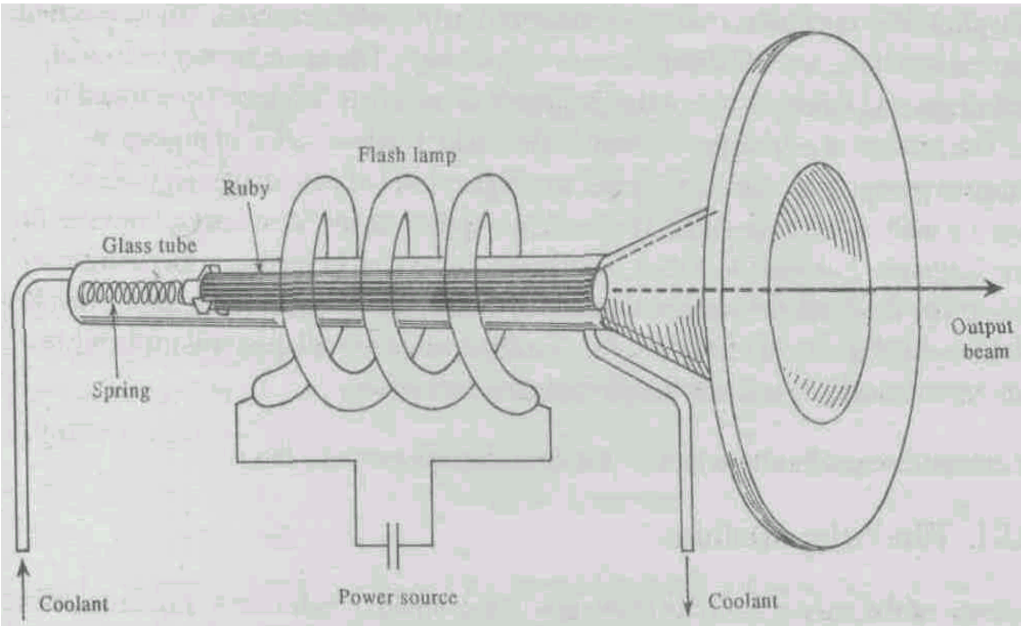

Fig. 2.5 The Ruby laser

2.4 Prerequisites For A Laser

A laser requires three prerequisites for operation. First, there should be an active medium which, when excited, supports population inversion and subsequently lasers. Secondly, we should ensure pumping mechanism that raises the system to an excited state. And lastly, in most cases, there is an optical cavity that provides the feedback necessary for laser oscillation. These are shown schematically in Fig. 2.6.

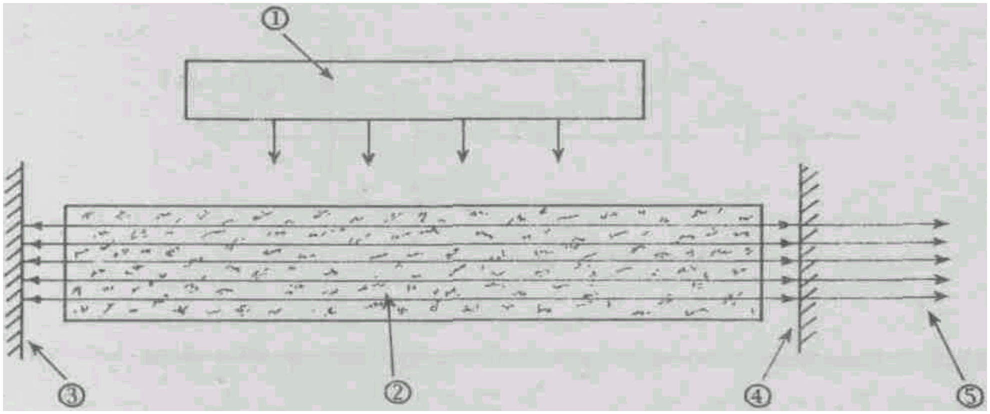

Fig. 2.6: Basic components of a laser oscillator: Energy source (1) supplies energy to active medium (2). Medium is contained between two mirrors (3 and 4). Mirror 3 is fully reflective while mirror 4 is partially transparent Laser radiation (5) emerges through partially transparent mirror.

In a typical laser operation, energy is transferred to the active material, which is raised to the excited state, and ultimately lasers in various ways. The medium may be a solid, liquid or gas and it may be one of the thousands of materials that have been found to laser. The process of raising the medium to the excited state is called **pumping,** in analogy to pumping of water from lower to a higher level of potential energy. Some lasers are built as laser amplifier. They need no optical cavity. Most lasers, however, are laser oscillators. For sustained laser oscillations, some kind of feedback mechanism is needed. The feedback mechanism is provided in the form of optical resonant cavity. In both laser amplifiers and oscillators, the first few quanta of radiation will probably be emitted spontaneously and will trigger stimulated emission.

Let us now discuss the above mentioned three components of a laser.

2.4.1 The Active Medium

The heart of the laser is a certain medium - solid, liquid or gaseous — called an active medium. Since Maiman's discovery of ruby, many new laser materials have been discovered. They include crystals other than ruby, glasses, plastics, liquids, gases and even plasma (the state of matter in which some of the atomic electrons are dissociated from the atoms). What should be the characteristics of an active medium? The only general requirement for an active medium is that it provides an upper energy state into which atoms can be pumped and a lower state to which they will return with the spontaneous emission of photons. The medium must also allow a population inversion between the two states. It may happen that the active species or centres, which provide lasing levels, constitute a small fraction of the medium. For example, in case of ruby, which is $A1_2O_3$ with some of the Al atoms replaced by Cr atoms, only the latter (Cr) is the active centre. Typical number of active species per cubic centimetre in solids and liquids is 10^{19} to 10^{20} and that for gaseous media their number is about 10^{15} to 10^{17}. How the light beam gets amplified when it passes through an active medium? To get the answer we examine the process of population inversion now.

Population Inversion

Why is the condition of population inversion between the lasing level necessary for the operation of lasers, i.e., for amplification of light to occur? We can investigate this by calculating the change in intensity of the light beam passing through an active medium. Refer to Fig. 2.7. A collimated beam of light having intensity I_ν travels along the x -axis through an active medium of thickness *dx*.

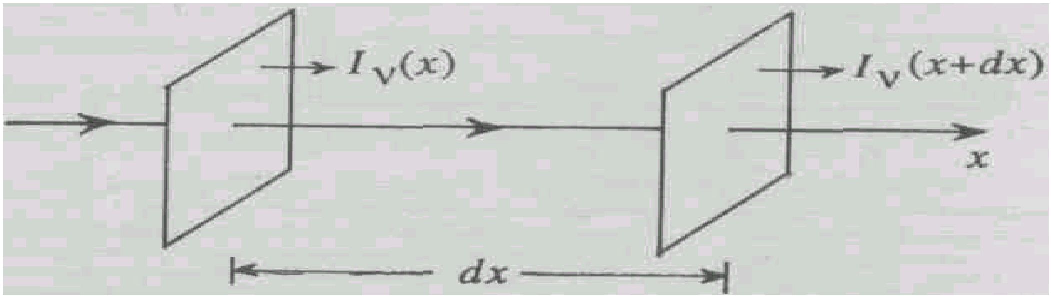

Fig. 2.7: Light beam of intensity I_ν passing through an active medium along the *x*-axis

If the cross-sectional area of each of the planes is *S*, the volume of the layer will be *Sdx*. Let $N_1(\nu)d\nu$ represent the number of atoms per unit volume which are capable of absorbing radiation whose frequency lies between ν and $\nu + d\nu$. *Tlie number of upwardly transmitted* ($E_1 \rightarrow E_2$) atoms per unit time in the layer of volume *Sdx* would be (refer to the last box)

$$N_1(\nu)d\nu B_{12}u(\nu)Sdx$$

In each transition, a photon of energy $h\nu$ is absorbed. Thus, energy lost per unit time from the incident radiation is

$$h\nu[N_1(\nu)d\nu B_{12}u(\nu)]Sdx$$

Similarly, let $N_2(\nu)d\nu$ represent the number of atoms per unit volume which are capable of undergoing stimulated emission by falling down to level E_1. The frequency of these photons lies between ν and $\nu + d\nu$. Then the number of stimulated photons emitted per unit time in the layer is

$$N_2(\nu)d\nu B_{21}u(\nu)Sdx$$

In each transition, photon of energy $h\nu$ is emitted and this reinforces the propagating beam. Thus the energy gain by the incident radiation per unit time is

$$h\nu[N_2(\nu)d\nu B_{21}u(\nu)]Sdx$$

You may have noticed that we have neglected spontaneous emission. It is so because a photon, emitted via spontaneous process, is in a random direction. And, as such, it does not contribute appreciably to the intensity of the beam.

As a result of above processes, will the intensity of the light beam increase or decrease with time? Since $u(\nu)d\nu Sdx$ represents the energy in the layer within frequency range ν and $\nu + d\nu$, we can write the rate of change of the energy with time as

$$\frac{\partial}{\partial t}(u(\nu)d\nu Sdx) = h\nu[-N_1(\nu)B_{12}u(\nu) + N_2(\nu)B_{21}u(\nu)]\, d\nu\, Sdx$$

or

$$\frac{\partial u(\nu)}{\partial t} = -h\nu[B_{12}N_1(\nu) - I_\nu(x)d\nu]u(\nu) \qquad (2.9)$$

If I_ν represents intensity, $I_\nu d\nu$ signifies the energy crossing a unit area per unit time whose frequency lies between ν and $\nu + d\nu$. Then

$$[I_\nu(x + dx)d\nu - I_\nu(x)d\nu]S$$

denotes the rate at which energy flows out of the layer. Since $u(\nu)d\nu\, Sdx$ represents radiation energy contained in the layer with frequency in the range ν and $\nu + d\nu$, we will have

$$[I_\nu(x+dx) - I_\nu(x)]d\nu S = \frac{\partial}{\partial t}[u(\nu)dv\, Sdx]$$

$$\frac{\partial u(\nu)}{\partial t} = \frac{I_\nu(x+dx) - I_\nu(x)}{dx} = \frac{\partial I_\nu}{\partial x} \quad (2.10)$$

From Eq. (2.9) and (2.10), we have

$$\frac{\partial I_v}{\partial x} = -h\nu[B_{12} N_1(\nu) - B_{21} N_2(\nu)]u(\nu)$$

But

$$I_v = u(\nu)v \quad (2.11)$$

where v = velocity of light in the active medium (= c/n ; n = refraction index of the medium). Thus, we get

$$\frac{\partial I_v}{\partial x} = -\frac{h\nu B}{v}(N_1 - N_2)I_\nu$$

where B (= $B_{12} = B_{21}$) denotes either Einstein's coefficient. Hence

$$\frac{\partial I_v}{I_v \partial x} = -\frac{h\nu n}{c}(N_1 - N_2)B \quad (2.12)$$

If the light beam is propagating in absorbing media, the loss of intensity, $-dI_\nu$, will be proportional to I_v and dx;

$$dI_v = -\alpha_\nu I_\nu dx$$

where α_ν is the absorption coefficient. We can rewrite it as

$$\frac{dI_\nu}{dx} = -\alpha_\nu I_\nu \quad (2.13)$$

On integration we find that

$$I_\nu = I_\nu(x=0)e^{-\alpha_\nu x} \quad (2.14)$$

If we compare Eqs. (2.12) and (2.13), we get the expression for absorption coefficient:

$$\alpha_{\nu} = \frac{h\nu\, n}{c}(N_1 - N_2)B \qquad (2.15)$$

At thermal equilibrium, $N_1 > N_2$, that is, the population of the ground state is greater than the population of the excited state and as can be seen from Eq. (2.15), α_{ν} is positive. Positive α_{ν} implies, (from equation 2.14) that the intensity of the beam decreases as it propagates through the material. The lost energy is used up in the excitation of atoms to higher energy states.

On the other hand, if we have a situation in which $N_2 > N_1$, α_{ν} will be negative and the intensity of the light beam would increase, that is, get amplified as it propagates through the material. This process is light amplification. Since this occurs when there is a higher population in excited state than in the ground (or lower energy) state, the material is said to be in the state of population inversion. Thus, the condition of population inversion is necessary for the amplification of intensity of light beam.

2.4.2 Excitation (or Pumping)

In the previous sub-section, you have learnt about the necessity of population inversion in the active medium for obtaining laser light. The process of obtaining population inversion is known as pumping or excitation. The aim of the pumping is to see that upper energy level is more intensely populated than the lower energy level. Alternatively, we can obtain the population inversion by depopulating lower energy level (other than ground state) faster than the upper energy level. There are several ways of pumping a laser and achieving the population inversion necessary for stimulated emission to occur. The most commonly used are the following:

1. Optical Pumping
2. Electric Discharge
3. Inelastic Atomic Collision
4. Direct Conversion

In **Optical Pumping,** a source of light is used to supply energy to the active medium. Most often this energy comes in the form of short flashes of light, a method first used in Maiman's Ruby Laser and widely used even today in **Solid-State Lasers.** The laser material is placed inside a helical xenon flash lamp of the type customary in photography. The xenon flash lamp for pumping is shown in Fig. 2.5.

Another method of pumping is by direct electron excitation as it occurs in an **electric discharge.** This method is preferred for pumping Gas lasers of which the argon laser is a good example. The electric field (typically several KV/m) causes electrons, emitted by the cathode, to be accelerated towards the anode. Some of the electrons will impinge on the atoms of the active medium (electron impact), and raise them to the excited state. As a result, population inversion is achieved in the active medium.

In the **inelastic atomic** collision method of pumping, the electric discharge provides the initial excitation which raises one type of atoms to their excited state or states. These atoms subsequently collide inelastically with another type of atoms. The energy transferred inelastically raises the later type of atoms to the excited states and these are the atoms which provide the population inversion. An example is Helium-Neon Laser, to be discussed later, in which such a pumping process is employed.

A direct conversion of electrical energy into radiation occurs in light emitting diodes. Such light emitting diodes (LED) are used for pumping by **direct conversion** in semi-conductor lasers.

These are some of the processes used for pumping atoms of the active medium to achieve population inversion. Atoms (or molecules) used as active centres often exhibit rather complex system of energy levels. However, for all the variety of these structures, the actual pumping schemes may be narrowed down to a few rather simple diagrams correctly showing the pumping process. Typically, these pumping schemes involve three to four levels. We think you would like to know about them.

Let us consider some of the pumping schemes. To do so, let us identify different energy states necessary to explain the pumping scheme as: the ground state as 0; the lower lasing state as 1; the upper lasing state as 2; and the pumping state as 3. We shall indicate pumping transition by upward arrow, the lasing transition by downward arrow and non-radiative fast decay by slanted arrows. Now let us consider a **three-level** pumping scheme as shown in Fig. 2.8a. Let us assume that by one of the pumping methods, more than half the number of atoms of active species have been pumped from the ground state to pumping state 3. The pumped atoms in state 3 decay non-radiatively to upper lasing state 2. This decay is very fast, (life time is typically of the order to 10^{-8} s). The upper lasing state 2 is generally a metastable state i.e. the life time of this state (~10^{-3} s) is much higher than the pumping state (or the excited state). Therefore, we have a situation of population inversion between lasing states 2 and 1 and hence lasing may take place. You may note that in this pumping scheme, the ground state (0) and the lower lasing state (1) are the same state. This feature of the pumping scheme proves too demanding for the pumping process because in normal circumstances, the ground state is highly populated. And, as you can appreciate, an ideal lower lasing state (1) should be empty or very thinly populated. How to get rid of this problem?

Atoms or molecules tend to occupy lowest energy state. Therefore, the population of the ground state (lowest energy state) is high.

According to the uncertainty principle, an energy state with longer life time will have a narrow frequency band.

