FUNDAMENTALS OF GEOMETRIC AND PHYSICAL OPTICS FOR UNDERGRADUATES

Book · May 2015

2 authors:

Chika Christian Onuchukwu
Chukwuemeka Odumegwu Ojukwu University

Ifeyinwa Amaka Ezenwa

CHAPTER ONE
GENERAL INTRODUCTION

1.1 HISTORY ABOUT THE STUDY OF LIGHT

Until the beginning of the seventeenth century, our understanding of the nature and properties of light evolved rather slowly, although, optical phenomena are straightforward to observe. The rectilinear propagation of light, which is most obvious in the shadows cast from a light source, has been known to the Greeks. From this observation, the ancient Greeks developed the concept of straight light rays. Another well-known fact was the equality of the angles of incidence and reflection for light rays, and Hero of Alexandria was able to attribute this law to the more general principle of a shortest path for light. Common optical devices were plane and curved mirrors and lenses.

Concerning the nature of light, different concepts were put forward. The "atomists" following Democritus (about 460–371 BC) believed that all objects consisted of atoms traveling through empty space. In this view, light rays were considered to be a flux of light particles traveling in straight lines and freely through empty space and which could penetrate transparent bodies. The different types of colors were explained by different shapes or sizes of these atoms of light. This can be regarded as the origin of the corpuscular theory of light. In clear contrast to these ideas, was the opinion of Aristotle (384–322 BC), who did not believe in the concept of empty space. For him, light was not a substance or body but a quality, with the primarily visible quality of objects being their color.

Throughout the Middle Ages, the atomistic philosophy was the common discussions. Ibn al Haitham (963–1039 AD) found that the assumption of the proportionality of the angles of incidence and reflection was an approximation that only holds for small angles and a precise description of the functioning of the human eye. Roger Bacon (1215–1294) knew quite well about the properties of lenses and concave mirrors and is regarded as the inventor of the camera, while Salvino degli Armati in 1929 was credited with the invention of spectacles.

From the seventeenth century, efforts for a better understanding of the nature of light saw but a few highlights. The breakthrough in the seventeenth century was initiated by the invention of new optical instruments. In 1600 Zacharias Janssen, built the first microscope, which allowed a glimpse into a world that remained out of reach for the naked eye. The telescope, was invented by the Dutch Hans Lippershey around 1608, Galileo Galilei in 1609 constructed his telescope. In 1611 Johannes Kepler (1571–1630) published the drawings of his telescope using a convex lens ocular. In 1621, Willebrord Snell (1591–1626) found the correct law of refraction. In its present form, the law of refraction was published by René Descartes in 1644 and also found the right explanation for the rainbow. In 1657, Pierre de Fermat derived the law of refraction from the principle of least time (least action), which assumed the propagation of light inside optically dense media to be slower than outside. After the discovery of the law of refraction, mathematicians like Carl Friedrich Gauss (1777–1855), William Rowan Hamilton (1805–1865) and Ernst Abbe (1840–1905) took over and continuously improved the theory of geometrical optics leading to the discovery of many fundamental optical phenomena. Francesco Maria Grimaldi (1618–1663) described the phenomenon of refraction; double refraction was explained in 1669 by Erasmus Bartholinus (1625–1698); and the phenomenon of polarization was discovered by Christian Huygens (1629–1695), who described the effect in 1690. However, in contrast to the phenomenon of double refraction, he was not able to explain it. Olaus Rømer (1644–1710), in 1676 determined the velocity of light from the delayed appearances of the eclipses of the moons of Jupiter during a period of increasing distance between Jupiter and the Earth.

Grimaldi light theory, comes much closer to a wave theory of light. He not only describes the phenomena of diffraction on small objects, including color fringes and the brightening in the center of the shadow, but he also mentions the phenomenon of interference (although he did not use this word) behind two closely neighboring apertures, and the colors appearing in the reflection of light from surfaces with densely spaced grooves. He also observed deviations from straight-line propagation of light and called them "diffractions". It was Robert Hooke (1635–1703) who formulated a real wave theory of light. He considered light as the propagation of a longitudinal wave in ether. He describes the appearance of color in thin films, and sought an explanation in the reflection of light at the front as well as the back side of the film. As he and his contemporaries were lacking the notion of interference, he was not able to complete this explanation. The appearance of color in the refraction of light is explained by wave fronts, which after refraction are no longer parallel to the direction of propagation. In general, colors were described as modifications of white light. Christian Huygens developed the well-known principle now bearing his name, according to which at every point of space the passing light excites elementary waves (or wavelets). Using this picture he derived a simple and convincing explanation of Snell's law of refraction. The most brilliant triumph of this model was the quantitative description of double refraction in calcite by assuming spherically and ellipsoidal shaped wavelets.

Isaac Newton's (1643–1724) is generally considered to be the representative for the corpuscular theory of light. The starting point of his investigations of prismatic colors was the problem of chromatic aberrations in lenses of telescopes. He was convinced that it was impossible to correct the chromatic aberrations of lenses, which led him to the construction of a mirror telescope. In 1672 Newton reported the decomposition of white light. His conclusion was that white light is composed of components of different colors and refractivities (refractive index). These components cannot be further decomposed by a prism and cannot be modified by other means. However, it is possible to compose white light from these components. The colors of bodies can be explained by their varying reflection or absorption of the different colors contained in white light. Newton’s at his time, considered the existence of ether. Although the ether is able to vibrate, we should not, according to Newton’s opinion, identify the nature of light with these ether undulations. The reason Newton mentions this is that for him, the straight propagation of light is not compatible with a wave theory. Indeed, this was a fundamental problem at a time when a mathematical treatment of wave theory, the principle of interference, and, in particular, the method of Fourier analysis were not at hand. It was Fresnel that found a way out of this dilemma using his method of zone constructions. Therefore, Newton saw himself forced to ascribe a corpuscular character to light. He explained the phenomena of refraction and total reflection by assuming that the light particles were dragged into spatial areas of diluted ether. According to his opinion, the velocity of light should be larger in transparent matter as compared to the ether between the bodies. In addition, violet light should consist of smaller particles because they are easier to deviate and therefore refracted more strongly.

In the eighteenth century, there are two important discovery in practical optics: first, the construction of achromatic lenses, which Newton considered to be impossible, first in 1733 by the amateur Chester Moor Hall and again in 1757 by John Dolland; and second, the discovery of stellar aberration by James Bradley in 1725. Stellar aberration refers to the apparent change in the position of a star due to the movement of the Earth, and it was considered to be a confirmation of the corpuscular theory of light, because it found a simple and straightforward explanation within this framework.

The nineteenth century saw the second period of a rapid evolution of optics, at the end of which the nature of light was understood to be a transverse electromagnetic wave phenomenon. In 1801, Thomas Young formulated the principle of interference of waves and explained, as an immediate application, the diffraction of light. Young proposed a wave theoretic explanation of Newton’s rings together with a determination of the wavelengths of light. In 1809 Etienne Louis Malus discovered the polarization of light in the reflections at mirrors, his observations led to the compelling conclusion that light, if really of wave-like character, must be transverse. Essential progress in wave theory is due to Augustin Jean Fresnel (1788–1827). His zone construction solved the long-standing problem of explaining the straight line propagation of light. Together with Dominique Jean Francois Arago (1786–1853), he showed in 1819 that two light rays with perpendicular polarization planes do not interfere. Starting from a theory of transverse waves, not only was Fresnel able to derive the formulas which today bear his name and which allow the exact determination of the intensities for the reflected and refracted parts of light, but also he completed the subject of crystal optics as a theory of propagating transverse waves in anisotropic crystals. The final decision in favor of wave theory was the measurement of the velocity of light in water by Jean Bernard Léon Foucault (1819–1868), which in 1850 definitely proved that the speed of light inside a medium is slower than in the vacuum. Based on the preliminary work of

Michael Faraday (1791–1867), James Clerk Maxwell (1831–1879) derived his fundamental equations of electrodynamics, which imply the existence of transverse electromagnetic waves propagating with a fixed velocity, the velocity of light. The final experimental detection of these waves by Heinrich Rudolf Hertz (1857–1894) in 1888 made optics a branch of electrodynamics. The theory of electrons developed by Hendrik Antoon Lorentz (1853–1928) allowed the explanation of optical properties of matter in terms of electromagnetic concepts. The derivation of Fresnel's equations from electrodynamics is also due to Lorentz. In addition, we owe him essential contributions to the solution of the ether problem. The famous experiment by Albert Abraham Michelson (1852–1931) and Edward Williams Morley (1838–1923) did not reveal any measurable motion of the Earth with respect to the ether, as one would have expected according to the ether hypothesis.

Just when the final victory of the wave theory of light seemed complete, Max Planck (1858–1947) explained, in 1900, the spectral energy distribution of a black body using his quantum hypothesis. In 1905, Albert Einstein (1879–1955) took up the concept of energy quantization and applied it to the hitherto unexplained photoelectric effect. In his interpretation, Einstein went far beyond the ideas of Planck in that he described light as consisting of single energy quanta, so-called photons, thus assigning particle-like properties to light. The development of quantum theory at the beginning of the twentieth century finally led to a deeper understanding of the nature and the properties of light. The discovery of the laser, the advance of the computer, the rapid development of holography and diffractive optics, the development of new materials, in particular materials with special nonlinear properties, as well as the sophistication and expansion of theoretical methods has led to a new revolution of quantum optics during the past few decades. At present, optics can be considered as a particularly strong and growing branch of physics and can be treated as geometric, physical (wave) or quantum optics.

Table 1.1: Important People and Events for the Evolution of Optics.

S/N	Name and Year	Event
1	Euclid (About 300 BC); Ibn Al Haitham (Ad 963–1039)	General Ideas About Optics
2	R. Bacon (1214–1294)	Discovery of The Camera Obscura
3	S. Degli Armati 1299	Discovery of Spectacles
4	Z. Janssen 1600	The First Microscope
5	H. Lippershey 1608; G. Galilei 1609; J. Kepler 1611	Construction of The First Telescope
6	W. Snell 1621	Formulation of The Law of Refraction
7	R. Descartes 1637	Theory of The Rainbow, Law of Refraction
8	P. De Fermat 1657	Derivation of The Law of Refraction, Principle of Temporally Shortest Path of Light
9	F. Grimaldi 1665	Discovery of Diffraction
10	E. Bartholinus 1670	Discovery of Double Refraction

11	C. Huygens 1678/90	Wavelets, Explanation of Double Refraction, Discovery of Polarization
12	R. Hooke 1665	Wave Theory, Colors of Thin Layers
13	I. Newton (1643–1727)	Construction of The Mirror Telescope. Component Theory of White Light, Colors of The Rainbow, Newton's Rings, Polarization, Diffraction, Corpuscular Theory
14	O. Rømer 1676	Measurement of The Speed Of Light
15	J. Bradley 1725/28	Stellar Aberration of The Light f Fixed Stars And Its Explanation
16	C. M. Hall 1733	Construction of Achromatic Lenses
17	J. Dolland 1757	Construction of Achromatic Lenses
18	F. W. Herschel 1800	Discovery of Infrared Radiation
19	J. W. Ritter; W. H. Wollaston 1801	Discovery of Ultraviolet Radiation
20	E. L. Malus 1809	Polarization By Reflection
21	D. Brewster 1815	Brewster's Angle
22	T. Young 1801,1807	Development Of Interferometry, Interpretation Of Light As A Transverse Wave
23	C. F. Gauss (1777–1855)	Geometrical Optics
24	J. Fraunhofer (1787–1826)	Fraunhofer Lines; Development Of Diffraction Theory
25	J. A. Fresnel (1788–1827)	Fresnel's Zone Construction, Fresnel's Equations, Development of Diffraction Theory of light
26	D. J. F. Arago(1786–1853)	Polarization, Color Phenomena, Interference
27	W. R. Hamilton (1805–1865)	Geometrical Optics, Conical Refraction
28	G. R. Kirchhoff (1811–1899); R. W. Bunsen, (1824–1889)	Spectral Analysis, Diffraction Theory
29	H. Fizeau 1849	Terrestrial Measurement of Velocity Of Light
30	J. B. L. Foucault 1850	Measurement of Velocity of Light In Media
31	H. Von Helmholtz (1821–1894)	Theory of Aberrations of Optical Instruments
32	E. Abbe (1840–1905)	Theory of Resolution of Optical Instruments
33	J. C. Maxwell (1831–1879)	Maxwell's Equations, Fundamentals of Electromagnetic Light Theory
34	L. G. Gouy (1854–1926)	Phase Change At Caustics
35	A. Somerfield (1868–1951)	Rigorous Solution Of Diffraction Problems
36	H. Hertz (1857–1894)	Detection of Electromagnetic Waves
37	W. C. Rontgen (1845–1923)	X-Rays Discovery
38	H. A. Lorentz(18531928)	Electron Theory of Optical Properties of Matter
39	M. Planck (1858–1947)	Theory Of Light Quanta
40	A. Einstein (1879–1955)	Theory Of Photon Effect, Special Relativity
41	A. A. Michelson (1852–1931)	Interferometer, Michelson–Morley Experiment
42	D. Gabor (1900–1979)	Holography
43	T. H. Maiman**;** C. H. Townes**; A**. M. Prochorow	For The Development Of Lasers
44	F. Zernike	Phase Contrast Microscope

1.2 LIGHT CAN TRAVEL THROUGH A VACUUM

Many people are confused by the relationship between sound and light. Although we use different organs to sense them, there are some similarities. For instance, both light and sound are typically emitted in all directions by their sources. One way to see that they are clearly different phenomena is to note their very different velocities. The speed of sound, however, can easily be observed just by watching a group of school children a $100\ m$ away as they clap their hands to a song or looking out for thundercloud and lightening, though both happens at the same instant and place, there is a time lag when we saw the light and heard the sound. The fundamental distinction between sound and light is that sound is an oscillation in air pressure, so it requires air (or some other material medium such as water) in which to travel. Today, we know that outer space is a vacuum, so the fact that we get light from the sun, moon and stars clearly shows that material media is not necessary for the propagation of light.

1.3 GEOMETRICAL OPTICS

Geometrical optics, describes light propagation in terms of rays of light. The ray in geometric optics is an abstraction (mathematical formulation) which can be used to approximately model how light will propagate. Light rays are defined to propagate in a rectilinear path (straight line) as they travel in a homogeneous (uniform density and orientation) medium. Rays bend (and may split in two) at the interface between two dissimilar media, may curve in a medium where the refractive index changes, and may be absorbed and reflected. The refraction, reflection, absorption and dispersion of light by the medium through which it propagates are generally referred to as scattering. Geometrical optics provides rules, which may depend on the color (wavelength/frequency) of the ray, for propagating these rays through an optical system. This mathematical simplification of the ray optics fails to account for optical effects such as diffraction and interference. It is an excellent approximation when the wavelength is very small compared with the size of structures with which the light interacts. Geometric optics can be used to describe the geometrical aspects of imaging, including optical aberrations.

As light travels through space, it oscillates in amplitude. In this description of the image caused by the light propagation, each maximum amplitude crest is marked with a plane to illustrate the wave front. The ray is the arrow perpendicular to these parallel surfaces. A light ray is a line or curve that is perpendicular to the light's wave fronts. A slightly more rigorous definition of a light ray follows from Fermat's principle, which states that the path taken between two points by a ray of light is the path that can be traversed in the least time. Geometrical optics is often simplified by making the paraxial approximation, or "small angle approximation." The mathematical behavior then becomes linear, allowing optical components and systems to be described by simple matrices (linear equations). This leads to the techniques of paraxial ray (rays of light parallel to the principle axis) tracing, which are used to find basic properties of optical systems, such as approximate image and object positions and magnifications.

