PHY 209	OPTICS 1

Course Code	PHY 209
Course Title	Optics I
Course Adaptor	Dr. Sanjay Gupta School of Science and Technology National Open University of Nigeria 14/16 Ahmadu Bello Way Victoria Island, Lagos
Programme Leader	Dr. Sanjay Gupta School of Science and Technology National Open University of Nigeria Victoria Island, Lagos

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria
Headquarters
14/16 Ahmadu Bello Way
Victoria Island
Lagos

Abuja Office
No. 5 Dar es Salaam Street
Off Aminu Kano Crescent
Wuse II, Abuja
Nigeria

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

Published by:
National Open University of Nigeria 2008

First Printed 2008

ISBN: 978-058-352-1

CONTENTS **PAGES**

Course Introduction

This course is intended to introduce light. From your previous physics courses, you may be familiar with some of the topics included here. But we have included such topics so to make the course self-contained. In Unit 1 we have shown that light is a transverse electromagnetic wave. The wave equations for **E** and **B** are derived from Maxwell's field equations. In Unit 2 we have discussed reflection and refraction of electromagnetic waves. You will also learn that all the laws of geometrical optics are inherent in Fermat's principle.

Perception of light by humans is discussed in Unit 3. You will learn that human vision involves a mix of physical and physiological processes. The role of eye as an image-forming device is discussed in detail. Theories of colour vision are also given in brief. Unit 4 discusses three polarisation states of light. You will learn that light can be polarised by reflection, refraction and selective absorption. Light propagation in anisotropic crystals and the phenomenon of birefringence are discussed in detail.

UNIT 1 NATURE OF LIGHT

CONTENTS

1.0 INTRODUCTION

You all know that light is responsible for our intimate contact with the universe through one of our sense organs. We are able to admire the wonders of the world and appreciate the beauty of nature only when there is light. The reds of the sun or the ruby, the greens of the grass or emerald and the blues of the sky or sapphire involve light. In a way light plays a vital role in sustaining life on earth. Even so, we are strangely unaware of its presence. We see not light but objects, (shapes, colours, textures and motion) as constructed by the brain from information received by it.

Have you ever thought: What is light? How light behaves when it reaches our eyes? And so on. These questions proved very difficult even for the genius of the class of Newton and Einstein. In fact, search for answers to these gave birth to a new branch of physics: **Optics,** which is extremely relevant to the modern world. It occupies a prominent place in various branches of science, engineering and technology. Optical studies have contributed to our understanding of the laws of nature. With the development of lasers, fibre optics, holography, optical communication and computation, optics has emerged as a fertile area of practical applications. It is therefore important for you to understand the language and vocabulary of optics very thoroughly.

In this unit you will learn some important facts and developments which were made to unfold the nature of light. In Sec. 3.1 you will learn about the corpuscular (particle) model of light. In Sec. 3.2 we have discussed the wave model of light, with particular reference to electromagnetic waves. You may now be tempted to ask: Does light behave like a particle or a wave? You will learn that it is like neither!

2.0 OBJECTIVES

After going through this unit you should be able to:

- name phenomena distinguishing corpuscular and wave models of light
- derive an expression for the velocity of electromagnetic waves
- specify the frequency ranges of different portions of electromagnetic spectrum, and
- explain the importance of Poynting Vector.

3.0 MAIN CONTENT

3.1 The Corpuscular Model

You must have read in your school physics course that corpuscular model is due to Newton. Contrary to this popular belief, the credit should be given to Descartes, although the earliest speculations about light are attributed to Pythagoras.

The speed of propagation of light has been measured by a variety of means. The earliest measurement made by Roemer in 1676, who made use of observations of the motion of the moons of Jupiter and apparent variations in the periods of their orbits resulting from the finite speed of propagation of light from Jupiter to earth. The first completely terrestrial measurement of the speed of light was made by Fizeau in 1849.

The corpuscular model is perhaps the simplest of the models of light. According to it, **light consists of minute invisible stream of particles called corpuscles**. A luminous body sends corpuscles out in all directions. These particles travel without being affected by earth's gravitation. Newton emphasized that corpuscles of different sizes stimulate sensation of different colours at the retina of our eye.

In your physics courses at school you must have learnt about evidences in favour of this model. Can you recall them? The two most important experimental evidences are:

(i) Light travels in straight lines. This rectilinear propagation of light is responsible for the formation of sharp (perfectly dark) shadows. If we illuminate a barrier in front of a white screen, the region of screen behind the barrier is completely dark and the region outside the barrier is completely lit. This suggests that light does not go around corners. Or does it?

(ii) Light can propagate through vacuum, i.e., light does not require any material medium, as does sound, for propagation.

We can also predict the correct form of the laws of reflection and refraction using the corpuscular model. However, a serious flaw in this theory is encountered in respect of the speed of light. Corpuscular model predicts that light travels faster in a denser medium. This, as you now recognise, contradicts the experimental findings of Fizeau. Do you expect the speed of light to depend on the nature of the source or the medium in which light propagate? Obviously, it is a property of the medium. This means that the speed of light has a definite value for each medium.

The other serious flaw in the corpuscular model came in the form of experimental observations like interference (re-distribution of energy in the form of dark and bright or coloured fringes), diffraction (bending around sharp edges) and polarization.

You may now like to answer an SAE.

SELF ASSESSMENT EXERCISE 1

Grimaldi observed that the shadow of a very small circular obstacle placed in the path of light is smaller than its actual size. Discuss how it contradicts corpuscular model.

In the experiment described in SAE 1, Grimaldi also observed coloured fringes around the shadow. This, as we now know, is a necessary consequence of the wavelike character of light. It is interesting to observe that even though Newton had some wavelike conception of light, he continued to emphasize the particle nature. You will learn about the wave model of light in the following section.

3.2 The Wave Model

The earliest systematic theory of light was put forward by a contemporary of Newton, Christian Huygens. Using the wave model, Huygens was able to explain the laws of reflection and refraction. However, the authority and eminence of Newton was so great that no one reposed faith in Huygens' proposition. In fact, wave model was revived and shaped by Young through his interference experiments.

Young showed that the wavelength of visible light lies in the range 4000 Å to 7000 Å (Typical values of wavelength for sound range from 15 cm for a high-pitched whistle to 3 m for a deep male voice.) This explains why the wave character of light goes unnoticed (on a human scale). Interference fringes can be seen only when the spacing between two light sources is of the order of the wavelength of light. That is also why diffraction effects are small and light is said to approximately travel in straight lines. (A ray is defined as the path of energy propagation in the limit of $\lambda \to 0$). A satisfactory explanation of diffraction of light was given by Fresnel on the basis of the wave model. An important part in establishing wave model was played by polarisation – a subtle property of light. It established that light is a transverse wave; the oscillations are perpendicular to the path of propagation. But what is it that oscillates? The answer was provided by Maxwell who provided real physical significance and sound pedestal to the wave theory. Maxwell identified light with electromagnetic waves. A light wave is associated with changing electric and magnetic fields. You will learn these details now.

3.3 Light as an Electromagnetic Wave

A varying electric field gives rise to a time and space varying magnetic field and vice-versa. This interplay of coupled electric and magnetic fields results in the propagation of three-dimensional electromagnetic waves. To show this, we first recall Maxwell's field equations:

$$\nabla \cdot \mathbf{D} = \rho \tag{1.1a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.1b}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \tag{1.1c}$$

and

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{1.1d}$$

where ρ and **J** denote the free charge density and the conduction current density, respectively. **E**, **D**, **B**, and **H** respectively represent the electric field, electric displacement, magnetic induction and the magnetic field. These are connected through the following constitutive relations:

$$\mathbf{D} = \varepsilon\, \mathbf{E} \tag{1.2a}$$

$$\mathbf{B} = \mu\, \mathbf{H} \tag{1.2b}$$

and

$$\mathbf{J} = \sigma\, \mathbf{E} \tag{1.2c}$$

where ε, μ and σ respectively denote the (dielectric) permittivity, magnetic permeability and the electrical conductivity of the medium.

For simplicity, we consider the field equations in vacuum so that $\rho = 0$ and $\mathbf{J} = 0$. Then, if we use connecting relations [Eqs. (1.2a-c)], Eq. (l.la-d) reduce to

$$\nabla \cdot \mathbf{E} = 0 \tag{1.3a}$$
$$\nabla \cdot \mathbf{H} = 0 \tag{1.3b}$$
$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \tag{1.3c}$$

and $$\nabla \times \mathbf{H} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{1.3d}$$

> The 3-D wave equation has the form
>
> $$\nabla^2 \psi = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2}$$
>
> where ψ is a physical quantity which propagates wavelike with speed v.

where μ_0 and ε_0 are the magnetic permeability and permittivity of free space. Taking the curl of Eq. (1.3c), we get

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E} &= -\mu_0 \nabla \times (\partial \mathbf{H}/\partial t) \\ &= -\mu_0 \frac{\partial}{\partial t}(\nabla \times \vec{H}) \end{aligned} \tag{1.4}$$

since $\frac{\partial}{\partial t}$ is independent of $\nabla \times$ operation.

To simplify the left hand side of this equation, we use the vector identity

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$$

Since $\nabla \cdot \mathbf{E} = 0$ in view of Eq. (1. 3a), we find that Eq. (1.4) reduces to

$$-\nabla^2 \mathbf{E} = -\mu_0 \frac{\partial}{\partial t}(\nabla \times \mathbf{H})$$

On substituting the value of $\nabla \times \mathbf{H}$ from Eq. (1.3d), we get

$$\nabla^2 \mathbf{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{1.5}$$

You can similarly show that

$$\nabla^2 \mathbf{H} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{1.6}$$

SELF ASSESSMENT EXERCISE 2

Prove Eq. (1.6).

Do you recognise Eqs.(1.5) and (1.6)? These are identical in form to the 3-D wave equation derived in the course Oscillations and Waves. This means that each component of **E** and **H** satisfies a wavelike equation. The speed of propagation of an electromagnetic wave in free space is given by

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{1.7}$$

This remarkably simple result shows that the speed of an electromagnetic wave depends only on μ_0 and ε_0. This suggests that all e.m. waves should, irrespective of frequency or amplitude, share this speed while propagating in free space. We can easily calculate the magnitude of v by noting that for free space

$$\varepsilon_0 = 8.8542\times 10^{-12} C^{-12} N^{-1} m^{-2}$$

and

$$\mu_0 = 4\pi \times 10^{-7} Ns^2 C^{-2}$$

Thus,

$$v = \frac{1}{[(8.8542\times 10^{-12} C^2 N^{-1} m^{-2})\times(4\times 10^{-7} Ns^2 C^{-2})]^{1/2}}$$

$$= 2.99794\times 10^8 ms^{-1}$$

This is precisely the speed of light! It is worthwhile to mention here that using the then best known value of ε_0, Maxwell found that electromagnetic waves should travel at a speed of 3.1074×10^8 ms'1. This, to his amusement, was very close to the speed of light measured by Fizeau (3.14858×10^8 ms^{-1}). Based on these numbers, Maxwell proposed the electromagnetic theory of light. In his own words,

> "This velocity is so nearly that of light, that it seems we have strong reason to believe that light itself is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws."

We cannot help but wonder at such pure gold having come out of his research on electric and magnetic phenomena. It was a rare moment of unveiled exuberance - a classic example of the unification of knowledge towards which science is ever striving. With this one calculation, Maxwell brought the entire science of optics under the umbrella of electromagnetism. Its significance is profound because it identifies light with structures consisting of electric and magnetic fields travelling freely through free space.

The direct experimental evidence for electromagnetic waves came through a series of brilliant experiments by Hertz. He found that he could detect the effect of electromagnetic induction at considerable distances from his apparatus. His apparatus is shown in Fig. 1.1. By measuring the wavelength and frequency of electromagnetic waves, Hertz calculated their speed. He found it to be precisely equal to the speed of light. He also demonstrated properties like reflection, refraction, interference, etc., and demonstrated conclusively that light is an electromagnetic wave.

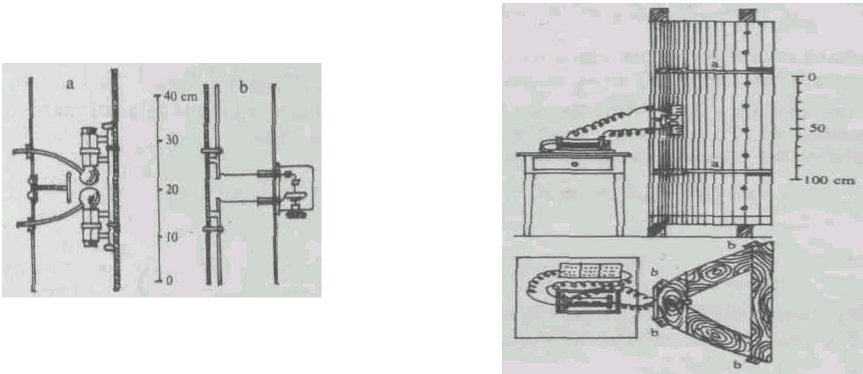


Fig.1.1 Hertz's apparatus for the generation and detection of electromagnetic waves

You now know that electromagnetic waves are generated by time varying electric and magnetic fields. So these are described by the amplitudes and phases of these fields. The simplest electromagnetic wave is the plane wave. You may recall that in a plane wave the phases of all points on a plane normal to the direction of propagation are same. And for a plane electromagnetic wave propagating along the $+z$-direction, the phase is $(kz - \omega t)$, where k is the wave number and ω is the angular frequency of electromagnetic plane wave. And the scalar electric and magnetic fields can be expressed as

$$\mathbf{E} = E_0 \exp[i(kz - \omega t)]$$
$$\mathbf{H} = H_0 \exp[i(kz - \omega t)]$$

where E_0 and H_0 are the amplitudes of **E** and **H**.

For a wave propagating along the $+z$-direction, the field vectors **E** and **H** are independent of x and y. Then, Eqs. (1.3a) and (1.3b) reduce to

$$\frac{\partial E_z}{\partial z} = 0 \tag{1.8a}$$

and

$$\frac{\partial H_z}{\partial z} = 0 \tag{1.8b}$$

By the same argument you will find that the time variation of E_z and H_z can be expressed as

$$\frac{\partial E_z}{\partial t} = 0 \tag{1.9a}$$

and

$$\frac{\partial H_z}{\partial z} = 0 \tag{1.9b}$$

> To arrive at Eqs. (1.9a,b), we write the z - component of Eqs. (1.3c) and (1.3d) as
>
> $$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\mu_0 \frac{\partial H_z}{\partial t}$$
>
> and $$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = -\varepsilon_0 \frac{\partial E_z}{\partial t}$$
>
> since **E** and **H** are independent of x and y, the LHS will be identically equal to zero.

What do these equations convey? Physically, these imply that the components of **E** and **H** along the direction of propagation of an electromagnetic wave ($+z$-direction in this case) does not depend upon time and the space coordinate z. So we must have

$$E_z = 0 = H_z \tag{1.10}$$

You should convince yourself why any other constant value of E_z and H_y would not represent a wave. We can now draw the following conclusions:

Plane electromagnetic waves have no longitudinal component. That is, they are transverse. This implies that if electric field is along the x- axis, the magnetic field will be along the y-axis so that we may write

$$\mathbf{E} = \hat{\mathbf{x}} E_0 e^{i(kz-\omega t)}$$

and $$\mathbf{H} = \hat{\mathbf{y}} H_0 e^{i(kz-\omega t)} \tag{1.11}$$

You may now ask: Are E_0 and H_0 connected? If so, what is the relation between them? To discover answer to this question you have to solve TQ2:

$$H_0 = \frac{k}{\mu_0 \omega} E_0$$

2. Since $\frac{k}{\mu_0 \omega}$ is a real number, the electric and magnetic vectors should be in phase. Thus if **E** becomes zero (maximum) at some instant, **H** must also necessarily be zero (maximum) and so on. This also shows that neither electric nor magnetic wave can exist without the other. An electric field varying in time sets up a space-time varying magnetic field, which, in turn, produces an electric field varying in space and time, and so on. You cannot separate them. This mutually supporting role results in the generation of electromagnetic waves. The pictorial representation of fields of a plane electromagnetic waves (propagating along the + z- direction) is shown in Fig. 1.2. You will note that electric and magnetic fields are oriented at right angles to one another and to the direction of wave motion. Moreover, the variation in the spacing of the field lines and their reversal from one region of densely spaced lines to another reflect the spatial sinusoidal dependence of the wave fields.

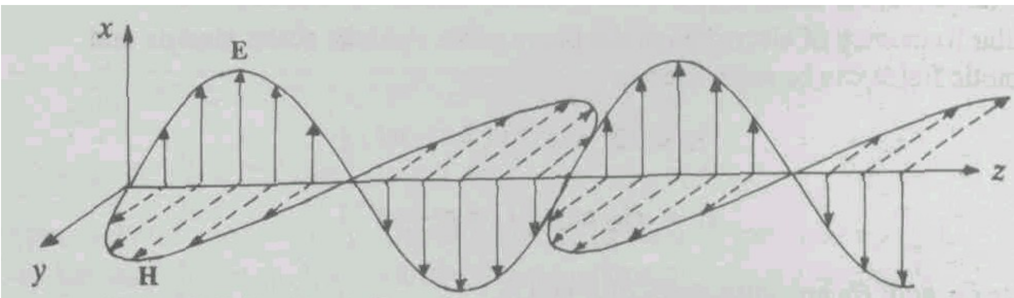


Fig.1.2 The electric and magnetic fields associated with a plane electromagnetic wave

3.3.1 Energy Transfer: The Poynting Vector

A general characteristic of wave motion is: *Wave carries energy, not matter*. Is it true even for electromagnetic waves? To know the answer, you should again consider the two field vectors (**E** and **H**) and calculate the divergence of their cross product. You can express it as

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{H} \cdot (\nabla \times \mathbf{E}) - \mathbf{E} \cdot (\nabla \times \mathbf{H}) \qquad (1.12)$$

Since $\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$

If you now substitute for the cross products on the right-hand side from Maxwell's third and fourth equations respectively for free space, you will get

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\mathbf{H} \cdot \mu_0 \frac{\partial \mathbf{H}}{\partial t} - \mathbf{E} \cdot \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

The time derivatives on the right-hand side can be written as

$$\mathbf{H} \cdot \mu_0 \frac{\partial \mathbf{H}}{\partial t} = \mathbf{H} \cdot \mu_0 \frac{\partial \mathbf{H}}{\partial t} = \frac{1}{2} \mu_0 \frac{\partial}{\partial t} (\mathbf{H} \cdot \mathbf{H})$$

and

$$\epsilon_o E \frac{\partial E}{\partial t} = \frac{1}{2} \epsilon_o \frac{\partial}{\partial t} (E.E)$$

so that $\nabla.(\vec{E} \, x \, \vec{H}) = -\frac{\partial}{\partial t} \frac{1}{2} (\epsilon_o \vec{E}.\vec{E} + \mu_o \vec{H}.\vec{H}) \qquad (1.13)$

Do you recognise Eq. (1.13)? If so, can you identify it with some known equation in physics? This equation resembles the equation of continuity in hydrostatics. To discover the physical significance of Eq. (1.13), you should integrate it over volume bound by the surface S and use Gauss' theorem. This yields

$$\int_V \nabla \cdot (\mathbf{E} \times \mathbf{H}) dV = -\frac{\partial}{\partial t} \int_V \frac{1}{2} (\varepsilon_0 \mathbf{E} \cdot \mathbf{E} + \mu_0 \mathbf{H} \cdot \mathbf{H}) dV$$

or

$$\int_S (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{A} = -\frac{\partial}{\partial t} \int_V \frac{1}{2} (\varepsilon_0 \mathbf{E} \cdot \mathbf{E} + \mu_0 \mathbf{H} \cdot \mathbf{H}) dV$$

Gauss' divergence theorem relates the surface integral of a vector function to the volume integral of the divergence of this same function:

$$\int_S \mathbf{D} \cdot d\mathbf{A} = \int_V \nabla \cdot \mathbf{D} dV$$

The surface integral is taken over the closed surface, *S* bounding the volume, *V*.

The integrand on the right hand side refers to the time rate of flow of electromagnetic energy in free space. You will note that both **E** and **H** contribute to it equally. The vector

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \qquad (1.14)$$

is called the **Poynting Vector.** It is obvious that **S**, **E** and **H** are mutually orthogonal. Physically it implies that **S** points in the direction of propagation of the wave since electromagnetic waves are transverse. This is illustrated in Fig. 1.3.

You may now like to know the time-average of energy carried by electromagnetic waves (light) per unit area. If you substitute for **E** and **H** in Eq. (1.14) and average over time, you will obtain

$$< \mathbf{S} > = \hat{\mathbf{z}} \frac{k}{2\mu_0 \omega} E_0^2 \qquad (1.15)$$

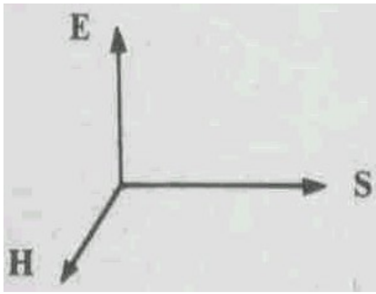


Fig. 1.3 The Poynting Vector

Before you proceed, you should convince yourself about the validity of this result. To ensure this we wish you to solve SAE 3.

SELF ASSESSMENT EXERCISE 3

Prove Eq. (1.15).

3.3.2 The Electromagnetic Spectrum

Soon after Hertz demonstrated the existence of electromagnetic waves in 1888, intense interest and activity got generated. In 1895, J.C. Bose, working at Calcutta, India produced electromagnetic waves of wavelengths in the range 25 mm to 5 m. (In 1901, Marconi succeeded in transmitting electromagnetic waves across the Atlantic Ocean. This created public sensation. In fact, this pioneering work marked the beginning of the era of communication using electromagnetic waves.) X-rays, discovered in 1898 by Roentgen, were shown in 1906 to be e.m.

waves of wavelength much smaller than the wavelength of light waves. Our knowledge of e.m. waves of various wavelengths has grown continuously since then. The e.m. spectrum, as we know it today, is shown in Fig. 1.4.

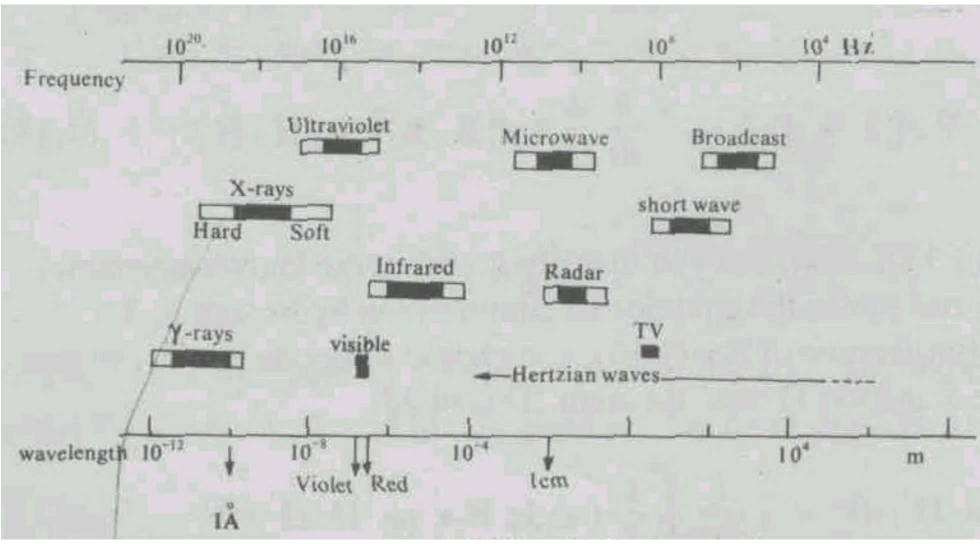


Fig. 1.4 The electromagnetic spectrum

The range of wavelengths (and their applications in modern technologies) is very wide. However, the boundaries of various regions are not sharply defined. The visible light is confined to a very limited portion of the spectrum from about 4000 Å to 7000 Å. As you know, different wavelengths correspond to different colours. The red is at the long wavelength-end of visible region and the violet at the short wavelength-end. For centuries our only information about the universe beyond earth has come from visible light. All electromagnetic waves from 1 m to 10^{6} m are referred to as radiowaves. These are used in the transmission of radio and television signals. The ordinary AM radio corresponds to waves with $\lambda = 100$m, whereas FM radio corresponds to 1m. The microwaves are used for radar and satellite communications ($\lambda \sim 0.5$m - 10^{-3} m).