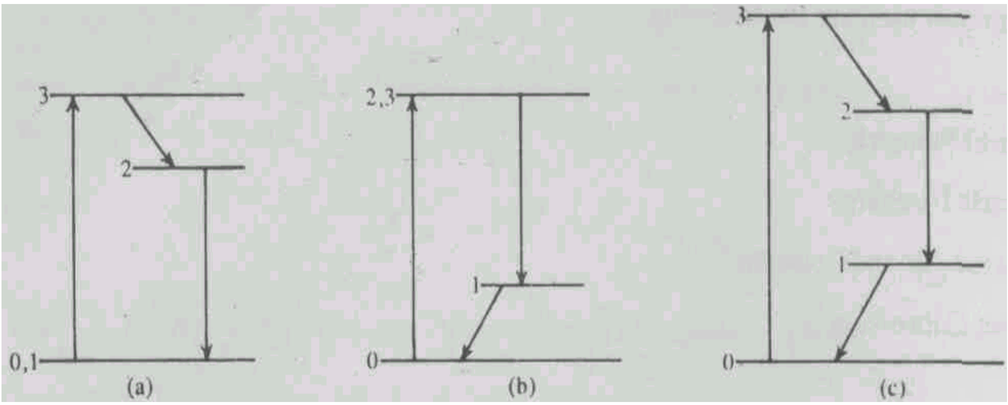

Fig. 2.8: Three level pumping schemes, (a) the ground state (0) and lower lasing state (1) are the same, (b) pumping state (3) and upper lasing state (2) are the same, (c) Four level pumping scheme

This problem can be taken care of if the pumping scheme is as shown in Fig. 2.8 b. As you can see, the atoms in the lower lasing state undergo non-radiative transition to the ground state (0). Since this transition is very fast (~10^{-8} s), the lower lasing level is empty for all practical purposes. You may, however, note that the same energy state acts as pumping state (3) and the upper lasing state (2). This state of affairs has its own shortcoming. If the pumping state has to act as upper lasing state, it must have a longer life time (metastable state) which implies that it must have very narrow frequency width. On the other hand, for proper utilisation of pumping energy, this state must have a wide frequency width so that more and more atoms get accommodated there. So, you see, it is a kind of conflicting requirements put on a single energy states.

The pumping scheme free from the shortcomings mentioned above with reference to three-level pumping scheme is what we call **four-level** pumping scheme shown in Fig. 2.8c. In this case, the pumping state (3) and the upper state (2) are separate; atoms in the pumping state undergo non-radiative transition to the upper lasing state. The four level pumping scheme, however, has some limitations. Substantial energy is lost during non-radiative transitions between pumping state (3) and the upper lasing state (2) and between the lower lasing state (1) and the ground state (0).

You may now ask: Which pumping scheme is better and preferred? Each pumping scheme has its own advantages and disadvantages. The choice of the pumping scheme in designing a laser depends upon the active media, the kind of use we want to put the laser light to, etc. We will discuss these aspects in the following sections. You may now like to answer an SAQ.

SAQ 2

If laser action occurs by the transition from an excited state to the ground state and it produces light of 693nm wavelength, what is the energy of the excited state. Take the energy of the ground state to be zero.

2.4.3 Feedback Mechanism: Optical Resonant Cavity

On the basis of the discussion in the previous sections, you now know that when a state of population inversion exists in an active medium, a light beam of particular frequency passing through it would get amplified. It happens because in such a situation, stimulated emission dominates spontaneous emission. This is the basic principle of optical amplifier. But a laser is much more than a simple optical amplifier. The laser, which produces a highly coherent beam of light, does not include a coherent light beam to initiate stimulated emission. Instead, it is the spontaneously emitted photon from the upper lasing state which stimulates the emission of new photons. Each spontaneous photon can initiate many other stimulated transitions which, in turn, may cause light amplification. Well, in this way, we do get amplification of light by stimulated emission. But, how is coherence of this amplified light ascertained? In other words, how can we ensure that the laser light has a very narrow bandwidth (monochromaticity) and a high degree of phase correlation? As such, the amplified light from laser is not coherent. It is because the spontaneous photons are independent of each other and travel in different directions. Therefore, the corresponding stimulated photons will also travel in different directions.

Can you suggest what we should do to obtain a highly coherent laser beam? To obtain a coherent light beam, we need to have a mechanism by which a condition is created such that spontaneous emission only in certain selected direction can develop stimulated emission. This mechanism is known as feedback mechanism. The spontaneous photons emitted in other directions leave the active medium without initiating much stimulated emission.

Now, you may ask; how do we actually achieve this favourable condition for spontaneously emitted photons in some preferred direction to further stimulate emission? Well, this is accomplished by means of an **optical resonator** – an essential component of a laser. Let us understand how an optical resonator works.

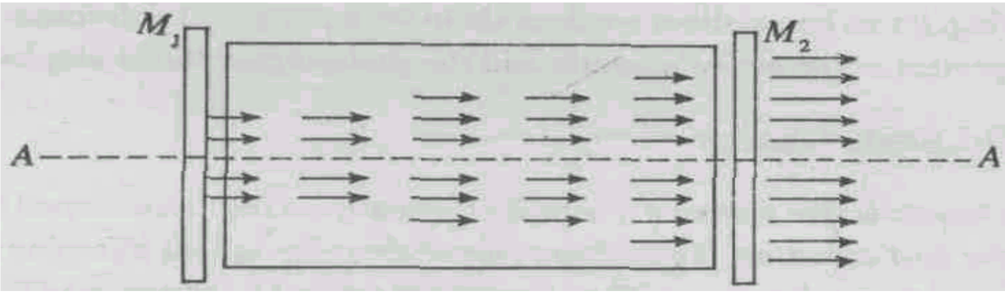

Fig. 2.9: Optical resonator consisting of two mirrors M_1 and M_2; M_1 is totally reflecting whereas M_2 is semitransparent; the axis of the mirrors is aligned with that of the active material

Optical cavity resonator can have many configurations. The schematic arrangement of a simple resonator is shown in Fig. 2.9. It consists of a pair of plane mirrors, M_1 and M_2, set on an optic axis which defines the direction of the laser beam. The active material is placed in between these mirrors. The photons emitted spontaneously along the *AA* direction or sufficiently close to it travel a relatively longer distance within the active material. It is so because photons travelling along *AA* will be reflected back and forth by the mirrors M_1 and M_2. You may notice that the direction of travel of these photons is quite fixed. Now, as a result of spending more time in the active material, these spontaneous photons will interact with more and more atoms in upper lasing level. Thus, the stimulated emission will add identical photons in the same direction, providing in ever-increasing population of coherent photons that bounce back and forth between the mirrors. On the other hand, spontaneous photons and the corresponding stimulated emission in other directions will traverse relatively shorter distances (and hence spend lesser time) in the active medium. Hence they will soon die out. Thus the optical resonant cavity provides the desired selectivity of propagation direction and thereby ensures the spatial coherence of the laser beam.

Now, what about monochromaticity of the laser light? Well, the laser light is highly monochromatic due to the very nature of its origin – the stimulated emission. It is so because the spontaneously emitted photons whose frequency do not match with the frequency difference between lasing levels will not give rise to stimulated emission. Thus, the band of wavelengths emitted during spontaneous emission is narrowed down. The monochromaticity of the laser light can further be enhanced by the optical resonant cavity. Suppose there are more than one upper lasing levels in a particular active medium. In that case, the laser output will consist radiations of more than one frequency. Now, if the mirrors of the resonant cavity are such that their reflectivity is a function of frequency, the radiations due to undesired lasing between levels will be damped out. Therefore, resonant cavity is the most vital component of the laser to obtain highly coherent light beam as output.

> You may recall that the spatial coherence is a measure of the uniformity of the phase across the optical wavefront. And the temporal coherence is a measure of the monochromaticity of the light.

In this section, you learnt the basic constituents of a laser. Since the invention of the ruby laser by Maiman in 1960, the research and development in this field has produced a variety of lasers. It is not possible to discuss all of them in detail here. However, we shall discuss some of them now.

2.5 Types of Lasers

Lasers can be classified in a variety of ways. One of these is in terms of their active media. As mentioned earlier, materials in all the three states of matter, namely, solid, liquid and gas, have been used as active medium to produce laser beam. Further, lasers have also been constructed using semi-conductors and plasma as active medium. In the following, let us know about some of them with particular reference to the physical properties of the active medium and the pumping methods employed.

2.5.1 Solid State Lasers

These lasers use an active material which is essentially an insulator doped with ions of impurity in the host structure. These lasers invariably use optical pumping to obtain the condition of population inversion. The sources for optical pumping may be discharge flashtubes, continuously operating lamps or even an auxiliary laser. The active centres in these lasers are transition element ions doped in the dielectric crystal. The host material for these active centres are generally oxide crystals. The most popular type of solid-state lasers are the **ruby** laser and **Nd:YAG** (neodymium: yttrium, aluminium, garnet) laser. Ruby is Al_2O_3 crystal (corundum) doped with triply ionised chromium atom (Cr^{3+}). You have learnt the functioning of this laser in section 2.2.

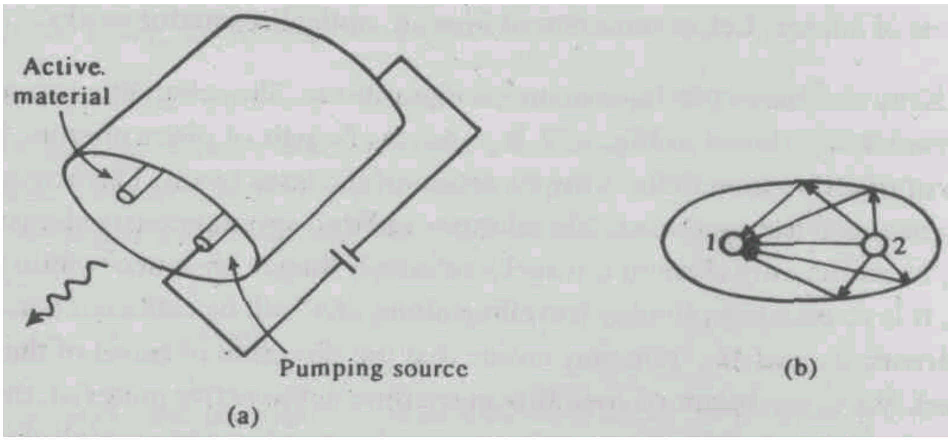

Fig. 2.10: Pumping arrangement for solid-state lasers

In solid-state lasers, the optical pumping is done by placing the active material (in the form of a rod) at one focus and the pumping source (in the shape of a right cylinder) at another focus of an elliptical reflector as shown in Fig. 2.10a. The advantage of such an arrangement is that any light leaving one focus of the ellipse will pass through the other focus after reflection from the silvered surface of the pump cavity. All of the pump radiation, therefore, is maximally focussed on the active material, as shown in Fig. 2.10b.

The Nd: YAG Laser

This laser, unlike ruby laser, employs four level pumping scheme. The energy levels of the neodymium (the active material) is shown in Fig. 2.11. In order to keep the discussion simple, we have not used the spectroscopic notations for different energy levels in Fig. 2.11. Rather, energy levels have been marked E_0, E_1, and so on. The optical pumping raises the Nd atoms in the ground state (E_0) to a few excited states (E_7, E_8). The energy levels marked E_4 and E_1 are the lasing levels. The pumped atoms in the excited states undergo non-radiative transition to the upper lasing level, E_4. Out of the group of lower lasing levels, the major portion of energy is emitted in the transition

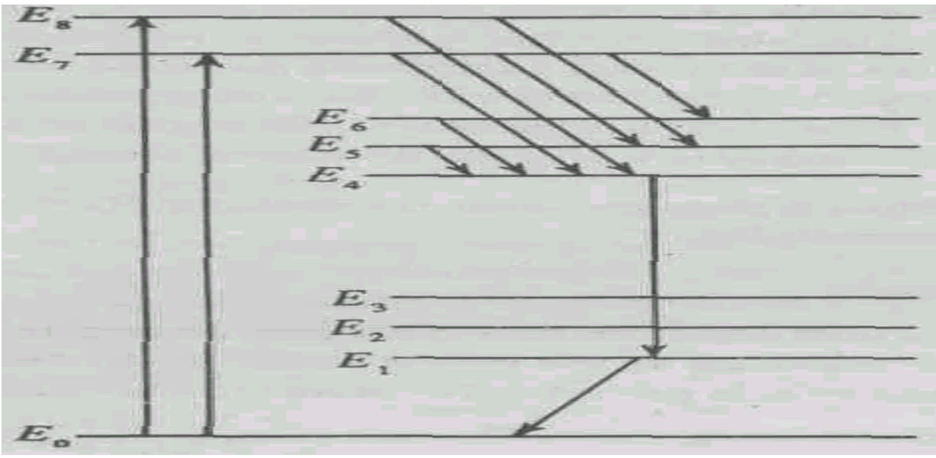

Fig. 2.11: Energy level diagram of Nd (neodymium) ion in Nd: YAG

E_4 □ E_1 . The Nd:YAG laser is an example of four-level laser.
This solid-state laser has two advantages: (a) it has a low excitation threshold and (b) has a high thermal conductivity. Due to high thermal conductivity, it can be used for generating light pulses at a high repetition rate or for continuous operation.

2.5.2 Liquid Lasers

In this class of lasers, as the name indicates, the active media are either the liquid solutions of organic dyes or specially prepared liquids doped with rare-earth ions viz., Nd^{3+}. However, the majority of liquid lasers use a solution of an organic dye as active medium and hence are also called **organic dye lasers.** Solvents used for the purpose are water, methanol, benzene, acetone etc. The liquid lasers are optically pumped. The energy states taking part in the lasing transition are the different vibrational energy states of different electronic energy states of the dye molecule. Since you may not be familiar with the vibrational energy states of molecules, we do not discuss the pumping scheme of this class of lasers.

In contrast to solids, liquids do not crack or shatter and can be made in sizes almost unlimited. Another advantage of liquid lasers is due to their (that of organic dyes) wide absorption bands in the visible and near ultraviolet portion of the electromagnetic spectrum. Therefore, liquid lasers are an ideal candidate for **tunable** laser, i.e., the frequency and hence energy of the output laser beam can be selected with ease.

2.5.3 Gas Lasers

The attractive feature of gas lasers in which rarefied gases are the active media, is that they can be designed to produce output beams over a wide range of wavelengths. Except for the caesium-vapour laser, gas lasers are pumped electrically rather than optically. Can you say why? It is because the conditions for amplification by stimulated emission, at one wavelength or another, are satisfied by an electrical discharge through almost any gas. Another reason for employing electrical pumping for gas lasers is that, unlike solids and liquids, the absorption lines of active centres in gaseous media exhibit substantially narrow widths. Therefore, optical pumping would prove very inefficient for gas lasers because the pump radiation obtained from optical sources do not have line spectrum of very narrow lines. In other words, the energy of optical pump radiation has a considerable spread in its value and since the gaseous active media will absorb radiation of almost single energy, most of the pump energy will go waste. Hence, optical pumping is not used for gas lasers. Further, gas lasers have advantage over solid state and liquid lasers in that they are free from local irregularities. Most gaseous systems have a high degree of optical perfection simply because the density of the gas is uniform.

We will now briefly describe a typical gas laser – the Helium-Neon gas laser. This was the first gas laser operated successfully.

The Helium-Neon Laser

In the helium-neon laser, a mixture of helium (He) and neon (Ne) gases is used as active medium. Lasing levels are provided by the exited states of the Ne atoms, whereas the He atoms play an important role in pumping Ne atoms to the excited states. The He-Ne laser is shown schematically in Fig, 2.12. The pumping is done by a stationary glow discharge fired by a direct current. When the potential difference between the anode and cathode is about 1000V, a glow discharge is initiated in the working capillary (containing He-Ne mixture) of a few millimetre diameter.