1.4 PROPERTIES OF LIGHT

1.4.1 The Rectilinear Propagation of Light

The rectilinear propagation of light is the terminology given to the principle that light travels in a straight line. This is manifested in the fact that objects can be made to cast shadows. When light is incident on objects that are not transparent to light, the shadow formed could be either total shadow (umbra) or partial shadow (penumbra) depending on whether the light is bright, extended and the size of the object being considered. In the pin-hole camera, used basically for stationary objects, the image is smaller, as the objects gets closer to the pin-hole. Also the image will be sharper, brighter depending on the amount of light leaving the object and the size of the pin-hole.

1.4.2 The Speed of Light

Ancient astronomers and philosophers believe that light travel with an infinite speed. The first person to prove that light's speed was finite, and to determine it numerically, was Ole Roemer, in a series of measurements around the year 1675. Roemer observed Io, one of Jupiter's moons, over a period of several years. From the variation in the period of Io around Jupiter Roemer estimated the speed of light to be approximately $2x10^8\ ms^{-1}$. This is mighty close to the modern measurement of $3x10^8\ ms^{-1}$, bearing in mind his crude method and approximate observations. We now know that light has a finite speed. It was A. A. Michelson that first established the modern speed of electromagnetic wave of all wavelengths. From the radio wavelengths to microwave, infrared, optical, ultraviolet, x-ray and γ $-$ray wavelengths, they all have the same finite speed $2.997925x10^8\ ms^{-1}$, differing only in wavelength (frequency).

1.4.3 The Speed of Light in Stationary Media

The Newton’s particle (corpuscular) theory of light predicted that light like mechanical wave should travel faster in denser media than in less dense media, while Huygens’s wave theory of light implies that light will travel faster in less dense medium. Various experiments by different physicist especially by A. A. Michelson showed that light travels slower in water than in air, and even slower in glass media than in ordinary air. This helped initially in defeating the corpuscular theory of light and the rise of wave theory of light. It is now known that the speed of light in air at normal atmospheric pressure and temperature is about $87x10^3\ ms^{-1}$ less than the speed in vacuum. For many practical purposes, this difference may be neglected and the speed of light in air taken to be ~$3x10^8\ ms^{-1}$.

1.4.4 The Refractive Index

The index of refraction or the refractive index, of any optical medium is defined as the ratio between the speed of light in vacuum and the speed of light in that medium.

$$Refractive\ index\ (\mu) = \frac{speed\ of\ light\ in\ vacuum}{speed\ of\ light\ in\ medium} = \frac{c}{v_m} \qquad 1.1$$

where c is the speed of light in vacuum and v_m is the speed of light in the particular optical medium and usually $\mu > 1$ defined as the absolute refractive index. For glass μ~1.5, for water, $\mu = 1.33$, and for dry air at atmospheric pressure and temperature, $\mu = 1.000292$, which can be approximated to $\mu = 1$ for most practical purposes.

1.4.5 Optical Density

The optical density of any transparent medium is a measure of its refractive index. A medium with relatively high refractive index is said to have a high optical density, while one with a low refractive index is said to have low optical density.

1.4.6 Optical Path

The path d a ray of light travels in any medium in the interval t is $d = vt$, but $v = c/\mu$, thus, we can write

$$d = \frac{ct}{\mu} \quad \rightarrow \quad \mu d = ct \rightarrow \delta = \mu d \qquad 1.2$$

The product μd is called the optical path. Thus, for the same time interval, the optical path of light is the same in all media, $\delta = \mu_1 d_1 = \mu_2 d_2 = \mu_i d_i$, but the path length d_i through each medium differs, due to differences in the velocity of light in the medium. In general, we define the optical path δ of light as the distance light travels in a vacuum in the same time interval it will travel a distance d in the given medium.

1.5 REFLECTION AND REFRACTION OF LIGHT

Smooth, polished and glossy surfaces such as mirrors reflect light in a simple and predictable way. This allows for production of reflected images that can be associated with an actual (real) or extrapolated (virtual) location in space. With such surfaces, the direction of the reflected ray is determined by the angle the incident ray makes with the surface normal, a line perpendicular to the surface at the point where the ray hits the boundary between the two media.

1.5.1 Rays of Light

A ray of light is a mathematical construction to help in the geometric analysis of propagation of light. It may also be considered as a line drawn in space corresponding to the direction of flow of radiant energy (light). Geometrically, it is a straight line with an arrow indicating the path and the direction of the propagation of light. A collection of rays of light is called a beam of light. There are different beams of light (i) Parallel Beam (ii) Convergent Beam (iii) Divergent Beam (iv) Scattered Beam.

1.5.2 Law of Reflection.

i. The incident and reflected rays and the normal all lie in a single plane,

ii. The angle between the reflected ray and the surface normal is the same as that between the incident ray and the normal i.e. the reflection angle is always equal to the incident angle

For flat mirrors, the law of reflection implies that images of objects are (i) upright (ii) the same distance behind the mirror as the objects are in front of the mirror (iii) same as the object size, the magnification of a flat mirror is equal to one (iv) laterally inverted, which is perceived as a left-right inversion

Images formed by mirrors with curved surfaces can be modeled by ray tracing and using the law of reflection at each point on the surface. For mirrors with parabolic surfaces, parallel rays incident on the mirror produce reflected rays that converge at a common focus. Other curved surfaces may also focus light, but with aberrations due to the diverging shapes, causing the focus to be smeared out in space. In particular, spherical mirrors exhibit spherical aberration. Curved mirrors can form images with magnification greater than or less than 1, and the image can be upright or inverted, real or virtual depending on the object's position with respect to the mirror. An upright image formed by reflection in a mirror is always virtual, while an inverted image is real and can be projected onto a viewing screen.

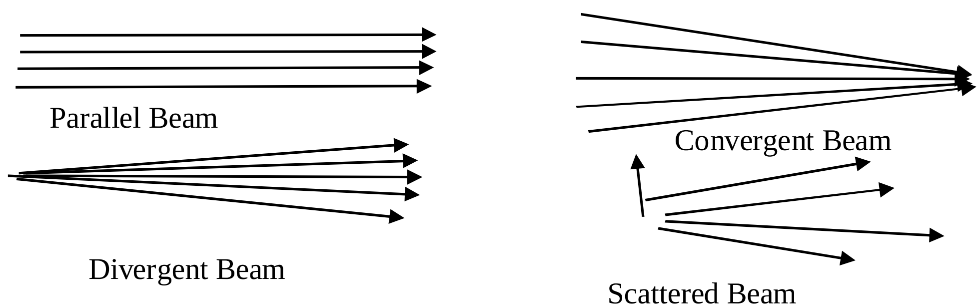


Fig 1.1 Different Beams of Light

1.5.3. Refraction

Refraction occurs when light travels through an area of space that has a changing index of refraction (from one medium to another). The simplest case of refraction occurs when there is an interface between a uniform medium with index of refraction μ_1 and another medium with index of refraction μ_2. In such situations, Snell's Law describes the resulting deflection of the light ray from its original line of propagation as

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_2 \quad 1.3$$

where θ_1 and θ_2 are the angles between the normal (to the interface) and the incident and refracted waves, respectively. This phenomenon is also associated with a changing speed of light as it propagates from one medium to another, as seen from the definition of index of refraction provided above which implies:

$$v_1 \sin\theta_1 = v_2 \sin\theta_2 \quad 1.4$$

where v_1 and v_2 are the light velocities through the respective media

Various consequences of Snell's Law include the fact that for light rays traveling from a material with a high index of refraction to a material with a low index of refraction, it is possible for the interaction with the interface to result in zero transmission. This phenomenon is called total internal reflection and allows for fiber optics technology. As light signals travel down a fiber optic cable, it undergoes total internal reflection allowing for essentially no light lost over the length of the cable. It is also possible to produce polarized light rays using a combination of reflection and refraction: When a refracted ray and the reflected ray form a right angle, the reflected ray has the property of "plane polarization". The angle of incidence required for such a scenario is known as Brewster's angle.

Snell's Law can be used to predict the deflection of light rays as they pass through "linear media" as long as the indexes of refraction and the geometry of the media are known. For example, the propagation of light through a prism results in the light ray being deflected depending on the shape and orientation of the prism. Additionally, since different frequencies of light have slightly different indexes of refraction in most materials, refraction can be used to produce dispersion spectra that appear as rainbows. Some media have an index of refraction which varies gradually with position and, thus, light rays curve through the medium rather than travel in straight lines. This effect is what is responsible for mirages seen on hot days where the changing index of refraction of the air causes the light rays to bend creating the appearance of specular reflections in the distance (as if on the surface of a pool of water). Material that has a varying index of refraction is called a gradient-index (GRIN) material and has many useful properties used in modern optical scanning technologies including photocopiers and scanners.

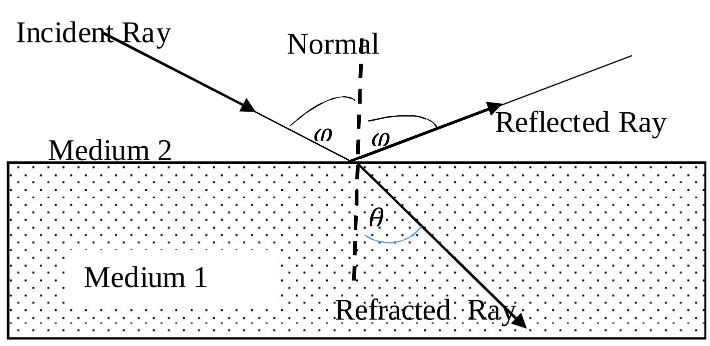


Fig 1.2 Reflection and Refraction of Light

1.6 FERMAT'S PRINCIPLE

The real path of a ray of light through several media with varying refractive index can be written as

$$\delta = \mu^1 d^1 + \mu^2 d^2 + \mu^3 d^3 + \cdots + \mu^r d^r \qquad 1.5$$

where d is the optical path. Fermat's principle is applicable to such type of variation of μ and hence contains within it the laws of refraction and reflection. It is stated as

The path taken by a light ray in going from one point to another through any set of media is such as to render its optical path a minimum.

It is readily shown mathematically that laws of reflection and refraction follow from Fermat's principle. From figure 1.3, the length of the optical path between the point Q in the medium with refractive index μ and the point Q' with refractive index μ' passing any point A is

$$\delta = \mu d + \mu' d', \qquad 1.6$$

where $d^2 = l^2 + (p - x)^2$ and $d'^2 = l'^2 + x^2$. Substituting these in equation (1.6), gives

$$\delta = \mu\sqrt{l^2 + (p - x)^2} + \mu'\sqrt{l'^2 + x^2} \qquad 1.7$$

According to Fermat's principle, the actual path must be a stationary point, thus, differentiating equation (1.7) and setting it equal to zero gives

$$\frac{d\delta}{dx} = \frac{\frac{1}{2}\mu(-2p+2x)}{\sqrt{l^2+(p-x)^2}} + \frac{\frac{1}{2}.2x\mu}{\sqrt{l^2+(p-x)^2}} = 0 \qquad 1.8$$

Simplifying equation (1.8) we have

$$\mu\frac{p-x}{d} = \frac{x}{d'}\mu' \qquad 1.9$$

But $\frac{p-x}{d} = \sin\theta$; $\frac{x}{d'} = \sin\varphi$, making these substitution, gives the required Snell's law.

$$\mu\sin\theta = \sin\varphi\,\mu' \qquad 1.10$$

The proof of law of reflection is obtained by similar diagram, and is left as an exercise for the student.

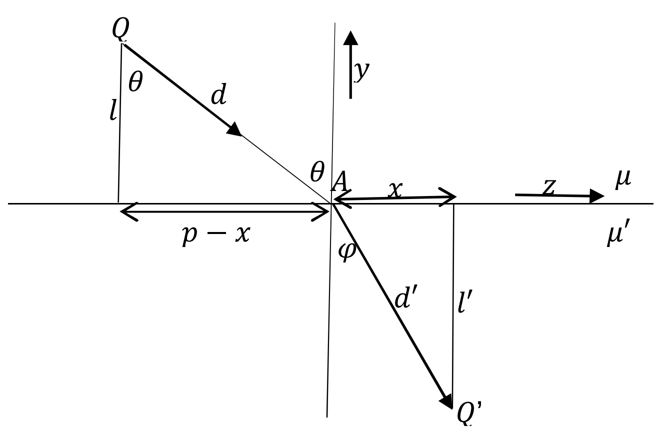


Fig 1.3 Geometry of a Refracted Ray illustrating Fermat's Principle

1.7 THE PRINCIPLE OF REVERSIBILITY OF LIGHT

If a reflected or a refracted ray is reversed in its direction, it will retrace its path. This is true of the passage of light through any optical system, however complicated. It follows, as a simple corollary, that if the refractive index from air to a medium is μ, the refractive index from the medium to air is $1/\mu$. This principle will make the light ray to travel through the same path from air-to-glass or from glass-to-air. If a light is incident on the boundary between air and another medium whose boundary is also parallel to another medium (say air-glass-water), before emerging into the air, the emergent ray is found to be parallel to the incident ray, and thus, the angle of incident is equal to the angle of emergence.

Using the following symbols defined as i_a = incident ray in air, r_a =reflected ray in air, i_g =incident ray in glass, r_g =reflected ray in glass, i_w =incident ray in water and r_w =reflected ray in water. We define the refractive index in the glass-water boundary as

$${}_g\mu_w = \frac{\sin i_g}{\sin r_w} \qquad 1.11$$

From reversibility of light, $i_a \equiv r_a$; and from alternate angle, $i_g = r_g$ and $i_w = r_w$, we can write equation (1.11) as

$${}_g\mu_w = \frac{\sin i_g}{\sin r_w} = \frac{\sin i_g}{\sin r_a} x \frac{\sin i_a}{\sin r_w} = {}_g\mu_a \; x \; {}_a\mu_w = \frac{{}_a\mu_w}{{}_a\mu_g} \qquad 1.12$$

since ${}_g\mu_a = \frac{1}{{}_a\mu_g}$. For instance, if we want to obtain the refractive index of glass-water boundary, we obtain the refractive index of air-water and that of air-glass, using equation (1.12), and the refractive index of glass-water boundary can be estimated. To estimate the refractive index of glass-water boundary, we note that ${}_a\mu_g = 1.5$; ${}_a\mu_w = 1.33 \rightarrow {}_g\mu_w = \frac{1.33}{1.5} = 0.89$. Generally, ${}_1\mu_3 = {}_1\mu_2 = {}_2\mu_3$. Recall that $\frac{\sin i_a}{\sin r_g} = {}_a\mu_g \rightarrow \sin i_a = {}_a\mu_g \sin i_g$, similarly $\sin i_a = {}_a\mu_w \sin i_w$, thus in general,

$${}_a\mu_g \sin i_g = {}_a\mu_w \sin i_w = \mu \sin i = constant \qquad 1.13$$

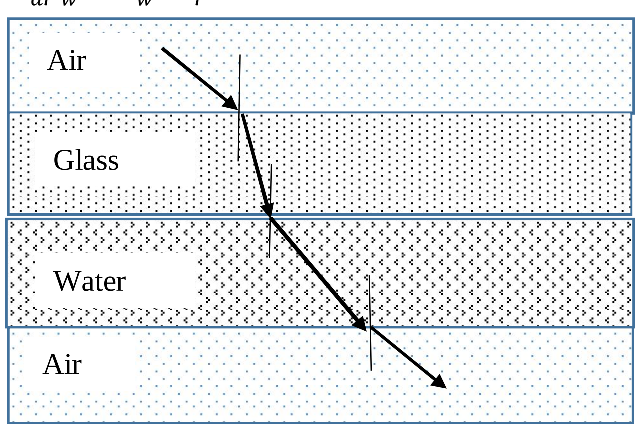


Fig 1.4. Illustrating the principle of Reversibility of Light

1.8 DISPERSION OF LIGHT

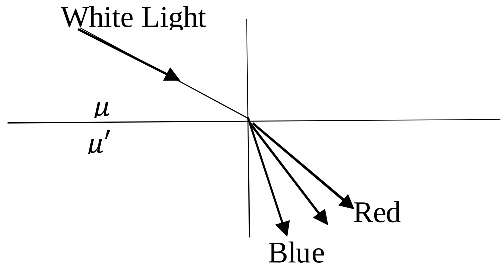


Fig 1.5 Dispersion of White Light into its spectrum upon Refraction

When white light is incident on a refractive medium, refracted rays of different colors (separation into spectrum) are formed. Each color has different value of refraction angle (φ), this implies that the refractive index of the medium varies with each color (i.e. the speed with which different colors of white light propagates in a medium varies, usually blue light travels faster than red light in dense media).