Between radio waves and visible light lies the infrared region. Beyond the visible region we encounter the ultraviolet rays, X-rays and gamma rays. You must convince yourself that all phenomena from radio waves to gamma rays are essentially the same; they are all electromagnetic waves which differ only in wavelength (or frequency). You may now be tempted to enquire: Why do we attribute different nomenclature to different portions of the electromagnetic spectrum? The distinction is a mere convenience while identifying their practical applications.

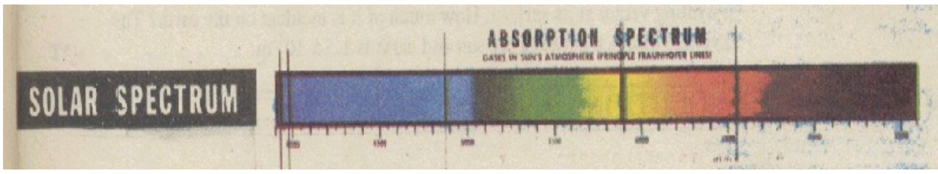


Fig. 1.5 The solar spectrum received on the earth

In our solar system, the sun is the major source of e.m. waves. If you closely examine the solar spectrum received on the earth, you will observe broad continuous spectrum crossed by Fraunhofer dark absorption lines (Fig. 1.5).
Let us now sum up what you have learnt in this unit.

4.0 CONCLUSION

In this unit you have learnt that light is a transverse electromagnetic wave. According to corpuscular model, light consists of minute invisible stream of particles called corpuscles. The wave equations for $\vec{E}$ and $\vec{B}$ are derived from Maxwell's field equations. An electric field varying in time sets up a spaced-time varying magnetic field, which in turn, produces an electric field varying in space and time, and so on.

An expression for pointing vector $\vec{S} = \vec{E} \times \vec{H}$ is obtained. $\vec{S}$ defines the direction of propagation of an electromagnetic wave. At the end of the unit, the electromagnetic spectrum of the waves is discussed. The visible light is confined to a very limited portion of the spectrum from about 4000 $\mathring{A}$ to 7000 $\mathring{A}$.

5.0 SUMMARY

- Light is an electromagnetic wave.
- The electric and magnetic fields constituting an electromagnetic wave satisfy the
  equations

$$\nabla^2 \mathbf{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}$$

and

$$\nabla^2 \mathbf{H} = \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{H}}{\partial t^2}$$

- For a plane electromagnetic wave propagating along the *z*- direction, the electric and magnetic fields can be expressed as

$$\mathbf{E} = \hat{\mathbf{x}} E_0 e^{i(kz-\omega t)}$$

and

$$\mathbf{H} = \hat{\mathbf{y}} H_0 e^{i(kz-\omega t)}$$

- The electromagnetic waves are transverse.
- The poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ defines the direction of propagation of an electromagnetic wave.

- The visible light is confined to a very limited portion (4000 Å - 7000 Å) of the electromagnetic spectrum.

6.0 TUTOR MARKED ASSIGNMENT

1. Derive the wave equation for the propagation of electromagnetic waves in a conducting medium.
2. Starting from Eqs. (1.3c) and (1.3d) show that

$$H_0 = \frac{k}{\mu_0 \omega} E_0$$

3. The energy radiated by the sun per second is approximately 4.0×10^{26} J s^{-1}. Assuming the sun to be a sphere of radius 7×10^8 m, calculate the value of the Poynting vector at its surface. How much of it is incident on the earth? The average distance between the sun and earth is 1.5×10^{11} m.

SOLUTIONS TO SELF ASSESSMENT EXERCISES

i) According to the corpuscular model, light travels in straight lines. As a result, the size of the shadow should be equal to the size of the object. Grimaldi's observation - the size of the shadow is smaller than the size of the obstacle – indicates that light bends around edges, contradicting corpuscular model.

ii) Taking the curl of Eq. (1.3d), we get

$$\begin{aligned} \nabla \times \nabla \times \mathbf{H} &= \varepsilon_0 \nabla \times \frac{\partial \mathbf{E}}{\partial t} \\ &= \varepsilon_0 \frac{\partial}{\partial t} (\nabla \times \mathbf{E}) \end{aligned}$$

Using the vector identity

curl curl $\mathbf{H}$ = grad div $\mathbf{H} - \nabla^2 \mathbf{H}$

we have

$$\nabla(\nabla\cdot\mathbf{H})-\nabla^2\mathbf{H}=\mu_0\varepsilon_0\frac{\partial}{\partial t}\left(-\frac{\partial\mathbf{H}}{\partial t}\right)$$

Since $\nabla\cdot\mathbf{H}=0$, we get

$$\nabla^2\mathbf{H}=\mu_0\varepsilon_0\frac{\partial\mathbf{H}}{\partial t^2}$$

iii) From Eq. (1.14), we have for Poynting vector

$$\mathbf{S}=\mathbf{E}\times\mathbf{H}$$

Taking only the real part of Eq. (1.11), the electric and magnetic field vectors can be represented as

$$\begin{aligned}
\mathbf{E} &= \hat{\mathbf{x}}E_0\cos(kz-\omega t)\\
\mathbf{H} &= \hat{\mathbf{y}}H_0\cos(kz-\omega t)\\
&= \hat{\mathbf{y}}\frac{k}{\mu_0\omega}E_0\cos(kz-\omega t) \qquad \left(\because\ H_0=\frac{k}{\mu_0\omega}E_0\right)
\end{aligned}$$

So,

$$\mathbf{E}\times\mathbf{H}=(\hat{\mathbf{x}}\times\hat{\mathbf{y}})\frac{k}{\mu_0\omega}E_0\cos^2(kz-\omega t)$$

or

$$\mathrm{S}=\hat{\mathrm{z}}\frac{\mathrm{k}}{\mu_0\omega}\mathrm{E}_0^2\cos^2(\mathrm{kz}-\omega\,\mathrm{t})$$

This gives the amount of energy crossing a unit area perpendicular to the z-axis per unit time. Typical frequency for an optical beam is of the order of $10^{15}\ s^{-1}$ and the cosine term will fluctuate rapidly. Therefore, any measuring device placed in the path would record only an average value. The time average of the cosine term, as you know, is 1/2. Hence

$$<\mathbf{S}>=\hat{\mathbf{z}}\frac{k}{2\mu_0\omega}\mathrm{E}_0^2$$

7.0 REFERENCES/FURTHER READINGS

Fundamental to Optics- Jenkins and White

Introduction to Modern Optics, Grant F. Fowles.

Optics; Hecht and Zajac

Optics, Smith and Thomson

UNIT 2 REFLECTION AND REFRACTION OF LIGHT

CONTENTS

1.0 INTRODUCTION

In the previous unit you have learnt that light is an electromagnetic wave. It is made up of mutually supporting electric and magnetic fields, which vary continuously in space and time. An interesting question related to e.m. waves is: What happens to these fields when such a wave is incident on the boundary separating two optically different media? You may recall from previous courses that when a wave passes from air to water or air to glass, we get a reflected wave and a refracted wave. Reflection of light from a silvered surface, a looking mirror say, is the most common optical effect. Reflection of e.m. waves governs the working of a radar. Reflection of radiowaves by the ionosphere makes signal transmission possible and is so crucial in the area of communication.

In your earlier school years, you have learnt that refraction explains the working of lenses and is responsible for seeing: our contact with surroundings. Even the grand spectacle of sunset or a rainbow can be explained in terms of refraction of light. Refraction of e. m. waves forms the basis of one of the greatest technological applications in signal transmission. In fact, electro-optics has seen tremendous growth via optical fibres for a variety of applications.

In Unit 7 of the course Oscillations and Waves, you learnt to explain reflection and refraction of waves on the basis of Huygen’s wave model. Now the question arises: Can we extend this analysis to electromagnetic waves, which include visible light, radiowaves, microwaves and X-rays? In Sec. 3.1 you will learn to derive the equations for reflected and transmitted fields (**E** and **B)** when an e.m. wave is incident normally as well as obliquely on the boundary of two media.

You are aware that many physical systems behave according to optimisation principle. When several fluids at different temperatures are mixed, the heat exchange takes place so that the total entropy of the system is maximum. A ball rolling on an undulating surface comes to rest at the lowest point. The profoundness of such situations and scientific laws governing them led Fermat to speculate: Does light also obey some optimization principle? And he concluded: **A ray of light chooses a path of extremum between two points.** This is known as Fermat's principle. Implicit in it are the assumptions:

(i) Light travels at a finite speed, and

(ii) The speed of light is lower in a denser medium.

In Sec. 3.3 you will learn about Fermat's principle. We have shown that all laws of geometrical optics are contained in it.

2.0 OBJECTIVES

After studying this unit you should be able to:

- explain reflection and refraction of e.m. waves incident normally and obliquely on the interface separating two optically different media
- apply Fermat's principle to explain the reflection and refraction of light, and
- solve problems based on reflection and refraction of e.m. waves.

3.0 MAIN CONTENT

3.1 Electromagnetic Waves at the Interface Separating Two Media

Consider a plane electromagnetic wave that is incident on a boundary between two linear media. That is, **D** and **H** are proportional to **E** and **B**, respectively, and the constants of proportionality are independent of position and direction. You can visualise it as light passing from air (medium 1) to glass (medium 2). Let us assume that there are no free charges or currents in the materials.

Fig. 2.1 shows a plane boundary between two media having different permittivities and permeabilities: ε_1, μ_1 for medium 1 and ε_2, μ_2 for medium 2. A uniform plane wave travelling to the right in medium 1 is incident on the interface **normal to the boundary.** As in the case of waves on a string, we expect a reflected wave propagating back into the medium and a transmitted (or refracted) wave travelling in the second medium. We wish (i) to derive expressions for the fields associated with reflected and refracted waves in terms of the field associated with the incident wave and (ii) know the fraction of the incident energy that is reflected and transmitted. To do so we need to know the **boundary conditions** satisfied by these waves at the interface separating the two media. We obtain these conditions by stipulating that Maxwell's equations must be satisfied at the boundary between these media. We first state the appropriate conditions. Their proof is given in the appendix to this Unit.

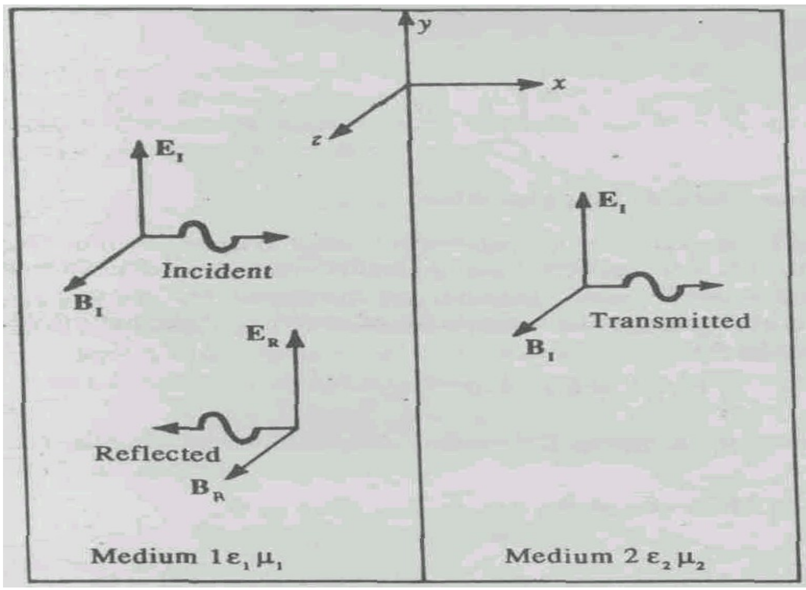


Fig.2.1 A uniform plane wave is incident normally on a plane boundary. The reflected and refracted (transmitted) waves are also shown. The angle of incidence is α and the angle of refraction is β

Boundary Conditions

The integral form of Maxwell's equations for a medium free of charges and currents is:

$$\varepsilon \int_S \mathbf{E} \cdot d\mathbf{S} = 0 \qquad (2.1a)$$

$$\int_S \mathbf{B} \cdot d\mathbf{S} = 0 \qquad (2.1b)$$

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = \frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} \qquad (2.1c)$$

and

$$\frac{1}{\mu} \int_C \mathbf{B} \cdot d\mathbf{l} = \varepsilon \frac{d}{dt} \int_S \mathbf{E} \cdot d\mathbf{S} \qquad (2.1d)$$

where S is a surface bound by the closed loop *C*.

The electric field can oscillate either parallel or normal to the plane of incidence. The magnetic field **B** will then be normal or parallel to the plane of incidence. We will denote these with subscripts || (parallel) and ⊥ (normal). The boundary conditions for normal and parallel components of electric and magnetic fields take the form (Appendix A).

$$\varepsilon_1 E_{1\perp} - \varepsilon_2 E_{2\perp} = 0 \tag{2.2a}$$

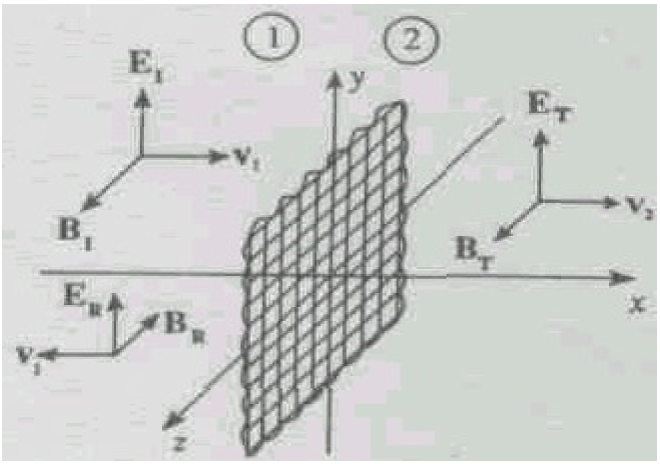
$$B_{1\perp} - B_{2\perp} = 0 \tag{2.2b}$$

$$E_{1\parallel} - E_{2\parallel} = 0 \tag{2.2c}$$

and

$$\frac{1}{\mu_1} B_{1\parallel} - \frac{1}{\mu_2} B_{2\parallel} = 0 \tag{2.2d}$$

We shall now use the boundary conditions expressed by Eqs. (2.2a- d) to study reflection and refraction (transmission) at normal as well as oblique incidence.



Flg.2.2 A sinusoidal plane e.m. wave incident normally at the boundary of two optically transparent media

3.1.1 Normal Incidence

Refer to Fig. 2.2. The yz-plane $(x = 0)$ forms the interface of two optically transparent (non-absorbing) media (refractive indices n_1 and n_2). A sinusoidal plane wave of frequency ω travelling in the x-direction is incident from the left. From Unit 7 of the Oscillations and Waves course you will recall that progressive waves are partially reflected and partially refracted at the boundary separating two physically different media. However, the energy of the reflected or transmitted e.m. waves depends upon their refractive indices.

The appropriate magnetic fields to be associated with electric fields are obtained from the equation

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

Let us suppose that the electric field is along the y-direction. Then the electric and magnetic fields associated with the incident wave are given by

$$\mathbf{E_I}(x,t) = E_{0I}\hat{\mathbf{j}}\exp[i(k_I x - \omega\, t) \qquad (2.3a)$$

and

$$\mathbf{B_I}(x,t) = \frac{E_{0I}}{v_1}\hat{\mathbf{k}}\exp[i(k_I x - \omega\, t) \qquad (2.3b)$$

The reflected wave propagates back into the first medium and can be represented by the following fields:

$$\mathbf{E_R}(x,t) = E_{0R}\hat{\mathbf{j}}\exp[-i(k_I x + \omega\, t) \qquad (2.4a)$$

and

$$\mathrm{B_R(x,t) = -\frac{E_{0R}}{v_1}\hat{k}\exp[-i(k_I x + \omega\, t)} \qquad (2.4b)$$

The minus sign in the exponents in Eqs. (2.4a,b) indicates that propagation of the wave is in the *-x* direction. But the negative sign with the amplitude in Eq. (2.4b) arises because of the transverse nature of e.m. waves and that the electric and magnetic field vectors should obey the relation

$$\mathbf{B_R} = \frac{1}{v_1}(\hat{\mathbf{k}}_\mathbf{I} \times \mathbf{E_R})$$

where $\hat{\mathbf{k}}_\mathbf{I}$ is unit vector along the direction of incidence.

If you visualise Eqs. (2.3) and (2.4) diagramatically, you will note that the electric vectors have been kept fixed in the same direction but the magnetic field vectors have been oriented. The orientation of the magnetic field vector ensures that the flow of energy is always along the direction of propagation of the wave (Poynting theorem).

The electric and magnetic fields of the transmitted wave, which travels to the right in medium 2, are given by

$$\mathbf{E_T}(x,t) = E_{0T}\hat{\mathbf{j}}\exp[i(\omega_T t + k_T x)] \tag{2.5a}$$

and

$$\mathbf{B_T}(x,t) = \frac{1}{v_2}[\hat{\mathbf{k}}_\mathbf{T} \times \mathbf{E_T}(x,t)] \tag{2.5b}$$

The phenomenon of reflection and refraction is usually analysed in two parts:

(i) To determine the relations between the field vectors of the reflected and refracted waves in terms of that of the incident wave. These relations determine the reflection and the transmission coefficients. In this derivation, we match the **E** and **B** fields in the two media at the interface with the help of appropriate boundary conditions there.

(ii) To establish relations between the angle of incidence and the angles of reflection and refraction we may emphasize that so far as the laws of reflection and refraction are concerned, explicit use of any boundary condition is not required.

Fresnel's Amplitude Relations

To derive expressions for the amplitudes of the reflected and the refracted waves in terms of the amplitude of the incident wave, we apply boundary conditions given by Eq. (2.2a-d) at **every point** on the interface at **all times.** At $x = 0$, the combined field to the left ($\mathbf{E}_I + E_R$ and $\mathbf{B}_I$ and $\mathbf{B}_R$) must join the fields to the right ($\mathrm{E_T}$ and $\mathrm{B_T}$). For normal incidence, there are no normal field components (perpendicular to the interface). But why? This is because neither **E** nor **B** field is in the x-direction. This means that Eqs. (2.2a,b) are trivial and only tangential components of the electric and magnetic fields should be matched at the plane $x = 0$. Thus

$$E_{0I} + E_{0R} = E_{0T} \tag{2.6a}$$

and

$$\frac{1}{\mu_1}\left(\frac{E_{oI}}{\upsilon_1} - \frac{E_{oR}}{\upsilon_1}\right) = \frac{1}{\mu_2}B_{0T}$$

or

$$\frac{1}{\mu_1}(B_{0I} + B_{0R}) = \frac{1}{\mu_2}\frac{E_{0I}}{v_2}$$

which, on simplification yields

$$E_{0I} - E_{0R} = \alpha\, E_{0T} \tag{2.6b}$$

where

$$\alpha = \frac{\mu_1 v_1}{\mu_2 v_2} = \sqrt{\frac{\mu_1 \varepsilon_1}{\mu_2 \varepsilon_2}} = \frac{\mu_1 n_2}{\mu_2 n_1} \tag{2.6c}$$

Solving Eqs. (2.6a) and (2.6b) for the reflected and transmitted electric field amplitudes in terms of the incident amplitude, you will find that

$$E_{0R} = \left(\frac{1-\alpha}{1+\alpha}\right) E_{0I} \tag{2.7a}$$

and

$$E_{0T} = \frac{2}{1+\alpha} E_{0I} \tag{2.7b}$$

For most optical media, the permeabilities are close to their values in vacuum $(\mu_1 \approx \mu_2 \approx \mu_o)$. In such cases $\alpha = \frac{v_1}{v_2}$ and we have

$$E_{0R} = \left(\frac{v_2 - v_1}{v_2 + v_1}\right) E_{0I}$$

and $$E_{oI} = \frac{2\upsilon_2}{\upsilon_2 + \upsilon_1} E_{o2} \tag{2.8}$$

This suggests that when $v_2 > v_1$, the reflected wave will be in phase with the incident wave and for $v_2 < v_1$, the reflected and incident waves will be out of phase. This is illustrated in Fig. 2.3.

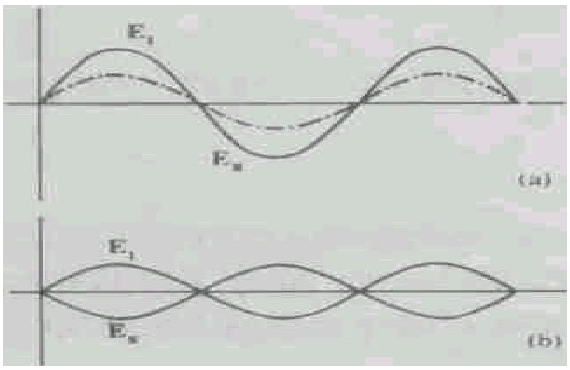


Fig.2.3 The phase relationship between reflected wave and the incident wave

In terms of the index of refraction $n\left(=\frac{c}{v}\right)$, we can rewrite Eq. (2.8) as

$$E_{0R} = \frac{n_1 - n_2}{n_1 + n_2} E_{0I}$$

and

$$E_{0T} = \frac{2n_1}{n_1 + n_2} E_{0I} \qquad (2.9)$$

When an e.m. wave passes from a rarer medium to a denser medium $(n_1 < n_2)$, the ratio $\frac{E_{0R}}{E_{0I}}$ will be negative. Physically, it means that the reflected wave is 180° out of phase with the incident wave. You have already learnt it in the case of reflection of sound waves in the course on Oscillations and Waves. When an e.m. wave is incident from a denser medium on the interface separating it from a rarer medium, ($n_1 > n_2$) the ratio $\frac{E_{0R}}{E_{0I}}$ is positive and no such phase change occurs.

We can now easily calculate the **reflection** and the **transmission coefficients,** which respectively measure the fraction of incident energy that is reflected and transmitted. The first step in this calculation is to recall that

$$R = \frac{I_R}{I_I}$$

and

$$T = \frac{I_T}{I_I}$$

where I_R, I_T and I_I respectively denote the reflected, transmitted and incident wave intensity. Intensity is defined as the average power per unit area, $(1/2)vE^2$. So you can readily show that

$$R = \frac{I_R}{I_I} = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \qquad (2.10a)$$

and

$$R = \frac{I_T}{I_I} = \frac{n_2}{n_1}\left(\frac{2n_1}{n_1 + n_2}\right)^2 \qquad (2.10b)$$

You can convince yourself that $R + T = 1$. For air $(n_1 = 1)$ - glass (n_2 = 1.5) interface, the R and T coefficients have the values R = 0.04 and T = 0.96. There is no energy stored (or absorbed) at the interface and you can now realise why most of the light is transmitted.

We will now repeat this exercise for the case of oblique incidence.