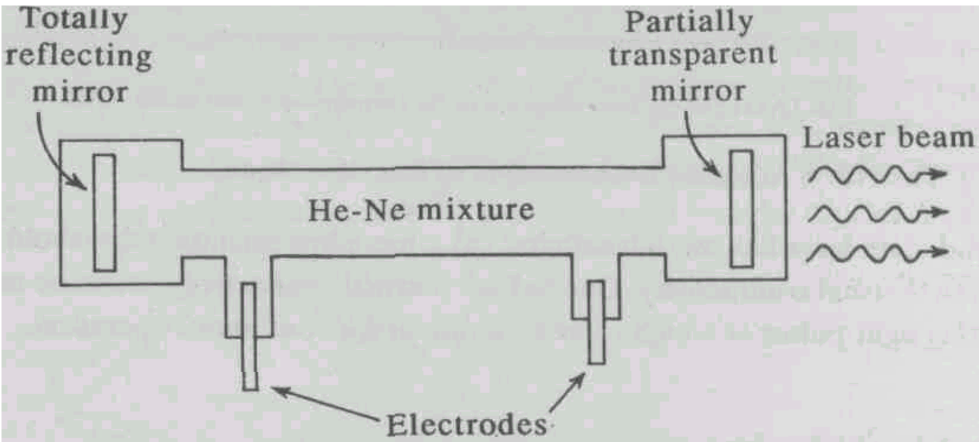

Fig. 2.12 The He-Ne Laser

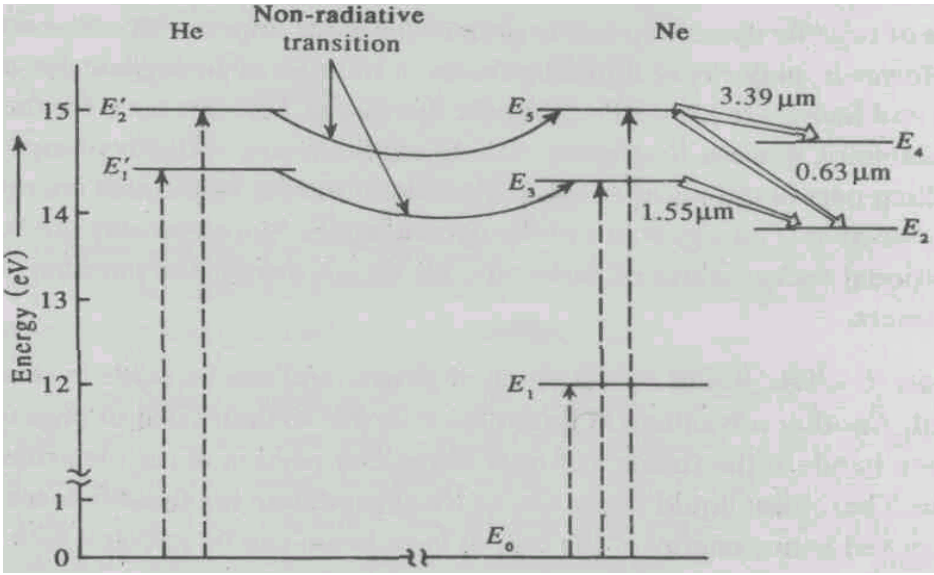

Fig. 2.13: Energy level diagram of helium-neon laser. Arrows (□) indicate the lasing transition

Now, let us look at the pumping scheme of the He-Ne laser. Refer to Fig. 2.13 which shows the energy level diagram of the He-Ne laser. When free electrons produced during the gas discharge pass through the He-Ne mixture, they collide with the He and Ne atoms and excite them by impact energy transfer. Such absorptive transitions due to electron impacts are shown by dashed arrows in Fig. 2.13. These excited states of He (i. e. E_1' and E_2') are metastable. Thus, He-atoms excited to these states stay there for a long time before losing energy by collision. The interesting feature of the He-Ne energy diagram is that the excited states of Ne, namely E_3 and E_5 have approximately the same energy as that of E_1' and E_2' of He atom. Therefore, when He-atoms in E_1' and E_2' collide with Ne-atoms in ground state, the He-atoms transfer their energy to Ne-atoms and raise them to the states E_3 and E_5. Such an exchange of energy is known as **resonant collision energy transfer**. Due to this energy transfer, He-atoms fall back to the ground state. As a result, the excited states E_3 and E_5 of Ne-atoms have a sizable population which is much more than that of states E_2 and E_4. Thus a condition of population inversion is achieved between the upper lasing levels E_5 (or E_3) and lower lasing levels E_4 (E_2). In such a situation, any spontaneously emitted photon can trigger laser action between these levels. The Ne atoms then drop down from the lower lasing levels E_2 and E_4, to the level E_1 through spontaneous emission.

The wavelength of transition between levels $E_5 \rightarrow E_4$, $E_5 \rightarrow E_2$, $E_3 \rightarrow E_2$ are 3.39 μ*m*, 0.63 μm and 1.55 μm respectively. As you can easily make out, radiations corresponding to 3.39 μm and 1.55 μm fall in the infrared region of the electromagnetic spectrum. The radiation corresponding to 0.63 μm, however, gives the red light - characteristic light of He-Ne laser. Proper selection of different frequencies may be made by choosing end mirrors of the resonant cavity which has high reflectivity over only the desired wavelength range.

Before we conclude our discussion about types of lasers, you must know that apart from those mentioned above, there are many other types of gas lasers. We may particularly mention **molecular** lasers (carbon dioxide laser), **chemical** lasers, **plasma** lasers, **semiconductor** lasers, etc. We have not discussed these here since for an understanding of their pumping schemes, you need to know molecular spectroscopy, semiconductor physics etc. It is, however, worth mentioning here that the essential principles, in so far as laser action is concerned, remain the same in all types of lasers.

The importance of lasers in contemporary physics lies in their so many and so varied applications. To give you a glimpse of these we now discuss some of the important applications of lasers.

2.5 Applications of Lasers

Applications of any device essentially stem from its unique features. What are the unique features of a laser? First and the foremost, laser light is highly coherent. This characteristic has enabled us to use lasers for data transmission and processing, precision measurements, photography (holography), etc. Secondly, laser light has unprecedented brightness (energy per unit area). Brightness of laser light, a by-product of its coherence, can be many orders of magnitude greater than the brightest of the light produced by conventional sources. Further, laser beams are highly directional.

Lasers and their Applications

In a typical laser, this directionality is limited only by the diffraction of the emerging beam by the laser aperture itself. The brightness and directionality of laser beam are exploited to produce targeted effects in materials. These applications include material working (such as heat treatment, welding, cutting, hole burning etc.), isotope separation, medical diagnostics, etc. In the following, you will learn some of these applications of lasers.

2.5.1 Communication

You may be aware that in a typical communication system, information is communicated (between the transmitter and the receiver) through electromagnetic waves, which are known as carrier wave. These are modulated by the desired signal (the oscillations of the information proper). Normally the signal frequency is appreciably lower than the frequency of the carrier wave. Moreover, the higher the carrier frequency, the wider frequency range it can modulate. In other words, the capacity of a communication channel is proportional to the frequency of the carrier wave. The frequency in the centre of the visible spectrum is about 100,000 times greater than the frequency of 6 cm waves used in microwave-radio relay systems. Consequently, the theoretical information capacity of a typical light wave is about 100,000 times greater than that of a typical microwave.

Long distance communication systems rely on the principle of multiplexing-the simultaneous transmission of many different messages (information) over the same pathway. The ordinary human voice (conversation) requires a frequency band from 200 to 4000 Hz, a band 3800 Hz wide. A telephone call, therefore, can be transmitted on any band that is 3800 Hz wide. It can be carried by a coaxial cable in the frequency band between 1,000,200 and 1,004,000 Hz, in the MHz range, or a He-Ne laser beam (638.8 nm, 4.738×10^{14}Hz) in the frequency range between 473,800,000,000,200 and 473,800,000,004,000 Hz. You may note here that the telephone message requires about 0.4 percent of the available co-axial carrier frequency. And, the same telephone message requires less than one-billionth of 1 percent of the available laser-beam frequency. Thus, the information carrying capacity could be enhanced tremendously if laser beams are employed as carriers. So, wait for some more time till laser trunk lines come into use in a big way and you may be saved from listening *"All the lines in this route are busy. Please dial after some time."*

Now you may ask: Light, as such, was available to us from time immemorial, then why is it that we are using (or planning to use!) it for communication purposes now? Is it related to the discovery of a laser in any way? Yes, it is. As we mentioned earlier, light from conventional sources may not be pure (that is, it may be non-monochromatic) and hence cannot be used for transmitting signals. Radio waves from an electromagnetic oscillator are confined to fairly narrow region of electromagnetic spectrum (i.e. it has a well defined frequency). These radio waves are, therefore, free from "noise" (considerable spread in frequency values) and hence can be used for carrying a signal. In contrast, all conventional light sources are essentially 'noise' generators, i.e., they simultaneously emit electromagnetic radiations of different frequencies and hence are not suitable as carrier waves. With the invention of lasers, however, the situation changed. As you know, the light produced by lasers is highly monochromatic and coherent which enable them to act as carrier waves in the communication systems.

Now, what is the medium through which laser beam travels while it carries information? The signal carrying laser beams can be transmitted through free (unguided) space, and by light guides. Light guides in the form optical fibres have found wide use in optical communication. You will learn about the details of fibre optics in Unit 4 of this course.

2.5.2 Basic Research

The discovery of the laser gave birth to an entirely new branch of optics known as nonlinear optics. Even at ordinary laser intensities, transparent materials (which are usually non-conductors), respond in an unusual manner. You may recall, for example, that the dielectric constant of a material depends on its nature as well as on the frequency of the light passing through it. But, it has been observed that when the ordinary light beam is replaced by a laser beam, the dielectric constant also depends on the instantaneous magnitude of the electric field component of the laser beam. In other words, the response of a material to high electric fields is non-linear. It is just one of the several non-linear effects that a laser beam produces when it interacts with matter. In fact, almost all the laws of optics are modified to some extent at the high intensities produced by pulsed lasers.

Another important application of lasers in basic research and development is in the field of thermonuclear fusion. As you know, for effective fusion to take place, an extremely high temperature ($\sim 10^8$ K) must be maintained. In principle, such high temperatures can be achieved by powerful laser beams.

Yet another remarkable application of lasers is in isotope separation. One of the basic requirements of harnessing nuclear energy from uranium is to have 2-3% of uranium isotope (^{235}U) in the fuel. In natural uranium, however, the percentage of ^{235}U is only 0.7. (The major constituent of natural uranium is ^{238}U.) Therefore, to have fuel enriched in ^{235}U, we can use laser beams. Each of these isotopes absorbs radiation of different frequency. So when a laser beam of particular frequency is passed through the mixture of ^{235}U and ^{238}U, the atoms of ^{235}U absorb the radiation and get excited. The excited atoms of the desired isotope are further excited so that they get ionised. Once ionised, it can easily be separated by applying a dc electric field. This is one of the several methods of using laser beam for isotope separation.

2.5.3 Medicine

A properly focussed laser beam is an excellent tool for surgery. The advantage of laser surgery is that it is bloodless since the beam not only cuts; it also "welds" blood vessels. It has a high sterility as no contact of tissues with surgical tools takes place. Also, the laser surgery is painless and operations are very fast. In fact there is not enough time for the patient to respond to the incision and sense pain. Laser beams are being widely used for performing eye and stone surgery.

A word of caution: any light can cause damage. Laser, in particular, can be highly damaging because it has spatial coherence, i.e., it can be focussed down to a high power densities. The maximum permissible exposure (MPE) is 0.0005 mJ cm $^{-2}$. For exposure time from 2×10^{-5} s to 10 s, the limit is MPE = $1.8t^{3/4}$ mJ cm^{-2}.

2.5.4 Industry

The invention of lasers has made it possible to develop sophisticated tools of material working (such as drilling, welding, etc) processes used in industry. With appropriate choice of lasers, a laser beam can be focussed into a light spot of diameter 10 – 10 μm! Can you imagine this dimension – it will be smaller than the dot you mark with your pen on a piece of paper! Due to this sharp focussing, a very high concentration of energy is available within a small spot on the surface of the material. For example, when a 1 kW output of a continuous wave (cw) laser is focussed a spot of 100 μm diameter, the resultant irradiance (intensity) will be 10 W cm. This makes laser an effective tool for drilling very fine hole through materials.

Laser cutting, as compared to other cutting processes, offers several advantages, e.g., the possibility of fine and precise cuts, minimal amount of mechanical distortion and thermal damage introduced in the material being cut, chemical purity of the cutting process, etc. Laser cutting is extensively used in industry. For example, in high-tech garment factories, CO_2 laser capable of 100W of continuous output is used for cutting cloth. The laser cuts 1m cloth in a second! And, laser cutting is also employed in the fabrication of spacecraft to cut the sheets of litanium, steel and aluminium. In cutting and most of the industrial applications, carbon-dioxide (CO_2) laser is used.

2.5.5 Environmental Measurements

You may be aware of the conventional technique of determining the concentration of various atmospheric pollutants such as gases (carbon monoxide, sulphur dioxide, oxides of nitrogen, etc) and a variety of material particles (dust, smoke, etc). In this method, the nature and concentration of pollutants is determined by chemical analysis. The major deficiency of this method is that it does not provide real-time data. The technique developed with lasers for measuring the concentration of pollutants is essentially the 'remote-sensing' technique which does not require sample to be analysed in laboratory. Since it provides information about the change in atmospheric composition with time, it can serve well for monitoring the environmental pollution.

For the determination of pollutants in the form of material particles, the technique is based on the scattering of light. The technique is known as LIDAR (light detection and ranging) and its operations are similar to those of a radar. In brief, a pulsed laser is passed through the location under investigation and the back-scattered light is detected by a photodetector. The time taken by the back scattered light to be detected gives information about the concentration of pollutant matter.

For the determination of gaseous pollutants, the basic principle involved is the absorption of light by the gaseous atoms or molecules. As different gases absorb at different wavelengths, passing laser beams of different wavelengths provides information about the gaseous constituents of the environment.

2.5.6 Photography: Holography

The conventional photographic process, as you know, consists of recording an illuminated three-dimensional object or scene as a two-dimensional image on a photosensitive surface. The light reflected from the object is focussed on the photosensitive surface by some kind of image forming device, which can be a complex series of lenses or simply a pinhole in an opaque screen.

The coherent nature of the laser beam has brought about a qualitatively new method of photography without a lens system. This new method, called holography, allows three-dimensional (that is, complete), pictures of a given object or a scene to be taken. Holography (also known as photography by wave-front reconstruction) does not, as such, record an image of the object being photographed; rather, it records the reflected light waves themselves. The photographic record so obtained is called hologram. The hologram bears no resemblance to the original object. It, however, contains - in a kind of optical code - all the information about the object that would be contained in an ordinary photograph. In addition, the hologram also contains information about the object that cannot be recorded by any other photographic process. Holography is the subject matter of the next unit (i.e. Unit 3).

2.6 Summary

- According to the Bohr's theory, if an atom makes a transition from an excited state (of energy E_i) to a state of lower energy E_f, emission of electromagnetic radiation (photons) take place. The energy of the emitted photon is $h\nu = E_i - E_f$.

- When electromagnetic radiation interacts with matter, three type of processes may occur:

  (i) Spontaneous Emission
  (ii) Absorption
  (iii) Stimulated Emission

- Light emitted by ordinary sources is due to spontaneous emission.

- The existence of stimulated emission of radiation was predicted by Einstein on the basis of thermodynamic considerations. If the population of the energy level E_1 be N_1 and that of E_2 be N_2 $(E_1 < E_2)$ then, the ratio of the population of the two states is given as

$$\frac{N_2}{N_1} = \frac{B_{12}u(\nu)}{A_{21} + B_{21}u(\nu)}$$

where, $u(\nu)$ is the energy density of radiation at frequency ν and B_{12}, B_{21} and A_{21} are Einstein coefficients.

- Einstein coefficients are related to each other through the relations

$$B_{21} = B_{12}$$

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h\nu^3}{c^3}$$

- Einstein's relation clearly indicates that stimulated emission may dominate spontaneous emission provided the condition of population inversion exists. And in an atomic system where a condition of population inversion exists, one may have amplification of light, that is, laser light.

- Einstein's prediction was first realised in the optical frequency range by Maiman who developed a laser using a ruby rod.

- There are three prerequisite for laser operation:

(i) Active medium
(ii) Pumping
(iii) Optical resonant cavity

- The change in intensity of a light beam passing through an active medium is given by

$$\frac{\partial I_\nu}{\partial x} = -I_\nu \frac{h\nu n}{c}(N_1 - N_2)B$$

where n is refractive index, B is Einstein's coefficient.

- This relation clearly indicates that for enhancement in the intensity of the light beam as it traverses the active medium, $N_2 > N_1$, i.e., a condition of population inversion must exist.

- There are a variety of methods for pumping, such as, optical pumping, electronic discharge, inelastic atomic collisions etc. The choice of pumping process mainly depends upon the nature of the active medium.

- There are two types of pumping schemes: three level and four-level.

- Optical resonant cavity helps in obtaining sustained laser light.

2.7 Terminal Questions

1. Assume that an atom has two energy levels separated by an energy corresponding to a frequency 4.7×10^{14} Hz, as in the He-Ne laser. Let us assume that all the atoms are located in one or the other of these two states. Calculate the fraction of atoms in the upper state at room temperature T = 300K.