1.9 SOURCES OF LIGHT

Sources of light are called luminous object, while non-luminous objects do not emit light. Substances that allows the transmission of light through them are called transparent objects, while substances that totally absorb light are called opaque objects, they do not allow the transmission of light through them. Some other substances like light tinted glass partially allow the transmission of light and are called translucent objects.

Here we list various sources of light, both natural and artificial sources and processes and devices that emit light. In this list, light is considered to be electromagnetic radiation that is visible to the human eye, and not blackbody radiation in the more general sense. The list is also limited to sources of light, as opposed to objects such as the Moon that provide light by reflection.

i. Astronomical objects - Sun, Starlight (Stars forming groups such as star clusters and galaxies. Deep sky objects including quasars, accretion discs around black holes, blazars, magnetars, pulsars, Supernova/nova, Milky Way, Cosmic rays.
ii. Atmospheric entry (via ionization and/or heating) - Meteors, Meteor showers, Bolide/Fireball, Earth-grazing fireball
iii. Lightning (Plasma physics) – lightning of various forms, Dry Lightning, Aurorae, Cerenkon radiation (from cosmic rays hitting atmosphere).
iv. Terrestrial – (a) Bioluminescence - Glowworms, Fireflies, and certain bacteria, Antarctic Krill, Parchment Worm, Foxfire, luminescent fungus etc. (b) Incandescence – Lava, Volcanic, Volcanic Eruption (lightning, heated material). (c) Radioluminescene (d) Triboluminescence (e) Piezolumniescence (f) Earthquake Light

v. Nuclear/High Energy – Sonoluminescence, Annihilation, Bremsstrahlung, Scintillation Cyclotron, Synchrotron, Cerenkov, etc.

vi. Direct Chemical – Chemoluminescence, Luminol, Florescence, Phosphorescence, Chemical explosives, Combustion Fires, etc.

vii. Incandescent lamps – Carbon Button Lamp, Conventional Incandescent lamps, Flashlights, Halogen Lamps, etc.

viii. Electron-Stimulated – Cathodluminescence, Cathode Ray Tubes, Electron Stimulated Luminescence, Crookes' Tube, Electroluminescent Lamps – (a) Light Emitting Diodes (LED) (b) Organic LEDs, (c) Polymer LEDs, (d) Solid State LEDs, (v) Light Emitting Electrochemical Cells (LECs), Electroluminescent Sheets, Electroluminescent Wires, Field Induced Polymer Electroluminescent (FIPEL), etc.

ix. Gas Discharge Lamps – Induction Lightning, Florescent Lamps – Compact Florescent Lamps, Tanning Lamp, Wood's Lamp, Geissler Tube, Moore Tube, Hollow-Cathode Tube, Excimer Lamp, Neon, Argon and Xenon Lamps, Nixie Tube, Plasma Lamp, etc.

x. Lasers – Ruby Laser Gas Laser, Semiconductor Laser, Chemical Laser, Dye Laser, Metal-Vapor Laser, Solid State Laser, Ion Laser, Quantum-Well Laser, Free-Electron Laser, Gas Dynamic Laser, etc.

Exercises One

1. Draw a ray diagram showing why a small light source (a candle, say) produces sharper shadows than a large one (e.g., a long fluorescent bulb).
2. A Global Positioning System (GPS) receiver is a device that lets you figure out where you are by receiving timed radio signals from satellites. It works by measuring the travel time for the signals, which is related to the distance between you and the satellite. By finding the ranges to several different satellites in this way, it can pin down your location in three dimensions to within a few meters. How accurate does the measurement of the time delay have to be to determine your position to this accuracy?
3. Estimate the frequency of an electromagnetic wave whose wavelength is similar in size to an atom (about a nm). Referring back to your electricity and magnetism text, in what part of the electromagnetic spectrum would such a wave lie (infrared, gamma-rays,)?
4. The Stealth bomber is designed with a smooth surfaces. Why would this make it difficult to detect via radar?
5. A man is walking at $1.0\ ms^{-1}$directly towards a flat mirror. At what speed is his separation from his image decreasing?
6. If a mirror on a wall is only big enough for you to see yourself from your head down to your waist, can you see your entire body by backing up?

7. Differentiate between luminous, opaque, translucent and transparent object. Give examples of each.
8. What is a difference between refraction and diffraction of light?
9. When light travels through different media, what is the differences between the path length and the optical length?
10. A ray of light transverse through a medium of refractive index $\mu = 1.56$, through a distance of 20 m, what is the optical path and the path length?
11. Explain why:
    a. An empty test-tube plunged in a beaker of water may present a silvery appearance.
    b. Fish are more easily seen from a bridge than from the bank of the river.
    c. Mirages occur frequently in hot deserts.
    d. Stars twinkle

CHAPTER TWO
REFLECTION AND REFRACTION AT PLANE SURFACES AND CURVED SURFACES

2.1 INTRODUCTION

The behaviors of beam of light upon reflection or refraction at a plane or curved surfaces are of great importance in geometrical optics. Plane surfaces and curved surfaces often occur in nature, e.g. the cleavage surfaces of crystals, diamond surfaces, surfaces of liquids, naturally shaped gem stones etc. Artificial planes or curved surfaces are used in various optical instruments to bring about deviations or lateral displacements of rays to create images as well as to break light into its colors to study their intensities and energy distributions.

Images

Images are formed due to real or apparent intersection of rays of light. There are two types of images (i) Real Images (ii) Virtual Images. Real Images are formed by the real intersection of two or more rays of light. They can be focused on a screen. Virtual images are formed by the apparent intersection of light rays, they cannot be focused on a screen. In complex optical instruments consisting of more than one optical device, the image of one device (whether real or virtual) forms the object of the second device. For example, the virtual image formed by an eyeglass is the object that the eye focused on to form an image.

2.2 REFLECTION AT PLANE MIRROR

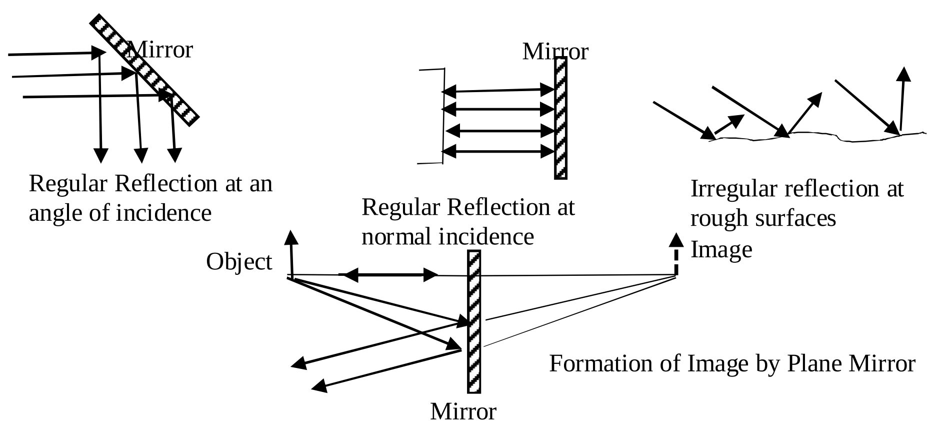


Fig 2.1 Reflection At Plane Surfaces And Image Formation

When one surface of a flat piece of glass is coated with highly reflecting material (like polished silver, gold), almost all the light on the opposite surface is reflected. Such coated glasses are called plane mirrors. Mirrors, smoothened surfaces, polished surface all produce regular reflection, since a parallel beam of light is reflected in one direction. Unpolished and rough surfaces, produce diffused, scattered or irregular reflection.

Applications of plane mirror include (i) cosmetic – use as looking glass (ii) driving mirror (iii) optical instruments – like Sextant, periscope, kaleidoscope.

When an object is viewed by reflection in a plane mirror, the apparent position of the image is determined by the direction in which the reflected rays enters the eye. The image so formed is (i) virtual (ii) erect (iii) formed behind the mirror (iv) same size as the object (v) as far behind the mirror as the object is in front of the mirror (vi) laterally inverted.

2.2.1 Inclined Mirrors

When an object is placed between two inclined mirrors, an observer can see several images whose number depends on the angle between the mirrors. The kaleidoscope is an optical instrument in which multiple images are formed by two mirrors usually inclined at $60°$ to each other. If the angle between the two mirrors is $60°$, five images are formed, if $90°$ three images are formed, when parallel, an infinite number of images are formed, at $45°$, the number of images is 7 images, at $30°$, we have 11 images. The number of images increases as the angle of inclination decreases.

A simple mathematical relation to estimate the number N of images formed by two inclined mirror at an angle θ is given by $N = \frac{360}{\theta} - 1$. This is applicable when, $\frac{360}{\theta}$ is an integer, else we apply $N = \frac{360}{\theta}$ and round up to the nearest integer value. When the two mirrors are parallel, the angle of inclination is 0, and theoretically, an infinite number of images would be observed.

2.2.2 Rotation of Reflected Ray

If a fixed ray is incident on a plane at an angle of incidence i, the angle of reflection $r = i$, so that the angle between the two rays will be $2i$. If the mirror is rotated through an angle θ, the angle of incidence will increase by θ, and so also will the angle of reflection. The total angle between the two rays is now $2i + 2\theta$, consequently, the angle of reflection will rotate through an angle of 2θ.

Consider a ray of light incident on a plane mirror at an angle of incidence i, before rotation. Form the law of reflection $i = r$. The mirror is now rotated through angle θ, the new angle of incidence will be $i' = i + \theta$, while the new angle of reflection will be $r' = i - \theta + \varphi$, where φ is the angle through which the reflected ray is rotated. Applying the law of reflection, $i + \theta = i - \theta + \varphi \rightarrow \varphi = 2\theta$. Thus, the reflected ray rotates through twice the angle of angle through which the mirror rotates. This principle is applied in the use of mirror galvanometer to detect very small currents, where the current flowing through cause the mirror to rotate and the angle of reflection through which the mirror rotates gives an indication of the amount of current flowing

2.3 REFLECTION AT CURVED MIRRORS

Basically there are three types of curved mirrors (i) Concave Mirror (ii) Convex Mirror (iii) Parabolic Mirror. Curved mirrors are made by depositing vaporized aluminum on a glass which is part of a sphere/parabola.

2.3.1 Formation of Images by Spherical Mirrors

In tracing ray diagram to obtain the image formed by spherical mirrors, four important rays are used, the intersection of any two of such rays, describes the nature and position of the image formed by the object. The rays are

i. All Paraxial rays are reflected from the mirror through the principal focus.
ii. All rays passing through the principal focus are reflected from the mirror parallel to the principal axis
iii. Ray incident on the pole making an angle θ with the principal axis, is reflected from the mirror with an angle θ from the principal axis.
iv. Ray of light passing through the center of curvature is reflected un-deviated from the mirror.

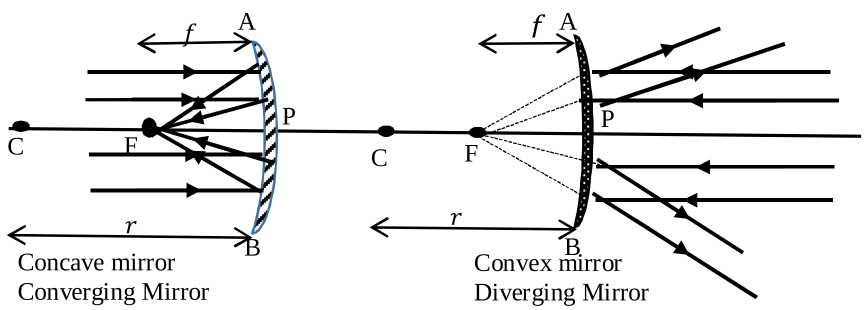


Fig 2.2 Spherical Mirrors. Note the side of the mirror that is polished

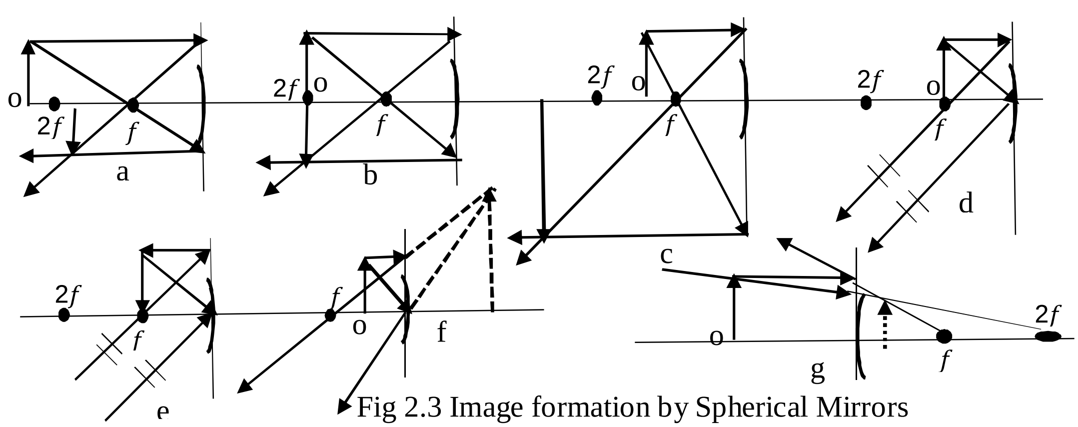


Fig 2.3 Image formation by Spherical Mirrors

The images formed by convex mirror is always virtual, diminished, and erect and is formed behind the mirror (see figure 2.3g). For concave mirror, the nature of image formed depends on the position of the object with respect to the mirror (see figure 2.3 a-f). ‘o’ represents the object.

a. Object beyond $2f$ – image is real, inverted, diminished, formed between f and $2f$.
b. Object at $2f$ – image is real, inverted, same size as the object, formed at $2f$.
c. Object between $2f$ and f – image is real, inverted, magnified, formed at beyond $2f$.
d. Object at f – image is formed at infinity.
e. Object at infinity – image is real, inverted diminished and formed at f.
f. Object between f and mirror – image is virtual, erect, magnified, at is formed behind the mirror.