3.1.2 Oblique Incidence

Refer to Fig. 2.4. A plane electromagnetic wave is incident at an angle θ_I. Let the angles of reflection and refraction be θ_R and θ_T. We can represent the fields associated with these three plane electromagnetic waves as

Incident Wave

$$\mathbf{E_I} = E_{0I}\exp[-i(\omega_I t - \mathbf{k}_I \cdot \mathbf{r})]$$
$$\mathbf{B_I} = \frac{1}{v_1}(\hat{\mathbf{k}}_\mathbf{I} \times \mathbf{E_I}) \tag{2.11a}$$

Reflected Wave

$$\mathbf{E_R} = E_{0R}\exp[-i(\omega_R t - \mathbf{k}_R \cdot \mathbf{r})]$$
$$\mathbf{B_R} = \frac{1}{v_1}(\hat{\mathbf{k}}_\mathbf{R} \times \mathbf{E_R}) \tag{2.11b}$$

Transmitted Wave

$$\mathbf{E_T} = E_{0T}\exp[-i(\omega_T t - \mathbf{k}_T \cdot \mathbf{r})]$$
$$\mathbf{B_T} = \frac{1}{v_2}(\hat{\mathbf{k}}_\mathbf{T} \times \mathbf{E_T}) \tag{2.11c}$$

You may recall that the boundary conditions must hold at every point on the interface at all times. If the boundary conditions hold at a point and at some time, they will hold at all points in space for all subsequent times only if the exponential parts in above expressions for each wave are the same, i.e.,

$$\omega_I t - \mathbf{k_I}\cdot\mathbf{r} = \omega_R t - \mathbf{k_R}\cdot\mathbf{r} = \omega_T t - \mathbf{k_T}\cdot\mathbf{r}$$

at the interface. This implies that for equality of phases at all times we must have

$$\omega_I = \omega_R = \omega_T = \omega \quad \text{(say)} \tag{2.12a}$$

That is, the frequency of an e.m. wave does not change when it undergoes reflection and refraction: all waves have the same frequency. Since the fields must satisfy Maxwell's equations, we must have for the wave vectors

$$\frac{k_I^2}{\omega^2} = \frac{1}{c^2} = \varepsilon_1 \mu_1 \qquad (2.13a)$$

$$\frac{k_T^2}{\omega^2} = \frac{1}{c^2} = \varepsilon_2 \mu_2$$
(2.13b)

$$\frac{k_R^2}{\omega^2} = \frac{1}{c^2} = \varepsilon_1 \mu_1 \qquad (2.13c)$$

Further, let k_{Ix}, k_{Iy} and k_{Iz} represent the x, y and z components of $\mathbf{k_I}$. We can use similar notation for $\mathbf{k_T}$ and $\mathbf{k_R}$. For the continuity conditions to be satisfied at all points on the interface, we must have

$$k_{Iy} = k_{Ty} = k_{Ry} \qquad (2.14a)$$

and

$$k_{Iz} = k_{Tz} = k_{Rz} \qquad (2.14b)$$

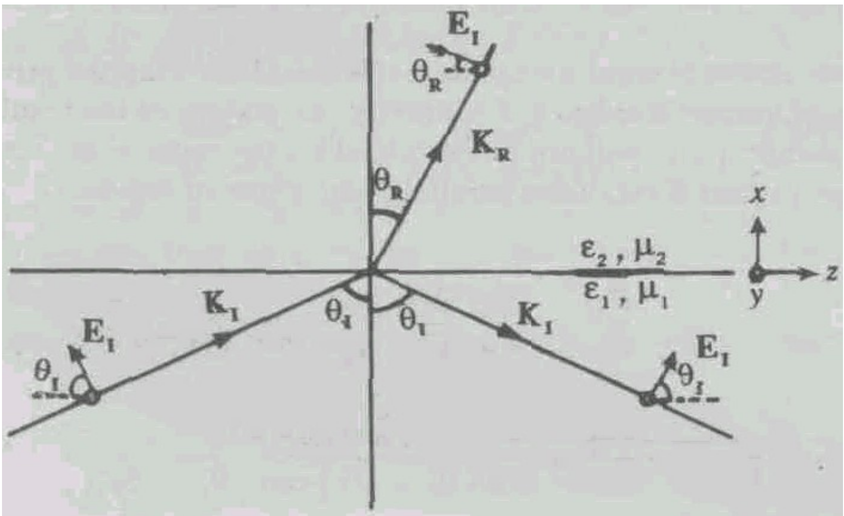


Fig. 2.4 The reflection of a plane wave with its electric vector parallel to the plane of incidence

Let us choose the y-axis such that

$$k_{Iy} = 0$$

(i.e., we assume $\mathbf{k_I}$ to lie in the x-z plane - see Fig. 2.4). Consequently

$$k_{Ty} = k_{Ry} = 0 \qquad (2.14c)$$

This result implies that the vectors $\mathbf{k_I}$, $\mathbf{k_T}$ and $\mathbf{k_R}$ will lie in the same plane. Further, from Eq. (2.14b) we get

$$k_I \sin\theta_I = k_T \sin\theta_T = k_R \sin\theta_R \tag{2.15}$$

Since $|\mathbf{k_I}| = |\mathbf{k_R}|$ (see Eq. 2.13a and c), we must have

$$\theta_I = \theta_R \tag{2.16}$$

That is, the angle of incidence is equal to the angle of reflection, which is the law of reflection. Further,

$$\frac{\sin\theta_I}{\sin\theta_T} = \frac{k_T}{k_I} = \frac{\omega\sqrt{\varepsilon_2\mu_2}}{\omega\sqrt{\varepsilon_1\mu_1}}$$

or

$$\frac{\sin\theta_I}{\sin\theta_T} = \sqrt{\frac{\varepsilon_2\mu_2}{\varepsilon_1\mu_1}} \tag{2.17}$$

If we denote the speeds of propagation of the waves in media 1 and 2 by $v_1\left(=\frac{1}{\sqrt{\varepsilon_1\mu_1}}\right)$ and $v_2\left(=\frac{1}{\sqrt{\varepsilon_2\mu_2}}\right)$, we find that Eq. (2.17) can be rewritten as

$$\frac{\sin\theta_I}{\sin\theta_T} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \tag{2.18}$$

where $n_1 = \frac{c}{v_1} = c\sqrt{\varepsilon_1\mu_1}$ and $n_2 = \frac{c}{v_2} = c\sqrt{\varepsilon_2\mu_2}$ represent the refractive indices of media 1 and 2 respectively. Do you recognise Eq. (2.18)? It is the well known Snell's law.

Eqs. (2.16) and (2.18) constitute the **laws of reflection and refraction** in optics.

You can now derive Fresnel's amplitude relations following the procedure outlined for the case of normal incidence. For brevity, we just quote the results without going into details. (You will not be examined for the same in the term-end examination.) When **E** oscillates parallel to the plane of incidence, we have

$$\frac{E_{R\parallel}}{E_{I\parallel}} = \frac{\tan(\theta_I - \theta_T)}{\tan(\theta_I + \theta_T)} \tag{2.19a}$$

$$\frac{E_{T\parallel}}{E_{I\parallel}} = \frac{2\cos\theta_I \sin\theta_T}{\sin(\theta_I + \theta_T)\cos(\theta_I - \theta_T)} \tag{2.19b}$$

When **E** oscillates normal to the plane of incidence, we have

$$\frac{E_{R\perp}}{E_{I\perp}} = -\frac{\sin(\theta_I - \theta_T)}{\sin(\theta_I + \theta_T)} \tag{2.20a}$$

$$\frac{E_{T\perp}}{E_{I\perp}} = \frac{2\sin\theta_T \cos\theta_I}{\sin(\theta_I + \theta_T)} \tag{2.20b}$$

You can easily verify that for normal incidence these equations reduce to Eq. (2.9).

The corresponding expressions for reflections and transmission coefficients for normal and parallel oscillations of E when a plane wave is incident obliquely are

$$R_{\parallel} = \frac{\tan^2(\theta_I - \theta_T)}{\tan^2(\theta_I + \theta_T)} \tag{2.21a}$$

$$T_{\parallel} = \frac{\sin 2\theta_I \sin 2\theta_T}{\sin^2(\theta_I + \theta_T)\cos^2(\theta_I - \theta_T)} \tag{2.21b}$$

$$R_{\perp} = \frac{\sin^2(\theta_I - \theta_T)}{\sin^2(\theta_I + \theta_T)} \tag{2.21c}$$

and

$$T_{\perp} = \frac{\sin 2\theta_I \sin 2\theta_T}{\sin^2(\theta_I + \theta_T)} \tag{2.21d}$$

As before, you can easily show that for normal incidence these equations reduce to Eq.(2.10a,b).

3.2 Idealization of Waves as Light Rays

So far you have learnt to explain reflection and refraction of plane electromagnetic waves at a plane interface. This signifies a relatively simple situation where the solutions of Maxwell's equations give the laws of propagation of light. It is not true in general, and we invariably seek approximations to describe a phenomenon well. One such approximation makes use of the smallness of the wavelength of light. You know that the wavelength of light is very small ($\sim 10^{-7}$ m). It is orders of magnitude less than the dimensions of optical instruments such as telescopes and microscopes. In such cases, the passage of light is most easily shown by geometrical rays. A ray is the path of propagation of energy in the zero wavelength limit ($\lambda \rightarrow 0$). The way in which rays may represent the propagation of wavefronts for some familiar situations is shown in Fig. 2.5. You will note that a plane wavefront corresponds to parallel rays and spherical wavefronts correspond to rays diverging from a point or converging to a point. You will agree that all parts of the wavefront take the same time to travel from the source.

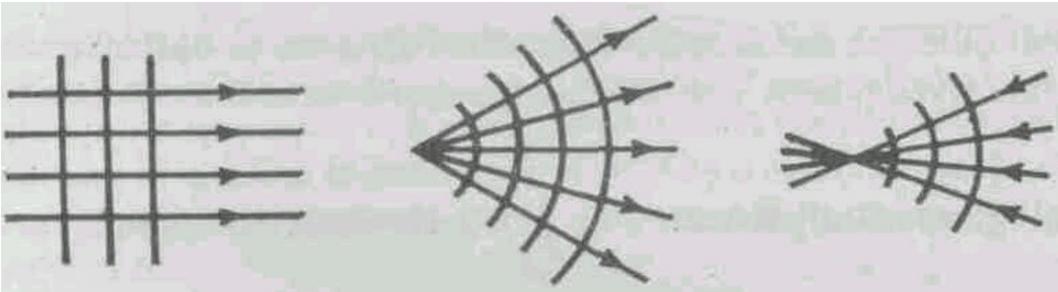


Fig.2.5: Ray representation of a plane, diverging spherical and converging spherical wavefronts moving from left to right

The laws of geometrical optics are incorporated in Fermat's principle. We will now discuss it in detail.

3.3 Fermat's Principle

In its original form, Fermat's principle may be stated as follows:

Any light ray travels between two end points along a line requiring the minimum transit time.

If v is the speed of light at a given point in a medium, the time taken to cover the distance *dl* is

$$dt = \frac{dl}{v} \tag{2.22}$$

In your earlier years you have learnt that the refractive index of a medium is defined as the ratio of the speed of light in vacuum to its speed in the medium, i.e.

$$n = \frac{c}{v}$$

Using this relation in Eq. (2.22), we get

$$dt = \frac{1}{c} n\, dl$$

> Huygens proposed that light propagates as a wavefront (a surface of constant phase) progresses in a medium perpendicular to itself with the speed of light. The zero wavelength approximation of wave optics is known as **geometrical optics.**

Hence, the time taken by light in covering the distance from point A to B is

$$\tau = \frac{1}{c}\int_A^B n\,dl$$

The quantity

$$L = \int_A^B n\,dl \tag{2.23}$$

has the dimensions of length and is called the **optical distance** or **optical path length** between two given points. You must realise that optical distance is different from the physical (geometrical) distance $\left(= \int_A^B dl\right)$. However, in a homogeneous medium, the optical distance is equal to the product of the geometrical length and the refractive index of the medium. Thus, we can write

$$\tau = \frac{L}{c}$$

This is Format's principle of least time. Let us pause for a moment and ask: Is there any exception to this law? Yes, there are cases where the optical path corresponds to maximum time or it is neither a maximum nor a minimum, i.e. stationary. To incorporate such situations, this principle is modified as follows:

Out of many paths connecting two given points, the light ray follows that path for which the time required is an extremum. In other words, the optical path length between any two points is a maximum, minimum, or stationary.

The essential point involved in Fermat's principle is that slight variation in the actual path causes a second-order variation in the actual path. Let us consider that light propagates from point A in the medium characterised by the refractive index n to the point B as shown in Fig. 2.6. According to this principle,

$$\delta\int_A^B n(x,y,z)\,dl = 0 \tag{2.24}$$

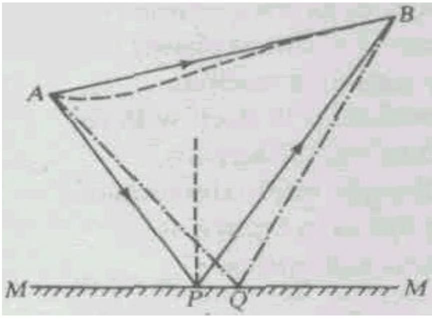


Fig. 2.6

For a homogeneous medium, the rays are straight lines, since the shortest optical path between two points is along a straight line.

In effect, Fermat's principle prohibits the consideration of an isolated ray of-light. It tells us that a path is real only when we extend our examination to the paths in **immediate neighbourhood** of the rays. To understand the meaning of this statement, let us consider the case of finding the path of a ray from a point A to a point *B* when both of them lie on the same side of a mirror *M* (Fig. 2.7). It can be seen that the ray can go directly from A to *B* without suffering any reflection.

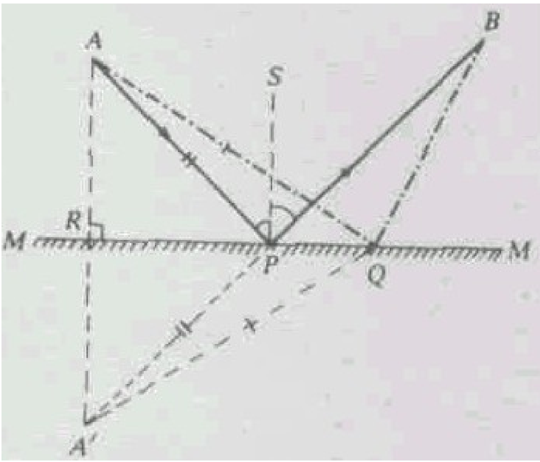


Fig. 2.7 Reflection of rays at a plane interface

Alternatively, it can go along the path *APB* after suffering a single reflection from the mirror. If Fermat's principle had asked for, say, an absolute minimum, then the *path APB* would be prohibited; but that is not the actual case. The path APB is also minimum in the neighbourhood involving paths like *AQB.* The phrase "immediate neighbourhood of path" would mean those paths that lie near the path under consideration and are similar to it. For example, the path *AQB* lies near *APB* and is similar to it; along both paths the ray suffers one reflection at the mirror. Thus Fermat's principle requires an extremum in the immediate neighbourhood of the actual path, and in general, there may be more than one ray path connecting two points.

All the laws of geometrical optics are incorporated in Fermat's principle. We now illustrate Fermat's principle by applying it to the reflection of light.

Example 1

Using Fermat's principle, derive the laws of reflection.

Solution

Let us first consider the case of reflection. Refer to Fig. 2.8. Light from a point A is reflected at a mirror *MM* towards a point *B. A* ray *APB* connects *A* and *B*. θ_I and θ_R are the angles of incidence and reflection, respectively. We have denoted the vertical distances of *A* and *B* from the mirror *MM by a* and *b.* From the construction in Fig. 2.8 and Pythagoras' theorem, we find that the total path length l of this ray from *A* to *MM* to *B* is

$$l = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2} \qquad (2.25)$$

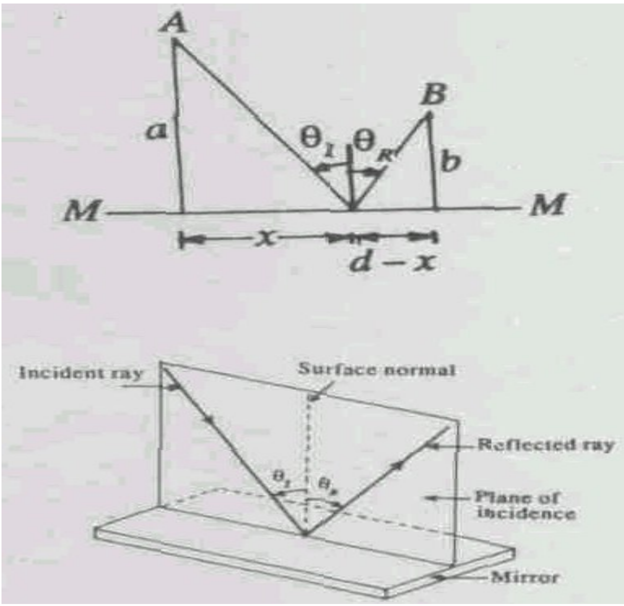


Fig. 2.8 Derivation of the laws of reflection using Fermat's principle

where *x* is the distance between the foot of the perpendicular from *A* and the point *P* at which the ray touches the mirror.

According to Fermat's principle, *P* will have a position such that the time of travel of light must be a minimum (a maximum or stationary). Expressed in another way, the total length l of the ray must be a minimum or maximum or stationary. In other words, for Fermat's principle to hold, the derivative of l with respect to *x* must be zero, i.e. $dl/dx = 0$. Hence, on differentiating Eq. (2.25) with respect to *x,* we get

$$\frac{dl}{dx} = \frac{1}{2}(a^2 + x^2)^{-1/2}(2x) + \frac{1}{2}[b^2 + (d-x)^2]^{-1/2} \times 2(d-x)(-1) = 0 \qquad (2.26)$$

which can be rewritten as

$$\frac{x}{(a^2 + x^2)^{1/2}} = \frac{d-x}{[b^2 + (d-x)^2]^{1/2}} \qquad (2.27)$$

By examining Fig. (2.8) you will note that this gives

$$\sin\theta_I = \sin\theta_R$$

or

$$\theta_I = \theta_R$$

which is (part of) the law of reflection. You will also note that the incident ray, the reflected ray and the normal to *MM* lie in the same incidence plane.

In the above example the time required or the optical path length can be seen to be minimum by calculating the second derivative and finding its value at x for which $dI/dx = 0$. The 2nd derivative turns out to be positive, showing it to be minimum. You can convince yourself by carrying out this simple calculation.

We now conclude what you have learnt in this unit.

4.0 CONCLUSION

Reflection and refraction of electromagnetic waves incident normally and obliquely on the interface separating two optically different media are explained in this unit. The appropriate boundary conditions satisfied by these waves at the interface separating the two media are stated. The reflection and transmission coefficients are calculated which respectively measure the fraction of incident energy that is reflected and transmitted. It is shown that all laws of geometric optics are contained in Fermat's principle. According to this principle, any light ray travels between two end points along a line requiring the minimum transit time.

5.0 SUMMARY

- When an e.m. wave is incident normally on the interface separating two optically different media, the reflected and transmitted electric field amplitudes are given by

$$E_{0R} = \frac{1-\alpha}{1+\alpha} E_{0I}$$

and

$$E_{0T} = \frac{2}{1+\alpha} E_{0I}$$

where $\alpha = \sqrt{\mu_1 \varepsilon_1 / \mu_2 \varepsilon_2}$ and E_{0I} is the amplitude of the incident electric field.

- The frequency of an e.m. wave is not affected when it undergoes reflection or refraction.

- Fermat's principle states that a ray of light travels between two given points along that path for which the time required is an extremum:

$$\delta \int_A^B n(x, y, z)\, dl = 0$$

6.0 TUTOR MARKED ASSIGNMENT

1. Derive Snell's law from Fermat's principle.

2. A collimated beam is incident parallel to the axis of a concave mirror. It is reflected into a converging beam. Using Fermat's principle show that the mirror is parabolic.

7.0 REFERENCES/FURTHER READINGS

Fundamental to Optics- Jenkins and White

Introduction to Modern Optics, Grant F. Fowles.

Optics; Hecht and Zajac

Optics, Smith and Thomson

APPENDIX – A

BOUNDARY CONDITIONS

Let us first consider the components of **E** and **B** fields that are normal to the boundary. We construct a thin Gaussian pillbox - extending just a little bit (hair-like) on either side of the boundary of the media, as shown in Fig. A. 1.

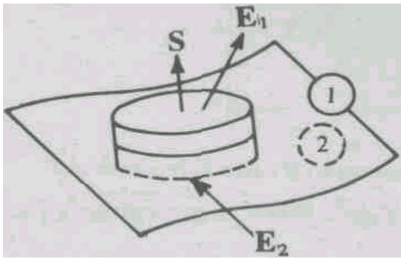


Fig A.1 The positive direction of S and E is from medium 2 towards medium 1

Eq. (2. la) implies that

$$\varepsilon\int \mathbf{E}\cdot d\mathbf{S} = \varepsilon_1\int \mathbf{E}_1\cdot d\mathbf{S} + \varepsilon_2\int \mathbf{E}_2\cdot d\mathbf{S} = 0$$

or

$$\varepsilon_1\mathbf{E}_1\cdot\mathbf{S} - \varepsilon_2\mathbf{E}_2\cdot\mathbf{S} = 0$$

In the limit thickness of the wafer goes to zero, the edges of the wafer do not contribute. Thus, the components of the electric fields perpendicular to the interface satisfy the condition

$$\varepsilon_1 E_{1\perp} - \varepsilon_2 E_{2\perp} = 0 \qquad \text{(A.1)}$$

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

By a similar argument for normal components of magnetic fields we obtain the following boundary condition from Eq. (2.1b):

$$B_{1\perp} - B_{2\perp} = 0 \qquad \text{(A.2)}$$

It may be emphasized here that only the normal components of **D** and **B** are equal on both sides of the boundary. Their total magnitudes may not be equal and their directions need not be the same. In fact, these fields may well be reflected or refracted and may also change directions.

We now consider the components of two fields parallel to the boundary and apply Eq. (2.1c) to a thin Amperian loop across the surface. This yields

$$\mathbf{E}_1 \cdot \mathbf{I} - \mathbf{E}_2 \cdot \mathbf{I} = -\frac{d}{dt}\int_S \mathbf{B} \cdot d\mathbf{S} = -\frac{d\phi_B}{dt}$$

where **B** is the magnetic flux. As the width of the loop goes to zero, the magnetic flux vanishes. Therefore,

$$(\mathbf{E}_1 - \mathbf{E}_2) \cdot \mathbf{I} = 0$$

which implies that

$$E_{1\|} - E_{2\|} = 0 \qquad \text{(A.3)}$$

That is, the components of **E** parallel to the interface are continuous across the boundary.

In the same way, from Eq. (2.1d) we find that the parallel components of the magnetic field are equal and continuous. Mathematically we write

$$\frac{1}{\mu_1} B_{1\|} = \frac{1}{\mu_2} B_{2\|} \qquad \text{(A.4)}$$

UNIT 3 PERCEPTION OF LIGHT

CONTENTS

1.0 INTRODUCTION

The sense of vision is one of our most prized possessions. It enables us to enjoy the splendours of nature, stimulates our thinking and enriches our lives in many ways. We become aware of the infinite variety of objects around us, especially their shapes, colours, textures and motion, etc., only due to our ability to see them. But have you ever thought: What makes us to see? It all begins with eyes but also depends on what happens behind the eye. Every object viewed is seen with light. The eye responds to illumination. We all know that all living species - from one celled amoeba to the great bald eagle - have developed special mechanisms for responding to light.

> The amoeba reacts only to extreme changes in light intensity such as light and darkness. The earthworms react to light through light sensitive cells present on their skin. This ability to sense only general level of light intensity is termed **photosensitivity.**

Human perception of light, i.e., vision, is a more developed process. It takes place almost spontaneously without anyone, other than the perceiver, knowing what is happening. Perception of light involves the formation of sharp images (in the visual apparatus) and their interpretation. Vision begins in the eye, but light is sensed by the brain. In fact, what we see is the world created by our visual apparatus inside our head. So we can say that vision involves a mix of physical and physiological phenomena. In this unit we will develop on what you already know. In Sec. 3.1 you will get an opportunity to review the internal eye structure and know how light is sensed. Sec. 3.2 is devoted to colour vision where you will learn about dimensions of colour, the trichromatic and opponent-colour theories.