2. A pulsed laser used for welding produces 100 W of power during 10 m. Calculate the energy delivered to the weld.

2.8 Solutions and Answers

SAQs

1. The ratio of the number of spontaneous to stimulated emission is given as

$$\frac{A_{21}}{B_{21}u(\nu)} = e^{h\nu/k_BT} - 1$$

The absolute temperature of an ordinary source of light has been given as

T = 10^3 K

Let us take the wavelength of light, λ = 6000 Å. Hence the corresponding frequency,

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^8\, m/s}{6000 \times 10^{-10} m} = 0.5 \times 10^{15} \text{ Hz}$$

Planck's constant $h = 6.6 \times 10^{-34}$ Js

Boltzmann constant $k_B = 1.38 \times 10^{-23}$ J K^{-1}

Hence,

$$\frac{A_{21}}{B_{21}u(\nu)} = \exp\left[\frac{6.6 \times 10^{-34}(J\,s) \times 0.5 \times 10^{15}(s^{-1})}{1.38 \times 10^{-23}\,(J\,K^{-1}) \times 10^3\,(K)}\right] - 1$$

$$= \exp(23) - 1$$

$$= 10^{10}$$

Thus, for ordinary sources of light, the number of spontaneous emission is much, much greater than the number of stimulated emission.

2. Let the energy of the excited state (upper lasing state) be E_2 and that of the ground state (lower lasing state) be E_1. The laser light is due to the atomic transitions from E_2 to E_1. Thus, the frequency of the laser light will be

$$\nu = \frac{E_2 - E_1}{h}$$

Now, as per the given problem,

$$E_2 = ?\ ,\ \ E_1 = 0 \text{ and } \lambda = 693 \text{ nm} = 693 \times 10^{-9} \text{ m}$$

Hence,

$$\nu = \frac{c}{\lambda} = \frac{3 \times 10^{8}\ m/s}{693 \times 10^{-9} m} = 3.1 \times 10^{14}\ s^{-1}$$

$$E_2 - E_1 = h\nu$$

$$E_2 = 6.6 \times 10^{-34} (J\,s) \times 3.1 \times 10^{14} (s^{-1})$$

$$= 20.46 \times 10^{-20} \text{ J}$$

$$= 12.77 \text{ eV}$$

TQs

1. Let the two energy levels be E_1 and E_2 (such that $E_1 < E_2$) and their population be N_1 and N_2 respectively. According to the Boltzmann distribution

$$\frac{N_2}{N_1} = e^{-(E_2 - E_1)/k_B T}$$

We know that

$$(E_2 - E_1) = h\nu$$

$$= 6.62 \times 10^{-34} (J\,s) \times 4.7 \times 10^{14} (s^{-1})$$

$$= 31.114 \times 10^{-20} \text{ J}$$

and

$$k_B T = 1.38 \times 10^{-23} (J/K) \times 300(K)$$

$$= 4.14 \times 10^{-21} \text{J}$$

Hence,

$$\frac{N_2}{N_1} = e^{-(E_2 - E_1)/k_B T}$$

$$= \exp\left[-\frac{31.114 \times 10^{-20}}{4.14 \times 10^{-21}}\right]$$

$= \exp(-75.1)$

$= 2.42 \times 10^{-33}$

2. Power = Energy per unit time

$$= \frac{\text{Energy}}{\text{Time}}$$

Given Power $= 100$ W $= 100$ (J/s)

Time $= 10$ ms $= 10 \times 10^{-3}$ s

$\therefore$ Energy $=$ Power $\times$ Time

$= 100 \text{ (J/s)} \times 10 \times 10^{-3} \text{ (s)}$

$= 1$ J

$\therefore$ Energy delivered to the weld is 1 Joule.

UNIT 3 HOLOGRAPHY

Unit Structure

3.1 Introduction

In the previous Unit, we pointed out that one of the revolutionary applications of lasers is in the development of a novel technique of photography, known as holography. This word is the combination of two Greek words - holos (complete) and graphos (writing). That is, holography is the technique of obtaining complete picture (as true as the object itself) of an object or a scene. In other words, it is a three-dimensional recording of an object or a scene. Well, you may be wondering as to what essentially differentiates this technique from the normal photography! In normal photography, a two-dimensional image of a three-dimensional object is recorded on a photosensitive surface. The photosensitive surface records the intensity distribution of light falling on it after reflection from the object. As a consequence, we obtain a permanent record of the intensity distribution that existed at the plane occupied by the photographic plate when it was exposed. Since the photosensitive surface is sensitive only to the intensity variation, the phase distribution existing in the plane of the photographic plate is completely lost and is responsible for the absence of the three-dimensional character in it. Holography is that technique of photography where not only the amplitude (and hence the intensity) but also the phase distribution can be recorded. As a result, pictures obtained by holographic technique possess three-dimensional form and are visually rich.

Holography was introduced by Dennis Gabor in 1948. He showed that one could indeed record both the amplitude and the phase of a wave by using interferometric principles. In Sec. 3.2, you will learn the basic concepts involved in the holographic technique. You will be able to appreciate the similarity between the hologram and the diffraction grating. The process of holography, i.e., how to obtain a hologram, how to obtain images from the hologram, etc., has been explained in Sec. 3.3. Due to the high cost of lasers (an essential requirement for holography), this technique is not being used extensively. The technique, however, has tremendous potential and some of the important applications have been explained in Sec. 3.4.

3.2 Objectives

After going through this unit, you should be able to

- differentiate between normal photography and holography
- explain the basic principle of holography
- describe how holograms are obtained, and 9 state some of the applications of holography.

Reference wave is the light wave falling directly on the photosensitive plate.

Object wave is the light wave reflected from the object and received at the photosensitive surface at the time of recording the hologram.

3.3 Holography: The Basic Principle

Holography is the process of recording the interference pattern produced by light waves reflected by an object and reference waves. This interference pattern of the object is unique and is called **hologram** (total recording). If you look at a hologram, you will realise that it does not even remotely resemble the object. However, when this recorded pattern is illuminated by a suitably chosen reconstruction wave, out of the many component waves emerging from the hologram, one wave completely resembles the object wave in both amplitude and phase. Thus, when you look at this wave, you perceive the object still being in position even though the object may not be present there. Since during reconstruction (that is, image production), the object wave itself is emerging from the hologram, the image has all the effects of three-dimensionality. You can indeed shift your viewing position and "look behind" the objects.

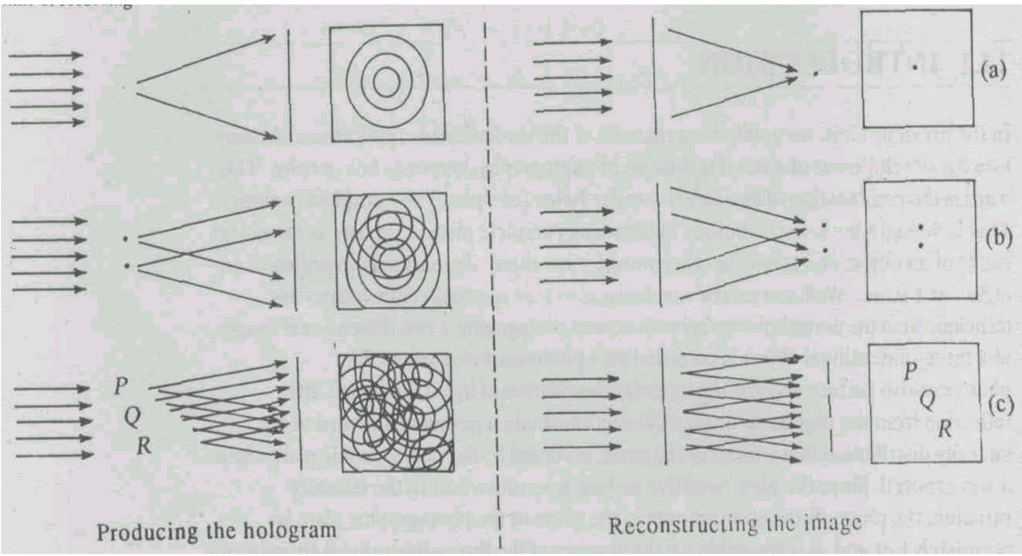

Fig. 3.1 The principle of holography: (a) Point object forming concentric diffraction rings as in zone plate; reconstruction of zone plate gives point image (top right). For two points and a more complex object, these features are shown in (b) and (c) respectively

Let us understand the basic concept involved in holography with the help of a simple example. Incident light, shown in Fig. 3.1 (a), is diffracted by a point object. It gives rise to a series of bright and dark concentric rings. The pattern is recorded photographically and made into a transparency. This pattern, called a **Gabor** zone plate, is similar to a Fresnel zone plate. In the second step (top right) light is incident on the ring pattern (i.e., the Gabor zone plate) and focussed by it into a point, as focussed by a zone-plate.

Now, refer to Fig. 3.1(b) in which the object consists of two points (pixels). The diffraction pattern then consists of two sets of concentric rings. When the pattern is illuminated, each of the two sets focus, and the image consists of two points. As the object is an aggregate of many pixels, its diffraction pattern is shown in Fig. 3.1(c). The intermediate recording is a continuum of superposed zone plates- an unrecognisable multiplicity of lines and rings. Each pixel in the object forms its own set of fringes. Within each set, the light interferes but between sets, there is no fixed phase relationship and hence no interference. In order to make the different signals compatible in phase, another wave called reference is added. Refer to Fig. 3.2 where the effect of adding a sufficiently strong reference beam to the random-phase signal is shown. As a result, the phase of the resultant of reference and the signal becomes similar to that of the reference alone. Thus contributions from different pixels produce an interference fringe pattern.

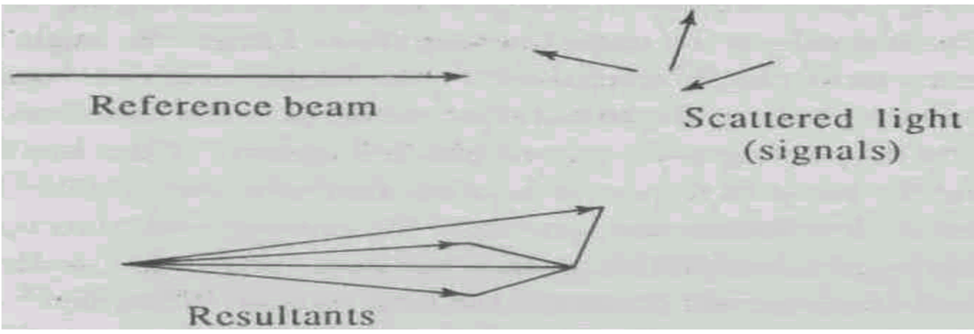

Fig. 3.2 Addition of a strong coherent reference beam (top left) with random-phase signals (top
right) gives similar resultant (bottom)

The essence of holography is that the process of image formation is being interrupted and split into two. In the first step, the object is transformed into a photographic record, called the hologram and in the second step called reconstruction; the hologram is transformed into image. No lens is needed in either step. You may now like to answer an SAQ.

SAQ 1

Using the size of the amplitude vectors drawn in Fig. 3.2 calculate
(a) the ratio of intensities, and (b) the contrast resulting from these intensities.

At this stage, you may say that in photography, what we essentially record is the light reflected from the object and not its diffraction pattern. Well, it is easy to extend the basic idea of holography, explained above in terms of Gabor's zone-plate, to the actual photography situations. Reflected light waves, like other waves, are described by their amplitude (or intensity) and their phase (or frequency). To capture the wave pattern completely (that is, to obtain the hologram) both the amplitude and the phase of the wave must be recorded at each point on the recording surface. As you are aware, recording of the amplitude portion of the wave is achieved in normal photography by converting it to corresponding variation in the opacity of the photographic emulsion. The photographic emulsion is, however, insensitive to phase relations. In holography (also known as wave-front reconstruction), the phase relations are rendered visible to the photographic plate through the technique of interferometry. You may recall from Optics II that interferometry converts phase relations into corresponding amplitude relations.

When two plane waves derived from a common source impinge at different angles on a screen, they produce a set of uniform, parallel interference fringes. The spacing of the fringes depends solely on the angle between the impinging waves (that is, on the path difference between them). A photographic recording of such a fringe pattern results in a grating-like structure. In case of holography, one of these waves is the one reflected from the object (called the object wave) and hence need not be a plane wave. The wavefront of the reflected wave will be highly irregular because of the unevenness of the object surface. When this irregular reflected wave pattern interferes with the reference wave, the resulting interference pattern will not be uniform. Rather, it will have irregular interference pattern – the irregularity of the impinging wavefronts. At places where the signal bearing waves (the object wave) have maximum amplitude, the interference fringes have the greatest contrast and vice-versa. Thus, variations in the amplitude of the object wave manifest as the variation in contrast of the recorded fringe pattern. Can you recall the implications of the spacing of the interference fringes? It is related to the path difference (and hence the phase difference) between the two interfering waves. And the path difference, in turn, depends on the angle between them. The larger the angle between the two interfering waves, the more closely spaced will be the fringes and vice-versa. Therefore, variations in the phase of the object wave manifest as the variations in the spacing of the fringes on the photographic record (the hologram). Thus, in a hologram, both the amplitude and the phase of the signal-bearing wave (the object wave) are preserved as variations in the contrast and spacing of the recorded interference fringes respectively. The hologram obtained in this manner has many properties similar to the diffraction grating about which we will discuss in the next section. When this hologram is illuminated by light of appropriate wavelength, a three- dimensional image of the object can be obtained.

3.4 Holography: The Process

As mentioned earlier, the process of image formation by holography is a two-step process. In the first step, the waves reflected from the object are recorded in such a way that complete information regarding the amplitude and phase variations is preserved. This recording of wavefront is called the hologram. The second step involves the reconstruction of an image of the object by illuminating the hologram by light wave called reconstruction wave (which is identical to the reference wave). In the following, we discuss these two steps and also mention some of the practical considerations about the holographic technique.

3.4.1 Production of a Hologram

Holograms can be produced in several ways depending upon the relative orientation of the reflected (or scattered) and the reference waves. For example, Gabor's zone-plate, which is nothing but a hologram, is the record of interference between the two waves travelling more-or-less in the same direction. This is easily done with objects that have enough open spaces between them, such as a wire mesh or opaque letters on a clear background (Fig. 3.1c). Signal and reference, in other words, travel in the same direction. Such a hologram is called Gabor hologram or in-line hologram. It was only after the invention of laser that this novel technique of photography became truly practical. With the help of lasers, N. Leith and Juris Upatnicks produced what is known as off-axis hologram. In the off-axis hologram, the reference beam and the object beam arrive at the recording plate from substantially different directions. This made possible holography of solid three-dimensional objects. Now, the question arises: How are holograms recorded? To understand this, refer to Fig. 3.3. A beam of coherent laser light (in which all points on the wavefront are in phase) is split into two beams. One beam illuminates the object to be recorded and the light reflected from this object falls on a photographic plate. The other beam, called the reference beam, is reflected from a mirror to the same photographic plate. Due to superposition of wavefronts of these two beams, an interference pattern is recorded on the photographic plate. The record on the photographic plate (hologram) is simply a pattern of interfering wavefronts and shows no resemblance to the recorded object. The hologram, however, contains "all the information" about the object.

Ordinarily, these interference fringes are very closely spaced and cannot be seen by unaided eye. Hence the hologram appears to be uniformly grey. When seen by microscope, however, a hologram is found to consist myriad of tiny "cells", each cell containing a series of fringes of various lengths and spacing. Further, a laser is needed for holography, merely because its coherence length exceeds the path difference due to unevenness of the object.

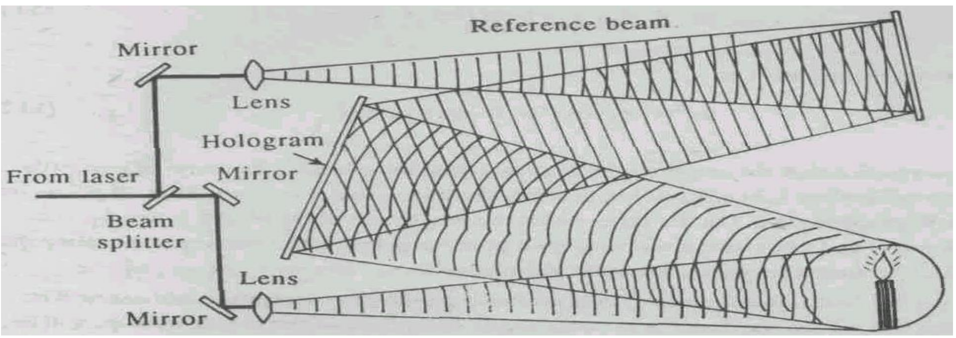

Fig. 3.3: Recording the hologram; microscope lenses broadens both beams without affecting their coherence

Now, having learnt how holograms are recorded, let us pause for a moment and think about the fundamental difference – in terms of technique as well as characteristics – of a hologram and a conventional photograph. This is the subject matter of TQ 1.