Application of concave mirror include its usage in reflecting telescopes, shaving mirrors, oculars, etc., while convex mirror are basically used as driving mirror due to their large field of view.

2.3.2 Mirror Formula

If an object is at a distance u from a curved mirror (concave or convex), of focal length f, and radius r, and the image is formed at a distance v form the mirror, it can be shown that (we will later develop the Lens maker’s equation), the distances are related by

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} = \frac{2}{r} \qquad 2.1$$

The magnification m is defined as (here h_I is the image height and h_O is the object height)

$$m = \frac{h_I}{h_O} = \frac{v}{u} = \frac{v}{f} - 1 \rightarrow \frac{1}{m} = \frac{u}{f} - 1 \qquad 2.2$$

In using the mirror formula, it is necessary to adhere to a particular convention – (i) Real is positive convention (ii) Cartesian convention (iii) Newtonian Convention (iv) Gaussian Convention. In this book, we shall adhere to real is positive convention, with the following rules

i. Real objects and images are assigned positive values. Virtual images are assigned negative values
ii. Converging systems with real focal length (concave mirror and convex lens) have their focal length assigned positive values. Diverging systems with virtual focal length (convex mirror and concave lens) have their focal length assigned negative values.

Example 2.1

An object is placed (i) $15\ cm$, (ii) $5\ cm$ in front of a concave mirror of radius of curvature $20\ cm$. Calculate the position, nature and magnification of the image in each case.

Solution

(i) $f = \frac{20}{2} = 10\ cm$, from $f = \frac{2}{r}$, $u = 15\ cm$; $\frac{1}{v} = \frac{1}{10} - \frac{1}{15} = \frac{1}{30} \rightarrow v = 30; m = \frac{v}{u} = \frac{30}{10} = 3$

Image is real, magnified and inverted

(ii) $u = 5\ cm; \frac{1}{f} = \frac{1}{u} + \frac{1}{v} \rightarrow \frac{1}{v} = \frac{1}{10} - \frac{1}{5} = -\frac{1}{10} \rightarrow v = -10; m = \frac{v}{u} = \frac{-10}{5} = -2$

Image is virtual, magnified and erect.

2.4 REFRACTION AT PLANE SURFACES

The apparent shortening of a spoon placed inside a glass of water; the apparent reduction in the depth of swimming pools; the rise in the letters of a word when a block of glass is placed over the letter, etc. are clear manifestation of refraction of light as it propagates from one medium optical density to another. When light rays propagate from optical less dense medium to denser medium, it is refracted towards the normal, but if it propagates from optical denser to less dense it is refracted away from the normal.

2.4.1. Apparent Depth

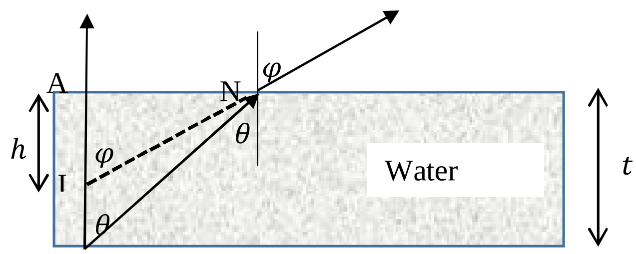


Fig 2.4 Apparent Depth

Usually the bottom of a pool of water appears nearer to the surface than it actually is. This is due to refraction of light when it passes from water to air. Using the principle of reversibility of light, the refractive index of water is

$$\mu = \frac{\sin\varphi}{\sin\theta} = \frac{\frac{AN}{NI}}{\frac{AN}{NO}} = \frac{NO}{NI} = \frac{AO}{AI} = \frac{t}{h} = \frac{real\ depth}{apparent\ depth} \qquad 2.3$$

We have used similar triangles for the equality $\frac{NO}{NI} = \frac{AO}{AI}$. The reduction in depth τ is given by

$$\tau = t - h = t - \frac{t}{\mu} = t\left(1 - \frac{1}{\mu}\right) \qquad 2.4$$

2.4.2 Critical Angle and Total Internal Reflection

When a ray of light passes from air to glass to air, at the air-glass boundary, some of the incident ray will be reflected, within the glass some will be absorbed and some will be transmitted but refracted toward the normal at the air-glass boundary but away from the normal at the glass-air boundary. As the angle of incidence is increased, the angle of refraction also increases. At a particular angle of incidence, the angle of refraction will be $90°$ i.e. the direction of the refracted ray will be parallel to the glass-air boundary. This angle of incidence is called the critical angle. Further increase in the angle of incidence will cause all the rays to be reflected – total internal reflection and none transmitted. Total internal reflection occurs when

i. Light is propagating from optically dense medium to less dense medium

ii. The critical angle is exceeded

At critical angle, incidence angle $i = i_c$, where i_c is the critical angle, and refracted angle $r = 90°$, thus using Snell's law (but noting that light is travelling from glass to air)

$$\frac{1}{\mu} = \frac{\sin i_c}{\sin 90} \rightarrow i_c = \sin^{-1} i_c \qquad 2.5$$

Total internal reflection is really total in the sense that theoretically no energy is lost upon reflection. However, practically, small losses occur due to absorption in the medium and to reflection at the surfaces where light enters and leave the medium. The glittering of diamond is due to total internal reflection. Other devices include the total reflecting prism (Porro prisms), usually employed to change the direction of light without loss to the energy/intensity of the light. Optical fibers employ total internal reflection to transport electromagnetic signals from one point to another without loss of signal strength.

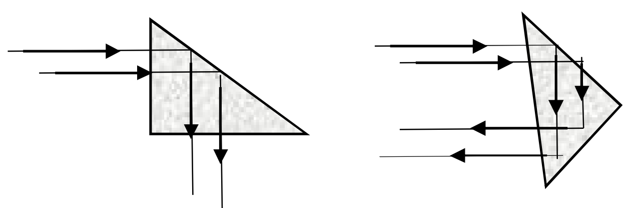


Fig 2.5 Reflecting Prisms using the Principle of Total Reflection

2.5 MIRAGE

A mirage is a naturally occurring optical phenomenon in which light rays are bent to produce a displaced image of distant objects of the sky as it propagates from denser medium to less dense medium. Cold air is denser than warm air and, therefore, has a greater refractive index. As light travels at a shallow angle along a boundary between layers of air at different temperature, the light rays bend towards the colder air. If the air near the ground is warmer than that higher up, the light ray bends upward, effectively being totally internal reflected just above the ground. Mirages can be categorized as Inferior, Superior and Fata Morgana.

The resting state of the Earth's atmosphere has a vertical temperature gradient of about $-1° \, Cm^{-1}$ of altitude. The value is negative because it gets colder as altitude increases. For a mirage to happen, the temperature gradient has to be much greater than that, the magnitude of the gradient needs to be at least $-2° \, Cm^{-1}$, and the mirage does not get strong until the magnitude reaches to about $-4° Cm^{-1}$to $-5° \, Cm^{-1}$. These conditions do occur with strong heating at ground level, for example when the sun has been shining on sand or asphalt, leading to the generation of an inferior image.

2.5.1 Inferior Mirage

Once the rays reach the viewer's eye, it is interpreted as if it traces back along a perfectly straight line of sight. However, this line is at a tangent to the path the ray takes at the point it reaches the eye. The result is that an "inferior image" of the sky above appears on the ground. The viewer may incorrectly interpret this sight as water, but it is actually the reflected image of the sky, which is, to the brain, a more reasonable and common occurrence. For exhausted travelers in the desert, an inferior mirage may appear to be a lake of water in the distance. An inferior mirage is called "inferior" because the mirage is located under the real object, other examples include the hot road mirage. Inferior images are not stable. Hot air rises, and cooler air (being more dense) descends, so the layers will mix, giving rise to turbulence. The image will be distorted accordingly. If there are several temperature layers, several mirages may mix, perhaps causing double images.

2.5.2 Superior Mirage

A superior mirage occurs when the air below the line of sight is colder than the air above it. This unusual arrangement is called a temperature inversion, since warm air above cold air is the opposite of the normal temperature gradient of the atmosphere. Passing through the temperature inversion, the light rays are bent down, and so the image appears above the true object, hence the name superior. Superior mirages are in general less common than inferior mirages, but, when they do occur, they tend to be more stable, as cold air has no tendency to move up and warm air has no tendency to move down. Superior mirages are quite common in Polar Regions, especially over large sheets of ice that have a uniform low temperature. Superior mirages also occur at more moderate latitudes, although in those cases they are weaker and tend to be less smooth and stable. For example, a distant shoreline may appear to *tower* and look higher (and, thus, perhaps closer) than it really is.

2.5.3 Fata Morgana

A Fata Morgana, is a very complex superior mirage. It appears with alternations of compressed and stretched zones, erect images, and inverted images. A Fata Morgana is also a fast-changing mirage. Fata Morgana mirages are most common in Polar Regions, especially over large sheets of ice with a uniform low temperature. While in Polar Regions, a Fata Morgana may be observed on cold days and, over oceans and lakes; in desert areas, a Fata Morgana may be observed on very hot days. For a Fata Morgana, temperature inversion has to be strong enough that light rays' curvatures within the inversion are stronger than the curvature of the Earth.

2.5.4 Mirage of Astronomical Objects

A mirage of an astronomical object is a naturally occurring optical phenomenon, in which light rays are bent to produce distorted or multiple images of an astronomical object. The mirages might be observed for such astronomical objects as the Sun, the Moon, the planets, bright stars, and very bright comets. The most commonly observed are sunset and sunrise mirages.

2.6 RAINBOW

A rainbow is an optical and meteorological phenomenon that is caused by reflection, refraction and dispersion of light in water droplets resulting in a spectrum of light appearing in the sky. It takes the form of a multicolored arc. Rainbows caused by sunlight always appear in the section of sky directly opposite the sun. Rainbows can be full circles; however, the average observer sees only an arc formed by illuminated droplets above the ground, and centered on a line from the sun to the observer's eye. Rainbows can be caused by many forms of airborne water – rain, mist, spray, and airborne dew. In a primary rainbow, the arc shows red on the outer part and violet on the inner side. This rainbow is caused by light being bent when entering a droplet of water, then reflected inside on the back of the droplet and refracted again when leaving it. In a double rainbow, a second arc is seen outside the primary arc, and has the order of its colors reversed, red facing toward the other one in both rainbows. This second rainbow is caused by light reflecting twice inside the water droplets.

A rainbow is not located at a specific distance from the observer, but comes from an optical illusion caused by any water droplets viewed from a certain angle relative to a light source. Thus, a rainbow is not an object and cannot be physically approached. Indeed, it is impossible for an observer to see a rainbow from water droplets at any angle other than the customary one of $42°$ from the direction opposite the light source. Light rays enter a raindrop from one direction (typically a straight line from the sun), reflect off the back of the raindrop, and fan out as they leave the raindrop. The light leaving the rainbow is spread over a wide angle, with a maximum intensity at the angles $40.89°–42°$. Between 2 and 100% of the light is reflected at each of the three surfaces encountered, depending on the angle of incidence.

2.7 REFRACTION THROUGH A PRISM

In a prism the two surfaces are inclined at some angle α called refraction angle, so that the deviation produced by the first surface is not annulled by the second but it further enhanced. This leads to chromatic dispersion of white light. Let d represent the total angle of deviation, we note the following equality from simple trigonometry (i) $\alpha + \beta = 180°$ - Opposite angle of a quadrilateral. (ii) $\beta + \gamma = 180°$ - Angle in a straight line, $\rightarrow \alpha = \gamma$ (iii) $\alpha = r_1 + r_2$ – External angle of a triangle is equal to the two interior opposite. Deviation of ray at the first surface $d_1 = i_1 - r_1$, and at the second surface, $d_2 = i_2 - r_2$, total deviation is given by

$$d = d_1 + d_2 = i_1 - r_1 + i_2 - r_2 = (i_1 + i_2) - (r_1 + r_2) = (i_1 + i_2) - \alpha \quad 2.6$$

Experiments show that as the angle of incidence is increased, the angle of deviation decreases until it reaches a minimum value d_{min} before it starts to increase again as i is increased to $90°$. To obtain the minimum deviation, we write from Snell's law

$$(\sin i_1 + \sin i_2) = \mu(\sin r_1 + \sin r_2) \quad 2.7$$

Using trigonometric relation equation (2.7) may be written as

$$\sin\left(\frac{i_1+i_2}{2}\right) = \mu \frac{\sin\left(\frac{r_1+r_2}{2}\right)\cos\left(\frac{r_1-r_2}{2}\right)}{\cos\left(\frac{i_1-i_2}{2}\right)} \quad 2.8$$

But $\alpha = r_1 + r_2$; $d = i_1 + i_2 - r_1 - r_2 \rightarrow \alpha + d = i_1 + i_2$ 2.9

Thus, equation (2.8) becomes

$$\sin\left(\frac{\alpha+d}{2}\right) = \mu \frac{\sin\left(\frac{\alpha}{2}\right)\cos\left(\frac{r_1-r_2}{2}\right)}{\cos\left(\frac{i_1-i_2}{2}\right)} \quad 2.10$$

For minimum deviation to occur, light must past symmetrically through the glass prism, which implies that $i_1 = i_2$; $r_1 = r_2$. Using this condition and re-arranging equation (2.10), we have

$$\mu = \frac{\sin\left(\frac{\alpha+d_{min}}{2}\right)}{\sin\left(\frac{\alpha}{2}\right)} \quad 2.11$$

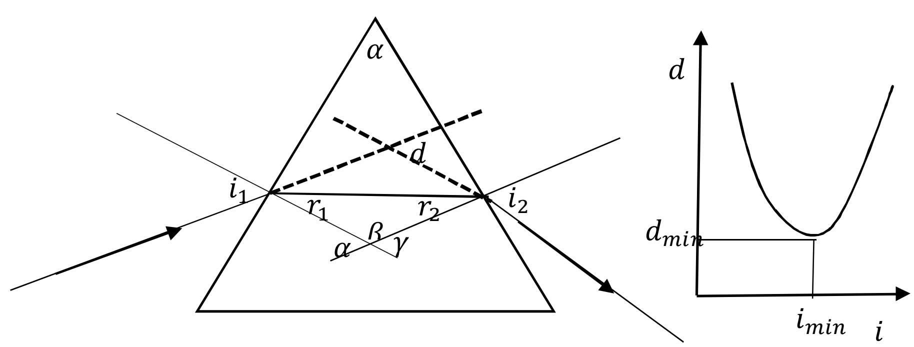


Fig 2.6 The Geometry Associated With Refraction by a Prism and A plot Showing Minimum Deviation

Equation (2.11) gives one of the methods one can use to accurately obtain the refractive index of a glass prism. The equations for the glass prism becomes much simpler when the refractive angle α becomes small enough that its sine and the sines of the angle of deviations d may be set equal to the angles themselves (especially for angular measure in radians). For such prism, equation (2.11) reduces to $= \frac{\alpha+d}{\alpha}$. The subscript in d has been dropped because such prisms are always used at their minimum deviation. It is customary to measure the power of a prism by the deflection it produces in cm at a distance of $1\ m$. The unit of prism power is the prism Diopter, defined as the deflection $1\ cm$ on a screen $1\ m$ away.