2.0 OBJECTIVES

After studying this unit, you should be able to:

- explain the functions of different parts of the eye
- list common eye defects and suggest remedial measures
- describe how the human eye responds to colour, and
- explain trichromatic and opponent-colour theories of colour vision.

3.0 MAIN CONTENT

3.1 Human Vision

Vision involves a mix of physical phenomena and physiological processes. We can understand how the image of an object is formed within the eye purely in terms of physical principles and processes. But from image formation to its perception by the brain, the process is physiological.

> Human vision also has a rich relationship with other senses. In fact, all our five senses cooperate and augment each other.

In this section our emphasis will be on the physics of vision. We shall also discuss very briefly the physiology of vision. Let us begin our study of human vision with the eyes - our windows to the external world.

> In medicine, the study of structure, functions and diseases of the eye is called **opthalmology.** Human eyes are very versatile and highly accurate. Their overall visual horizon is broad. But they are less acute than a hawk's eyes and less wide-seeing than those of a deer. Moreover, human eyes are not ideally suited for seeing underwater, nor are they very efficient at night. Even in twilight, eyes lose all perceptions.

Our eyes are very versatile. They possess a staggering degree of adaptability and precision. They are capable of extremely rapid movement. That is why we can instantaneously shift the focus from a book in hand to a distant star, adapt to bright or dim light, distinguish colours, scan space, estimate distance, size and direction of movement. You may now ask: How does vision begin in the eye? What is the internal structure of the eye? How does the brain interpret images? The answers to such questions have fascinated man for thousands of years.

Physiologists say that the human eye is an image-making device. (In a way, the human eye has striking similarities to a camera of automatic intensity and focal control.) To know the details of the mechanism of vision, some knowledge of the visual apparatus is necessary. You will now learn about the structure of eye and how it works as an optical instrument.

3.1.1 Viewing Apparatus: The Eye

Our eyes, as you know, are located in the bony sockets and are cushioned in fatty connective tissue. The adult human eye measures about 1.5 cm in diameter. Now refer to Fig. 3.1. It shows a labelled diagram of human eye.

The **sclera** or 'white' of the eyeball is an opaque, fibro-elastic capsule. It is fairly tough and gives shape to the eyeball, protects its inner parts and withstands the intraocular pressure in the eye. The muscle fibres which control eyeball movement are inserted on the sclera. The **cornea** is a tough curved front membrane that covers the **iris,** the coloured circular curtain in the eye. The cornea acts as transparent window to the eye and is the major converging element.

The cornea is followed by a chamber filled with a transparent watery liquid, the **aqueous humor**, which is produced continuously in the eye. It is mainly responsible for the maintenance of intraocular pressure. Besides this, aqueous humor is the only link between the circulatory system and the eye-lens or cornea. (Neither the lens nor the cornea has blood vessels.) The intraocular pressure maintains the shape of the eye, helps to keep the retina smoothly applied to the choroid and form clear images. Near the rear of this chamber is the iris. The iris is opaque but has a small central hole (aperture), called **pupil.** In our common observation, the pupil appears more like a black solid screen. Why? This is because behind the opening is the dark interior of the eye. The size of the pupil in normal eye is about 2 mm. The light enters the eyeball through this area. The iris is suspended between the cornea and the lens. The principal function of the iris is to regulate the intensity of light entering the eyeball. When the light is bright, the iris contracts and the size of the pupil decreases and vice versa.

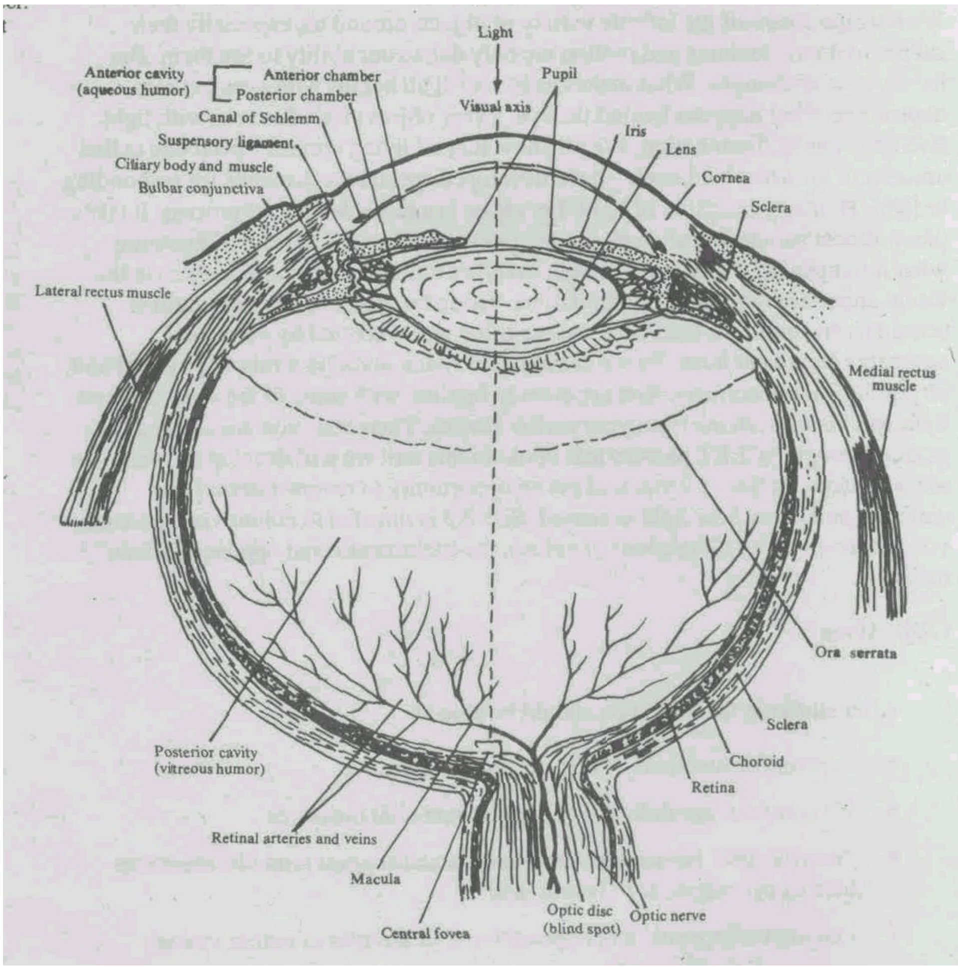


Fig.3.1 A schematic labelled diagram of human eye

Thread-like suspensory ligaments hold the biconvex crystalline eye-lens, which is just behind the pupil and iris. The muscle responsible for changes in the shape of the lens for near as well as far vision is called the **ciliary muscle.** The eye-lens is an elastic structure made of protein fibres arranged like the layers of an onion. It is perfectly transparent and its focal length is about 3 cm.

The crystalline lens is followed by a dark chamber, which is filled with **vitreous humor.** It is a transparent jelly-like substance. It augments the functions of aqueous humor and helps the eye hold its shape. The rear boundary of this chamber is the **retina,** where the image of the object is formed. The microscopic structure of the retina is shown in Fig. 3.2(a). It consists of a nervous layer and a pigmented layer. Apart from sensing the shape and the movement of an object, the retina also senses its colour. The retina consists of five types of **neuronal cells:** the photoreceptors, bipolar, horizontal, amacrine and ganglion neurons. A magnified view of the arrangement of neuronal cells in the retina is shown in Fig. 3.2(b).

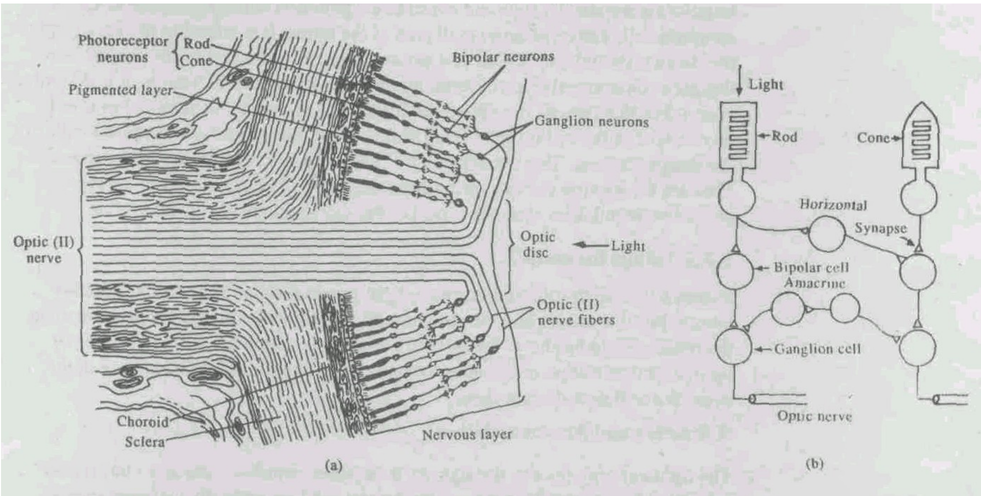


Fig.3.2: (a) Microscopic structure of retina (b) A Magnified view of arrangement of neuronal cells in the retina

The photoreceptor neurons are of two types: **rods** and **cones.** (This nomenclature is due to their geometrical shapes.) It is estimated that about 130 million rods and cones are found lining the retina. Of these, about six million are cones and about twenty times as many are rods. The light sensitive pigments of photoreceptors are formed from vitamin A.

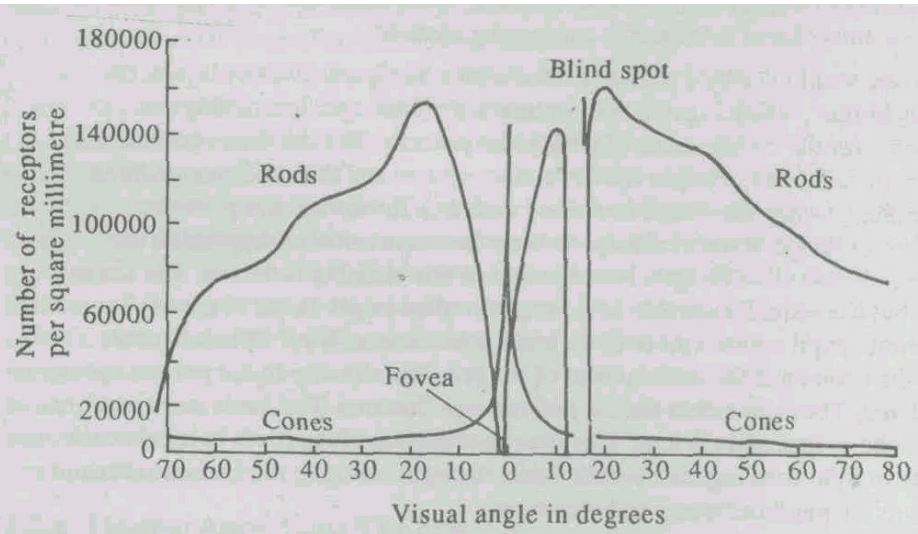


Fig. 3.3 Distribution of rods and cones in the retina of the human eye

At the very centre of the retina is a small yellowish depression, called **fovea.** This small valley (of about 5mm diameter) contains a large number (~110,000) of cones and no rods. The distribution of rods and cones across the human retina is shown in Fig. 3.3. The horizontal axis shows the distances in degrees of visual angle from the fovea located at 0°.

Rods are highly specialized for vision in dim light. They enable us to discriminate between different shades of dark and light, see shapes and movements. That is, rods provide a high sensitivity. Cones contain light sensitive pigments which make colour vision and sharpness of vision (high visual acuity) possible.

When light is absorbed by photoreceptor cells, the light sensitive pigments are broken up by specific wavelengths of light. The bipolar nerve cells carry nerve impulses generated by rods and cones to the ganglion cells. The axons of the ganglion cells converge on a small area of the retina. It is lateral to the fovea and is free from rods and cones. Can you say anything about its ability for vision? Since this area contains only nerve fibres, no image is formed on it. That is, it is devoid of vision. For this reason, it is called the **blind spot.** You may be tempted to ask: Is there a spot in the eye for maximum vision? Certainly yes, the fovea is the valley of the sharpest vision. This remarkable perceptive ability is provided by the cones. Muscles for moving the eye spring from the sclera. The conjunctiva - a supple protective membrane - joins the front of the eye to the inside of the eyelids.

3.1.2 Image Formation

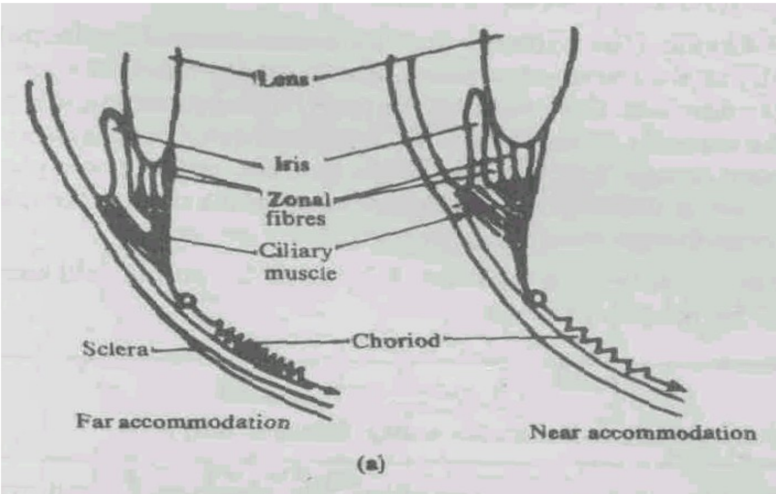
Before stimulating rods and/or cones, light passes through the cornea, aqueous humor, pupil, eye-lens and vitreous humor. For clear vision, the image formed on the retina should be sharp. Image formation on the retina involves the refraction of light, accommodation of the eye-lens, constriction of the pupil, and convergence of the eyes. We will now discuss these.

Refraction and Accommodation

The light entering the eye through the transparent window - **cornea** - undergoes refraction four times. This is because the eye has four optically different media: cornea (n = 1.38), aqueous humor (n = 1.33), eye-lens (n = 1.40), and vitreous humor (n = 1.34). Most of the refraction occurs at the air-cornea interface. Can you say why? This is because the cornea has a considerably larger refractive index than air (n = 1.0). Moreover, due to the curved shape, the cornea bends the light towards the retina. Additional bending is provided by the eye-lens, which is surrounded on both sides by eye-fluids (Fig. 3.1). However, the power of the lens to refract light is less than that of the cornea. So the main function of the lens is to make fine adjustments in focussing. The focussing power of eye lens depends on the tension in the ciliary muscle. When the ciliary muscle is relaxed, the lens is stretched and thinned. When a visual object is 6m or more away from the eye, the cornea receives almost parallel light rays. When the eye is focussing an object nearer than 6m, the ciliary muscles contract. As a result, the lens shortens, thickens and bulges and its focussing power increases. These features are illustrated in Fig. 3.4. The great value of the lens lies in its unique ability to automatically change its focal control. This ability is called **accommodation.** Since accommodation means work for the muscles attached to the eye lens, viewing an object nearer than 6m for a long time can cause eye-strain.

> While a healthy cornea is transparent, disease or injury may result in blindness. But eye surgeons have now acquired competence in replacing damaged cornea with clear one from human donors. Any imperfection in the shape of the cornea may cause distortion in visual images.

> The eye-lens of elderly people tends to be less flexible and loses ability to accommodate. This condition is called prestyopia. For extra focussing power, they use glasses (spectacles or contact lens).



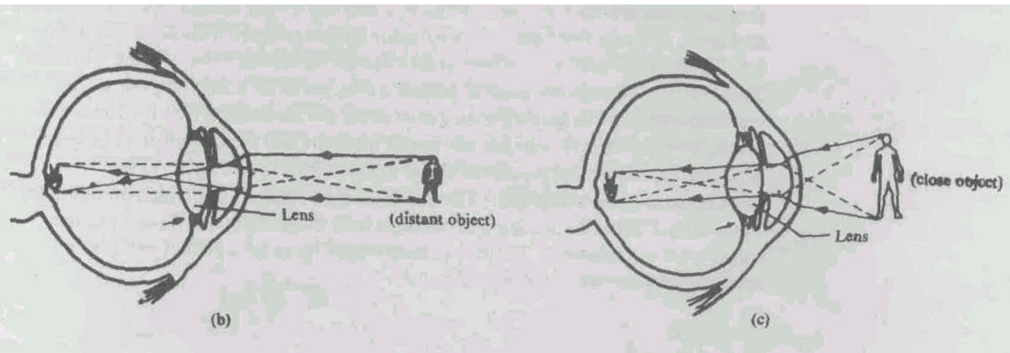


Fig. 3.4: Far and near accommodation (a) In the diagram on the left, the ciliary muscle Is relaxed. This causes the eye-lens to curve less. In the other diagram, the ciliary muscle Is contracted. This causes the lens to curve more, (b) Accommodation for far vision (6m or more away), (c) Accommodation for near vision

Constriction of the Pupil

Constriction of the pupil means the narrowing down of the diameter of the hole through which light enters the eye. This action occurs simultaneously with accommodation of the eye-lens and prevents the entry of light rays through the periphery of eye-lens, which can result in blurred vision. The pupil also constricts in bright light to protect the retina from sudden or intense stimulation. (When the level of illumination is low, the pupil dilates so that the retina can receive enough light.)

Convergence

Human beings have single binocular vision. This signifies that both eyes focus on only one set of objects. When we stare straight ahead at a distant object, the incoming light rays are directed at both pupils, get refracted and are focussed at identical spots on the two retinas. Suppose that we move close to the object and keep our attention on the same stationary object. Our common sense suggests that even now images should form on the same points (in both retinas). It really does happen and our eyes automatically make adjustments by radial movement of two eyeballs. This is referred to as **convergence.**

Refer to Fig. 3.4 again. You will note that the images formed on the retina are inverted laterally as well as up-side-down. But in reality we do not see a topsy-turvy world. You may now ask: How does this happen? The solution to this apparent riddle lies in the capacity of the brain which automatically processes visual images. This suggests that though vision begins in the eye, perception takes place in our brain. Its proof lies in that severe brain injury can blind a person completely and permanently, even though the eyes continue to function perfectly.

You may now like to reflect on what you have read. So you should answer the following SAE before you proceed.

SELF ASSESSMENT EXERCISE 1

Human beings are unable to see under water. Discuss why?

By now you must be convinced that mechanically speaking, the human eye is an optical instrument resembling a camera. (A better analogy exists between the eye and a closed circuit colour TV system.) The eyeball has a light focussing system (cornea and lens), aperture (iris) and a photographic screen (retina). This is shown in Fig. 3.5. There are of course very important differences between our eye and a camera. The engineering sophistication of human eye is yet to be achieved even in the costliest camera. The cameraman has to move the his camera lens for change of focus, whereas the eye-lens has automatic intensity and focal control. (The brain constantly analyses and perceives visual images. This is analogous to the development of a photograph.) The image on the retina is not permanent but fades away after l/20th of a second and overlaps with the next image. This gives the impression of continuity. There is of course no film in the eye that records the images permanently as a photo film does.

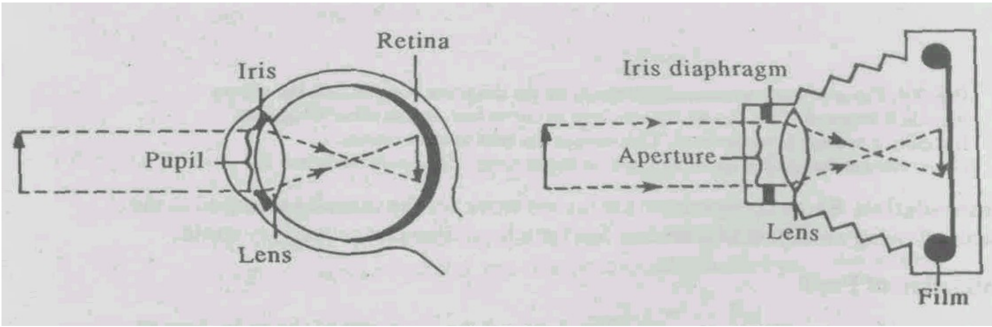


Fig.3.5 The similarity between the eye and the camera

3.1.3 Information Processing

As soon as light impulses impinge on the retina (and an image is formed), these are absorbed by rods and cones, which contain four kinds of photosensitive substances. These **visual pigment molecules** undergo structural (chemical) changes. It is believed that each rod cell contains about seventy million molecules of a purple-coloured photosensitive pigment, **rhodopsin.** Like rods, cones contain violet - coloured photosensitive pigment, **iodopsin.**

Rhodopsin has a molecular weight of about 4 x 10^4 dalton. It consists of the **scotopsin** protein and the chromophore retinene, a derivative of vitamin A in the form called cis-retinene. Any deficiency of vitamin A causes **night-blindness.** Fig. 3.6 shows the absorption curve of rhodopsin.

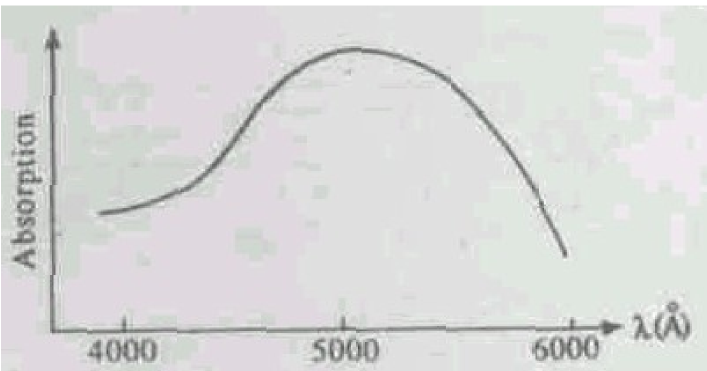


Fig. 3.6 The absorption curve of rhodopsin

Each pigment molecule consists of two components: a colourless protein, **opsin,** and a coloured chromophore, **retinene.** Opsin is different for each of the four visual pigments and determines the frequency of light to which each pigment responds.

Let us now understand as to what happens to rhodopsin in rods. (The same basic changes occur in the visual pigments in cones.) Refer to Fig. 3.7, which depicts the rhodopsin cycle. The first step in this process is the absorption of photon by rhodopsin, which then undergoes a chemical change. Its cis-retinene portion changes to **all-trans-retinene.** On referring to Fig. 3.8 you will note the rotation that occurs around the carbon numbered 12. This change triggers decomposition of rhodopsin (into scotopsin and all-trans-retinene) by a multi-stage process known as bleaching action. The pigment loses colour and the visual excitory event is believed to occur. Then rhodopsin is resynthesized in the presence of vitamin A. In this process, an enzyme, **retinene isomerase,** plays the most vital role.