3.4.2 Reconstruction of Image

As mentioned above, the hologram of an object is the recording of the interference pattern, on a photographic plate, produced by the object and the reference waves. The hologram, when viewed with unaided eye, does not even remotely resemble the object photographed. The process of obtaining the image of the object is known as **reconstruction**. In the reconstruction process, as shown in Fig. 3.4, the hologram is illuminated by the light beam (which is similar to the reference beam) alone and the reconstructed wavefronts appear to diverge from the image of the object. Let us investigate the process analytically.

Let us represent the wave reflected (or scattered) from the object when it reaches the photographic plate as (Fig. 3.5)

$$\square_1 = A_1(x, y) \cos[\square\, t + \square_1 (x, y) \qquad (3.1)$$

and the reference wave as

$$\square_2 = A_2 \cos[\square\, t + \square_2 (x, y)] \qquad (3.2)$$

You may notice that the amplitude of the reference wave is not a function of x or y (the photographic plate is in the xy-plane) indicating, therefore, that it is constant at all points on the photographic plate. On the other hand, the amplitude of the object wave, A_1 is a function of x and y because it will vary from point to point on the photographic plate due to reflection from the object. Similarly, the phase of the reference wave $\square_2$ will be constant if it (the reference wave) falls normally on the photographic plate and will be a function of x and y if the incidence is at some angle. The phase of the object wave $\square_1$ will be, however, a function of x and y. When these two waves arrive at the photographic plate, the total field distribution will be

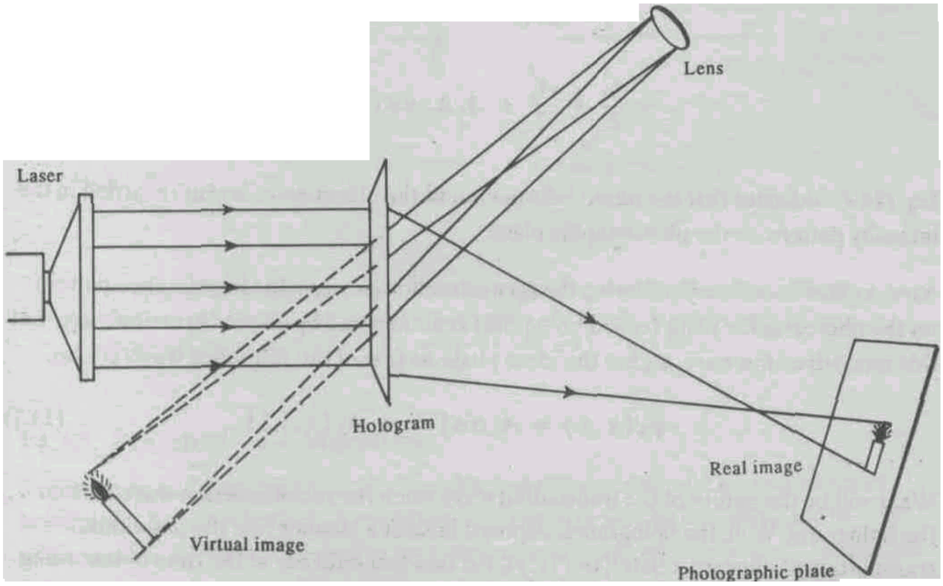

Fig. 3.4 Reconstruction process of an Image in holography

$$\psi_{total} = \psi_1 + \psi_2$$
$$= A_1(x, y)\cos[\omega t + \phi_1(x, y)] + A_2\cos[\omega t + \phi_2(x, y)] \qquad (3.3)$$

As you know, the photographic plate responds only to the intensity. Thus, to get the intensity distribution on the photographic plate, we must take the time average of (3.3)

$$I(x, y) = \langle(\psi_{total})\rangle$$
$$= \langle [A_1(x, y)\cos[\omega t + \phi_1(x, y)] + A_2\cos[\omega t + \phi_2(x, y)]]^2 \rangle$$
$$= A_1^2\langle\cos^2(\omega t + \phi_1)\rangle + A_2^2\langle\cos^2(\omega t + \phi_2)\rangle$$
$$+ 2A_1A_2\langle\cos(\omega t + \phi_1)\cos(\omega t + \phi_2)\rangle$$
$$= \frac{A_1^2}{2} + \frac{A_2^2}{2} + 2A_1A_2\,\frac{1}{2}\langle\cos(2\omega t + \phi_1 + \phi_2) + \cos(\phi_2 - \phi_1)\rangle$$
$$(\because \cos(A + B) + \cos(A - B) = 2\cos A\cos B)$$

$$= \frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1A_2\cos(\phi_2 - \phi_1) \qquad (3.4)$$

Eq. (3.4) indicates that the phase information of the object wave is also recorded in the intensity pattern on the photographic plate.

Now, as mentioned earlier, during the reconstruction process, the interference pattern on the photographic plate (called hologram) is illuminated by a reconstruction wave. Let this reconstruction wave, ψ_3 have the same phase as that of the reference wave, ψ_2. So,

$$\psi_3(x, y) = A_3 \cos[\omega t + \phi_2(x, y)] \tag{3.5}$$

What will be the nature of the transmitted wave when the reconstruction wave falls on the hologram? Well, the hologram is exposed in such a manner that the amplitude transmittance is linearly related to $I(x, y)$, the incident intensity at the time of recording. So, we have, the transmitted wave

$$\psi_4 \propto \psi_3(x, y) I(x, y)$$

$$\psi_4 = \left[\frac{(A_1^2 + A_2^2)}{2}\psi_3 + \frac{A_1 A_2 A_3}{2}\cos(\omega t + \phi_1) + \frac{A_1 A_2 A_3}{2}\cos(\omega t + 2\phi_2 - \phi_1)\right] \tag{3.6}$$

SAQ 2

Starting from the relation $\psi_4 \propto \psi_3(x, y) I(x, y)$, derive Eq.(3.6) using Eqs. (3.4) and (3.5).

The transmitted wave represented by Eq. (3.6) consists of three terms. What do these term signify physically? The first term is the reconstruction wave (ψ_3) with its amplitude modulated by the amplitude of the object wave (A_1). It is so because A_1 is a function of x and y whereas the reference wave amplitude A_2 is a constant. As a result, this part of the transmitted wave will travel, with slight attenuation, in the direction of the reconstruction wave. The second term is identical to the object wave (ψ_1) except for the constant term $(A_2 A_3)/2$. Here lies the beauty of holography. **The hologram and the reconstruction wave have generated a wave which is in every way identical to the wave which originated from the real object itself while recording the hologram**. This part of the transmitted wave forms a virtual image of the object. The third term which is similar to the object wave forms a real image of the object. As a result, a three-dimensional picture of the object can be obtained by placing a camera in the position of real image. The reconstruction process along with various parts of the transmitted wave is shown in Fig. 3.5. You may note that the object is not present when image is reconstructed. However, one of the evolving beam, resulting due to the reconstruction process, is identical to the beam reflected by the object at the time of recording the hologram.

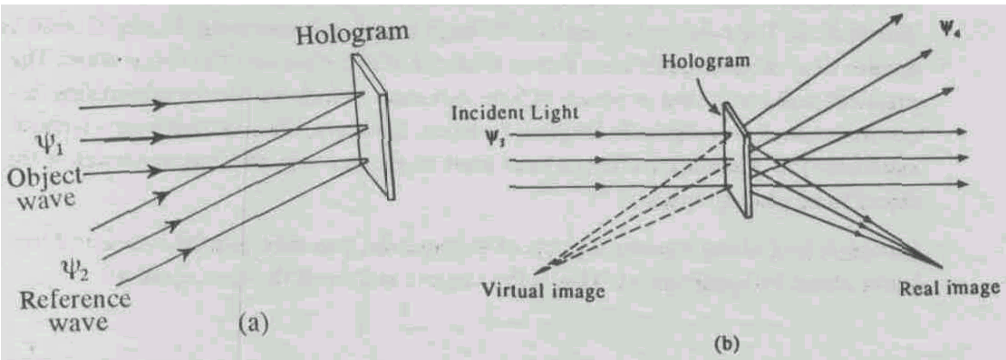

Fig. 3.5: (a) Recording the hologram: Wave reflected from an object interferes with the reference wave, (b) Reconstruction: The hologram diffracts the reconstruction wave, resulting in transmitted wave which produces a real and a virtual image

3.4.3 Practical Considerations

So far, we have discussed physical principles and the experimental arrangements of holography. Suppose you are in the actual process of producing holograms and its subsequent reconstruction to obtain a three-dimensional image of the object. What are the important aspects of the process, and components used therein, about which you should be careful? Well, there are several practical considerations in holography, which are essentially related to the photographic film, the stability and the coherence condition. Let us have a closer look on these practical considerations.

As far as the photographic film is concerned, hologram must be recorded on films of high resolvance. Look again at Fig. 3.3. You may notice that the reference wave, (the light reflected by the mirror), and the signal (the light reflected by the object) subtend certain angle at the photographic plate. If this angle is too large, more than a few degrees, the fringes formed between the signal and reference are very closely spaced and even the best emulsion cannot resolve them. To obtain high resolution, extremely fine-grain film has to be used. But fine-grain films are very slow and hence require larger exposure time (a few minutes). And, if during this exposure time the object moves, the recording of hologram will not be proper. What is the way out of this problem? The way out of the situation is to use high power laser beam to compensate for the exposure time.

Further, the whole system of recording the hologram should be highly stable, i.e., it should be completely free from vibration. Can you say why? It is because the density of the fringes on the photographic film is extremely high. For example, if the angle between the signal and the reference wave is 30^0 (Refer Fig. 3.3) and the wavelength of the laser light is 633nm, the fringe frequency (Refer to Optics II):

= 1/ *d* ; where *d* is fringe width

$$= \frac{\sin\theta}{\lambda} = \frac{\sin 30^0}{633\times10^{-9}} \text{ lines per metre}$$

= 7×10^5 lines per metre.

Can you imagine the smallness of this separation! The fringe width will typically be a thousandth of a millimeter. Therefore, if any component of the holographic set-up moves during recording, the whole fringe pattern will disappear. To meet this stability requirement, the film exposure time should be kept minimum (by using very high power laser) and the holographic system should be isolated from outside vibrations.

The most important and obvious consideration in holography is to use coherent illumination. The coherence length of the laser used for illuminating the object must be greater than the path difference between the reference wave and the object wave. The practical problem is that as the power of laser increases (which we use for minimising the exposure time), its coherence length reduces. Similarly, the coherence area (spatial coherence) of illumination from a laser must be greater than the transverse size of the object to be photographed.

Having learnt about various aspects of holography, you may now be interested to know about its applications. This is the subject matter of the next section.

3.5 Applications of Holography

There are many aspects of holography. Its influence on interferometry, photography, microscopy, astronomy, pattern recognition and even art has only begun to bear fruit. We will now discuss these in brief.

Holographic Interferometry

You will appreciate that, in most of the cases, one of the first areas to benefit from the new technique was the area that gave rise to it. Similar was the case with holography which introduced a new range of powerful methods to interferometry. Interferometry is generally used for precise measurement and comparison of wavelengths, for measuring very small distances or thicknesses (of the order of wavelengths of light), etc. Testing for stresses, strains and surface deformation is one of the most useful practical applications of holographic interferometry.

In the double-exposure technique of holographic interferometry for measuring deformation in object due to strain, two exposures are made of the object – one before loading, and the other after (i.e., under strain). The original object and the object after deformation are recorded holographically on the same photographic plate. The hologram thus obtained is a double exposure, with the second pattern of wave fronts superposed on the first. When this hologram is reconstructed by illuminating it with the reference wave, both images are viewed simultaneously. Since they are slightly different due to deformation, the two images interfere. Thus, any distortion of the object will show in the form of fringes. Like other kinds of interferometry, the technique readily detects changes that produce optical-path difference of the order of a fraction of the wavelength of light. And unlike normal interferometry, however, it is possible to perform experiment quite readily with almost any type of material.

Holographic Microscope

Microscopy has been the primary area of application of holography. In fact, Gabor's discovery of this technique was the outcome of his attempt to enhance the resolving power of an electron microscope. In contrast to a conventional high power microscope, a holographic microscope has an appreciable depth of field and it need not be focussed at all. To see how a holographic microscope functions, refer to Fig. 3.6. The light beam from a laser is split into two. One beam is passing through the specimen and through the microscope, and the other beam is led around it. The two beams interfere on the Film, producing a hologram. The reconstructed image can be viewed in any desired cross-section. The observer merely looks at the cross-section he or she wishes to see, moving back and forth throughout the depth of the image without the object being present at all.

Information Storage

Information can be stored and retrieved more efficiently in the form of holograms than in the form of real images. Further, it is the characteristic of the hologram that it will only reconstruct the holographic image if the reconstruction beam is incident on the hologram at the correct angle. Due to this property, several holograms can be recorded on the same holographic plate by using a slightly different angle between the object and the reference beams for each hologram. Thus, on reconstruction, depending upon the angle of incidence of reconstruction beam, a particular holographic image will be visible. Perhaps this is how information is stored in the brain. If that is the case, it would help explain why attempts to locate certain centres in the brain never met with much success and why brain injury often does not lead to predictable circumscribed defects.

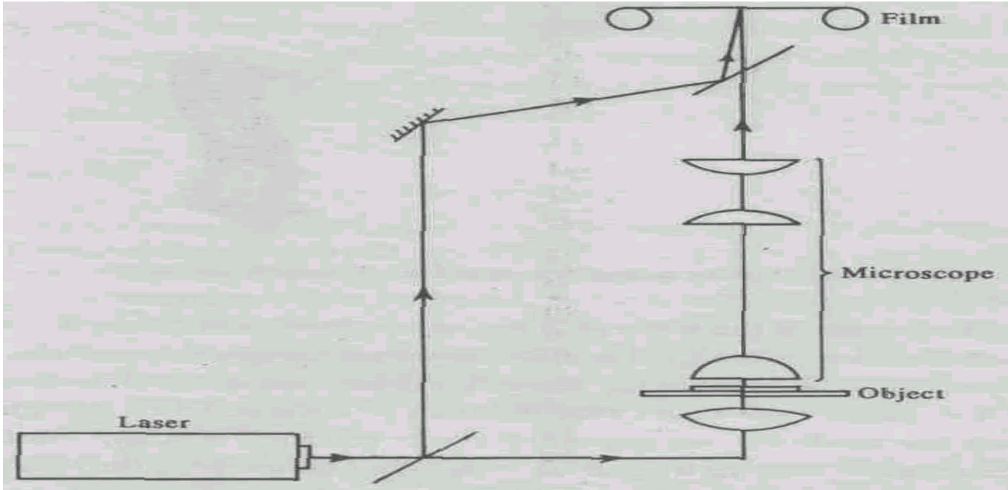

Fig. 3.6 Holographic microscope

Pattern Recognition

One of the most exciting applications of holography is in pattern recognition, also called the character recognition. Early pattern recognition systems, before holography came on the scene, were based on geometrical optics. Consider, for example, that we want to read the letter *A* (Fig. 3.7). *A* set of characters *A*, *B*, *C* ... are printed on a strip of film and this film is moved through the image plane. If the character to be read matches the character on the film, the output from a photo detector is zero, triggering a printer. But, in reality, this does not work. The character and the negative must be aligned perfectly, both in position and size, which is an unrealistic requirement.

Modern pattern recognition systems are based on holography. In place of a mark containing the real image of the letter *A* we may use the hologram of the letter *A*.