2.8 DISPERSION OF WHITE LIGHT BY A PRISM

When white light falls on a glass prism and the emergent beam focused on a screen, different colors made up of Red-Orange-Yellow-Green-Blue-Indigo-Violet are seen. This collection of colors is called the spectrum of white light. Usually the colors are not distinctly separated but overlap, thus impure spectrum is formed. Using two or more glass prisms properly arranged as in a spectrum analyzer produces pure spectrum. For a small angled prism, we can obtain an expression for the deviation. In case of monochromatic light, the refractive index of that color of light is $\mu = \frac{\sin i}{\sin r}$, and for small angle $\sin\theta \approx \tan\theta = \theta$ in radian. This implies we can write $i_j = \mu r_j$; where $j = 1,2,3$ .... The deviation has been defined as

$$d = i_1 - r_1 + i_2 - r_2 = \mu r_1 - r_1 + \mu r_2 - r_2 = (r_1 + r_2)(\mu - 1) = (\mu - 1)\alpha \quad 2.12$$

Equation (2.12) defines the magnitude of the deviation given by a prism for any small angle of incidence. We note that the deviation is independent of the value of the incident angle. The angular dispersion between the emergent red and blue light is defined as the angle between the two rays, and is given by

$$\theta = (\mu_b - 1)\alpha - (\mu_r - 1)\alpha = \alpha(\mu_b - \mu_r) \quad 2.13$$

where μ_b and μ_r is the refractive indices for blue light and red light respectively. For white light, the mean deviation is the deviation of the yellow light, since it is the color approximately in the middle of the spectrum. The mean refractive index of the material is also quoted for yellow light. The amount of dispersion of any material is measured by its dispersive power defined as

$$w = \frac{angular\ dispersion\ between\ blue\ and\ red\ rays}{mean\ deviation} = \frac{d_b - d_r}{\alpha(\mu-1)} = \frac{(\mu_b - \mu_r)}{(\mu-1)} \quad 2.14$$

2.9 FIBER OPTICS

When light in an optically dense medium approaches the boundary of a less dense medium at an angle greater than the critical angle i_c, it is totally internally reflected. Theoretically, there is no loss in energy, this is applicable in fiber optics and cable for the transmission of large data using electromagnetic waves without loss. Bundles of tiny fibers of clear glass arranged in ordered arrays can be used to transmit light over long distances and through bends and corners without loss of data.

2.10 REFRACTION AT CURVED SURFACES – THIN LENSES

Lenses consist of pieces of glass of varying thickness from the middle to the edges, bounded by spherical surfaces on one or both sides. There are various types of lenses – the converging lenses and the diverging lenses, thick or thin lenses. We shall be concentrating on thin lenses, defined as a lens whose thickness is considered small in comparison with the distances generally associated with its optical properties like the radii of curvature, the primary focus and secondary focal lengths, object and image distances. In figure 2.7, we show the diagrams of several standard forms of lenses.

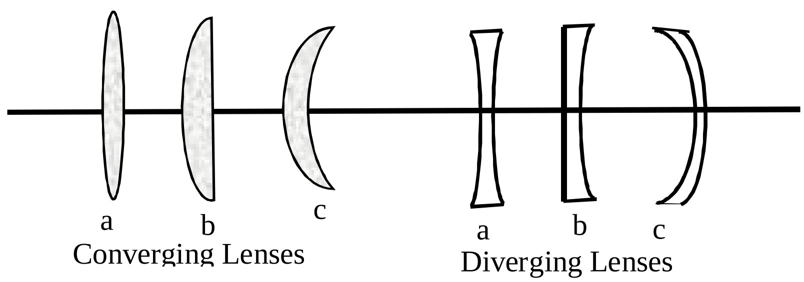


Fig 2.7 Different Types of Lenses

The three converging or positive lenses (convex lenses) which are thicker at the center than at the edges are (a) equiconvex/bi-convex (b) plano-convex (c) positive or converging meniscus. The three diverging or negative lenses (concave lenses) which are thinner at the center than at the edges are (a) equiconcave/bi-concave (b) plano-concave (c) negative or diverging meniscus. These lenses are made of glass, quartz, fluorite, rock salt, and plastics that are free from any form of inhomogeneity.

Thin lenses have two principal focal points. The primary focal point is a point on the principal axis having the property that any ray coming from it or proceeding towards it travels parallel to the principal axis after refraction. The secondary focal point is an axial point having the property that any incident ray travelling parallel to the principal axis will after refraction, proceed towards or appear to come from it. A convex lens has positive focal length, and paraxial rays are refracted converging to the principal focus after passing through the lens. Concave lens has a negative focal length and paraxial rays are refracted seemingly to be diverging from the principal focus from the lens. The optical center is the center of a lens, the distance between the optical center and the principal focus defines the focal length.

2.10.1 Image Formation by Convex Lens

Three particular classes of rays are used in geometrical construction to locate the image formed by a converging lens:

i. Rays parallel to the principal axis pass through the principal focus after refraction by the lens.
ii. Rays through the principal focus emerges parallel to the principal axis after refraction from the lens.
iii. Rays through the optical center appear un-deviated

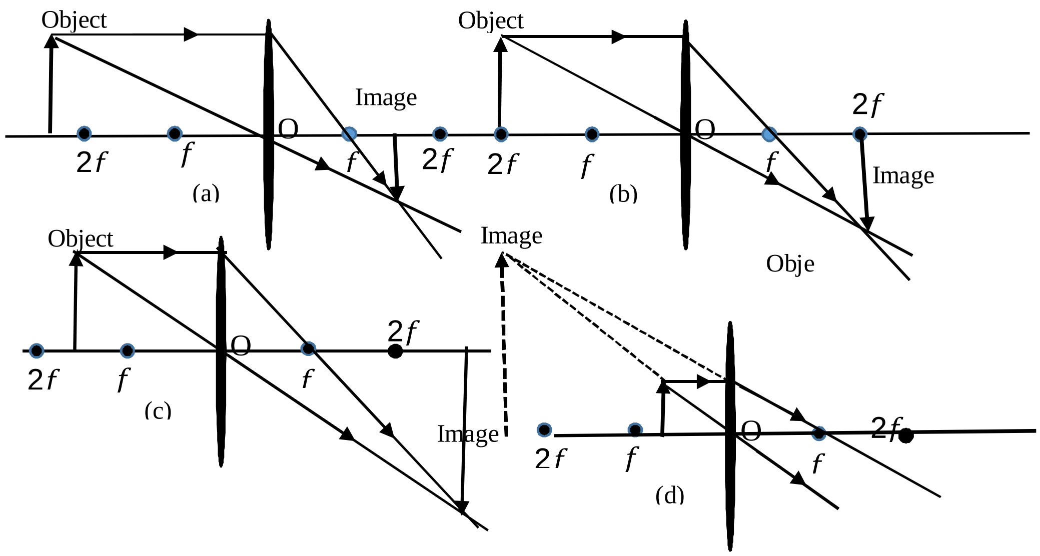


Fig 2.8 Image Formation By Convex

Images formed by convex lens depends on the objects distance with respect to the lens. Figure 2.8 show the formation of images by convex mirror.

i. Object beyond $2f$ – (a) Image is real (b) inverted (c) diminished and (d) formed at the opposite side of the lens between f and $2f$.
ii. Object at $2f$ – (a) Image is real (b) inverted (c) same size as the object and (d) formed at the opposite side of the lens at $2f$.
iii. Object between f and $2f$ – (a) Image is real (b) inverted (c) magnified and (d) formed at the opposite side of the lens beyond $2f$.
iv. Object f – Image is at infinity
v. Object between f and lens – (a) Image is virtual (b) erect (c) magnified and (d) formed at the side of the lens.
vi. Object at infinity – (a) Image is real (b) inverted (c) diminished and (d) formed at the opposite side of the lens at f

For concave lens, the image is always (a) virtual (b) diminished (c) formed between the lens and the object and (d) on the same side of the object.

2.10.2 Relationship between Image and Object Distance for the Lens

The image of the object O can be obtained by refraction at the curve AC. Assuming O is in the medium of refractive index μ_1, a ray OC passes through the medium μ_2. Since

OC is normal to the surface AC, the ray OA, is refracted through the angle i_1 assuming that $\mu_1 < \mu_2$. OC and the ray BI form the image of O at I. If OA and BI make an angle i_1 and $i_2$with the normal AE at A, then we can write using Snell's law

$$\mu_1 \sin i_1 = \mu_2 \sin i_2 \rightarrow \mu_1 i_2 = \mu_2 i_2 \quad i_j \; is \; small \; (radian) \qquad 2.15$$

But $i_1 = \theta_1 + \theta_2$, $i_2 = \theta_2 - \theta_3$ (from Triangular rules). Substituting in equation (2.15), we have

$$\mu_1(\theta_1 + \theta_2) = \mu_2(\theta_2 - \theta_3) \rightarrow \mu_1\theta_1 + \mu_2\theta_3 = \theta_2(\mu_2 - \mu_1) \qquad 2.16$$

Using $\sin\theta = \tan\theta = \theta \; (radian)$ when θ is small, we can write $\theta_1 = \frac{h}{OC}$, $\theta_2 = \frac{h}{DE}$, $\theta_3 = \frac{h}{DI}$, using this in equation (2.16) we have

$$\mu_1\frac{h}{OC} + \mu_2\frac{h}{DI} = \frac{h}{DE}(\mu_2 - \mu_1) = \frac{\mu_1}{u} + \frac{\mu_2}{v} = \frac{\mu_2 - \mu_1}{r_1} \qquad 2.17$$

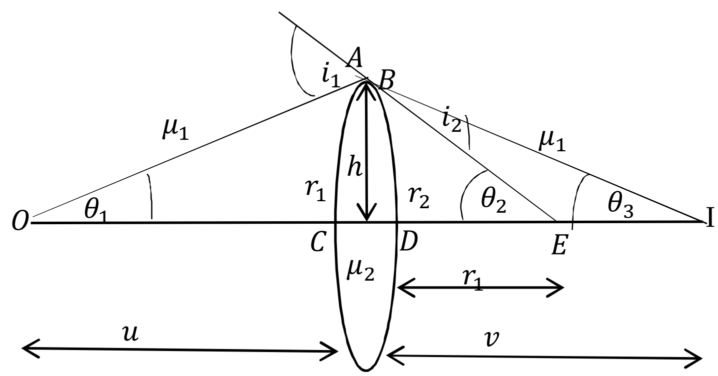


Fig 2.9 Geometry For Lens Formula

Equation (2.17) gives the relation between the object distance and the image distance from the middle of the refracting surfaces if the radius of curvature r and the refractive indices of the media are known. The quantity $\frac{\mu_2-\mu_1}{r}$ is defined as the power of the refractive surface. It is positive for converging systems and negative for diverging system. If we now assume that the lens thickness is negligibly small compared with the object distance, we note that the image distance for the first surface becomes the object distance for the second surface, thus $u = -v$, so we can write

$$\frac{\mu_2}{-v} + \frac{\mu_1}{v} = \frac{\mu_2 - \mu_1}{r_2} \qquad 2.18$$

Adding equations (2.18) and (2.17), we have the general lens formula as

$$\frac{\mu_1}{u} + \frac{\mu_1}{v} = \mu_2 - \mu_1\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \rightarrow \frac{1}{u} + \frac{1}{v} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \qquad 2.19$$

For parallel rays incident on the lens, $u = \infty$, $v = f$, thus we have

$$\frac{1}{f} + \frac{1}{\infty} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \rightarrow \frac{1}{f} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \qquad 2.20$$

Thus in general, comparing equations (2.19) and (2.20)

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \qquad 2.21$$

Equations (2.20) and (2.21) applies to both converging and diverging lens provided that the sign convection are obeyed. Real object and image distances are assigned positive values, while virtual objects and images distances are assigned negative values. Focal length of converging system are positive, while that of diverging system, negative.

Note that the foal length of the lens depends on the refractive index of its material and the refractive index of the medium which it is placed in as well as the radii of the curved surfaces. For air, $\mu_1 = 1$, thus we can write

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \qquad 2.22$$

We have used the absolute refractive index of the lens material and have dropped the subscript. For bi-convex lens r_1, r_2 are positive; for Plano-convex, $r_1 \sim +ve$, $r_2 \sim \infty$, bi-concave, r_1, r_2 are negative.

2.10.3 Combined Focal Length of Two Thin Lenses

Assuming two lenses with focal lengths f_1, f_2 respectively are in contact, then if f is their combined focal length, then we can write

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \qquad 2.23$$

If they are separated by a small distance d, then their combined focal length becomes

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_2 f_1} \qquad 2.24$$

Equation (2.24) holds for any two thin lenses in contact, whatever is the combination – converging and converging, converging and diverging and diverging and diverging systems.

2.10.4 The Power of a Thin Lens

The power of a thin lens in diopters is given as the reciprocal of the focal length in meters.

$$p = \frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \qquad 2.25$$

where r_1 and r_2 are measured in meters. For combined thin lenses in contact, the power of the combined lens is equal to the sum of the power of the individual lenses. Converging system have positive values assigned to the power rating while diverging systems have negative value assigned to its power rating.

Example 2.2

The end of a glass rod of index 1.50 is grounded and polished to a hemispherical surface of radius $1\ cm$. A small object is placed in air on the axis $4\ cm$ to the left of the vertex. Find the image position. Take $\mu_{air} = 1$.

Solution

$$\frac{\mu_{air}}{u} + \frac{\mu_{glass}}{v} = \frac{\mu_{glass} - \mu_{air}}{r} \rightarrow \frac{1}{4} + \frac{1.5}{v} = \frac{1.5 - 1}{1} \rightarrow v = 6.0\ cm$$

Example 2.3

The radii of both surfaces of a bi-convex lens of refractive index 1.60 are equal to 8.0 cm. Find its power.

Solution

$p = \frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \rightarrow p = (1.6 - 1)\left(\frac{1}{0.08} + \frac{1}{0.08}\right) = +15\ Diopter$

Example 2.4

An object is located 6.0 cm in front of a converging lens of focal length 4.0 cm. Describe the image.

Solution

$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \rightarrow \frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{4} - \frac{1}{6} = \frac{1}{12} \rightarrow v = 12\ cm, m = \frac{v}{u} = 2$

The image is real, inverted, magnified and is located beyond $2f$

Example 2.5

If an object is located 6.0 cm in front of a converging lens of focal length 10 cm, where will the image be formed?

Solution

$\frac{1}{v} = \frac{1}{f} - \frac{1}{u} = \frac{1}{10} - \frac{1}{6} = -\frac{2}{30} \rightarrow v = -15.0\ cm,\ m = -\frac{15}{6} = -2.5$

The image is virtual, erect, magnified and is located on the same side of the lens as where the object is located.

Example 2.6

An object is placed 12.0 cm in front of a diverging lens of focal length 6.0 cm. Describe the image formed.

Solution

$\frac{1}{f} = \frac{1}{v} + \frac{1}{v} \rightarrow \frac{1}{v} = -\frac{1}{6} - \frac{1}{12} = -\frac{1}{4} \rightarrow v = -4.0\ cm;\ m = -\frac{4}{12} = -\frac{1}{3}$

The image is virtual, erect, diminished and is located on the same side of the lens as where the object is located, this is the general characteristics of virtual images formed by diverging systems.

Example 2.7

A plano-convex lens having a focal length of 25.0 cm is to be made of a glass $\mu = 1.520$. Calculate the radius of curvature of the lens.

Solution

$r_1 = \infty$ since one surface is plane, its radius of curvature is infinite,

$r_2 = ?, f = 25.0\ cm, \mu = 1.520$

$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \rightarrow \frac{1}{25} = (1.520 - 1)\left(\frac{1}{\infty} - \frac{1}{r}\right) \rightarrow r = -25x0.520 = -13.0\ cm$

Example 2.8

A concave surface with radius of 4 cm separates two media of refractive index $\mu = 1.00, \mu' = 1.50$. An object is located in the first medium at a distance of 10 cm from the vertex. Find (a) the primary focal length (b) the secondary focal length, and (c) the image distance.