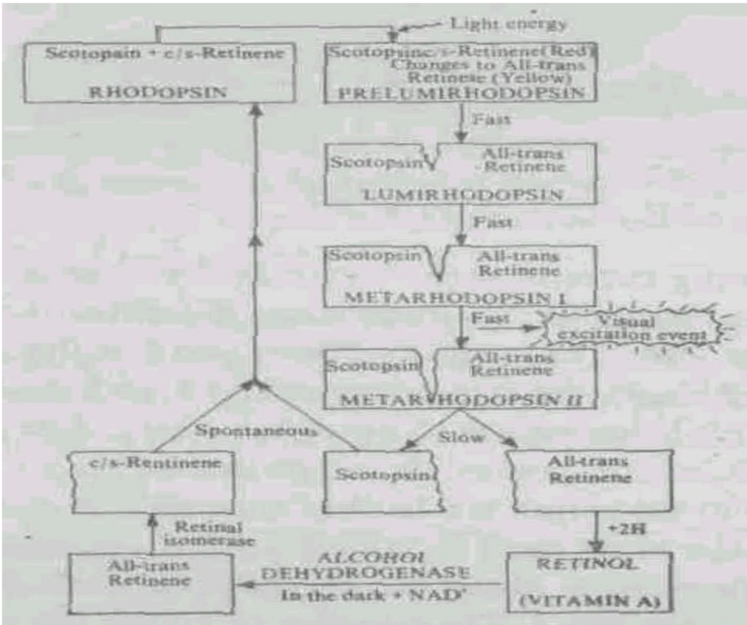


Fig. 3.7 The rhodopin cycle: Bleaching action

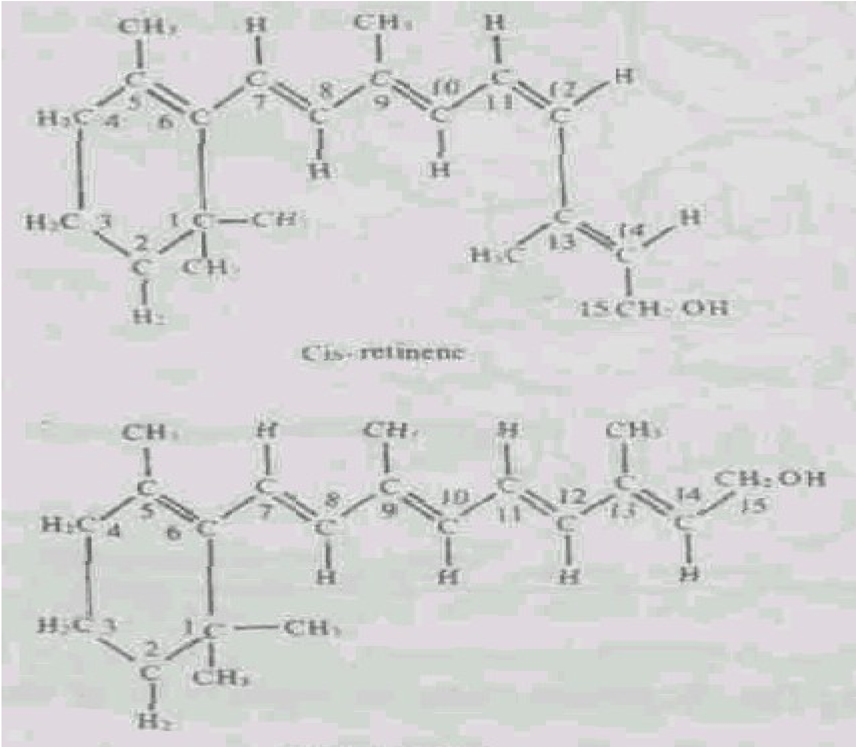


Fig.3.8 Structures of cis-retinence and all-trans- retinence

The rods respond even to poor illumination such as twilight. Rhodopsin is highly sensitive to even small amounts of light. Their responses to light generate colourless images and objects are seen only in shades of grey. It is for this reason that you see a red flower black in the evening. On the other hand, the pigments of the cones are much less sensitive to light and require bright illumination to initiate decomposition of chromophore. Visual acuity or ability to see clearly and to distinguish two points close together is very high and their responses produce coloured images.

The information received in terms of light is converted into electrical signals in the retina. The potential of the cell membranes of the photoreceptor cells undergoes a change even on brief illumination. This occurs through a complex chemical process involving a flow of calcium ions and sodium ions across the membrane. The change in membrane potential, ΔV_m, is governed by the following equations in time and space:

$$\Delta V_m(t) = I_m R(1 - e^{-t/\tau}) \quad (3.1)$$

and

$$\Delta V_m(x) = V_0 e^{-x/L} \quad (3.2)$$

where I_m is the membrane current, R the membrane resistance, τ is the membrane time constant. V_0 is the change in the membrane potential at $x = 0$ (*x* being the distance away from the site of current injection) and L

is the length constant. As can be seen, the spread of ΔV_m in space is governed by L (whose values fall in the range of about 0.1 to 1 mm). It is important to note that while slow potentials are generated in most cells, action potentials are produced only in the ganglion cells. The signals generated in the retina are further transmitted to the higher centres in the visual pathway of the brain such as lateral geniculate nucleus and visual cortex. In this way, precise information about the image projected on the retina is conducted accurately to the brain. The transfer of visual information in a typical retinal circuit is shown (Fig. 3.9).

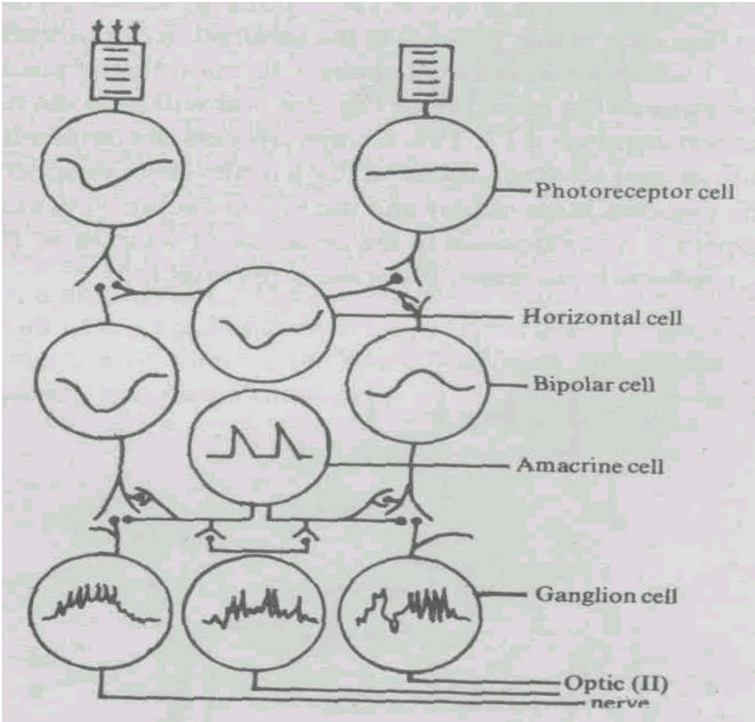


Fig.3.9 Retinal circuit showing the electrical links between cells of the retina: Action potential

We hope that now you have a reasonable idea of how we perceive the world around us. You may now like to know the factors that hamper vision.

3.1.4 Defects of Vision

Sometimes the eye loses its power of accommodation. When this happens, we are unable to see objects clearly and vision becomes blurred. These are corrected by using contact lenses or spectacles.

In one kind of such a defect, human beings can see nearby objects clearly but it is difficult to see objects at long distances, in such a (defective) eye, the image of distant objects is formed in front of the retina (Fig. 3.10a) rather than on the retina. This defect of the eye is known as **short-sightedness** or **myopia.** It is frequently observed in children. In short-sightedness, the eyeball gets elongated. It can be corrected by using a concave (divergent) lens (Fig. 3.10b) of appropriate focal length which moves the image on to the retina.

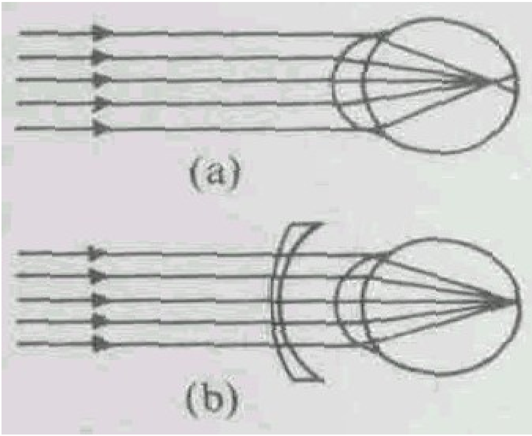


Fig. 3.10 (a) Short sightedness (b) its correction

In another eye defect, eyeball gets shortened. Though distant objects are seen clearly, nearby objects look blurred. In this case the image is formed behind the retina (Fig. 3.11a). This defect is known as **long-sightedness** or **hypermetropia.** It is normally observed in elderly people. It can be corrected by using a convex (convergent) lens of appropriate focal length (Fig. 3.lib).

Sometimes a person may suffer from both myopia and hypermetropia. Such people often use bifocal lenses, in which one part of the lens acts as a concave lens and the other part as a-convex lens. The third type of defect of vision is called **astigmatism,** wherein distorted images are formed. The corrective lenses are used to restore proper vision.

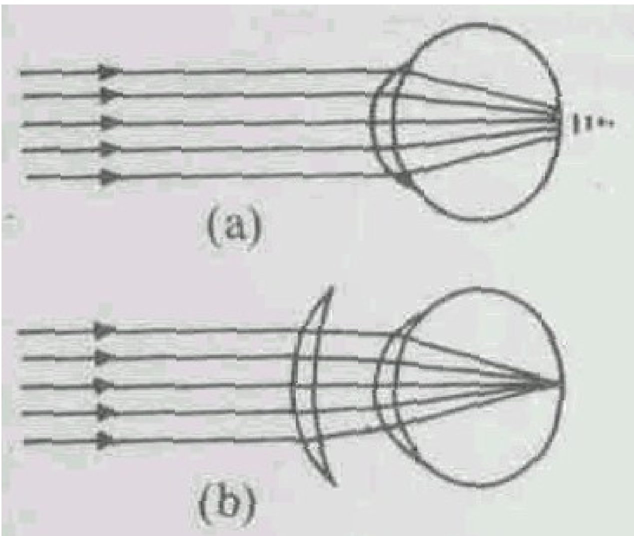


Fig. 3.11 (a) Long sightedness (b) its correction

3.2 Colour Vision

You all know that human beings have remarkable sense to adore the varied creations of nature. This is particularly because colour is an automatic part of our perception. In fact, the phenomenon of colour vision has added real charm to life. Can you now realize what vision is like without colour? You will learn that the colour is a perceptual experience; a creation of the eye and the mind.

One of the earliest observations about colour perception was made in 1825 by Purkinje. He observed that at twilight, blue blossoms on flowers in his garden appeared more brilliant than the red. To understand this you must know the mechanism of colour vision. The process of colour perception is influenced by the physiology of the eye and the psychology of the person. Before we plunge into these details, it is important to know the dimensions of colour, i.e., the parameters with which colour may be defined.

3.2.1 The Dimensions of Colour

The most important physical dimension of colour is the wavelength of light. For most light sources, what we perceive is the dominant colour, which we call the **hue.** It is hue to which we give the names like red, blue or greenish yellow. In fact, the terms colour and hue are frequently used interchangeably. You may therefore conclude that **hue is the perceptual correlate for variations in wavelength.**

The second dimension relevant to colour vision is **illuminance,** which refers to the amount of light reaching the eye directly from the source. Illuminance, therefore, characterizes the perceived brightness of a coloured light. This relationship (between illuminance and brightness) is fairly complex because perceptual sensitivity varies with the wavelength of light. Every individual with normal eye possesses maximum sensitivity to light between the green and yellow parts of the spectrum (500nm - 600 nm). And the sensitivity to predominantly blue light (400 - 500 nm) is rather low.

> Intensity is defined as the amount of energy reaching a receiver of given cross-sectional area every second.

Another physical dimension associated with colour is the degree **of purity** of spectral composition. That is, purity characterises the extent to which a colour (hue) appears to be mixed with white light. This is responsible for variations in the perceived saturations of the colour. For example, when we add white light in a spectrally pure blue, the light begins to look sky-blue. On progressive addition of white light you may eventually observe it as white.

We may therefore conclude that

> Colour, as a perceptual phenomenon, is three-dimensional and is characterised by hue, saturation and brightness.

Thinking logically, you may now ask: Is there any other alternative expression for the dimensions of colour? The answer to this question is: Yes, there is. It is based on the observation that colour depends on the intensity of light. Let us now learn about it in some detail.

Trichromacy

You must have realised sometimes that when the intensity of light is low, we see no colours. You also know that by varying the wavelengths and/or intensities of lights of different colours, it is possible to produce light of a desired colour. In your school you must have learnt that all the colours of the visible spectrum can be produced by mixing lights of just three different wavelengths: red, green and blue. These are known as the primary colours. The explanation for this trichromacy lies in the mechanism for colour vision. You will learn about it in the next sub-section.

Another important phenomenon associated with colour vision is complementarity of colours, i.e., pairs of colours, when mixed, seem to annihilate one another. For example, when we mix suitable proportions of a monochromatic blue light ($\lambda \sim 470$ mm) with a monochromatic yellow light ($\lambda \sim 575$ nm), we obtain a colourless grey.

Reflecting on this observation, Hering suggested that complementary pairing is an indicator for pairing in the mechanisms responsible for signalling colour in the visual system. The complementary relationships among pairs of colours can be well represented as shown in Fig. 3.12. To locate the complementary colour in this figure all that you have to do is to choose any point and draw a line passing through the centre of the circle. A suitably adjusted mixture of two complementary colours will appear grey.

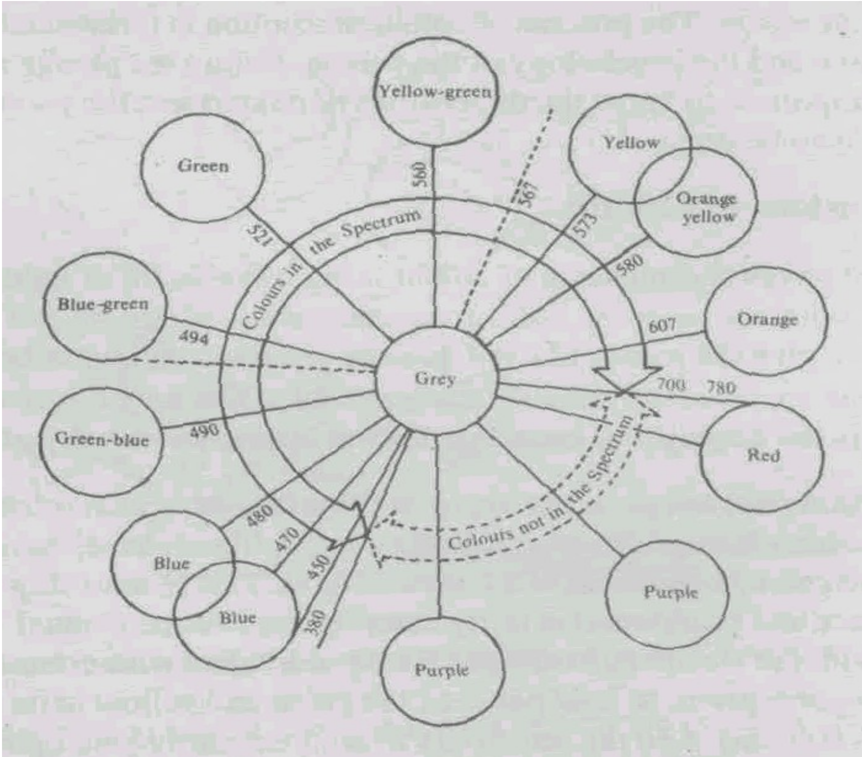


Fig. 3.12 The complementary colour circle

Before you proceed further, we want you to pause for a while and answer the following SAE.

SELF ASSESSMENT EXERCISE 2

How would you indicate brightness and saturation in Fig. 3.12?

Note the presence of 'purple' hues. You may recall that dispersion of white light by a prism does not reveal this hue. Then the question arises: What is their significance in the colour circle? The complementary circle will remain incomplete without them.

You may also note that though colour circle represents colours as a continuum, primary colours are perceptually quite distinct. The phenomena of primary colours and trichromacy led Young to propose three different types of receptors (cones) for colour vision. You will learn the details as you proceed.

Colour Blindness

You now know that a single monochromatic light can be produced by combining two primary colours. The measurements made to know the amounts of these colours required to match a given monochromatic colour gave fairly standard results. That is, when we ask a group of people to match a test colour, experience tells that they mix the same proportions of primary colours. But colour-mixing requirements for some individual may be anomalous. In fact, some individuals may need only two, rather than three, primary colours to match all the monochromatic hues. These anomalies are indicative of varying degrees of **colour blindness.** People who show anomalous colour-matching requirement do not see the same colours as individuals having normal vision. The most common defect is in the proportions of red and green lights required to match a monochromatic yellow. The manifestation of this in everyday life is a limited ability to distinguish between red and green.

3.2.2 Colour Receptors

In the above paragraphs you have learnt that trichromatic theory led Young to propose that the eye possesses three types of cones, each containing a different pigment. And three types of pigments in the cones correspond to three primary colours (three-dimensional colour vision). The absorption curves for these pigments are shown in Fig. 3.13. You will note that the curves show substantial overlap. Moreover, the blue mechanism is markedly less sensitive than the other two.

The argument leading to this conclusion is rather subtle and needs closer analysis. To understand this, let us ask: How do humans distinguish such a large number of colours? Do we need a different type of receptor to discriminate each colour? Since the colours are numerous, the number of receptors available for a particular colour will be a small fraction of the total number of colour receptors. When monochromatic light reaches our eye, only the corresponding class of receptors will respond. And since the total number of responding receptors is comparatively small, the ability to see a monochromatic light will be much less than the ability to see white light. But in practice, this is not true. This led Young to conclude that only a few different types of receptors are present, which by working in combination give rise to all the different colours we perceive. His experience with colour mixing led him to conclude that the number of receptor types is only three.

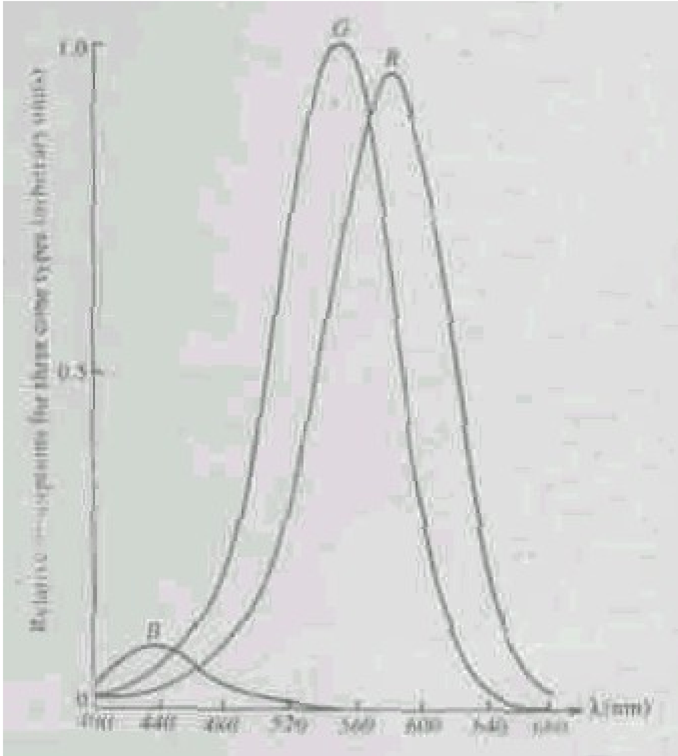


Fig. 3.13 Spectral absorption curves for three different cone pigments

This theory was proposed even before very little was known about the physiology of the visual system. The outputs from the three types of receptors are transmitted separately to the brain which combines the information and constructs certain abstractions to which we give names like hue, saturation, yellow, blue etc.

We all know that yellow gives a sensation independent of red, blue and green, i.e. it seems as much of a primary colour. But no coding system is postulated for yellow in the trichromatic theory. Such feelings, though subjective, led Hering to propose an alternative theory of colour based on four colours: red, yellow, green and blue. This is known as **opponent-colour theory.** These colours are associated in pairs: red-green and blue-yellow. The members of a, pair are thought to act in opposition adding up to white. Hering also specified a third pair of black and white to represent the varying brightness and saturation of colours. (The perception of brightness of the colour also depends on the mood of the perceiver.) You must appreciate that the most important difference between this theory and the trichromatic theory lies not in the number of postulated receptor types, but in the way their outputs are signalled to the brain. Fig. 3.14 depicts a simple version of the opponent-process theory. Three basic receptor types are indicated by *X, Y* and Z. A mixture of *Y* and *Z* is perceived as yellow. White is obtained by mixing X, Y and *Z.*

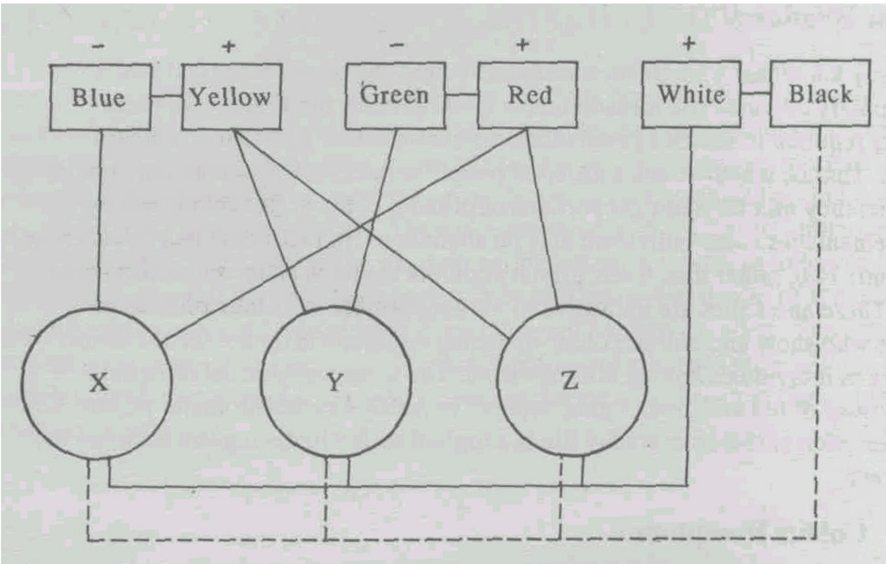


Fig.3.14 Opponent-process theory based on Bering's postulate: X, Y and Z denote basic receptor types

According to the model shown in Fig. 3.14, three different receptor types are each sensitive to a range of wavelengths. The mode of operation is such that the activity level increases in response to a predominant input about one colour. You may ask: What happens in response to the input about the complementary colour? We expect it to decrease. To illustrate it, let us consider that the input to the blue-yellow system is predominantly in the yellow region of the spectrum. Then, there is an increase in activity (over a spontaneous level) about yellow colour. On the other hand, if the input is predominantly blue, there is a decrease in activity. Activity in the black-white mechanism is based on outputs from all three receptor types.

Even though trichromatic theory and the opponent-process theories appear conflicting, recent studies show evidences that they are compatible. Research at the Johns Hopkins University (US) provides evidence in favour of the trichromatic theory. However, the cones do not send 'color signals' directly to the brain. Cone signals pass through a series of neurons which are colour specific.

Vision is an endlessly fascinating area. We here conclude saying: The eye is not merely an instrument for survival; it is a means for enrichment of life.

We will now like you to answer the following SAE.

SELF ASSESSMENT EXERCISE 3

Name the regions of the retina specialised for (a) colour and detailed vision at high levels of illumination and (b) non-colour vision at low levels of illumination.

4.0 CONCLUSION

In this unit, the internal human eye structure and the function of each part of the eye is discussed in detail. It is mentioned that human eyes are image making devices. The image of an object is formed at the retina.

The eyes have striking similarities to a camera. There are however differences in detail. The distribution of rods and cones in the retina of human eye are explained. The image formation of the retina involves refraction of light, accommodation of eye-lens, constriction of pupil and convergence of the eyes. The two type of defects of the eye are Myopia (short-sightedness) and hypermetropia (long-sightedness) which can be corrected by using a concave and convex lens respectively.

The mechanism of colour vision is also discussed in this unit. The process of colour, as a perceptual by the physiology of the eye and the psychology of the person. Colour, as a perceptual phenomenon, is three dimensional(3-D) and is characterised by hue, saturation and brightness.