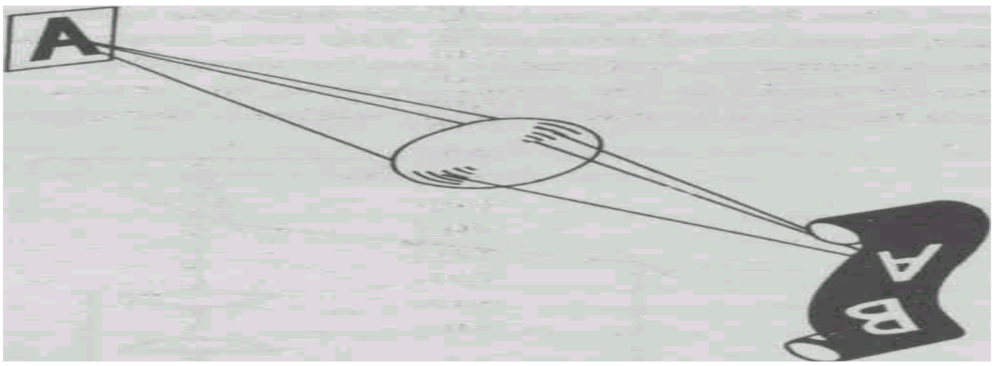

Fig. 3.7 Pattern recognition based on geometric optics

As in holography, the hologram of the letter *A* is the superposition of two sets of wavefronts, the signal and the reference. The signal is diffracted by an original *A* and the reference is a beam of collimated light. Subsequently, when the hologram of *A* is illuminated with light from another *A*, plane wavefronts arise that can be focussed onto a bright spot (Fig. 3.8 top). The spot can easily be recognised by eye or photoelectrically. On the other hand, if the wavefronts are coming from *B* or from other characters, they do not transform into perfectly plane wavefronts and do not produce a focussed spot. Instead, a diffuse patch of light (centre) is produced. Hence, we can scan a given matrix of characters and determine whether or not a particular character is present (bottom).

The holograms shown in Fig. 3.8 appear to be amplitude filters. But because they are generated by interference between signal and reference, they in fact represent both amplitude and phase of light. They are called "complex", "matched", or "vander Lugt filters."

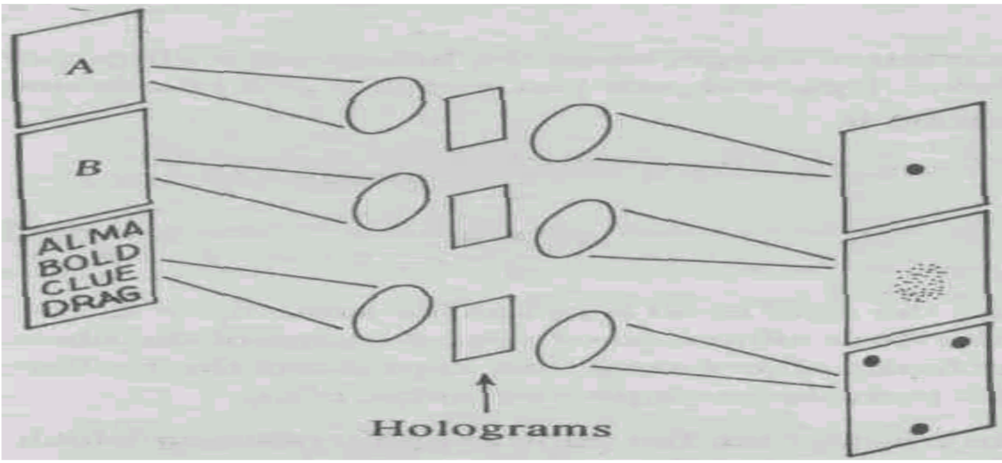

Fig. 3.8 Pattern recognition by holographic vander Lugt filter. (The holograms are seen
between the lenses.)

Form reading machine are a distant reality. Some letters and words are "inside" others. For example *F* is inside *E, P* is inside *R* and *B, T, L* have the same horizontal and vertical lines and 'arc' is inside 'search'. Clearly, the more alike are two characters, the less will be the power of discrimination. Another problem will be to 'teach' the machine to recognise the "meaning" of a letter set in different typeface. The letter *A* can be written in an infinite number of variations possible when it comes to handwriting. However, pattern-recognition using holographs is being extensively used in developing fingerprints library which stores the fingerprints of individuals with dubious character.

3.6 Summary

- Holography, discovered by Gabor, is a novel technique of photography by which a three-dimensional picture of an object or a scene can be obtained. In holography, the interference pattern produced by the light reflected from the object and a reference beam is recorded. Such recording on the photographic plate is called the hologram.

- The three-dimensional picture of the object is obtained by illuminating the hologram by a reconstruction light beam, which in most cases is identical to the reference beam. Holography is, therefore, also known as wavefront reconstruction photography.

- Hologram is produced by splitting a beam of coherent light from a laser into two. One beam is directed, with the help of mirror(s), towards the object and the other is made to fall directly on the photographic plate. The light reflected from the object reaches the photographic plate and interferes with the reference beam. The recorded interference pattern on the photographic plate is the hologram.

- If $\psi_1 = A_1(x, y)\cos[\omega t + \phi_1(x, y)]$ and $\psi_2 = A_2\cos[\omega t + \phi_2(x, y)]$, respectively represents the object wave (wave reflected from the object being photographed) and the reference wave, the intensity distribution on the photographic plate is given as

$$\frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1 A_2 \cos(\phi_2 - \phi_1)$$

- During reconstruction of image, when the hologram is illuminated by the reconstruction wave, $(\psi_3 = A_3\cos[\omega t + \phi_2(x, y)]$ the transmitted wave through the hologram is

$$\psi_4 = \left[\frac{(A_1^2 + A_2^2)}{2}\right]\psi_3 + \frac{A_1 A_2 A_3}{2}\cos(\omega t + \phi_1) + \frac{A_1 A_2 A_3}{2}\cos(\omega t + 2\phi_2 - \phi_1)$$

- The second term on the right hand side has the same form as the object wave and it represents the three-dimensional virtual image of the object. The third term is also similar to the object wave and represents the real image of the object which can be recorded on a photographic plate.

- In order to obtain a hologram, the photographic plate on which the hologram is to be obtained must be of high resolution. This is required because the density of interference fringes in the hologram is extremely high. Also, the whole arrangement of holography – recording the hologram as well as its subsequent reconstruction – must be highly stable, i.e., it should be free from even the slightest mechanical vibration. And of course, we must use coherent light for recording the hologram as well as reconstructing the image.

- Holography has varied applications. Holographic interferometry is a distinct improvement over normal interferometry because the former can be used for any kind of material. Holographic microscopy has enormous magnification and it also offers appreciable depth of field. Holography finds extensive use in information storage and pattern recognition.

3.7 Terminal Questions

1. (a) How is the process of holography different from ordinary photography? (b) Discuss some of the salient features of a hologram?

2. Following Gabor, assume that amplitudes of signals and reference are in ratio 1:10. Suppose that the two beams when they combine may be completely out of phase or in phase. What is the maximum ratio of their intensities?

3. If the angle subtended at the hologram by the signal and the reference beam is 15°, what is the spacing of the fringes provided the wavelength is 492 nm?

3.8 SOLUTIONS AND ANSWERS

SAQs

1. (a) The least possible amplitude (when signal and reference are out of phase, pointing in opposite directions) is 4.36 – 1 = 3.36.

This is because, measuring the lengths of vectors Fig. 3.2, we find that the ratio of signal versus reference is 1: 4.36.

The highest possible amplitude (when signal and reference are in phase) and pointing in the same direction is 4.36 + 1 = 5.36. The ratio of the amplitudes = 3.36/5.36. Thus, the ratio of intensities is

= $(3.36/5.36)^2$ = 0.39

(b) The contrast is given as

$$\frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

$$\frac{(5.36)^2 - (3.36)^2}{(5.36)^2 + (3.36)^2} = 0.44$$

which is high enough to make the reconstruction visible.

2. The transmitted wave is linearly proportional to the incident intensity $I(x,y)$ at the time of recording the hologram and the reconstruction wave, i.e.,

$$\psi_4 \propto \psi_3(x,y)I(x,y)$$

$$\propto \psi_3\left[\frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1A_2\cos(\phi_2 - \phi_1)\right] \quad \text{(using egn 3.4)}$$

$$\propto \frac{(A_1^2 + A_2^2)}{2}\psi_3 + \psi_3[A_1A_2\cos(\phi_2 - \phi_1)]$$

$$\propto \frac{(A_1^2 + A_2^2)}{2}\psi_3 + [A_3\cos(\omega t + \phi_2)][A_1A_2\cos(\phi_2 - \phi_1)]$$

(using equation 3.5)

$$\propto \frac{(A_1^2 + A_2^2)}{2}\psi_3 + A_1A_2A_3[\cos(\omega t + \phi_2)\cos(\phi_2 - \phi_1)]$$

$$= \frac{(A_1^2 + A_2^2)}{2}\psi_3 + A_1A_2A_3 \frac{1}{2}[\cos(\omega t + \phi_2 - \phi_2 + \phi_1) + \cos(\omega t + \phi_2 + \phi_2 - \phi_1)]$$

(using $\cos(A + B) + \cos(A - B) = 2A \cos A \cos B$)

$$\psi_4 = \left[\frac{(A_1^2 + A_2^2)}{2}\psi_3 + \frac{A_1A_2A_3}{2}\cos(\omega t + \phi_1) + \frac{A_1A_2A_3}{2}\cos(\omega t + 2\phi_2 - \phi_1)\right]$$

which is equation (14.6)

TQs

1. (a) The technique of holography (photography by wave front reconstruction) differs from that of ordinary photography in three aspects. First, in ordinary photography, the light reflected from the object is received on the photographic plate with the help of lenses or other image forming device. The amplitude of the light wave, reflected from each point of the object, is recorded at corresponding point on the photographic plate. On the other hand, in holography, no lens or other image-forming device is needed. As such, no image is formed on the hologram. What essentially is obtained is the interference pattern due to the light reflected from the object and the reference beam. Secondly, for obtaining a hologram, coherent light is used whereas in the case of normal photography, no such source of light is needed. The requirement of coherent light is due to the fact that the hologram is an interference pattern. Thirdly, in holography, a set of mirrors is used to render the reference and object beam on the photographic plate.

(b) Hologram has several interesting properties. Some of them are given below:

(i) The image obtained from the hologram has three-dimensional character unlike normal photographs, which are two-dimensional. Due to the three-dimensional character of the image obtained in holography, you can observe different perspective of the object by changing the viewing position. Also, if a scene has been recorded, you can focus at different depths.

(ii) We do not obtain negative in holography. The Hologram itself, however, can be considered as negative in so far as obtaining the positive is concerned. Otherwise, there is no similarity between the typical negative of the ordinary photographs and the hologram. You may have noticed that when the negative of an ordinary photograph is seen through, we do get a feel of the object or the scene photographed. On the other hand, when we look at a hologram we observe a hodgepodge of blobs and whorls; it has no resemblance whatsoever with the original object.

2. Let the amplitude of the signal (or the object wave) be A_1 and that of the reference wave be A_2 , then, as per the problem

$$\frac{A_1}{A_2} = \frac{1}{10}$$

When these two waves are out of phase, their resultant amplitude will be (10 -1) = 9. On the other hand, when they are in phase, the resultant amplitude will be (10 + 1) = 11. Thus, the ratio of their intensities,

$$\frac{I_{\min}}{I_{\max}} = \frac{(9)^2}{(11)^2} = 0.67$$

3. The spacing of the fringes is given as

$$d = \frac{\lambda}{\sin\theta}$$

$$= \frac{492 \times 10^{-9}}{\sin 15^0} \text{ m}$$

$$= 1.8\ \mu\text{m}$$

UNIT 4 FIBRE OPTICS

Unit Structure

4.1 Introduction

You might have seen advertisement displays (made of glass or plastic rods) and illuminated fountains. While looking at these, you might also have noticed that light seems to travel along curved path. In the above-mentioned cases, most of the incoming light is contained within the boundaries of the medium (glass or plastic or water) due to the phenomenon of total **internal reflection.** And since the medium itself has a curved shape, the light travelling through it appears to travel along a curved path. Optical fibre, which is made of transparent glass or plastic, also transmit light in a similar fashion. These fibres are thread like structure and a bundle of it can be used to transmit light around corners and over long distances. Since optical fibre can transmit light around corners, it is being used for obtaining images of inaccessible regions e.g. the interior parts of human body. The real potential of the optical fibres was, however, revealed only after the discovery of lasers.

You may recall from Unit 2 of this course that the discovery of lasers – a source of coherent and monochromatic light – raised the hope of realising communication at optical frequencies. Since increase in the frequency of the carrier wave enables it to carry more information, communication at optical frequencies (~ 10^{15} Hz) has obvious advantages over communication at radio wave (10^{6} Hz) and microwave (~10^{9} Hz) frequencies. But, early attempts at communication at optical frequencies faced a major problem. When optical radiation travels through the Earth's atmosphere, it is attenuated by dust particles, fog, rain etc. Thus, a need for an optical wave guide was felt and the answer was the optical fibres. Optical fibres are an integral part of optical communication – transmission of speech, data, picture or other information – by light. In this unit, you will study optical fibres, especially in the context of optical communication.

In Sec. 4.2, you will learn the physical principles involved in the transmission of light through fibres. Types of fibres used in optical communication has also been described. General considerations about the optical communication through fibres has been discussed in Sec. 4.3. In the same section, you will also learn about the requirements which must be met by optical fibres so that efficient optical communication may take place. The area of optical fibre is relatively new and an exciting field of activity. A thorough understanding demands rather sophisticated mathematical background on the part of the student. It has, therefore, been attempted here to keep the mathematical aspects to a bare minimum and the underlying physical principles have been highlighted.

4.2 Objectives

After going through this unit, you should be able to:

- explain light transmission through fibre
- distinguish between step-index and GRIN fibres
- derive expression for pulse dispersion in fibres, and
- solve simple problems on optical fibres.

4.3 Optical Fibres

An optical fibre consists of a cylindrical glass core surrounded by a transparent cladding of lower refractive index. This assembly is further covered by a plastic coating to protect it against chemical attack, mechanical impact and other handling damages. Fig. 4.1 shows the geometry of a typical optical fibre. The core diameter is in the range 5 μm to 125 μm with the cladding diameter usually in the range 100 μm to l50 μm. The plastic coating diameter is around 250 μm.

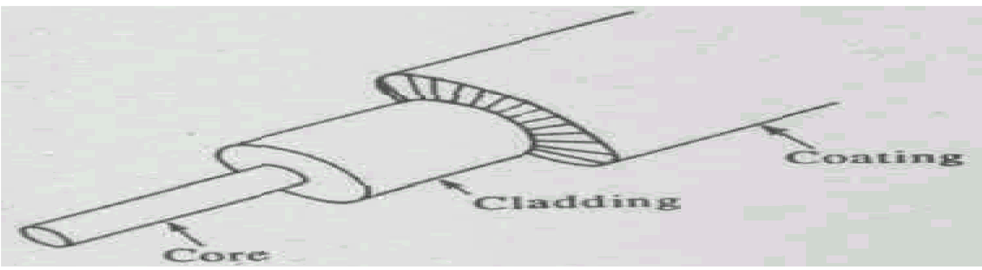

Fig. 4.1 Optical Fibre

In order to understand why the incoming light does not come out through the cylindrical surface of the fibre, you should recall the phenomenon of total internal reflection. You are aware that when light travels from an optically denser medium to a rarer medium, it bends away from the normal as shown in Fig. 15.2(a). If the refractive indices of the two media are n_1 and n_2 such that $n_1 > n_2$ and θ_1 and θ_2 the angle of incidence and the angle of refraction respectively, then, from Snell's law

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} \qquad (4.1)$$

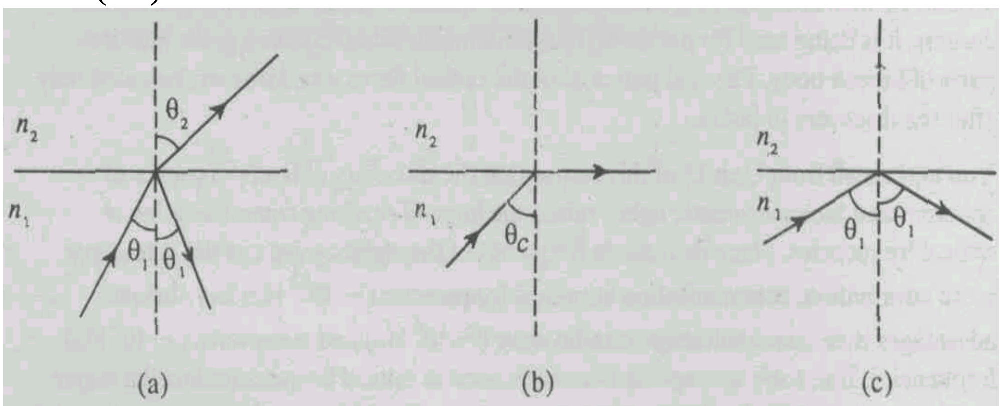

Fig. 4.2 Total Internal reflection

As the angle of incidence is increased, the refracted ray will further bend away form the normal. Ultimately, when the angle of incidence reaches the critical value – known as critical angle, θ_c – the refracted ray travels along the interface separating the two media, as shown in Fig. 4.2 (b). And, when the angle of incidence is increased beyond θ_c, there is no refracted ray and the incident ray undergoes total internal reflection into the optically denser medium, Fig. 4.2(c). This phenomenon is known as total internal reflection and the critical angle, θ_c is given as, from Eq.(4.1)

$$\frac{n_1}{n_2} = \frac{\sin(\pi/2)}{\sin\theta_c} \quad \Rightarrow \quad \theta_c = \sin^{-1}\left(\frac{n_2}{n_1}\right) \qquad (4.2)$$

Transmission of light, based on above principle, through an optical fibre of core refractive index n_1 and cladding refractive index n_2 with $n_1 > n_2$ is shown in Fig. 4.3(a).