Solution

The primary focal length is given by $\frac{\mu_1}{f_1} = \frac{\mu_2-\mu_1}{r_1}$ and the secondary focal length $\frac{\mu_2}{f_2} = \frac{\mu_2-\mu_1}{r_1}$

Combining these two equations gives $\frac{f_2}{f_1} = \frac{\mu_2}{\mu_1}$, by implication one may write

$\frac{\mu_1}{u} + \frac{\mu_2}{v} = \frac{\mu_1}{f_1}; \frac{\mu_1}{u} + \frac{\mu_2}{v} = \frac{\mu_2}{f_2}$

$\frac{1}{f_1} = \frac{1.5-1.0}{-4} \to f = -8.0\ c;\ \frac{1.5}{f_2} = \frac{1.5-1.0}{-4} \to f = -12.0\ cm,$

$\frac{1.0}{10} + \frac{1.5}{v} = \frac{1.0}{8.0} \to v = -6.66\ cm$

Example 2.9

One end of a plastic rod of index 1.5 is grounded and polished to a radius of +2.0 cm (convex). If an object in air is located on the axis 12.0 cm from the vertex (edge), what is the image distance?

Solution

$\frac{\mu_1}{u} + \frac{\mu_2}{v} = \frac{\mu_2-\mu_1}{r_1} \to \frac{1}{12} + \frac{1.5}{v} = \frac{1.5-1}{2} \to v = 9.0\ cm$

Example 2.10

One end of a glass rod of $\mu = 1.50$, is grounded and polished with a convex spherical surface of radius $10\ cm$. An object is placed in the air on the axis $40\ cm$ to the left of the vertex. Find (i) the power of the surface and (ii) the position of the image.

Solution

(i) Here $r_2 = \infty$. This is the case, since one end of the glass rod is plane while the other end is polished into conical shape $p = \frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \frac{\mu-1}{r} = \frac{1.5-1}{0.1} = +5.0\ D$;

(ii) $\frac{\mu_{air}}{u} + \frac{\mu_{glass}}{v} = \frac{\mu_{glass}-\mu_{air}}{r} \to \frac{1}{40} + \frac{1.5}{v} = \frac{0.5}{10} \to \frac{1.5}{v} = \frac{0.5}{10} - \frac{1}{40} = \frac{1}{40} \to v = +60\ cm$

Example 2.11

An object is at a distance of $20\ cm$ from a converging lens of focal length $12\ cm$. What is the image distance from the lens?

Solution

$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \to \frac{1}{v} = \frac{1}{12} - \frac{1}{20} \to v = 30\ cm$

Example 2.12

A converging beam of light is incident on a diverging lens of focal length $18.0\ cm$. If the beam is directed to a point $6.0\ cm$ behind the lens, find the position of the image.

Solution

The object is a virtual object (behind the lens) $\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \to \frac{1}{v} = \frac{1}{-18} - \frac{1}{-6} \to v = 9.0\ cm$

Example 2.13

An object is placed at a distance of $7.0\ cm$ from a lens and a virtual image which is magnified 8 times is produced. What is the focal length of the lens and the image distance? What type of lens is it?

Solution

$m = \frac{v}{u} \rightarrow v = mu = 7x - 8 = -56\ cm$

$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} = \frac{1}{7} - \frac{1}{56} = \frac{56-8}{56} \rightarrow f = 8.0\ cm$, the lens has a positive focal length, thus a converging lens.

Example 2.14

A converging lens has a focal length of $20\ cm$. At what distance from the lens must an object be placed so that its linear magnification is of magnitude (i) 2 (ii) -2

Solution

For $m = 2$, we have $\frac{1}{m} = \frac{u}{f} - 1 \rightarrow \frac{1}{2} = \frac{u}{20} - 1 \rightarrow u = 30\ cm$

For $m = -2$, we have $-\frac{1}{2} = \frac{u}{20} - 1 \rightarrow u = 10\ cm$

Example 2.15

A biconvex glass with refractive index $\mu = 1.50$ has a faces of radii of curvature $10.0\ cm$. What is its focal length when placed in water of $\mu = 1.33$?

Solution

$\frac{1}{f} = \left(\frac{\mu_2}{\mu_1} - 1\right)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = \left(\frac{1.5}{1.3} - 1\right)\left(\frac{1}{10} + \frac{1}{10}\right) = 39.1\ cm$

Example 2.16

A converging lens of focal length 10.0 cm, is placed on a plane mirror, and the space between the lens and the mirror is filled with water of refractive index $\mu = 1.33$. Find the radius of curvature of the lower surface of the glass lens.

Solution

For two lenses in contact $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \rightarrow \frac{1}{f_2} = \frac{1}{15} - \frac{1}{10} \rightarrow f_2 = -30.0\ cm$,

the focal length of the water lens is $-30.0\ cm$

$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right)$; $r_2 = \infty$ since the lower surface of the lens is flat

$-\frac{1}{30} = \frac{1.33-1}{r} \rightarrow r = -9.9\ cm$, the radius of curvature of the lower surface of the glass lens is $9.9\ cm$.

Example 2.17

Calculate the position and magnification of the image formed by two coaxial converging lenses, each of focal length $10.0\ cm$ and $30.0\ cm$ apart, if the object is on the axis $12.5\ cm$ from one of the lenses.

Solution

For the first lens, $\frac{1}{f_1} = \frac{1}{u_1} + \frac{1}{v_1} \rightarrow \frac{1}{v_1} = \frac{1}{f_1} - \frac{1}{u_1} = \frac{1}{10} - \frac{1}{12.5} \rightarrow v_1 = 50.0\ cm$

Hence the object for the second lens lies $u_2 = 50.0 - 30.0 = 20.0\,cm$, beyond it on the other side of the lens and is real (we are using real is positive convention). Hence

$\frac{1}{v_2} = \frac{1}{f_2} - \frac{1}{u_2} = \frac{1}{10} - \frac{1}{20} \rightarrow v_2 = \frac{20}{3}\ cm$

$m = m_1 m_2 = \frac{v_1}{u_1}\frac{v_2}{u_2} = \frac{50}{12.5} x \frac{6.667}{20} = \frac{4}{3}$

Example 2.18

A biconvex lens has a thickness of 2.0 cm along its axis, and has spherical surfaces each 10.0 cm radius. The refractive index of the glass is $\mu = 1.5$. Obtain the position and magnification of an image formed by an object on the axis 30.0 cm from the nearest surface. Take refractive index of air to be $\mu = 1.0$

Solution

$\frac{\mu_1}{u_1} + \frac{\mu_2}{v_1} = \frac{\mu_2 - \mu_1}{r_1} \rightarrow \frac{1.5}{v} + \frac{1}{30} = \frac{1.5-1}{10} \rightarrow v_1 = 90.0\,cm.$

The image formed by the first surface acts as an object for the second surface, from which the distance is $90.0 - 2.0 = 88.0\,cm$. The object distance $u_2 = -88.0\,cm$, virtual object.

$\frac{\mu_1}{u_2} + \frac{\mu_2}{v_2} = \frac{\mu_2 - \mu_1}{r_1} \rightarrow \frac{1}{v_2} - \frac{1.5}{88} = \frac{1.5-1}{10} \rightarrow v_2 = 14.9\,cm$

$m_1 = \frac{\mu_1}{\mu_2}\frac{v_1}{u_1} = \frac{1}{1.5}\frac{90}{30} = 2.00$

$m_2 = \frac{\mu_1}{\mu_2}\frac{v_2}{u_2} = \frac{1.5}{1}\frac{14.9}{88} = 0.25$

$m = m_1 m_2 = 2.00 x 0.25 = 0.50$

Example 2.19

Find the position of the image of a distant object formed by a sphere of radius a and refractive index μ. Consider paraxial rays only.

Solution

$\frac{\mu_1}{u} + \frac{\mu_2}{v} = \frac{\mu_2 - \mu_1}{r_1}$

At the first surface, $u = \infty, \mu_1 = 1, \mu_2 = \mu, r = +a$, substituting gives

$\frac{1}{\infty} + \frac{\mu}{v} = \frac{\mu - 1}{a} \rightarrow v = \frac{\mu a}{\mu - 1}$

At the second surface, $\mu_1 = \mu, \mu_2 = 1, r = +a, u = -\left(\frac{\mu a}{\mu - 1} - 2a\right)$, i.e. the image distance from the lens minus the diameter of the lens, substituting we have

$\frac{1}{v} - \frac{1}{\left(\frac{\mu a}{\mu - 1} - 2a\right)} = \frac{\mu - 1}{a} \rightarrow v = \frac{a(2-\mu)}{2(\mu - 1)}$

The image is $\frac{a(2-\mu)}{2(\mu-1)}$ beyond the farther surface of the sphere.

Exercises Two

1. If the index for a piece of glass is 1.525, calculate the speed of light in the glass.
2. Calculate the difference between the speed of light in vacuum and the speed of light in air if the refractive index of air is 1.0002340.
3. If the moon's distance from the earth is $3.840x10^5\ km$, how long will it take a microwave to travel from earth to the moon and back?
4. How long does it take light from the sun to reach the earth? Assume the earth's distance from the sun to be $1.50x10^8\ km$.
5. A beam of light passes through a block of glass $10.0\ cm$ thick, then through water for a distance of $30.5\ cm$, and finally through another block of glass $5.0\ cm$ thick. If the refractive index of both glass is 1.5250 and of water is 1.3330, find the total optical path.
6. A water tank is $62.0\ cm$ long inside and has glass ends which are each $2.50\ cm$ thick. If the refractive index of water is 1.330 and of glass is 1.6240, find the overall optical path.
7. A ray of light in air is incident on the polished surface of a block of glass at an angle of $10°$ (a) If the refractive index of the glass is 1.5258, find the angle of refraction. (b) Use the incident angle in radians (and assume that the angle in radians equal the sine of the angle), what is the angle of refraction? (c) Find the percentage error. (d) Repeat the same problem this time with incident angle of $45°$.
8. An object $15\ cm$ from a lens produces a real image $30\ cm$ from the lens. What is the focal length of the lens?
9. A lens produces a virtual image which is 3 times the height of the object at a distance of $15.0\ cm$ from the lens. Find the object distance and the focal length of the lens.
10. A diverging lens produces a virtual image half of the object when the object is at a distance of $10.0\ cm$ from the lens. What is the focal length of the lens and the image distance?
11. What is wrong with this statement 'a diverging mirror produced a magnified real image of an object'?
12. Determine the position and the characteristics of the image of an object which is at a distance of $80\ cm$ from a diverging mirror of focal length $20.0\ cm$.
13. An illuminated object and a screen are placed 90 cm apart. What lens is required in order to produce an image which is twice the height of the object on the screen? What are the possible positions of the lens with respect to the object?
14. If in a camera the distance from the lens to the film, when a distant object is clearly focused, is $4\ cm$, how far will the lens have to be moved, and in which direction, in order to focus clearly an object at a distance of $8\ m$?

15. An object placed in front of a converging lens gives a real image with magnification 5; when the object is moved $6\ cm$ along the axis of the lens, a real image of magnification 2 is obtained. What is the focal length of the lens?
16. A metal plate containing an illuminated circular hole is placed at one end of an optical bench, and a screen at the other. By means of a convex lens an image of the hole is formed on the screen, the diameter of the image being $2.25\ cm$. If the lens is moved $20\ cm$ along the bench an image of the hole again appears on the screen, its diameter being now $1.00\ cm$. What is the real size of the hole, and how far is it from the screen?
17. A thin concave lens has a focal length of $30\ cm$. Determine the position of the image which it forms of an object $10\ cm$ away (a) if the object is real, (b) if it is virtual. In each case show the path of a pencil of rays by which an eye, suitably placed, may see a non-axial point on the image.
18. The distance between an object and its image formed by a converging lens is $40\ cm$ and the magnification is 3. If the magnification were 5, what would be the distance apart of object and image?
19. Show that the image formed by a converging lens of a virtual object in any position is real, erect and diminished, and nearer to the lens than the principal focus.
20. A small object lies along the axis of a converging lens. Show that the longitudinal magnification (i.e. along the axis) is equal to the square of the transverse magnification.

CHAPTER THREE
OPTICAL INSTRUMENTS

3.1 INTRODUCTION

An optical instrument either processes light waves to enhance an image for viewing, or analyzes light waves (or photons) to determine one of a number of characteristic properties of the light. The eye being a natural optical instrument, it is the most complicated, no artificial optical instrument can yet match its ability for accommodation and dynamic range. The first optical instruments were magnifying glasses and telescopes. Telescope were used for magnification of distant images, and microscopes used for magnifying very tiny images. Since the days of Galileo, these instruments have been greatly improved and extended into other portions of the electromagnetic spectrum. The binocular device is a generally compact instrument for both eyes designed for mobile use. A camera could be considered a type of optical instrument, with the pinhole camera being very simple examples of such devices. Other classes of optical instrument used for the analyses of the properties of light or optical materials include:

i. Interferometer for measuring the interference (wave propagation) properties of light waves
ii. Photometer for measuring light intensity
iii. Polarimeter for measuring dispersion or rotation of polarized light
iv. Reflectometer for measuring the reflectivity of a surface or object.
v. Refractometer for measuring refractive index of various materials
vi. Spectrometer or monochromator for generating or measuring a portion of the optical spectrum, for the purpose of chemical or material analysis
vii. Autocollimator which is used to measure angular deflections
viii. Vertometer which is used to determine refractive power of lenses such as glasses, contact lenses and magnifier lens
ix. Polarization Controller
x. DNA Sequencers can be considered optical instruments as they analyze the *color* and intensity of the light emitted by a fluorochrome attached to a specific nucleotide of a DNA strand
xi. Surface Plasmon resonance-based instruments use refractometry to measure and analyze bio-molecular interactions

In this chapter, we shall be interested in analysis of the working of simple optical instruments due to the combination of simple thin lenses and mirrors.

3.2 THE EYE

Eyes are the organs of vision. They detect light and convert it into electro-chemical impulses in neurons. In higher organisms, the eye is a complex optical system which collects light from the surrounding environment, regulates its intensity through a diaphragm, focuses it through an adjustable assembly of lenses to form an image, converts this image into a set of electrical signals, and transmits these signals to the brain through complex neural pathways that connect the eye via the optic nerve to the visual cortex and other areas of the brain. Eyes with resolving power have come in fundamentally different forms, and 96% of animal species possess such a complex optical system. The simplest "eyes", such as those in microorganisms, do nothing but detect whether the surroundings are light or dark. For the more complex eyes, retinal photosensitive cells send signals to the brain through the neurons for adjustments of the light intensity, image formation and interpretations. Complex eyes can distinguish shapes and colors. The visual fields of many organisms, involve large areas of binocular vision to improve depth perception and to maximize the field of view.

3.2.1 The Human Eye

The human eye is an organ that reacts to light and has several purposes. As a sense organ, the eye allows vision. Rod and cone cells in the retina allow conscious light perception and vision and interpretation, including color differentiation and the perception of depth. The human eye can distinguish about 10 million colors, for those that can enjoy 'normal' vision with overall power as $p \sim 58.64\ diopters$. Yet in some instances like *Optical Illusion*, our perception of vision cannot be relied upon. In spite of any imperfection in human vision, we can perceive beauty, form and motion with illumination from white light and its constituent colors.