5.0 SUMMARY

- The perception of light involves the formation of sharp images in the eye and their interpretation in the brain. That is, vision involves a mix of physical and physiological phenomena.
- Human eyes are image-making devices. They have striking similarities to a camera of automatic intensity and focal control. There are however differences in details.
- The cornea is the major converging element in the eye.
- The image of an object is formed on the retina. It consists of five types of neuronal cells: photoreceptors, bipolar, horizontal, acrine and ganglion neurons.
- The photoreceptor neurons are of two types: rods and cones. Rods are specially suited for vision in dim light and provide high-sensitivity. Colour vision and sharpness are possible due to cones.
- Image formation on the retina involves refraction of light, accommodation of eye-lens, constriction of pupil and convergence of the eyes.

- Information processing involves structural changes in photosensitive pigment rhodopsin by bleaching action.
- Two common defects of the eye are myopia (short sightedness) and hypermetropia (long sightedness). These are corrected by using a concave and a convex lens respectively.
- Colour, as a perceptual phenomenon, is three-dimensional: hue, illuminance and purity.
- According to Young's trichromacy theory, colour vision requires three types of receptors (cones) for three primary colours.
- According to Bering's opponent-colour theory, colours are associated in pairs: red-green, blue-yellow and add up to white. The brightness and saturation are determined by a black-white pair.

6.0 TUTOR MARKED ASSIGNMENT

1. List the differences between the human visual system and a camera.
2. When we enter a dark room, we feel blinded. Gradually we become dark-adapted. The dark adaptation curve shown here shows a kink. Can you suggest an explanation in terms of rod and cone-adaptation?

SOLUTIONS TO SELF ASSESSMENT EXERCISES

i. The refractive indices of water and cornea are 1.33 and 1.38, respectively. Due to small difference in these values, the cornea is unable to bend light towards the retina. This is why humans are unable to see under water.

ii. An arrow originating at the centre and directed towards the circumference would indicate increasing colour saturation. Brightness does not depend on hue and saturation. So a line drawn normally out of the page (towards you) would represent increasing brightness.

iii. The first description applies to the fovea whereas the second description applies to the peripheral regions.

7.0 REFERENCES/FURTHER READINGS

Fundamental to Optics- Jenkins and White

Introduction to Modern Optics, Grant F. Fowles.

Optics; Hecht and Zajac

Optics, Smith and Thomson

UNIT 4 POLARISATION OF LIGHT

CONTENTS

1.0 INTRODUCTION

In Unit 1 of this course, you learnt that light is a transverse electromagnetic wave. In your school physics curriculum you have learnt that while every wave exhibits interference and diffraction, polarisation is peculiar only to transverse waves. You may even be familiar with the basics of polarisation in your PHY 124 like: What distinguishes the polarised light from unpolarised light? Is light from an ordinary (or natural) source polarised? How do we get polarised light? And so on. In this unit we propose to build upon this preliminary knowledge.

You must have seen people using antiglare goggles and also antiglare windshields for their cars. Do you know that polarisation of light has something to do with these? Polarisation of light also plays a vital role in designing sky light filters for cameras and numerous optical instruments, including the polarising microscope and the polarimeter.

In Sec. 3.1 we have discussed as to what is polarisation. In Sec. 3.2, you will learn about simple states of polarised light. Sec. 3.3 is devoted to ideal polarisers and Malus' law. In this section you will also learn about double refraction or optical birefringence - a property of materials helpful in producing polarised light. In Sec. 3.4, you will learn some techniques of producing circularly and elliptically polarised light.

2.0 OBJECTIVES

After going through this unit you should be able to:

- explain what is linearly, circularly or elliptically polarised state of light;
- describe how can light be polarised by reflection;
- solve simple problems based on Malus' law and Brewster's law;
- explain how optical birefringence helps in production of polarised light; and
- explain the production of linearly polarised light by dichroism.

3.0 MAIN CONTENT

3.1 What is Polarisation?

What is polarisation? Why light, not sound, waves are known to polarise? These are some of the basic questions to which we must address ourselves. Polarisation is related to the orientation (oscillations) of associated fields (particles). Refer to Fig. 4.1, which depicts a mechanical wave (travelling along a string). From Fig. 4.1(a) you will note that the string vibrates only in the vertical plane. And vibrations of medium particles are confined to just one single plane. Such a wave is said to be (plane) polarised. How would you classify waves shown in Fig. 4.1(b) and (c)? The wave shown in Fig. 4.1(b) is plane polarised since vibrations are confined to the horizontal plane. But the wave in Fig. 4.1(c) is unpolarised because simultaneous vibrations in more than one plane are present. However, it can be polarised by placing a slit in its path as in Fig. 4.1(d). When the first slit is oriented vertically, horizontal vibrations are cut off. This means that only vertical vibrations are allowed to pass so that the wave is linearly polarised. What happens when a horizontal slit is placed beyond the vertical slit in the path of propagation of the wave? Horizontal as well as vertical components (of the incident wave) will be blocked. And the wave amplitude will reduce to zero.

Let us now consider visible light. The light from a source (bulb) is made to pass through a polaroid (P), which is just like slit one in Fig. 4.1. The intensity of light is seen to come down to about 50%. Rotating P in its own plane introduces no further change in light intensity. Now if a second identical Polaroid (A) is introduced in the path of light so that it is parallel to P, the intensity of light from the bulb remains unaffected.

But rotating A in its own plane has a dramatic effect! For 90^0 rotation, the light is nearly cut-off.

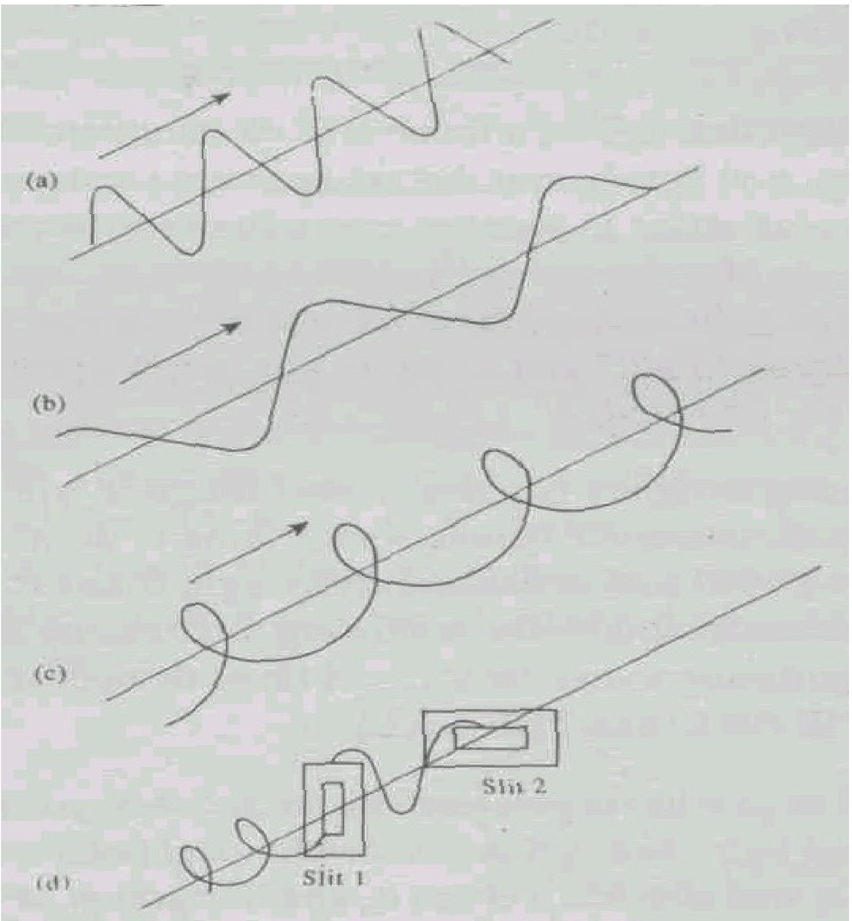


Fig. 4.1:(a) A vertically plane polarised wave on a string (b) A horizontally plane polarised wave (c) an unpolarised wave, (d) The wave in (c) becomes plane polarised after passing through slit one; the wave amplitude reduces to zero if another slit oriented perpendicular to slit one is introduced.

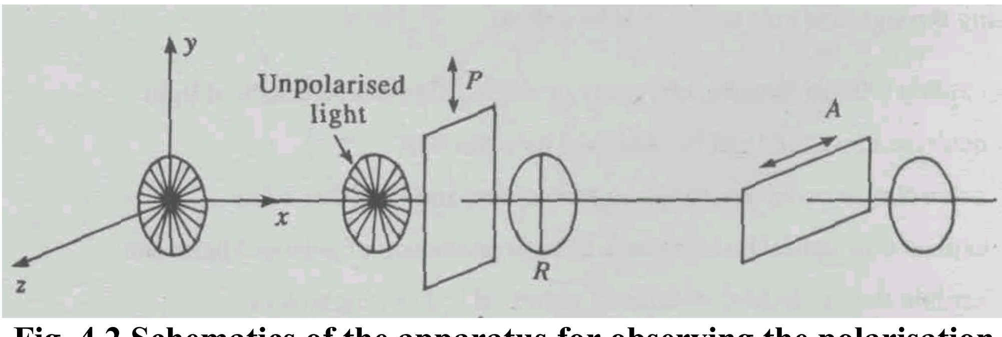


Fig. 4.2 Schematics of the apparatus for observing the polarisation of light

You can analyse this result in terms of electromagnetic theory, which demands complete description of associated electric vector and the way it oscillates with respect to the direction of propagation. For the arrangement shown in Fig. 4.2, the electric field vector at the source has all orientations in the yz- plane. The wave propagates as such till it reaches the polaroid P, which allows essentially unhindered passage of electric vector oriented parallel to its transmission axis. If the transmission axis is along y-axis, the electric field along y-direction (E_y) passes through it unaffected. In addition, the y-components of electric field vectors inclined to the y-axis can also pass through *P*. Thus, after passing through the polaroid *P,* the electric vectors oriented only along the y-axis will be present. **When electric vector oscillates along a straight line in a plane perpendicular to the direction of propagation, the light is said to be plane polarised.** The plane polarised wave further travels to the polaroid A, which is identical to *P*. When A is at 90 with respect to *P,* it can allow only the z-components of **E** to pass. Since only y-components of **E** are present in the wave incident on A, no light is transmitted by *A*.

We may now conclude that

1. No polarisation of longitudinal waves occurs as the vibrations are along the line of transmission only.
2. The transverse nature of light is responsible for their polarisation.

An important manifestation of this result arises in TV reception. You may have seen that the TV antenna on your roof tops are fixed in horizontal position. Have you ever thought about it? This is because the **TV signal transmission in our country** is through horizontally oriented transmitting antenna. The explanation for this lies in the observation that the pick up by the receiving antenna is maximum when it is oriented parallel to the transmitting antenna. This is illustrated in Fig. 4.3 for a vertical (dipole) transmitting antenna.

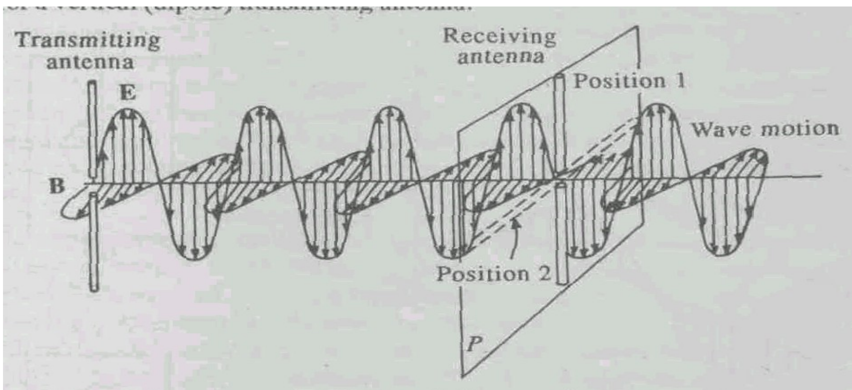


Fig. 4.3: Polarisation of an electromagnetic wave. The antenna responds to the vertical electric field strength of the wave. Reception is maximum in Position 1 and minimum in Position 2.

You may now like to know: Do natural (or ordinary) light sources emit polarised light? Answer to this question is 'yes' as well as 'no'! Is this answer not funny? You know that emission of light involves a large number of randomly oriented atomic (or molecular) emitters. Every individual excited atom radiates polarised waves for about 10^{-8} s. These waves form a resultant wave of given polarisation which persists for the lifetime of the excited atom. At the same time, other atoms (molecules) also emit waves, whose resultant states of polarisation may be quite different. Because of this randomness, every orientation of electric vector in space is equally probable. That is, electric vectors associated with light waves from a source are oriented in all directions in space and thus there is a completely unpredictable change in the overall polarisation. Moreover, due to such rapid changes, individual resultant polarisation states become almost indiscernible. The light is then said to be **unpolarised.**

In practice, visible light does not correspond to either of these extremes. The oscillations of electric field vectors are neither completely regular nor completely irregular. That is, light from any source is partially polarised. We ascribe a degree of polarisation to partially polarised light. The degree of polarisation is one for completely polarised light and zero for unpolarised light.
The next logical step perhaps would be to know various types of polarised light. Let us learn about this aspect now.

3.2 Simple States of Polarised Light

You now know that in e.m. theory, light propagation is depicted as evolution of electric field vector in a plane perpendicular to the direction of transmission. For unpolarised light, spatial variation of electric field at any given time is more or less irregular. For plane polarised light, the tip of electric vector oscillates up and down in a straight line in the same plane. The space variation of E for linearly polarised wave is shown in Fig, 4.4 (a). The diagram on the left shows the path followed by the tip of the electric vector as time passes. You will know that the tip of **E** executes one full cycle as one full wavelength passes through a reference plane. There are two other states of polarisation: circular polarisation and elliptical polarisation. The path followed by the tip of E, as the time passes, for these is shown in Fig. 4.4 (b) and (c), respectively.

In a right handed coordinate system if a right handed screw is turned so that it rotates the x -axis towards the y -axis, the direction of advance of the screw represents the positive z -axis.

The yz -plane (or $x = 0$ plane) in Fig. 4.4 is the plane of polarisation of the wave. We can identify other states of polarisation **by looking at the** trajectories **of the tip of the electric field vector** as the wave passes through the reference plane. You should always look at the reference plane from the side away from the source (looking back at the source) for the definitions to be unique.

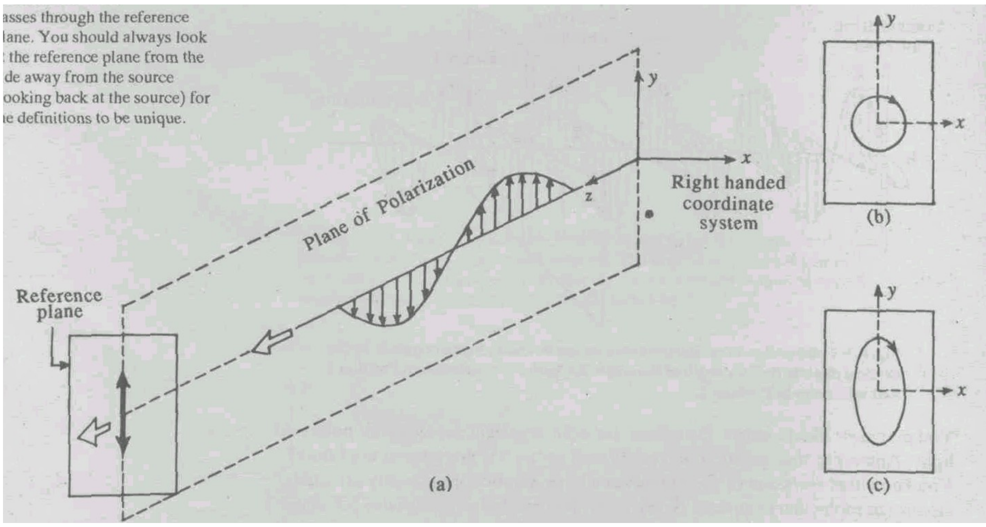


Fig. 4.4: Spatial variation of electric field vector for (a) linearly polarised light. The diagram on the left show the path taken by the tip of the electric vector as time varies. (b) and (c) show the path taken by the tip of the electric vector for circularly and elliptically polarised light

Let us now mathematically analyse how the superposition of two plane polarised light waves of the same frequency moving in the same direction gives rise to linearly, circularly or elliptically polarised light.

3.2.1 Linear Polarisation

Suppose that two light waves are moving along the z-direction. Let their electric field vectors be mutually perpendicular, i.e., we choose these along the x and y axes and can represent them respectively in the form

$$\mathbf{E}_1(z,t) = \hat{\mathbf{e}}_x E_{01}\cos(kz - \omega t) \tag{4.1}$$

and

$$\mathbf{E}_2(z,t) = \hat{\mathbf{e}}_y E_{02}\cos(kz - \omega t + \varphi) \tag{4.2}$$

$\mathbf{E}_2(z,t)$ lags $\mathbf{E}_1(z,t)$ for $\varphi > 0$ and vice versa.

Here $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ are unit vectors along the x and y-axes respectively. (These are also called polarisation vectors.) ϕ is the phase difference between the two waves.

We expect that the nature of the resultant wave will be determined by the phase difference between them and the value of the ratio E_{02}/E_{01}. Mathematically, we can write the vector sum of these as

$$\begin{aligned}\mathbf{E}(z,t) &= \mathbf{E}_1(z,t) + \mathbf{E}_2(z,t)\\ &= \hat{\mathbf{e}}_x E_{01}\cos(kz - \omega t) + \hat{\mathbf{e}}_y E_{02}\cos(kz - \omega t + \phi)\end{aligned} \tag{4.3}$$

Let us first take the simplest case where ϕ is zero or an integral multiple of $\pm 2\pi$. That is, when in-phase waves are superposed, Eq. (4.3) takes the form

$$\mathbf{E}(z,t) = (\hat{\mathbf{e}}_x E_{01} + \hat{\mathbf{e}}_y E_{02})\cos(kz - \omega\, t) \qquad (4.4)$$

The amplitude $\sqrt{E_{01}^2 + E_{02}^2}$ and the electric field oscillations in the reference frame make an angle $\theta = \tan^{-1}(E_{02}/E_{01})$ with the x-axis.

For the special case of in-phase waves of equal amplitude $(E_{01} = E_{02} = E_0)$, the resultant wave has amplitude $\sqrt{2}E_0$ and the associated electric vector is oriented at 45 with the x- axis. So we may conclude that **when two in-phase linearly polarised light waves are superposed, the resultant wave has fixed orientation as well as amplitude.** That is, it is also linearly polarised, as depicted in Fig. 4.5 (a). In the plane of observation, you will see a single resultant **E** oscillating cosinusoidally in time along an inclined line (Fig. 4.5 (b)). The **E** - field progresses through one complete cycle as the wave advances along the z-axis through one wavelength.

If we reverse this process, we can say that **any linearly polarised light can be visualised as a combination of two linearly polarised lights with planes of polarisation parallel to x = 0 and y = 0 planes.** (This is similar to resolving a vector in a plane along two mutually perpendicular directions.) In the subsequent sections, you will use this result frequently.

If the phase difference between two plane polarised light waves is an odd integral multiple of $\pm\pi$, the resultant wave will again be linearly polarised:

$$\mathbf{E}(z,t) = (\hat{\mathbf{e}}_x E_{01} - \hat{\mathbf{e}}_y E_{02})\cos(kz - \omega\, t) \qquad (4.5)$$

What is the orientation of the resultant electric vector in the reference plane? To know the answer of this question, work out the following SAE.

SELF ASSESSMENT EXERCISE 1

Depict the orientation of the electric vector defined by Eq. (4.5) in the reference (observation) plane.

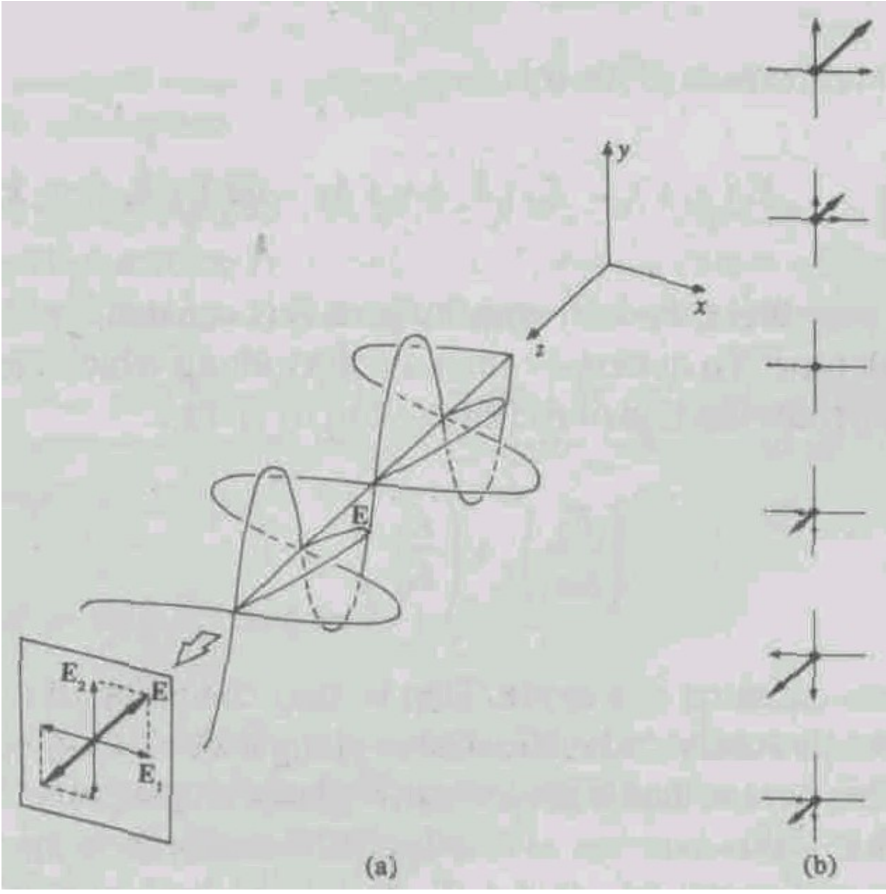


Fig. 4.5 Schematic representation of a plane polarised light wave

3.2.2 Circular Polarisation

We now investigate the nature of the resultant wave arising due to superposition of two plane polarised waves whose amplitudes are equal $(E_{01} = E_{02} = E_0)$ but phases differ by $\pi/2$, i.e. their relative phase difference $\phi = (2n-1)\frac{\pi}{2}$, $n = 0, \pm 1, \pm 2,...$ For n = 0, we can rewrite Eqs. (4.1) and (4.2) as

$$\mathbf{E}_1(z,t) = \hat{\mathbf{e}}_x E_0 \cos(kz - \omega t) \quad (4.6a)$$

$$\mathbf{E}_2(z,t) = \hat{\mathbf{e}}_y E_0 \sin(kz - \omega t) \quad (4.6b)$$

The resultant wave is given by

$$\mathbf{E}(z,t) = E_0[\hat{\mathbf{e}}_x \cos(kz - \omega t) + \hat{\mathbf{e}}_y \sin(kz - \omega t)] \quad (4.7)$$

You may note that the scalar amplitude of **E** is constant $(= E_0)$ but its orientation varies with time. To determine the trajectory along which the tip of **E** moves, we can readily combine Eqs.(4.6a) and (4.6b) to yield

$$\left(\frac{E_1}{E_0}\right)^2 + \left(\frac{E_2}{E_0}\right)^2 = 1 \qquad (4.8)$$

which is the equation of a circle. That is, the orientation of the resultant electric field vector changes continuously and its tip moves along a circle as the wave propagates (time passes). This means that **E** is not restricted to a single plane. The question now arises: What is the direction of rotation? Obviously there are two possibilities; Clockwise and counter-clockwise. To know which of these is relevant here, you should tabulate **E** at different space points at a given time, $t = 0$ say:

Location In space	$z = 0$	$z = \frac{\lambda}{8}$	$z = \frac{\lambda}{4}$	$z = \frac{3\lambda}{8}$	$z = \frac{\lambda}{2}$	$z = \frac{5\lambda}{8}$	$z = \frac{3\lambda}{4}$	$z = \frac{7\lambda}{8}$	$z = \lambda$
Electric field	$\hat{\mathbf{e}}_x E_0$	$\frac{\hat{\mathbf{e}}_x + \hat{\mathbf{e}}_y}{\sqrt{2}} E_0$	$\hat{\mathbf{e}}_y E_0$	$\frac{\hat{\mathbf{e}}_x - \hat{\mathbf{e}}_y}{\sqrt{2}} E_0$	$-\hat{\mathbf{e}}_x E_0$	$-\frac{\hat{\mathbf{e}}_x + \hat{\mathbf{e}}_y}{\sqrt{2}} E_0$	$-\hat{\mathbf{e}}_y E_0$	$-\frac{\hat{\mathbf{e}}_x - \hat{\mathbf{e}}_y}{\sqrt{2}} E_0$	$\hat{\mathbf{e}}_x E_0$

These are depicted in Fig. 4.6. If you position yourself in the reference plane and observe the evolution of **E** from $z = \lambda$ to $z = 0$ (backward towards source), you will find that the tip of **E** rotates clockwise. Such a light wave is said to be **right circular wave.** The electric field makes one complete rotation as the wave advances through one wavelength.