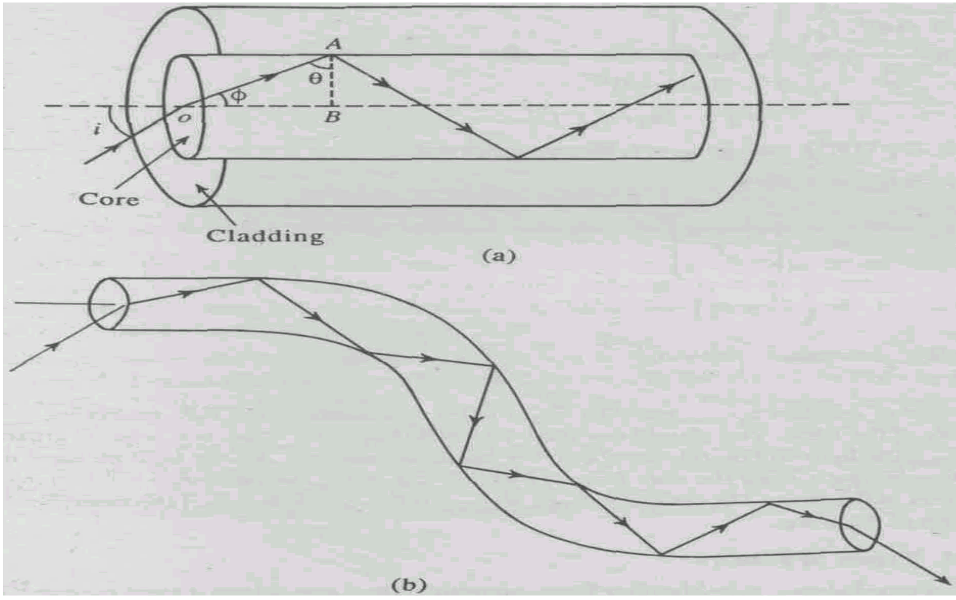

Fig. 4.3: (a) Light propagation through a fibre by total internal reflection. (b) Light propagation through a bent fibre

When the ray of light is incident at angle θ ($> \theta_c$) at the core-cladding interface, it undergoes total internal reflection. Due to the cylindrical symmetry of the fibre, the ray undergoes total internal reflection at subsequent incidences at the core-cladding interface and hence gets trapped inside the fibre. Due to this "guiding" property, optical fibres are also called **"Optical Waveguides."** Fibres in the bent form can also guide the light, as indicated in Fig. 4.3(b), provided that, even at curved portions, the angle of incidence is greater than θ_c. Do you know why cladding material is needed? The need for a cladding material of lower refractive index is due to two reasons, First, to achieve total internal reflection at the core-cladding interface. Secondly, when light undergoes total internal reflection, a part of it penetrates into the cladding material (region of lower refractive index). This may lead to leakage of light, and it may also couple with the light travelling in adjacent fibres. The use of sufficiently thick cladding material prevents this type of loss.

You may note, from Eq. (4.2), that the critical angle for the incident ray depends on the refractive indices of the core and the cladding material. In Fig. 4.3(a), θ is the angle at which incident light falls on the core-cladding interface and this angle is different from the angle, i , at which light is incident at the entrance aperture of the fibre. It is so because the entrance aperture is an air (refractive index $n_0 \sim 1$)-glass (refractive index n_1) interface. Thus, according to Snell's law, (refer to Fig. 4.3(a))

$$n_0 \sin i = n_1 \sin\phi \qquad (4.3)$$

Now, if this ray has to undergo total internal reflection at the core-cladding interface, from Eq.(4.2)

$$\sin\phi \geq n_2 / n_1$$

from $\triangle OAB$,

$$\sin\theta = \sin(90^0 - \phi) = \cos\phi$$

$$= (1 - \sin^2\phi)^{1/2}$$

$$= [1 - (n_2 / n_1)^2]^{1/2}$$

Hence, Eq. (4.3), taking $n_0 = 1$, may be written as,

$$\sin i_{max} = n_1 \sin\theta$$

$$= n_1 \left[\frac{n_1^2 - n_2^2}{n_1^2}\right]^{1/2}$$

$$= \left(n_1^2 - n_2^2\right)^{1/2}$$

$$i_{max} = \sin^{-1}\left[n_1^2 - n_2^2\right]^{1/2} \qquad (4.4)$$

The angle of incidence, i_{max}, given by Eq. (4.4) is a measure of the light gathering capacity of the fibre. You should convince yourself that if the incidence angle is greater than i_{max}, the light will be refracted into the cladding material. All the light incident on the fibre aperture along the core formed by $i = 0$ to $i = i_{max}$ will undergo total internal reflection in the fibre. The quantity $(n_1^2 - n_2^2)^{1/2}$ in Eq. (15.4) is called the numerical aperture of the fibre.

4.3.1 Types of Fibres

As mentioned above, in its simplest form, an optical fibre consists of a glass core and a cladding (also of glass) of lower refractive index. This type of fibre in which there is a sudden change in the refractive index at the core-cladding interface is called **Step-index fibre**. The variation of the refractive index with the radius of such a fibre is shown in Fig. 4.4.

Further, when light travels through optical fibres, there are different types of losses as well as a broadening of the pulse. These aspects of the optical fibres are of vital importance for optical communications and have been discussed in the next section. In order to overcome some of the inherent deficiencies of the step-index fibres, another type of fibre in use is called **GRadient-INdex Fibre** (or **GRIN-fibre**).

In optical communication, signal is transmitted through the fibre in the form of pulses.

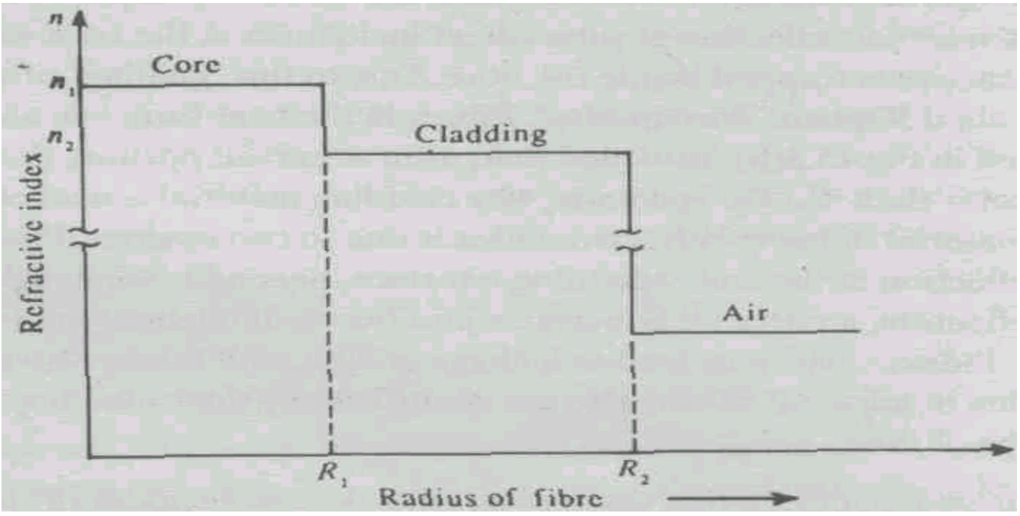

Fig. 4.4: Refractive index profile of a step-index fibre

In the GRIN fibre, the refractive index of the core material decreases continuously along the radius, nearly in parabolic manner, from a maximum value at the centre of the core to a constant value at the core-cladding interface. The variation of the refractive index, with radius, of a GRIN - fibre is shown in Fig. 4.5(a).

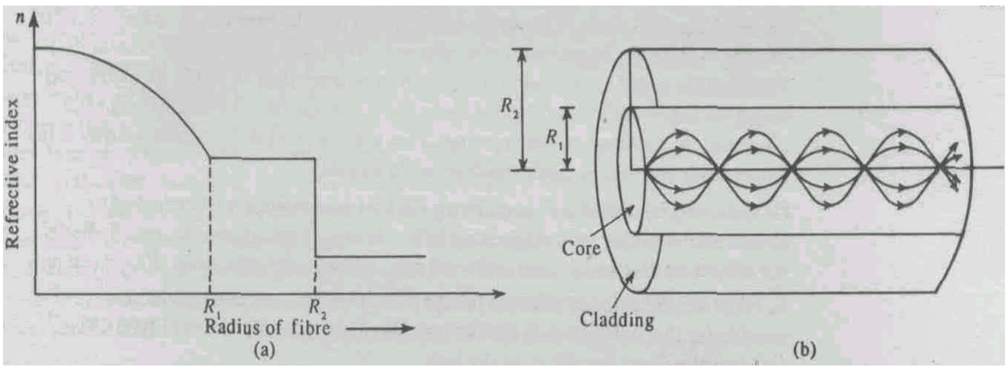

Fig. 4.5 (a) The refractive index profile of a Gradient-INdex fibre; (b) Ray paths in such a fibre

Since the refractive index gradually decreases as one moves away from the axis of the fibre, a ray that enters the fibre is continuously bent towards the axis of the fibre as shown in Fig. 4.5(b). Can you explain why this happens? This smooth bending of the ray towards the axis is again a consequence of Snell's law. As the ray moves away from the centre, it encounters media of lower and lower refractive indices and hence bends towards the axis of the fibre. Can you name a natural phenomenon which results due to the atmospheric gradient of refractive index? You guessed rightly – the Mirage, which is observed while looking across an expanse of hot desert or on a tarmac on a hot, sunny day is one such example.

SAQ1

What will happen if the refractive index of the cladding material is higher than that of the core?

Having learnt about the basic principles involved in the transmission of light in optical fibres, let us study some of its important features as a component of an optical communication system. But before we do that, let us see what uses optical fibres have been put to.

4.3.2 Applications of Optical Fibres

The most elementary application of optical fibres is in the transmission of light.

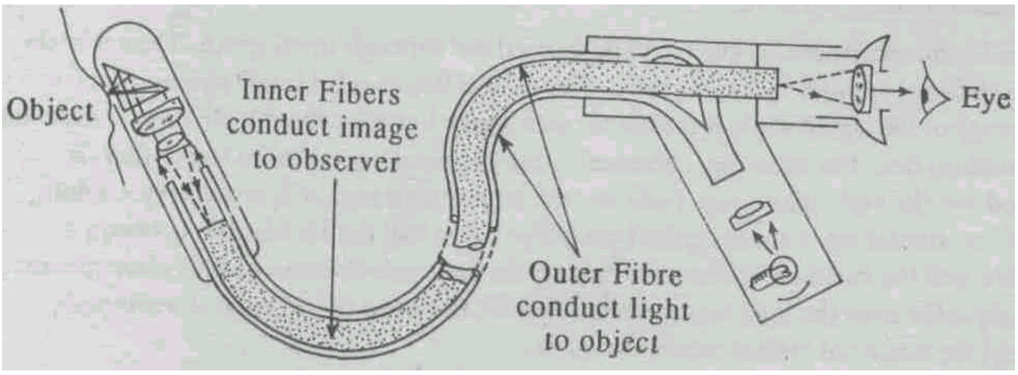

Fig. 4.6 Flexible fibrescope

An example of transmission of images using optical Fibres is in the flexible **fibrescope.** As shown in Fig. 4.6, some of the fibres conduct light into the cavity to be examined, while the others carry the image back to the observer. The image conducting fibres, up to 140000 of them, are by necessity very thin, often no more than 10 □ *m* in diameter, and the entire fibre-bundle has diameter of the order of a few mm. Fibrescopes are used extensively in medicine and engineering. They make it possible to inspect a cavity in the human body and to look inside the heart while it beats.

Of increasing interest is the use of fibre guides for communication. Compared to electrical conductors, optical fibres are lighter in weight, less expensive, equally flexible, not subject to electrical interferences and more secure to interceptions. Fibres can now be made which have losses as low as 0.2dB/km. This is a remarkable achievement considering that only two decades ago the best fibres had losses in excess of 1000 dB/km and 20 dB/km was thought to be the limit.

4.4 Optical Communication Through Fibre

As mentioned earlier, optical communication refers to the transmission of speech, data, picture or other information by light. You may recall from Unit 2 that the replacement of radiowaves and microwaves by light waves is especially attractive because of the enhanced information carrying capacity of the latter. Optical frequencies are some five orders of magnitude higher than, say, microwave frequencies. Therefore, larger volume of information can be transmitted through fibre cable compared to that through copper coaxial cable (used for microwave communication) of similar size. Further, in contrast with metallic conduction techniques (e.g. through copper cables), communication by light offers the possibility of complete electrical isolation, immunity to electromagnetic interference and freedom from signal leakage. In a typical optical communication system, the information-carrying signal originates in a transmitter, passes through an optical fibre link or an optical channel and enters a receiver, which reconstruct the original information. In order to minimize the distortion, the signal is encoded into digital form before transmission. In this way, retrieval of the signal at some distance down the line depends only on the recognition of either the presence or the absence of a pulse representing a binary (0 or 1) digit. Minor distortion and noise may therefore be tolerated as long as pulses can be detected and regenerated, free from distortion.

You may be wondering that with above advantages, why light was not used for communication purposes. It is not as if these advantages of using light as carrier of information were not known. Rather, it was the unavailability of a suitable source of light, which could be modulated. Light from lasers, being highly monochromatic, can effectively be modulated by the information carrying signals. The laser light, acting as the carrier wave, responds, either directly or indirectly, to the electrical signal say, from telephone. These signals can, therefore, modulate the carrier wave, which then travels through the optical fibre (the optical waveguide). At the receiving end of the fibre, a photodetector receives and demodulates these optical signal into sound waves. For long distance optical transmission line, yet another component, called **repeater** is used in optical communication system. A repeater essentially amplifies and reshapes the signal and retransmits it along the fibre.

Optical communication, as such, can be carried out through open space. Then why do we need fibres to carry optical signals? The reason lies in substantial attenuation (or damping) of the signal while it travels in open space between the information source and information use. For example, communication between one satellite to another is carried out through open space because the intervening region is essentially a vacuum. However, similar open space optical communication will not be feasible between a satellite and the earth or between two places on the earth because the earth's atmosphere strongly influences the transmission of light. Hence, the need for an optical waveguide (fibres) for terrestrial optical communication.

Well, you have learnt in the previous section how light is transmitted through optical fibres. But, is this property of fibres enough for transmitting information carrying signals from one point to another? No, the optical fibre must have some additional characteristics if at all it has to serve as an effective optical signal-carrying medium. The optical fibre should be, as much as possible, free from pulse dispersion in order to carry large volume of information. Pulse dispersion arises because different light rays take different times to travel a fixed length in the fibre. Secondly, as we know, even the light from lasers may have a spread in its wavelength. That is, even laser light is not completely monochromatic. And since the refractive index of fibre material is a function of wavelength, light of different wavelengths will travel with different velocities. This inherent property of material is yet another cause of pulse dispersion and is known as material dispersion. Further, the optical radiation will be attenuated by the material of the fibre due to scattering and other phenomenon. In the following you will learn how these problems can be tackled.

4.4.1 Pulse Dispersion in Fibres

You may recall from Sec. 4.2 that rays of light incident at the core-cladding interface at an angle greater than the critical angle θ_c undergoes total internal reflection and propagates through the fibre as shown in Fig. 4.7.