The eye is like a fine camera, with a shutter, iris, and lens system on one side and a light sensitive film called retina on the other side. The lens focuses the real but inverted diminished image of the object on the retina, the iris acts like a diaphragm which opens wide for faint light and reduces its aperture for bright light. The iris also determines the color of the eyes. The retina contains hundreds of light sensitive nerves called cones and rods that changes light pulses into electric signal for transmission to the brain for interpretation. While the cones responds to bright light and are responsible for distinction in colors, rods are sensitive to faint light, to motions and variation in intensity. Our perception of light may be divided into two parts – (i) the optical components leading to the formation of image on the retina and (ii) the property of the nerve canal and brain to interpret the electrical impulse produced.

3.2.2 Structure

The eye is not shaped like a perfect sphere, rather it is a fused two-piece unit. The smaller frontal unit, more curved, called the cornea is linked to the larger unit called the sclera. The corneal segment is typically about $8\ mm$ in radius. The sclerotic chamber constitutes the remaining larger part of the eye with radius typically about $12\ mm$. The cornea and sclera are connected by a ring called the limbus. The iris – the color of the eye – and its black center, the pupil, are seen instead of the cornea due to the cornea's transparency. The area opposite the pupil called the fundus, shows the characteristic pale optic disk, where vessels entering the eye pass across and optic nerve fibers depart the globe. The eye includes a lens similar to lenses found in optical instruments such as cameras and the same principles can be applied. The pupil of the human eye is its aperture; the iris is the diaphragm that serves as the aperture stop. Refraction in the cornea causes the effective aperture (the entrance pupil) to differ slightly from the physical pupil diameter. The entrance pupil is typically about $4\,mm$ in diameter, although it can range from $2\,mm$ in a brightly lit place to $8\,mm$ in the dark. The latter value decreases slowly with age; older people's eyes sometimes dilate to not more than $5 - 6\,mm$.

Summarizing, the eye being one of the most intricate and sensitive organ has the following basic structures

i. **Eye Lens** – This is what focuses the light entering the eye.
ii. **Ciliary Muscles** – This are attached to the eye lens surfaces and is used in altering the focal length of the eye lens
iii. **Retina** – The light sensitive area of the eye where the image of the object is formed.
iv. **Yellow Spot** – The most sensitive spot on the retina.
v. **Iris** – The colored circle round the eye lens.
vi. **Cornea** – The thick transparent protective covering in front of the lens which also acts as refractive medium
vii. **Vitreous Humor and Aqueous Humor** – Liquid behind and in front of the eye lens in which the eye lens floats.
viii. **Blind Spot –** This is the region where the optic nerves enter the eye and is insensitive to light. Image formed at the blind spot cannot be perceived.
ix. **Pupil** – The circular opening or diaphragm in the iris through which light passes.

3.2.3 Size

The dimensions differ among adults by only one or two millimeters; it is remarkably consistent across different ethnicities. The vertical measure, generally less than the horizontal distance, is about $24\,mm$ among adults, at birth about $16–17\,mm$. The eyeball grows rapidly, increasing to $22.5–23\,mm$, by three years of age. By age 13, the eye attains its full size. The typical adult eye has an anterior to posterior diameter of $24\,mm$, a volume of $6\,cm^3$ and a mass of $7.5\,g$.

3.2.4 Components

The component of the eye is made up of three coats, enclosing three transparent structures. The outermost layer, known as the fibrous tunic, is composed of the cornea and sclera. The middle layer, known as the vascular tunic, consists of the choroid, ciliary body, and iris. The innermost is the retina, within these coats are the aqueous humor, the vitreous body, and the flexible lens. The aqueous humor is a clear fluid that is contained in two areas: the anterior chamber between the cornea and the iris, and the posterior chamber between the iris and the lens. The lens is suspended to the ciliary body by the suspensory ligament, made up of fine transparent fibers. The vitreous body is a clear jelly that is much larger than the aqueous humor present behind the lens, and the rest is bordered by the sclera, and lens. They are connected via the pupil

3.2.5 Field of View

The approximate field of view of an individual human eye is $95°$ away from the nose, $75°$ downward, $60°$ toward the nose, and $60°$ upward, allowing humans to have an almost 180-degree forward-facing horizontal field of view. With eyeball rotation of about $90°$, horizontal field of view is as high as $270°$. About $12–15°$ temporal and $1.5°$ below the horizontal is the optic nerve or blind spot which is roughly $7.5°$ high and $5.5°$ wide.

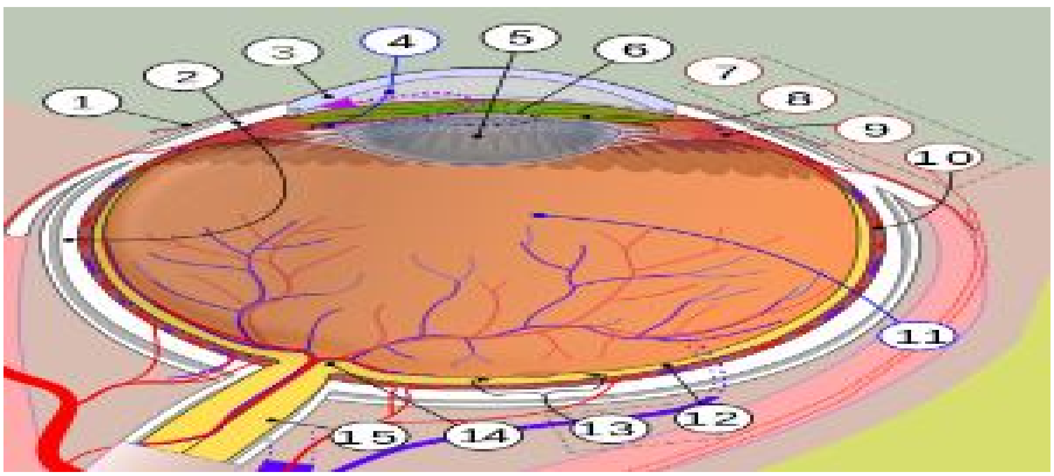


Fig 3.1. Diagram of a human eye (Adapted from Wikipedia) (**1. Conjunctiva, 2. Sclera, 3. Cornea, 4. Aqueous humor, 5. Lens, 6. Pupil, 7. Uvea with 8. Iris, 9. Ciliary body and 10. Choroid; 11. Vitreous humor, 12. Retina Width 13. Macula; 14. Optic Disc → Blind Spot, 15. Optic Nerve.)**

3.2.6 Dynamic Range

The retina has a static contrast ratio of around $100:1$ (by implication it can differentiate between two objects that are 100 times brighter than each other). As soon as the eye moves it re-adjusts its exposure both chemically and geometrically by adjusting the iris which regulates the size of the pupil. The process is nonlinear and multifaceted, so an interruption by light merely starts the adaptation process over again. The human eye can detect a luminance range of 10^{14}. At the low end of the range is the absolute threshold of vision for a steady light across a wide field of view, about $10^{-6}\ cdm^{-2}$, while the upper end of the range is given in terms of normal visual performance as $10^8\ cdm^{-2}$ (cdm^{-2} is defined as candela per square meter – the unit of luminance).

3.2.7 Accommodation of the Lens

Accommodation is achieved by changing the curvature of the lens is carried out by the ciliary muscles surrounding the lens. They narrow the diameter of the ciliary body, relax the fibers of the suspensory ligament, and allow the lens to relax into a more convex shape. A more convex (more contracted and thicker) lens refracts light more strongly and focuses divergent light rays onto the retina allowing for closer objects to be brought into focus. On the other hand, distant objects are brought into focus by relaxing the curvature of the lens by making them more extended and thinner.

3.2.8 Binocular Vision

The world appears to us as three-dimensional because the images seen by our two eyes are slightly different; the left eye sees more of the left side of the object than the right, and vice-versa, enabling the solidity of an object to be appreciated. The possession of two eyes also helps to judge distances. When looking at near objects, the lines of vision of the two eyes must converge and the muscular effort involved helps to give an idea of the distance.

3.2.9 Dark Adaptation

One of the outstanding properties of the eye is the enormous range of its sensitivity to light. When we enter a darkened room, from the daylight, our eyes take some time to become accustomed to the darkness and we may be quite unable to find anything without assistance initially. The process of becoming accustomed to the darkness is known as dark adaptation of the eye and is believed to involve the manufacture of extra quantities of visual purple. It is slow during the first 10 minutes and rapid in the next 20-30 minutes. Night-blind people are very slow in becoming adapted and may take up to an hour to get adapted.

3.2.10 The Purkinje Effect

In 1825 J.E. Purkinje, observed that, in twilight, red flowers appear darker than blue ones, although they may appear equally bright in sunlight. Eventually, at dusk, the red flowers look black and the blue flowers grey. This change of the spectral sensitivity of the eye with a decrease in light intensity is known as Purkinje Effect. Vision at low intensity, when all color disappear, is known as scotopic vision; vision at high intensity is known as photopic vision.

3.2.11. Color Adaptation

If the eye is exposed to a bright red light (as an example) for some minutes its sensitivity to red light is depressed. Consequently, the colors of objects appear different to an eye which has been so exposed than they do so to a normal eye. This phenomenon is known as color adaptation.

3.2.12 After-Image

If one looks for a second or two at a bright object and then closes one's eyes, a bright image of the object will be seen, known as a positive after-image. The after-image may change color and the phenomenon is very complex. If one looks at a bright red source of light for a short time and then transfers one's gaze to a brightly illuminated white sheet of paper, a blue-green image of the source of light will be seen. This is known as negative after-image. And its color is complimentary to that of the original image. Usually, it is explained by assuming that the sensitivity of the red receptors, in that part of the retina where the original image was formed, is depressed, and hence the white paper stimulates chiefly the blue and the green receptors giving rise to a blue-green image. The phenomenon is sometimes called successive contrast. Another phenomenon the enhancement of colors of an object when placed beside other colored objects is called simultaneous contrast. In general, every colored object tends to modify and enhance the colors of neighboring objects in the direction of its own complementary color.

3.3. DEFECTS OF THE EYE

There are many diseases, disorders, and age-related changes that may affect the eyes and surrounding structures. As the eye ages, certain changes occur that can be attributed solely to the aging process. Most of these eye defect which are anatomic and physiologic processes follow a gradual decline. With aging, the quality of vision worsens due to reasons independent of diseases of the aging eye. While there are many changes of significance in the non-diseased eye, the most functionally important changes seem to be a reduction in pupil size and the loss of accommodation. Other eye defect include the presbyopia, hypermetropia, astigmatism.

3.3.1 Myopia

Myopia commonly known as near-sightedness or short-sightedness is a condition of the eye where the light that comes in does not directly focus on the retina but in front of it, causing the image that one sees when looking at a distant object to be out of focus, but in focus when looking at a close object. It may also be corrected by refractive surgery, though there are cases of associated side effects or with corrective lenses that have a negative optical power (concave lenses) which compensates for the excessive convergence of the myopic eye.

3.3.2 Hypermetropia

Hyperopia or hypermetropia commonly known as being farsighted or longsighted is a defect of vision caused by an imperfection in the eye. Occurring when the eyeball is too short or the lens cannot become round enough, causing difficulty focusing on near objects, and in extreme cases causing a sufferer to be unable to focus on objects at any distance. As an object moves toward the eye, the eye must increase its optical power to keep the image in focus on the retina. If the power of the cornea and lens is insufficient, as in hyperopia, the image will appear blurred. The causes of hyperopia are typically genetic and involve an eye that is too short or a cornea that is too flat, so that images focus at a point behind the retina. It may be corrected with convex lenses in eyeglasses or contact lenses. Convex lenses have a positive power value, which causes the light to focus closer than its normal range.

3.3.3 Lack of Accommodation or Presbyopia

Presbyopia is a condition associated with aging in which the eye exhibits a progressively diminished ability to focus on objects. Presbyopia's exact mechanisms are not fully understood; research evidence most strongly supports a loss of elasticity of the crystalline lens, although changes in the lens's curvature from continual growth and loss of power of the ciliary muscles (the muscles that bend and straighten the lens) have also been postulated as its cause. The first signs of presbyopia – eyestrain, difficulty seeing in dim light, problems focusing on small objects and/or fine print – are usually first noticed between the ages of 40 and 50. Corrective lenses that provide a range of vision correction, are used in treatment, including varifocal or bifocal lenses to eliminate the need for a separate pair of reading glasses, newer bifocal or varifocal spectacle lenses attempt to correct both near and far vision with the same lens.

3.3.4 Astigmatism

Astigmatism is an optical defect in which vision is blurred due to the inability of the optics of the eye to focus a point object into a sharp focused image on the retina. Astigmatism causes difficulties in seeing fine detail resulting in blurred vision, for example, the image may be clearly focused on the retina in the horizontal plane, but not in the vertical plane. This may be due to an irregular curvature of the cornea or lens. The refractive error of the astigmatic eye stems from a difference in degree of curvature refraction of the two different meridians (i.e., the eye has different focal points in different planes). The two types of astigmatism are regular and irregular. Three options exist for the treatment of astigmatism: glasses, contact lenses (either hard contact lenses or toric contact lenses), and refractive surgery. Irregular astigmatism is often caused by a corneal scar or scattering in the crystalline lens, and cannot be corrected by standard spectacle lenses, but can be corrected by contact lenses. The more common regular astigmatism arising from either the cornea or crystalline lens can be corrected by cylindrical eyeglasses or toric lenses.

3.3.5 Color Blindness

This is the absence or paralysis of the fibers of the retina which are sensitive to the color one is blind to. There are various types: (i) Protanopia – Absence/Failure of Red receptors (ii) Deuteranopia – Red and Green (iii) Tritanopia – Blues and Greens confused (iv) Monochromatism – Total color blindness (v) Anomalous Trichromatism – Though may appear normal but require different amounts of Green and Red color in their color mixture than normal people.

3.4 THE PINHOLE CAMERA.

A pinhole camera is a simple camera without a lens and with a single small aperture, a pinhole – and is effectively a light-proof box with a small hole in one side and a light sensitive material (film) on the opposite side. Light from a scene passes through this single point and projects an inverted image on the opposite side of the box. It is completely dark on all the other sides of the box (usually painted black), to absorb all diffused rays of light. Up to a certain point, the smaller the hole, the sharper the image, but the dimmer the projected image, else the bigger the hole, the brighter the image but blurred.

Optimally, the size of the aperture should be 1/100 or less of the distance between it and the projected image. A method of calculating the optimal pinhole diameter was first attempted by Jozef Petzval. The crispest image is obtained using a pinhole size determined by the formula

$$d = \sqrt{2f\lambda} \qquad 3.1$$

where d is pinhole diameter, f is focal length (distance from pinhole to image plane) and λ is the wavelength of light. For standard black-and-white film, a wavelength of light corresponding to yellow-green ($550\,nm$) should yield optimum results. For a pinhole-to-film distance of $25\,mm$, this works out to a pinhole $0.17\,mm$ in diameter. While Young indicated that the optimal image is obtained if

$$d = 2\sqrt{f\lambda}. \qquad 3.2$$

Young obtained this expression by also considering the resolution of the image.