In case the phase difference $\phi = (2n+1)\frac{\pi}{2}$, $n = 0, \pm 1, \pm 2, \ldots$, Eq. (4.7) is modified to

$$\mathbf{E}(z,t) = E_0[\hat{\mathbf{e}}_x \cos(kz - \omega t) - \hat{\mathbf{e}}_y \sin(kz - \omega t)] \qquad (4.7)$$

Alternatively, we may fix an arbitrary point $z = z_0$ and observe the evolution of **E** as time passes. The figure below depicts what is happening at some arbitrary point z_0 on the axis.

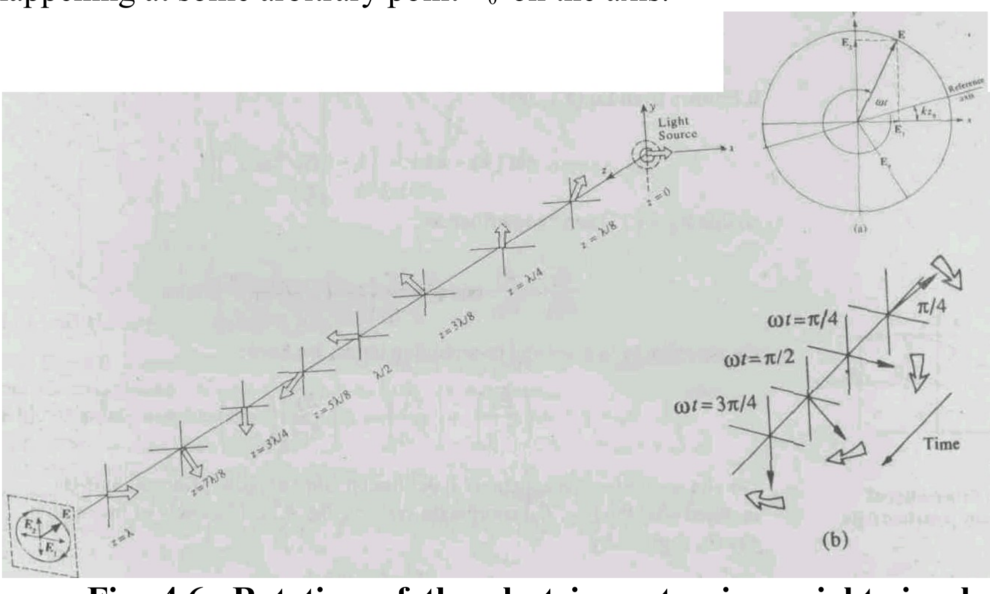


Fig. 4.6: Rotation of the electric vector in a right-circular wave. For consistency, we have used a

right handed system

It shows that the **E**-vector rotates counter-clockwise in the reference frame. (Before proceeding further you should convince yourself by tabulating the values of **E** at t = 0 for different space point.) Such a wave is referred to as **left-circular wave.**
Can you now guess as to what will happen if two oppositely polarised circular waves of equal amplitude are superposed? Mathematically, you should add Eqs. (4.8) and (4.9). Then you will find that

$$\mathbf{E} = 2E_0\hat{\mathbf{e}}_x \cos(kz - \omega t) \tag{4.10}$$

This equation is similar to Eq. (4.1) which represents a linearly polarised light wave. Thus, we may conclude that the **superposition of two oppositely polarised circular waves (of same amplitude) results in a linearly or plane polarised light wave.**

3.2.3 Elliptical Polarisation

Let us now consider the most general case where two orthogonal linearly polarised light waves of unequal amplitudes and having an arbitrary phase difference ϕ are superposed. Physically we expect that beside its rotation, even the magnitude of resultant electric field vector will change. This means that the tip of **E** should trace out an ellipse in the reference plane as the wave propagates. To analyse this mathematically, we write the scalar part of Eq. (4.2) in expanded form:

$$\frac{E_2}{E_{02}} = \cos(kz - \omega t)\cos\phi - \sin(kz - \omega t)\sin\phi$$

On combining it with Eq. (4.1) we find that

$$\frac{E_2}{E_{02}} = \frac{E_1}{E_{01}}\cos\phi - \sin(kz - \omega t)\sin\phi$$

or

$$\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}}\cos\phi = -\sin(kz - \omega t)\sin\phi \tag{4.11}$$

It follows from Eq.(4.1) that

$$\sin(kz - \omega t) = [1 - (E_1 / E_{01})^2]^{1/2}$$

so that Eq. (4.11) can be written as

$$\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}}\cos\phi = -[1-(E_1/E_{01})^2]^{1/2}\sin\phi$$

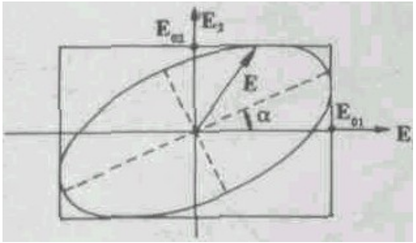


Fig 4.7 Schematics of elliptically polarised light

On squaring both sides and re-arranging terms, we have

$$\left(\frac{E_2}{E_{02}}\right)^2 + \left(\frac{E_1}{E_{01}}\right)^2 - 2\left(\frac{E_2}{E_{02}}\right)\left(\frac{E_1}{E_{01}}\right)\cos\phi = \sin^2\phi \qquad (4.12)$$

Do you recognise this equation? It defines an ellipse whose principal axis is inclined with the (E_1, E_2) coordinate system (Fig. 4.7). The angle of inclination, say α, is given by

$$\tan 2\alpha = \frac{2E_{01}E_{02}\cos\phi}{E_{01}^2 - E_{02}^2} \qquad (4.13)$$

For $\alpha = 0$ or equivalently $\phi = \pm\pi/2$, $\pm 3\pi/2$, …, Eq. (4.12) reduces to

$$\left(\frac{E_2}{E_{02}}\right)^2 + \left(\frac{E_1}{E_{01}}\right)^2 = 1 \qquad (4.14)$$

which defines an ellipse whose principal axes are aligned with the coordinate axes. We would now like you to solve an SAE.

SELF ASSESSMENT EXERCISE 2

Starting from Eq. (4.12) show that linear and circular polarisation states are special cases of elliptical polarisation.

Now that you understand what polarised light is, the next logical step is to know techniques used to get polarised light. You will learn some of these now.

3.3 Principles of Producing Linearly Polarised Light

The most important optical device in any polarised light producing arrangement is a polariser. It changes (the input) natural light to some form of polarised light (output). Polarisers are available in several configurations. An ideal polariser is one which reduces the intensity of an incident unpolarised light beam by exactly 50 percent. When unpolarised light is incident on an ideal polariser, the outgoing light is in a definite polarisation state (P-state) with an orientation parallel to the transmission axis of the polariser. That is, the polariser somehow discards all except one particular polarisation state. How do we determine whether or not a device is a linear polariser? The law which provides us with the necessary tool is Malus' law.

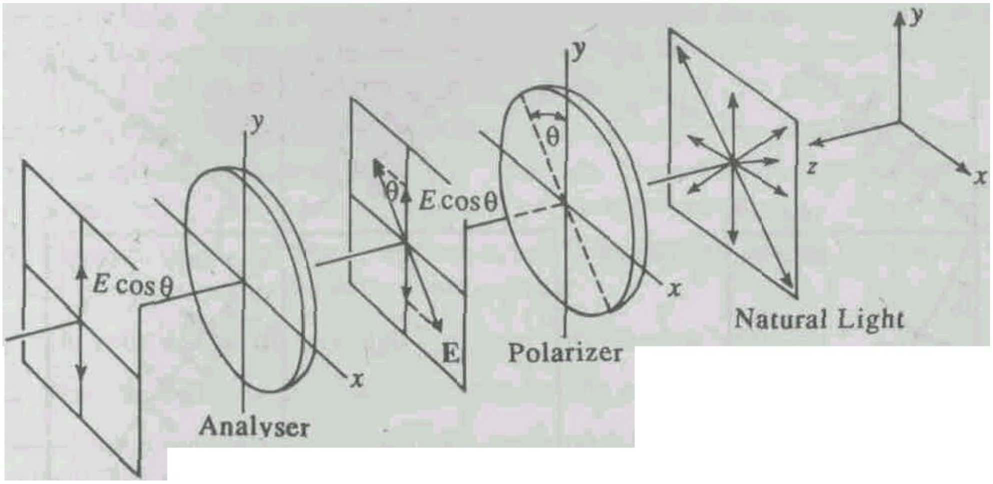


Fig. 4.8 A linear polariser

3.3.1 Ideal Polariser: Malus' Law

Refer to Fig. 4.8. Unpolarised light is incident on an ideal polariser, whose transmission axis makes an angle θ with the y-axis. For this arrangement, only a P-state parallel to the transmission axis of the polariser will be transmitted. This light is incident on an identical ideal polariser, called analyser, whose transmission axis is vertical. Suppose that there is no absorption of light. Then, if **E** is the electric field transmitted by the polariser, only its component $E\cos\theta$ parallel to the transmission axis of the analyser would pass through. The intensity of the polarised light reaching the detector is given by

$$I(\theta) = I(0)\cos^2\theta \qquad (4.15)$$

where θ is the angle between the transmission axes of the polariser and the analyser. The maximum intensity $I(0)$ occurs when the transmission axis of the polariser and the analyser are parallel.

Eq. (4.15) constitutes what is known as **Malus' law.** To use it to check whether an optical device is an ideal linear polariser or not, you may like to solve an SAE.

SELF ASSESSMENT EXERCISE 3

Unpolarised light falls on two polarising sheets placed one over another. What must be the angle between their transmission axes if the intensity of light transmitted finally is one-third the intensity of the incident light? Assume that each polarising sheet acts as an ideal polariser.

So far we have confined ourselves to a linear ideal polariser. Polarisers are available in several configurations. (We can have circular or elliptical polarisers as well.) They are based on one of the following physical mechanisms: reflection, birefringence or double refraction, scattering and dichroism or selective absorption. You will now learn about some of these in detail.

3.3.2 Polarisation by Reflection: Brewster's Law

Reflection of light from a dielectric like plastic or glass is one of the most common methods of obtaining polarised light. You may have noticed the glare across a windowpane or the sheen on the surface of a billiard ball or book jacket. It is due to reflection at the surface and the light is partially polarised. To understand its theoretical basis we will consider laboratory situations.

> This effect was studied by Malus. One evening he was examining a calcite crystal while standing at the window of his house. The image of the Sun was reflected towards him from the windows of Luxembourg Palace. When he looked at the image through the calcite crystal, he was amused at the disappearance of one of the

Suppose that an unpolarised light wave is incident on an interface between two different media at an angle θ_i as shown in Fig. 4.9.

The reflection coefficients when the electric vector of the incident wave is perpendicular to the plane of incidence or when it lies in the plane of incidence are given by Fresnel's equations (Eqs. (2.2la) and (2.21c)):

$$R_{\perp} = \frac{\tan^2(\theta_i - \theta_r)}{\tan^2(\theta_i + \theta_r)} \tag{4.16a}$$

and

$$R_{\perp} = \frac{\sin^2(\theta_i - \theta_r)}{\sin^2(\theta_i - \theta_r)} \tag{4.16b}$$

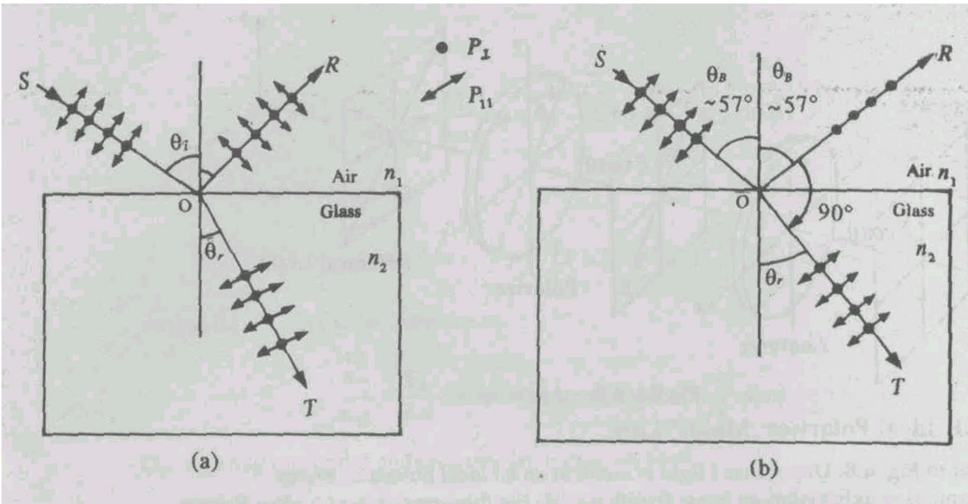


Fig. 4.9: (a) Polarisation by reflection: the unpolarised light beam has been represented as $\leftarrow * \rightarrow$ which indicate two electric field vibrations. '*' indicates electric field vibration perpendicular to the page ($\mathbf{P}_\perp$) and '$\leftrightarrow$' indicates electric field vibration in the plane of the paper ($\mathbf{P}_\parallel$). (b) At Brewster's angle, the reflected light is plane polarised

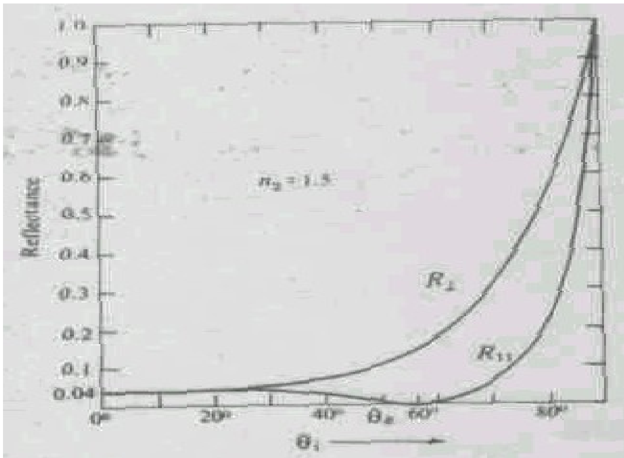


Fig. 4.10 Variation of reflectance with angle of Incidence

where θ_r is the angle of refraction. These equations show that whereas $R_\perp$ can never be zero, $R_\parallel$ will become zero when $\theta_i + \theta_r = \frac{\pi}{2}$. (The case $\theta_i = \theta_r$ is trivial as it implies continuity of optically identical media.) That is, there will be no reflected light beam with **E** parallel to the plane of incidence. The angle of incidence for which light is completely transmitted is called **Brewster's angle.** Let us denote it by θ_B. A plot of $R_\perp$ and $R_\parallel$ versus θ_B is shown in Fig. 4.10 for the particular case of an air-glass interface.

We can represent an incoming unpolarised light as made up of two orthogonal, equal amplitude P-states with the electric field vector parallel and perpendicular to the plane of incidence. Therefore, when the unpolarised wave is incident on an interface and the angle of incidence is equal to the Brewster's angle, the reflected wave will be linearly polarised with **E** normal to the incident plane. This provides us with one of the most convenient methods of producing polarised light. To elaborate, we recall from Snell's law that

$$n_1 \sin\theta_B = n_2 \sin\theta_r$$

where n_1 and n_2 are the refractive indices of the media at whose interface light undergoes reflection. Since $\theta_r = \frac{\pi}{2} - \theta_B$, it readily follows that

$$n_1 \sin\theta_B = n_2 \cos\theta_B$$

or

$$\tan\theta_B = \frac{n_2}{n_1} \qquad (14.17)$$

That is, the tangent of Brewster's angle is equal to the ratio of the refractive indices of the media at whose interface incident light is reflected. When the incident beam is in air ($n_1 = 1$) and the transmitting medium is glass ($n_2 = 1.5$), the Brewster angle is nearly 56°. Similarly, θ_B for an air-water interface, like the surface of a pond or a lake is 53°. This means that when the sun is 37° above the horizontal, the light reflected by a calm pond or lake should be completely linearly polarised.

We, however, encounter some problems in utilizing this phenomenon to construct an effective polariser on account of two reasons:

(i) The reflected beam, although completely polarised, is weak.

(ii) The transmitted beam, although strong, is only partially polarised.

These shortcomings are overcome using a **pile of plate** polarisers. You can fabricate such a device with glass plates for the visible, silver chloride plates for the infrared, and quartz for the ultraviolet region. It is an easy matter to construct a crude arrangement of this sort with a dozen or so microscope slides (Fig. 4.11). The beautiful colours that appear when the slides are in contact is due to interference.

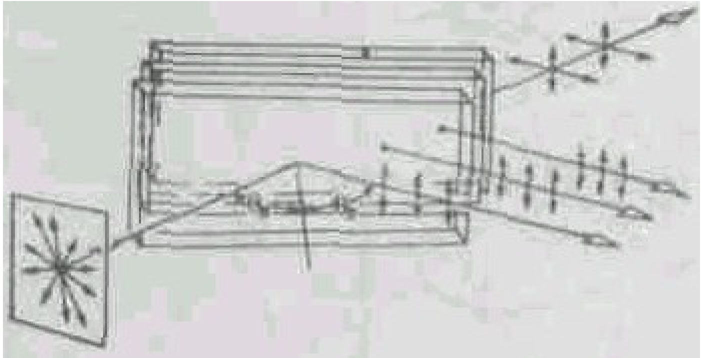


Fig. 4.11 Polarisation of light by a pile of plates

You may now like to solve an SAE.

SELF ASSESSMENT EXERCISE 4

A plate of flint glass is immersed in water. Calculate the Brewster angles for internal as well as external reflection at an interface.

Having studied as to how reflection of light can be used to produce polarised light, you may be tempted to know whether or not the phenomenon of refraction can also be used for the same? Refraction of light in isotropic crystals like NaCl or non-crystalline substances like glass, water or air does not lead to polarisation of light. However, refraction in crystalline substances like calcite or cellophane is optically anisotropic because it leads to what is known as double refraction or birefringence. This is because anisotropic crystals display two distinct principal indices of refraction, which correspond to the E-oscillations parallel and perpendicular to the optic axis. Let us now learn how birefringence can be used to produce polarised light.

3.3.3 Polarisation by Double Refraction

Mark a **black dot** on a piece of paper and observe it through a glass plate. You will see only one dot. Now use a calcite crystal. You will be surprised at the remarkable observation: instead of one, **two grey dots** appear, as shown in Fig. 4.12. Further rotation of the crystal will cause one of the dots to remain stationary while the other appears to move in a circle about it. Similarly, if you place a calcite crystal on your book, you will see two images of each letter. It is because the calcite crystal splits the incident light beam into two beams. This phenomenon of splitting of a light beam into two is known as double refraction or birefringence. Materials exhibiting this property are said to be birefringent. We bring you the excitement of Bartholinus, who discovered birefringence, in his words:

Greatly prized by all men is the diamond, and many are the joys which similar treasures bring, such as precious stones and pearls ... but he, who, on the other hand, prefers the knowledge of unusual phenomena to these delights, he will, I hope, have no less joy in a new sort of body, namely, a transparent crystal, recently brought to us from Iceland, which perhaps is one of the greatest wonders that nature has produced. As my investigation of this crystal proceeded there showed itself a wonderful and extraordinary phenomenon: objects which are looked at through the crystal do not show, as in the case of other transparent bodies, a single refracted image, but they appear double.

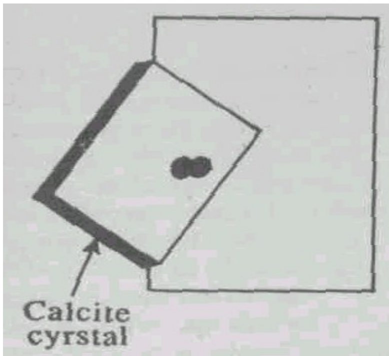


Fig. 4.12 Double refraction of a light beam by calcite crystal

> In some of the textbooks, you may find that ordinary and extraordinary rays are being denoted by bold letters O and E. We have used small letters (o-and e-) to avoid confusion with the notation for the electric field.

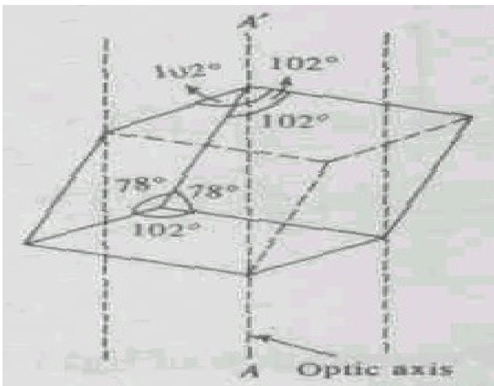


Fig. 4.13 A Calcite crystal. The line AA' shows the direction of the optic axis. For the calcite crystal, the direction of the optic axis is determined by joining the two blunt corners of the crystal

Before we discuss polarisation of light by double refraction in detail, you should familiarise yourself with some of the concepts related to this phenomenon. The two refracted beams into which incident light splits have different angles of refraction. The distinguishing feature of these two refracted light beams is that one of these obeys the Snell's law. It is called the ordinary ray (o-ray) in accordance with the nomenclature given by Bartholinus. The other beam does not obey Snell's law and is called the extraordinary ray (e-ray). That is, a birefringent crystal displays two distinct indices of refraction. Another important concept is that of the optic axis, which signifies some special direction in a birefringent crystal along which two refractive indices are equal (i. e. both o-and e-rays traval in the same direction with the same velocity). When unpolarised light is incident perpendicular to these special directions, both the o-and the e-rays travel in the same direction with different velocities. You may now like to know: Does optic axis refer to any particular line through the crystal? The answer to this question is: It refers to a direction. This means that for any given point in the crystal, an optic axis may be drawn which will be parallel to that for any other point. For example, AA' and broken lines parallel to AA' show the optic axis for a calcite crystal as shown in Fig. 4.13.

Birefringent crystals which posses only one optic axis are called uniaxial crystals. Similarly, crystals having two optic axes are called biaxial crystals. Calcite, quartz and ice are examples of uniaxial crystals and mica is a biaxial crystal. Most of the polarisation devices are made of uniaxial crystals. Further, the uniaxial crystal for which the refractive index o-ray (n_o) is more than the refractive index for the e-ray (n_e) is called negative uniaxial crystal. On the other hand, if $n_e > n_o$, we have a **positive** uniaxial crystal. Values of n_0 and n_e for some of the birefringent crystals are given in Table 4.1. The difference $\Delta n = n_e - n_0$ is a measure of birefringence.