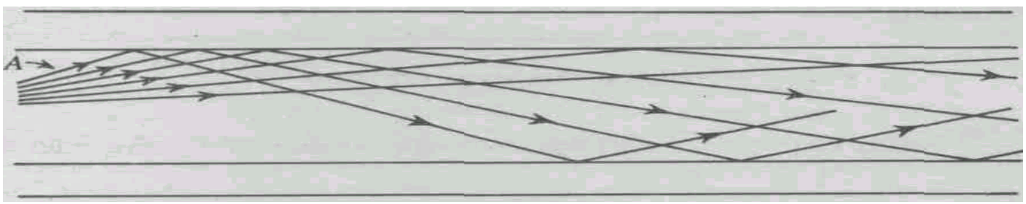

Fig. 4.7 Rays of light passing through a fibre

However, the ray, marked *A* in Fig. 4.7, which is incident on the core-cladding interface at the largest angle will travel a longer optical path as compared to other rays incident at smaller angles. As a result, different rays will take different times in traversing a given length of the fibre. This causes broadening of the information carrying pulses, as shown in Fig. 4.8. What effect does the pulse broadening have on the signal transmission capacity of the fibre?

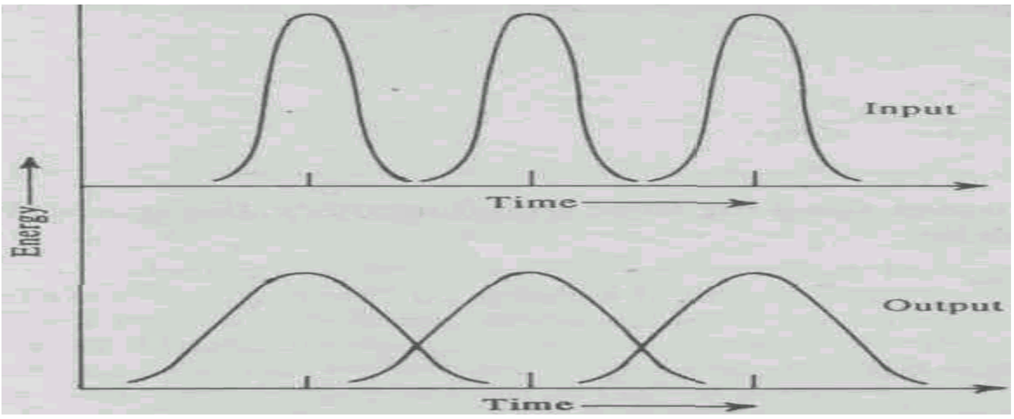

Fig. 4.8 Pulse dispersion: (a) At the Input, the information carrying pulses are well resolved, (b) At the output, due to broadening, pulses overlap and are irresolvable

Well, pulse broadening severely restricts the transmission capacity of the fibre. It is so because the pulses which are well resolved (Fig. 4.8a) at the input may overlap at the output (Fig. 15.8b) due to pulse broadening. To avoid this overlap, the time delay between two consecutive pulses must be increased. Therefore, the number of pulses that can be transmitted per unit time through the fibre will go down, that is, the transmission capacity of the fibre will be reduced.

To have a quantitative idea about the pulse dispersion in case of propagation through step-index fibre, refer to Fig. 4.9. Let a ray of light be incident at an angle *i* with the axis of the fibre. The time taken by this ray to travel a distance *PR*

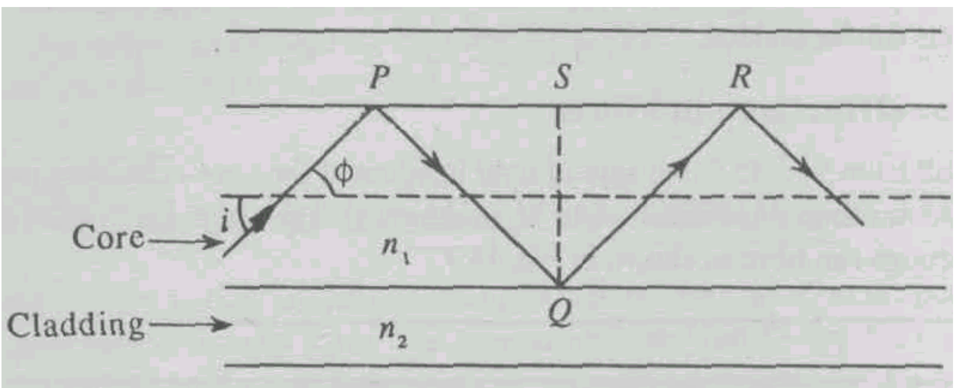

Fig. 4.9 Ray of light passing through a step-index fibre

$$t = \frac{PQ + QR}{c / n_1}$$

where c / n_1 = velocity of light in the core medium (refractive index n_1)

$$t = \frac{n_1}{c}\frac{1}{\cos\theta}(PS + QR)$$

$$= \frac{n_1(PR)}{c\cos\theta}$$

What does this relation indicate? It indicates that the time taken by the ray of light in travelling a distance through the fibre depends on the angle it makes with the axis of the fibre. Thus, for a fixed length *L* of the fibre, minimum time will be taken by a ray which travel along the axis of the fibre ($\theta = 0$) i.e.,

$$t_{min} = n_1 L / c$$

and the maximum time will be taken by the ray for which θ is equal to $(\pi/2 - \theta_c)$; where θ_c is the critical angle at the core-cladding interface. Thus, $\theta = \cos^{-1}(n_2 / n_1)$ and the maximum time

$$t_{max} = \frac{n_1 L}{c(n_2 / n_1)} = \frac{n_1^2 L}{cn_2}$$

Thus, if all the input rays travel along the fibre simultaneously, the spread in time in traversing a distance *L* will be

$$\Delta t = t_{max} - t_{min}$$

$$= \frac{n_1 L}{cn_2}(n_1 - n_2) \qquad (4.5)$$

SAQ2

If the core and cladding refractive indices for a step-index fibre is 1.47 and 1.46 respectively, what will be the broadening of a pulse after a distance of 5 km?

Due to the pulse dispersion represented by Eq. (4.5), the signal transmission capacity of optical fibres is severely restrained. Therefore, an efficient optical fibre should have least possible pulse dispersion so that it can carry larger number of pulses per unit time.

Now the question is: Do we have any method of minimizing the pulse broadening in optical fibres? Yes, there are methods by which we may minimize the pulse broadening. One of them is to use gradient-index (GRIN) fibre. In the following, you will learn how GRIN-fibres help in reducing pulse broadening.

4.4.2 Pulse dispersion: GRIN Fibres

You may recall form Sec. 4.2 that the core of the GRIN-fibre offers gradually decreasing refractive index environment to light rays as it moves away from the axis of the fibre. Let us see how this parabolic refractive index profile of the GRIN-fibre (Fig. 4.5(a)) helps in reducing the pulse dispersion. Refer to Fig. 4.10 in which two rays *A* and *B* are shown to enter the core axis at different angles. As the rays move towards the core-cladding interface, they encounter decreasing refractive index environment. As a result, both of them will bend away from the normal and hence towards the axis of the core. The paths taken by rays are not straight lines as in the case of step-index fibre; rather, it is sinusoidal. It is because in the core, refractive index is a continuously decreasing function of the core radius. Now, ray *A* which makes the smaller angle with the axis travels smaller distance through the core whereas ray *B* travels a longer distance. However, the time taken by both of them, separately, in traversing a fixed distance along the fibre will be the same. Can you say why? It is so because ray *A* which travels a shorter distance, does so in the region of higher refractive index. Hence the velocity of light along the path taken by ray *A* will be smaller (velocity of light = *c/n).* On the other hand, ray *B* which travels a relatively larger distance, does so in the region of lower refractive index and hence with higher velocity. The net result is that the rays making different angles with the core axis take equal time in propagating through the fibre. Due to this reason, the pulse broadening is reduced in GRIN-fibre.

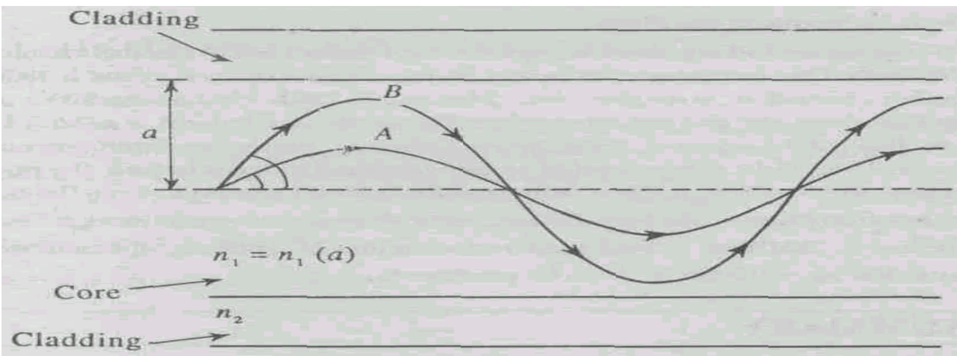

Fig. 4.10 Two rays A and B travelling through a GRIN-fibre

The volume of information which may be transmitted through GRIN-fibre is more or less free from pulse broadening due to the above reason. The information carrying capacity of such fibre is only limited by material dispersion about which you will learn in the following.

4.4.3 Material Dispersion

Above, we discussed about the pulse broadening in optical fibres arising because of the fact that light rays incident at different angles at the core-cladding interface take different times to traverse a fixed length of fibre. We also discussed how to reduce this dispersion by using GRIN-fibre. Now, suppose that the light beam travelling through the fibre is free from the pulse broadening due to the above mentioned-reason. Does it mean that the beam is free from pulse broadening? No, there is yet another source of pulse broadening known as material dispersion. Material dispersion arises due to the variation of refractive index with wavelength, i.e. the velocity of light in the medium is dependent upon its wavelength. You are aware that light even from a highly pure source (like laser which gives highly monochromatic light) will have a spread in its wavelength. Therefore, different wavelengths, within the range, will travel with different velocities and hence will arrive at the end of the fibre at different times and cause broadening of the pulse. You may note that material broadening is an intrinsic physical property of the fibre material.

Although glass is transparent to electromagnetic radiation in the visible range, it does absorb a part of it due to several processes. As a result, the input power of the light beam will suffer a loss while traversing the length of the fibre. In the following, we briefly discuss some of the processes causing power loss in fibres.

4.4.4 Power Loss

When electromagnetic radiation interacts with matter, it may lose energy via different mechanisms. In the case of optical fibre material, silica, major loss in energy or power is caused due to the absorption of photons by impurity atoms. Therefore, to minimise this loss, the fibre material should be of high purity. Secondly, the photons may also lose energy by exciting the atoms of oxygen and silicon (the building blocks of silica, SiO_2). Thirdly, silica being an amorphous material, offers randomly varying refractive index. Due to this, the propagating light beam may get scattered and its direction of propagating may change drastically. These loss-causing mechanisms are taken care of by proper design and synthesis technique of the fibres.

The power loss we are talking about is expressed in terms of bel or decibel, which are comparable units. One bel means that power in one channel or at one time is 10 times that in another channel, or at another time. 2 bel means 100 □, 3 bel means 1000 □, and so on. For practical use, the unit bel is too large. Hence the decibel, dB, is used. 1 bel = 10 dB. A decibel (dB) is equal to $10 \log_{10}(p_2/p_1)$ where p_1 and p_2 are input and output power levels. Thus, if the power level of an optical signal reduces by half, the power loss in decibels will be $10\log_{10}(1/2) = -3$dB. In optical fibre communication, the power loss is expressed as dB/km. In long distance optical communication through fibres, the permissible loss is 20dB/km. With modern techniques of synthesis, optical fibres with power loss as low as ~ 0.2dBkm can be produced.

4.5 Summary

An optical fibre consists of a transparent glass core and a cladding of lower refractive index. Since the refractive index of the cladding material is lower than that of the core, much of the light launched into one end will emerge from the other end due to a large number of total internal reflections.

In the step-index fibre, the refractive index changes suddenly at the core-cladding interface. On the other hand, in the gradient-index (GRIN-) fibres, the refractive index decreases continuously from the core axis as a function of radius.

The maximum entrance core angle, also known as acceptance angle, is a measure of the light gathering capacity of the fibre and is given as

$$\sin i_{max} = \frac{1}{n_0}\left[n_1^2 - n_2^2\right]^{1/2}$$

The term $\left[n_1^2 - n_2^2\right]^{1/2}$ is known as the **numerical aperture** of the fibre.

In optical communication, information is transmitted in the form of pulses. While travelling through the fibre, these pulses broaden because rays incident at different angles at the core- cladding interface take different times in traversing a fixed length of the fibre. Pulse broadening due to this reason in a step-index fibre of length L is given as,

$$\Delta t = \frac{n_1 L}{c n_2}[n_1 - n_2]$$

Pulse broadening can be greatly reduced if, instead of step-index fibre, we use a GRIN-fibre. It is so because in GRIN-fibre, though different rays traverse different optical paths in the core, they all take same time in travelling through a given length of the fibre.

Material dispersion is yet another cause of pulse broadening. Material dispersion arises because the refractive index (and hence the velocity of light) in a medium is a function of the wavelength of light. And, even highly monochromatic light has a spread in its wavelength.

4.6 Terminal Questions

1. Suppose you have two optical fibres *A* and *B.* The refractive indices of the core (n_1) and the cladding (n_2) materials is

$$(n_1)_A = 1.52,\ (n_2)_A = 1.41,\ (n_1)_B = 1.53,\ (n_2)_B = 1.39$$

Which of the two fibres will have higher light gathering capacity?

2. A step-index fibre 6.35×10^{-5} m in diameter has a core of refractive index 1.53 and a cladding of refractive index 1.39. Determine (a) the numerical aperture for the fibre; (b) the acceptance angle (or maximum entrance cone angle).

4.7 Solutions and Answers

SAQs

1. If the refractive index of the cladding material is higher than that of the core material of the fibre, the incoming light will not undergo total internal reflection. It is so because when the light travels from a rarer to denser medium, it bends towards the surface normal. Thus, the light ray incident on the core-cladding interface will, instead of coming inside the core, get refracted in the cladding material (refer to Fig. 4.2).

2. The pulse broadening is given as

$$\Delta t = \frac{n_1 L}{c n_2}[n_1 - n_2]$$

As per the problem,

$L = 5 \times 10^3$ m, $n_1 = 1.47$, $n_2 = 1.46$ and $c = 3 \times 10^8$ m/s

So,

$$\Delta t = \frac{1.47 \times 5 \times 10^3 (m)}{3 \times 10^8 (m/s) \times 1.46}(1.47 - 1.46)$$

$$\Delta t = \frac{1.47 \times 5 \times 10^3 (m)}{3 \times 10^8 (m/s)}(0.01)$$

$$= 0.17\ \mu s$$

TQs

1. Refer to Fig. 4.3. The maximum angle of incidence, i_{max}, of the light beam at air-core interface is the measure of the light gathering capacity of the fibre. The sine of this angle of incidence is given as

$$\sin i_{max} = \frac{1}{n_0}\left[n_1^2 - n_2^2\right]^{1/2}$$

where n_0, n_1 and n_2 are the refractive indices of air, core and cladding respectively.

n_0 = refractive index of air = 1

For fibre *A*,

$n_1 = 1.52$ and $n_2 = 1.41$

$$\sin i_{max} = \left[(1.52)^2 - (1.41)^2\right]^{1/2}$$
$$= [0.3223]^{1/2}$$

$$(i_{max})_A = \sin^{-1}(0.57) \approx 35^0$$

For fibre *B*,

$n_1 = 1.53$ and $n_2 = 1.39$

$$\sin i_{max} = \left[(1.53)^2 - (1.39)^2\right]^{1/2}$$

$$(i_{max})_B = \sin^{-1}(0.64) \approx 40^0$$

2 (a) The numerical aperture of the fibre is given as

$$N.A = (n_1^2 - n_2^2)^{1/2}$$

$$= [(1.53)^2 - (1.39)^2]^{1/2}$$

$$= 0.64$$

(b) The acceptance angle or the maximum entrance angle, i_{max}, corresponds to θ_c, the critical angle for total internal reflection at the core-cladding interface.

$$\sin i_{max} = \frac{1}{n_0}(N.A)$$

$$= 0.64$$

$$\therefore \quad i_{max} = \sin^{-1}(0.64)$$

$$\approx 40^0$$

4.8 References/Further Reading

Fedor Mitschke. (2016). Fiber Optics Physics and Technology. Springer Berlin.

Richard A Dunlap. (2019). The basic physics of lasers. Morgan & Claypool Publishers.

OCW/Videos- https://ocw.mit.edu/courses/8-03sc-physics-iii-vibrations-and-waves-fall-2016/pages/part-iii-optics/