Table 4.1: Refractive indices of some uniaxial birefringent crystals for light of wavelength 5893 Å

Crystal	n_o	n_e
Tourmaline	1.669	1.638
Calcite	1.6584	1.4864
Quartz	1.5443	1.5534
Sodium Nitrate	1.5854	1.3368
Ice	1.309	1.313

Let us now enquire how unpolarised light incident on a uniaxial crystal gets polarised? We know that when unpolarised light beam enters a calcite crystal, it splits into the o-and the e-rays. The electric field vector of e-ray vibrates in the plane containing the optic axis and the electric field vector of o-ray vibrates perpendicular to it, as shown in Fig. 4.14. We may, therefore, conclude that due to double refraction, the unpolarised light beam splits into two components which are **plane polarised.**

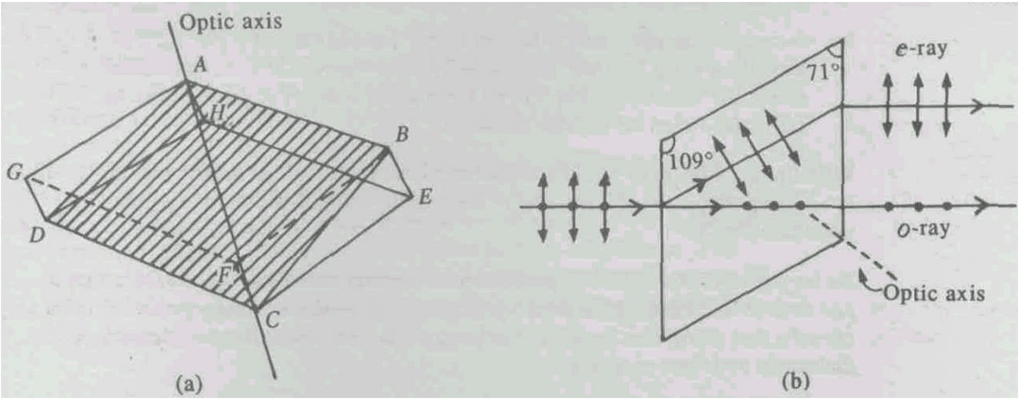


Fig. 4.14: (a) *ABCD is* one of the principal sections of the calcite crystal; it contains the optic axis and is normal to the cleavage faces *BECF* and *AHDG.* (b) Unpolarised light beam passing through a principal section of the calcite crystal.

Huygens explained many aspects of double refraction in calcite on the basis of the wave theory. Since the o-ray obeys Snell's law, it propagates with uniform velocity in all directions in the crystal. As a result, the wave surfaces are spherical. However, the e-ray propagates with different velocities in different directions in the crystal and hence the resulting wave surface is an ellipsoid of revolution, i.e., a spheroid. Further, to reconcile with the fact that both the o-and e-rays travel with the same velocity along the optic axis, both the wave surfaces were assumed to touch each other at the two extremities of the optic axis. These features are depicted in Fig. 4.15. You may now like to know the nature of wave surfaces for o-and e-waves in positive uniaxial crystals. This is subject matter of TQ 1.

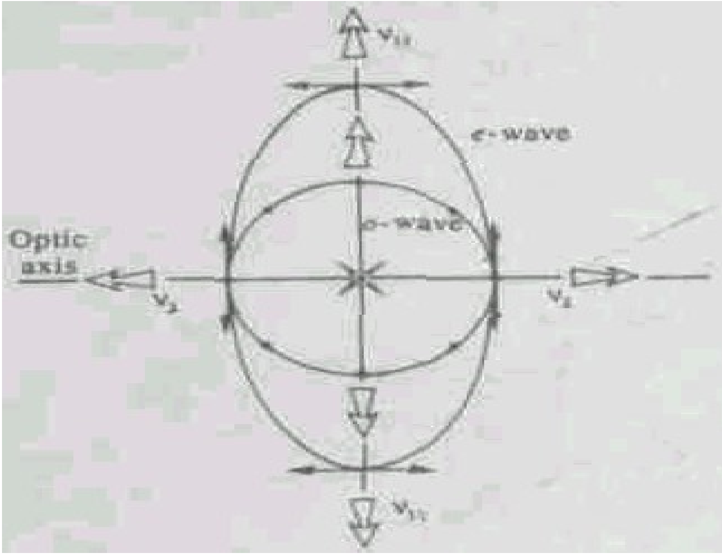


Fig. 4.15 o-and e-wave surfaces in negative uniaxial crystal (calcite)

From the above discussion it follows that in double refraction, an unpolarised light wave splits into o-and e-components with their E-vibrations perpendicular to each other. By selective absorption of one of the P-states, we can produce linearly polarised light. This is readily done by a device, called Nicol prism, by removing the o-ray through total internal reflection. It was designed by William Nicol in 1828. You will learn about it now.

Nicol prism

The Nicol prism is made from a naturally occurring crystal of calcite. The length of the crystal is three times its width and the smaller faces *PQ* and *RS* and ground from 71° to a more acute angle of 68° (Fig. 4.16). The crystal is then cut along PS by a plane passing through *P* and S and perpendicular to the principal section *PQSR.* The cut surfaces are polished to optical flatness and then cemented together with a layer of (nonrefringent material) Canada balsam.

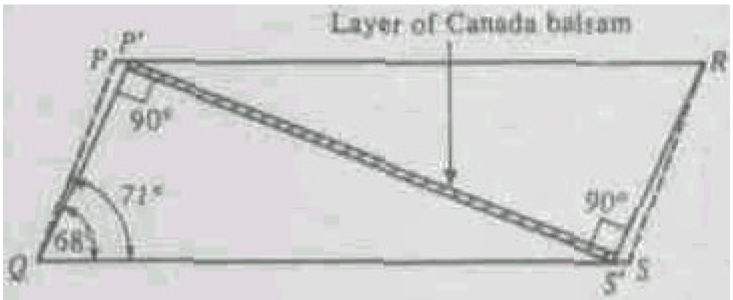


Fig. 4.16 Nicol Prism

Can you guess why Canada balsam is used as cementing material? Well, for sodium light, the refractive index of Canada balsam is 1.552, which is midway between the refractive indices for o-ray (n_o = 1.658) and the e-ray (n_e = 1.486) in calcite. Thus, it is an optically rarer medium with respect to an ordinary ray and denser for an extraordinary ray. The critical angle for total internal reflection of o-ray is $\sin^{-1}\dfrac{1.552}{1.658} = 69^0$. So, when incident unpolarised light splits into two rays inside the crystal, the o-ray gets totally reflected at the Canada balsam surface when it is incident on it at an angle of 69°. (It is for this reason that the end faces of the crystal are ground so as to make the angles 68° from 71°.) The emergent light will, therefore, be made up only of plane polarised e-component.

Some of the limitations of Nicol prism as polariser are:

i. It can be used for polarisation of visible light only.
ii. e-ray also can get totally reflected by the Canada balsam surface if it is travelling along the optic axis. Why? It is so because in this situation the refractive index for e-ray will be same as for o-ray (i.e., greater than the refractive index for Canada balsam).

With time, a number of modifications have been incorporated in the basic design of the Nicol prism to overcome some of these limitations. However, we will not go into these details.

So far you have studied about the production of linearly polarised light by reflection and double refraction. Other methods employed to produce linearly polarised light are selective absorption (or dichroism) and scattering. We will here discuss only dichroism and that too in brief.

3.3.4 Selective Absorption: Dichroism

As you know, unpolarised light wave can be regarded as made up of two orthogonal, linearly polarised waves. Many naturally occurring and man made materials have the property of selective absorption of one of these; the other passes through without much attenuation. This property is known as **dichroism.** Materials exhibiting this property are said to be **dichroic materials.** The net result of passing an unpolarised light through dichroic material is the production of linearly polarised light beam. A particularly simple dichroic device is the so-called Wire-Grid polariser. You will learn about it now.

The Wire-Grid Polariser

The wire-grid polariser constists of a grid of parallel conducting wires, as shown in Fig. 4.17. Suppose that unpolarised light is incident on the grid from the right. It can be thought as made up of two orthogonal P-states; P_x and P_y in the reference plane R_z. The y-component of the electric field drives the electrons of each wire and generates a current. It produces (Joule) heating of the wire. The net result is that energy is transferred from the field to the wire grid. In addition, electrons accelerating along the y-direction radiate in the forward as well as backward directions. The incident wave tends to be cancelled by the wave re-radiated in the forward direction. As a result, transmission of the y-component of field is almost blocked. However, the x-component of the field is essentially unaltered as it propagates through the grid and the light coming out of the wire-grid is linearly polarised. The wire-grid polariser almost completely attenuates the P_y component when the spacing between the wires is less than or equal to the wavelength of the incident wave. You must realise that this restriction is rather stringent for the fabrication of a wire-grid polariser for visible light ($\lambda \sim 5 \times 10^{-7}$ m).

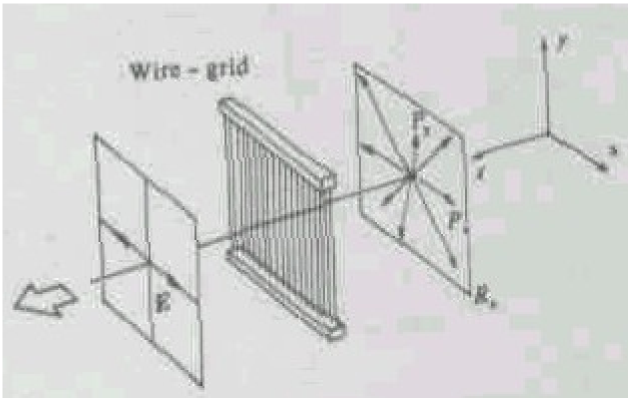


Fig. 4.17 The Wire-grid Polarizer

An easy way out of this difficulty in the fabrication of the grid polariser is to employ lone chain polymer molecules made up of atoms which provide high electrical conductivity along the length of the chain. These chains of polymer molecules behave similar to the wires in the wire-grid polariser. The alignments of these chains are almost parallel to each other. Because of high electrical conductivity, the electric vector of unpolarised light parallel to the chain gets absorbed. And the P-state perpendicular to these chains passes through. These chemically synthesized polarisers are fabricated in the form of plastic sheets and are known as **polaroids.** Since the spacing between these molecular chains in a polaroid is small compared to the optical wavelength, such polaroids are extremely effective in producing linearly polarised light.

Dichroic Crystals

Some naturally occurring crystalline materials are inherently dichroic due to anisotropy in their structure. One of the best known dichroic materials is **tourmaline**, a precious stone often used in jewellery. Tourmalines are essentially boron silicates of differing chemical composition .The component of **E** perpendicular to the principal axis is strongly absorbed by the sample. The thicker the crystal, the more complete will be the absorption. A plate cut from a tourmaline crystal parallel to its optic axis acts as a linear polariser. This is illustrated in Fig. 4.18.

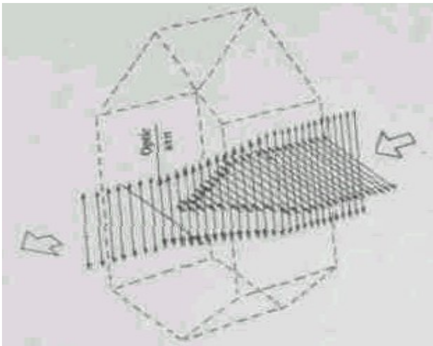


Fig. 4.18 Tourmaline crystal polariser

We shall now consider a class of optical elements known as wave plates which serve to change the polarisation of the incident wave. A wave plate introduces a phase lag between the two P-states by a predetermined amount. That is, the relative phase of the two emerging components is different from its initial value. This concept can be used to convert a given polarisation state into any other and in so doing it is possible even to produce circular or elliptic polarisation as well. This is the subject matter of the next section.

3.4 Wave Plates: Circular and Elliptic Polarisers

Consider a plane wave incident on a calcite crystal. It splits into o-and e-waves. Since calcite is a negative uniaxial crystal, $n_o > n_e$ and $v_{\parallel}$ (velocity of e-wave) $> v_{\perp}$ (velocity of o-wave) implying that the e-ray travels faster than the o-ray. After traversing the calcite crystal of thickness d , the path difference between them is given by

$$\Delta = d(n_o - n_e)$$

> In case of positive uniaxial crystals, $n_e > n_o$ and hence the path difference will be $d(n_e - n_o)$. In fact the general expression for the path difference is $d(|n_e - n_o|)$.

and the relative phase difference between o- and e-rays is

$$\delta = \frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}(n_o - n_e)d \qquad (4.18)$$

though while entering, both the components were in phase.

The **state of polarisation** of the emerging light depends on δ, apart from the amplitudes of incoming orthogonal field components. Let us now consider some specific cases:

(i) When the phase difference, $\delta = 2m\pi$, where m is an integer, the relative path difference is $m\lambda$. A device which induces a path difference between the two orthogonal field vibrations in integral multiples of λ is called the full wave plate. It introduces no observable effect on the polarisation of the incident beam. That is, the field vibrations of the emergent light will-be identical with the field vibrations of the incident light.

(ii) When $\delta = (2m+1)\pi$, the relative path difference will be $\left(m + \frac{1}{2}\right)\lambda$. Such crystals are called **half-wave plates.**

(iii) When $\delta = (2m+1)\frac{\pi}{2}$, the relative path difference will be $\left(m + \frac{1}{2}\right)\frac{\lambda}{2}$. Such a birefringent sheet is called **quarter-wave plate.** When linearly polarised light traverses a quarter-wave plate, the emergent light will, in general, be elliptical and the axes of the ellipse will coincide with the privileged directions of the thin plate. However, half-wave or full-wave plate leave the state of polarisation unchanged.

Thus, we may conclude that the **path difference between the o- and e-waves in a birefringent device depends on its thickness.**

You should now solve the following SAE.

SELF ASSESSMENT EXERCISE 5

Calculate the thickness of a quarter wave-plate for light of wavelength 5890 Å. The refractive indices for o - and e -rays are 1.55 and 1.50 respectively.

We now conclude what you have learnt in this unit.

4.0 CONCLUSION

Polarisation is related to the orientation of associated fields. When electric vector oscillates to the direction of propagation, the light is said to be plane polarised. There is no polarisation of longitudinal waves occurs as the vibrations are along the line of transmission only. Hence, the transverse nature of light is responsible for their polarisation. It is also mathematically analyse how superposition of two plane polarised light waves of same frequency moving in the same direction gives rise to linearly circularly or elliptically polarised light. According to Malu's law

$$I(\theta) = I(o)\cos^2\theta$$

Where θ is the angle between the transmission axes of the polariser and the analyser.

The maximum intensity I(o) occurs when the transmission axis of the polariser and the analyser are parallel. The angle of incidence for which light is completely transmitted is called Brewster's angle θ_B.

$$\tan\theta_B = \frac{n_2}{n_1}$$

The tangent of Brewster angle is equal to the ratio of the refractive indices of the media at whose interface incident light is reflected. Light propagation in anisotropic crystals and phenomenon of birefringence are also discussed. The technique of producing circularly and elliptically polarised light are also mentioned.

5.0 SUMMARY

- Visible light can be linearly, circularly or elliptically polarised. All these polarisation states arise on superposition of two linearly (or plane) polarised light waves characterised by different amplitudes and phases.
- The electric field vectors of two linearly polarised light beams propagating along z-axis can be represented as

$$\mathbf{E}_1(z,t) = \hat{\mathbf{e}}_x E_{01}\cos(kz - \omega t)$$
$$\mathbf{E}_2(z,t) = \hat{\mathbf{e}}_y E_{02}\cos(kz - \omega t)$$

  where E_{01} and E_{02} are the amplitudes of the two waves and ϕ is the phase difference between them. Superposition of these two polarised waves will result in

Linearly polarised light if $\phi = 0$ or an integral multiple of $\pm\ 2\pi$.
Circularly polarised light if $\phi = \pi/2$ and $E_{01} = E_{02}$
Elliptically polarised light if $\phi = \pi/2$ and $E_{01} \neq E_{02}$

- According to Malus, when the transmission axes of polariser and the analyser are at an angle θ, the intensity of the polarised light reaching the detector is given by $I(\theta) = I(0)\cos^2\theta$, where $I(0)$ is the intensity of the polarised light when $\theta = 0$.
- When natural light strikes an interface at Brewster's angle $\theta_B = \tan^{-1}(n_2/n_1)$, where n_1 and n_2 are the refractive indices of medium of incidence and transmission, the reflected light is linearly polarised.
- When light falls on a calcite crystal, it splits into two. The phenomenon is known as double refraction or birefringence. These two refracted beams are known as o- and e-rays. Snell's law holds for o-rays (ordinary rays).
- In a birefringent material, the o- and the e-rays travel in the same direction with same velocity along the optic axis. However, in a direction perpendicular to the optic axis, they travels with different velocities. The electric field vibrations for o- and the e-rays are mutually perpendicular.
- The phenomenon of double refraction produces linearly polarised light. Nicol prism works on this principle. In the Nicol prism, the o-ray undergoes total internal reflection at the interface and the transmitted beam consists of only electric field vibrations corresponding to e-ray and hence the transmitted beam is linearly polarised.
- Selective absorption (or dichroism) of the electric field component with particular orientations by material can also be used for producing linearly polarised light. Tourmaline is an example of dichroic material.
- For a calcite crystal of thickness *d* the path difference between o- and e- rays is given by $\Delta = \mathrm{d}(|\mathrm{n_o} - \mathrm{n_e}|)$.

  The corresponding phase difference

  $$\delta = \frac{2\pi}{\lambda}\Delta = \frac{2\pi}{\lambda}d(|n_o - n_e|)$$

  When the phase difference $\delta = 2m\pi$, where *m* is an integer, the relative path difference between the o- and e-rays will be $m\lambda$. Such crystals are called full-wave plates. When $\delta = (2m+1)\pi$, the path difference will be $\lambda/2$ and such a crystal acts as a full-wave plate. And when $\delta = (2m+1)\lambda/2$, the path difference will be $\lambda/4$ (for *m* = 0) and such a crystal is called a quarter-wave plate.

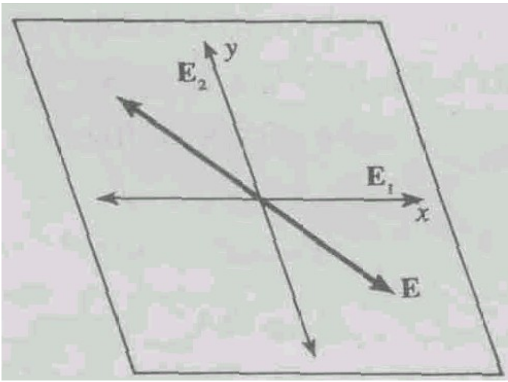
6.0 TUTOR MARKED ASSIGNMENT

1. In sub-section 4.4.3, you studied the propagation of o- and e-waves in a negative uniaxial crystal (calcite). Draw a diagram and describe the propagation of o- and e-waves in a positive uniaxial crystal (quartz) for normal incidence.

2. For a certain crystal, $n_o = 1.5442$ and $n_e = 1.5533$ for light of wavelength 6×10^{-7} m. Calculate the least thickness of a quarter-wave plate made from the crystal for use with light of this wavelength.

SOLUTIONS TO SELF ASSESSMENT EXERCISE

i. The plane of vibration of the electric vector defined by Eq. (4.5) is rotated with respect to that shown in the Fig. 4.5. This is signified by the negative sign before $\hat{\mathbf{e}}_y$ in the parentheses and is depicted below.



ii. We know from Eq. (4.12) that

$$\left(\frac{E_2}{E_{02}}\right)^2 + \left(\frac{E_1}{E_{01}}\right)^2 - 2\left(\frac{E_2}{E_{02}}\right)\left(\frac{E_1}{E_{01}}\right)\cos\phi = \sin^2\phi \qquad \text{(i)}$$

If we choose $\phi = \pi =$ in (i), we get

$$\left(\frac{E_2}{E_{02}}\right)^2 + \left(\frac{E_1}{E_{01}}\right)^2 - 2\left(\frac{E_2}{E_{02}}\right)\left(\frac{E_1}{E_{01}}\right) = 0$$

which can be written in a compact form:

$$\left(\frac{E_2}{E_{02}} - \frac{E_1}{E_{01}}\right)^2 = 0$$

or

$$E_2 = \frac{E_{02}}{E_{01}} E_1$$

This defines a straight line $(y = mx)$ with slope E_{02}/E_{01}. In other words, elliptically polarised light reduces to linearly polarised light for $\phi = n\pi$ ($n = 0, \pm 1, 2, ...$).

When $\phi = \pi/2$ and $E_{01} = E_{02} = E_0$, Eq. (4.12) reduces to

$$\left(\frac{E_2}{E_0}\right)^2 + \left(\frac{E_1}{E_0}\right)^2 = 1$$

which defines a circle $(x^2 + y^2 = a^2)$ of radius E_0.

iii. Since both polarising sheets are ideal, the intensity of the incident unpolarised beam, I, will reduce to half after passing through one of them as shown in the Fig.4.19. After passing through the second polarising sheet, we are told that the intensity reduces to one third of original value.

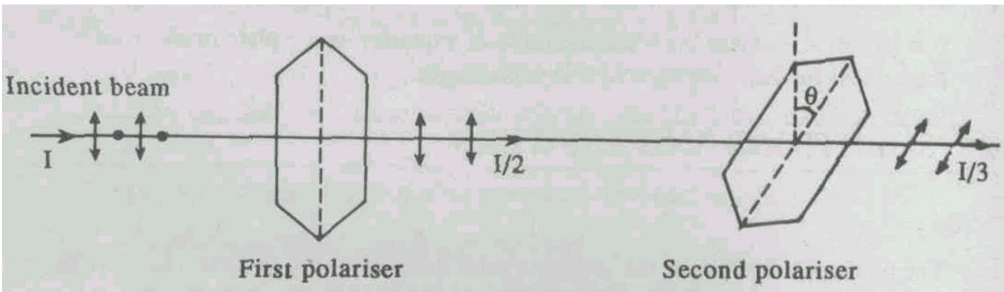


Fig. 4.19 Unpolarised light beam of intensity I passing through two polarisers

From Malus' Law we know that

$$I(\theta) = I(0)\cos^2\theta$$

Here, $I(\theta) = I/3$ and $I(0) = 1/2$.

Therefore, $\cos^2\theta = (2/3) = 0.666$

or $\theta = \cos^{-1}(0.666)^{1/2}$

$= 35.3^0$

That is, the angle between the transmission axes of two polarisers is about 35^0.

iv. For external reflection

$$\tan i_B = \frac{n_2}{n_1} = \frac{1.67}{1.33}$$

$$\Rightarrow \qquad i_B = \tan^{-1}\left(\frac{1.67}{1.33}\right)$$

or $\qquad i_B = 51.47^0$

For internal reflection

$$\tan i_B = \frac{n_1}{n_2} = \frac{1.33}{1.67}$$

$$i_B = 38.53^0$$

v. The path difference produced between the o- and e- rays of birefringent crystal of thickness *d* is

$$\Delta = d(|n_o - n_e|)$$

And the corresponding relative phase difference is given by

$$\delta = \frac{2\pi}{\pi}\Delta$$

$$= \frac{2\pi}{\lambda} d(|n_o - n_e|)$$

The phase difference produced by a quarter-wave plate

$$\delta = \pi / 2$$

On comparing the above expressions for the phase difference, we have

$$d = \frac{\lambda}{4}(n_o - n_e)$$

$$= \frac{5890 \times 10^{-10}}{4}(1.55 - 1.50)$$

$$= 73.63 \text{ Å}$$

$$= 74 \text{ Å}$$

7.0 REFERENCES/FURTHER READINGS

Fundamental to Optics- Jenkins and White

Introduction to Modern Optics, Grant F. Fowles.

Optics; Hecht and Zajac

Optics, Smith and Thomson