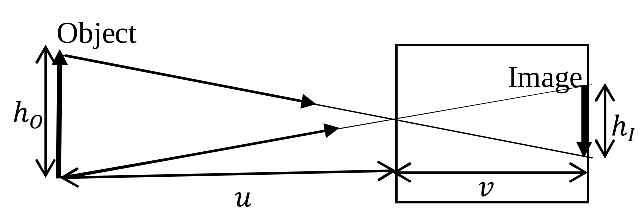


Fig 3.2 The Pinhole Camera

If h_O is the object height, h_I the image height, u the object distance from the pinhole and v the image distance from the pinhole, then from similar triangle, it is easily shown that

$$m = \frac{h_I}{h_O} = \frac{v}{u} \qquad 3.3$$

where m is the magnification.

3.5 THE CAMERA

A camera is an optical instrument that records images that can be stored directly, transmitted to another location, or both. These images may be still photographs or moving images such as videos. The term *camera* comes from the word *camera obscura*, an early mechanism for projecting images. The functioning of the camera is very similar to the functioning of the human eye. A camera may work with the light of the visible spectrum or with other portions of the electromagnetic spectrum. A still camera is an optical device which creates a single image of an object or scene, and records it on an electronic sensor or light sensitive film. All cameras use the same basic design: light enters an enclosed box through a converging system of lenses and an image is recorded on a light-sensitive medium. A shutter mechanism controls the length of time that light can enter the camera. Most photographic cameras have functions that allow a person to view the scene to be recorded, allow for a desired part of the scene to be in focus, and to control the exposure so that it is not too bright or too dim. A display, often a liquid crystal display (LCD) or light emitting diode display (LED), permits the user to view scene to be recorded and settings such as speed, exposure, and shutter speed. A video camera operates similarly to a still camera, except it records a series of static images in rapid succession, commonly at a rate of 24 frames per second. When the images are combined and displayed in order, the illusion of motion is achieved.

3.5.1 Speed of Lenses

Quality pictures of stationary objects can be taken with relative ease with several time exposures, but when objects are in relative motion, extreme short exposure time and large aperture are needed to make good picture. Thus, the most important feature of a good camera is the objective lens system with relative large aperture range capable of covering as large an angular field as possible. Unfortunately, a lens of large aperture is subject to many forms of aberrations, requiring a compromise in camera designs. The parameter that can be used to measure the quality and quantity of light allowed by the lens of a camera for a given time depends on the focal length of the lens and its diameter and is called the speed of the lens (focal ratio or $\boldsymbol{f-value}$).

Objects are photographed by emitted or reflected light. The amount of light per unit area reflected or emitted by an object is called its brightness or luminance $\mathcal{B}$, while the amount of light per unit area falling on a receiver (photographic film etc.) is called the illuminance $\mathcal{L}$. The illuminance $\mathcal{L}$ depends on several factors - (i) the brightness $\mathcal{B}$ of the object (ii) the area of the entrance pupil of the lens (aperture - $\pi A^2/4$) and (iii) the focal length f of the lens. The light entering a camera is proportional to the brightness of the object and the aperture size of the lens, but is inversely proportional to the square of the focal length, thus we can write

$$\mathcal{L} = \frac{k\mathcal{B}}{f^2}\frac{\pi A^2}{4} \rightarrow \mathcal{L} \propto \frac{A^2}{f^2} \qquad 3.4$$

Here k is the proportionality constant. Equation (3.4) indicates that for a given object with brightness $\mathcal{B}$, if f is doubled, light will spread over 4 times the area, reducing illuminance $\mathcal{L}$, but if the lens aperture diameter is doubled, the illuminance is quadrupled. In practice, one talks about the focal ratio

$$f - value = \frac{f}{A}. \qquad 3.5$$

Various designs of camera lens include

i. The Meniscus Lens – This is a single converging meniscus lens with a fixed stop. It suffers considerable spherical aberration and is usually designed with an $f - value \sim f/11$.
ii. The Symmetrical Lens – This lens design consists of two identical set of thick lenses with a stop midway between them. In general, each half of the lens is corrected for lateral aberration, and by putting them together, curvature of field and distortions are eliminated. Spherical aberration limits the aperture to designs with $f - value \sim f/8$.
iii. Triplet Anastigmats – This lens design consists of two converging lens and a diverging lens in between. The spaces between them is design to give optimal positive power and to eliminate astigmatism, curvature of field and generally minimize aberrations. Designs with $f - value \sim f/2$ are readily available and suitable for motion pictures.

3.5.2 Image Capture

Traditional cameras capture light onto photographic plate or photographic film. Video and digital cameras use an electronic image sensor, usually a charge coupled device (CCD) to capture images which can be transferred or stored in a memory card or other storage inside the camera for later playback or processing. However these categories overlap as still cameras are often used to capture moving images in special effects work and many modern cameras can quickly switch between still and motion recording modes.

3.5.3 Lens

The lens of a camera captures the light from the subject and brings it to a focus on the sensor. The design and manufacture of the lens is critical to the quality of the photograph being taken. Camera lenses are made in a wide range of focal lengths. They range from the extreme wide angle lens to the standard lens and the medium telephoto. Each lens is best suited to a certain type of photography. The extreme wide angle may be preferred for architecture because it has the capacity to capture a wide view of a building. The normal lens, because it often has a wide aperture, is often used for street and documentary photography. The telephoto lens is useful for sports and wildlife though it is more susceptible to camera shake.

3.5.4 Focus

The distance range in which objects appear clear and sharp, called depth of field, can be adjusted by many cameras. This allows for a photographer to control which objects appear in focus, and which do not. Due to the optical properties of photographic lenses, only objects within a limited range of distances from the camera will be reproduced clearly. The process of adjusting this range is known as changing the camera's focus. There are various ways of focusing a camera accurately. The simplest cameras have fixed focus and use a small aperture and wide-angle lens to ensure that everything within a certain range of distance from the lens is in reasonable focus. Rangefinder cameras allow the distance to objects to be measured by means of a coupled parallax unit on top of the camera, allowing the focus to be set with accuracy. Modern cameras often offer autofocus systems to focus the camera automatically by a variety of methods including post focusing. Post focusing means that the pictures are first taken, and then focusing later at the processing unit.

3.5.5 Exposure Control

The size of the aperture and the brightness of the scene controls the amount of light that enters the camera during a period of time, and the shutter controls the length of time that the light hits the recording surface. Equivalent exposures can be made using a large aperture size with a fast shutter speed and a small aperture with a slow shutter.

3.5.6 Shutters

Although a range of different shutter devices have been used during the development of the camera only two types have been widely used and remain in use today. The Leaf shutter is a shutter contained within the lens structure, often close to the diaphragm consisting of a number of metal leaves which are maintained under spring tension and which are opened and then closed when the shutter is released. The exposure time is determined by the interval between opening and closing. In this shutter design, the whole film frame is exposed at one time. Disadvantages of such shutters are their inability to reliably produce very fast shutter speeds

The focal-plane shutter operates as close to the film plane as possible and consists of cloth curtains that are pulled across the film plane with a carefully determined gap between the two curtains (typically running horizontally) or consisting of a series of metal plates (typically moving vertically) just in front of the film plane. The focal-plane shutter is primarily associated with the single lens reflex type of cameras, since covering the film rather than blocking light passing through the lens allows the photographer to view through the lens at all times *except* during the exposure itself.

3.5.7 Camera Designs

Various camera designs exist, from the earliest plate camera, to the folding camera, box camera, the rangefinder camera, instant picture camera, single-lens camera (SLR), twin-lens reflex camera (TLR), large-format camera, medium-format camera (35 mm camera), subminiature/miniature camera, water-submersible camera, auto-focus camera, movie/video camera and the digital camera. A digital camera is a camera that encodes digital images and videos digitally and stores them for later reproduction. Most cameras sold today are digital, and digital cameras are incorporated into many devices ranging from mobile phones (called camera phones) to vehicles. Digital and film cameras share an optical system, typically using a lens with a variable diaphragm to focus light onto an image pickup device (basically a CCD – Charge Coupled –Device).

3.6 SIMPLE MICROSCOPES

A simple microscope is a converging lens (with positive power), that function to increase the angular size of the object. Usually the object is placed between the focal length and the lens, thus a magnified, erect but virtual image of the object can be seen. The lens is moved until the image is seen distinctly at the near point (about 25 cm) from the normal eye. The image subtends an angle θ' at the eye, so that the image looks much larger than the object.

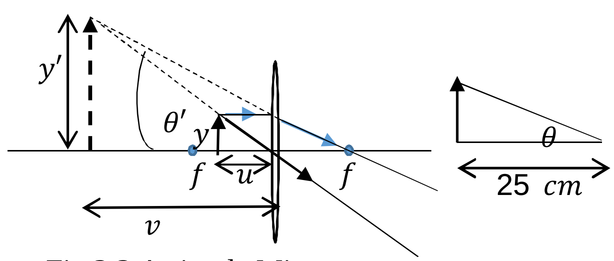


Fig 3.3 A simple Microscope

Let, θ be the angular dimension of the object, θ' the image angular size, u the object's distance from the lens, v the image's distance, y the object's height and y' the image height, the angular magnification is given by

$$m_\theta = \frac{\theta'}{\theta} \qquad 3.6$$

Using the standard defined nearest point (25 cm – a point closest to the eye at which a sharp image can still be obtained), the object distance is ($v = -25$, virtual image)

$$\frac{1}{u} = \frac{1}{f} + \frac{1}{25} = \frac{25+f}{25f} \qquad 3.7$$

For small angle, $\tan\theta = \theta\ (radian)$, thus we can write $\theta = \frac{y}{25}, \theta' = \frac{y}{u} = y\left(\frac{25+f}{25f}\right)$, substituting in equation (4.5), we have

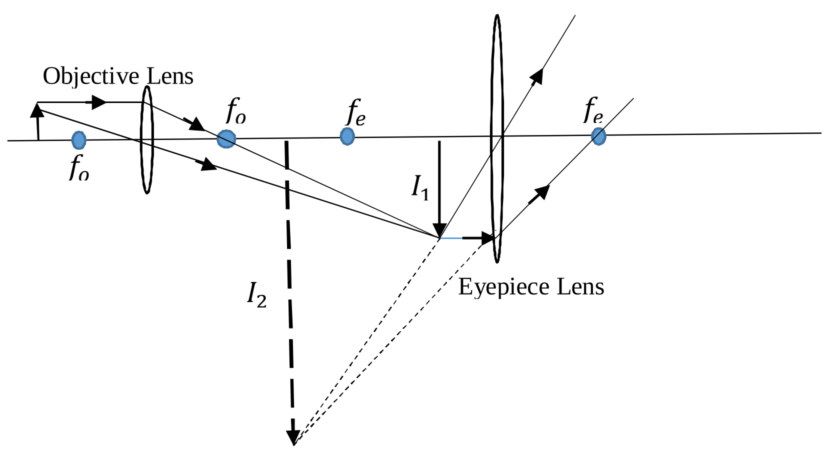
$$m_\theta = \frac{\theta'}{\theta} = \frac{y\left(\frac{25+f}{25f}\right)}{\frac{y}{25}} = \frac{25}{f} + 1 \qquad 3.8$$

3.7 THE COMPOUND MICROSCOPE

A compound microscope is a microscope which uses a lens close to the object being viewed to collect light (called the objective lens) which focuses a real image of the object inside the microscope (image 1). That image is then magnified by a second lens or group of lenses (called the eyepiece) that gives the viewer an enlarged inverted virtual image of the object (image 2). The use of a compound objective/eyepiece combination allows for much higher magnification, reduced chromatic aberration and exchangeable objective lenses to adjust the magnification. Optical microscopy is used extensively in microelectronics, nanophysics, biotechnology, pharmaceutical research, mineralogy and microbiology.

As shown in figure 3.4, a compound microscope has two converging lenses, the eyepiece and the objective. Usually, the focal length of the eyepiece f_e is greater than the focal length of the objective f_o i.e. $f_e > f_o$. The object is placed just outside the principle focus of the objective, this produces a real, magnified, inverted image (image 1). The arrangement is such that image 1 fall with the focal length of the eyepiece and becomes its object, a virtual, erect, magnified image is formed (image 2). The total magnification m of a compound microscope is the product of the magnification of the eyepiece m_e and the objective m_o given by

$$m = m_o m_e \qquad 3.9$$



3.8 THE ASTRONOMICAL TELESCOPE

An optical telescope gathers and focuses light mainly from the visible part of the electromagnetic spectrum. Optical telescopes increase the apparent angular size of distant objects as well as their apparent brightness. In order for the image to be observed, photographed, studied, and sent to a computer, telescopes work by employing one or more curved optical elements, usually made from glass lenses and/or mirrors, to gather light and to bring that light or radiation to a focal point. There are two main types:

- The refracting telescope which uses lenses to form an image.
- The reflecting telescope which uses an arrangement of mirrors to form an image.

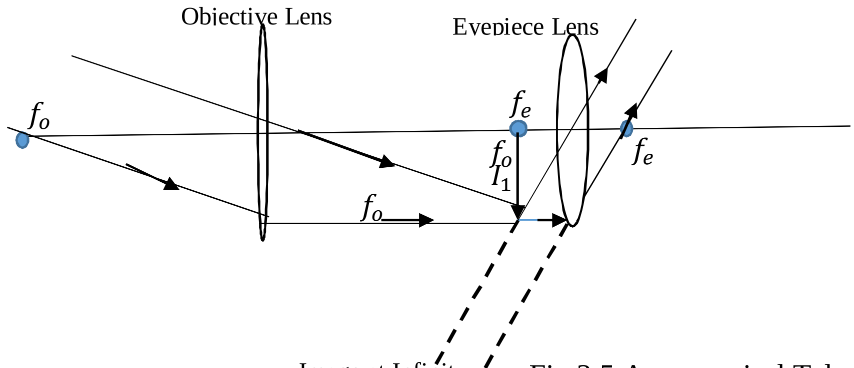


Fig 3.5 Astronomical Telescope

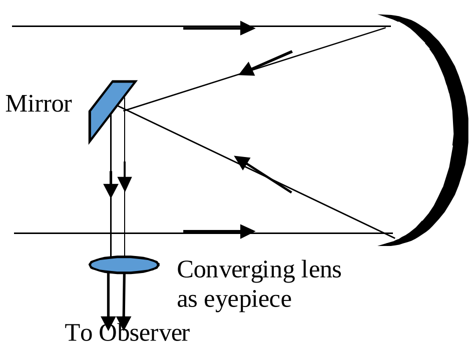


Fig 3.6 Reflecting Astronomical Telescope

The refracting telescope consist of two converging lenses, an objective lens of long focal length and an eyepiece of short focal length, so $f_o > f_e$. Rays from astronomical sources are nearly parallel on reaching the objective lens, so it forms a real, inverted diminished image I_1, of the object at f_o. $I_1$is at f_e, so the eyepiece forms a virtual, erect magnified image formed at infinity I_2. Unlike compound microscope, the lenses are aligned so that their focal plane and focal lengths coincides.

The reflecting telescope has a spherical/parabolic mirror as its objective. This is necessary for large telescopes, since the larger the objective's diameter, it can gather more light, and the longer the focal length, the higher the magnification. Large refracting telescope suffer from spherical and chromatic aberration as well as high engineering cost and designs. Parallel rays from astronomical sources are reflected by the converging mirror at the objective, these reflected rays are intercepted by a small mirror before they form real image I_1. this image is magnified in the usual way by an eyepiece of a converging lens. The magnification is given by

$$m = \frac{f_o}{f_e} = \frac{D}{d} \qquad 3.10$$

where D is the diameter of the objective and d the diameter of the eyepiece.

3.9 THE TERRESTRIAL TELESCOPE

The image formed by astronomical telescope is always inverted, magnified and virtual with respect to the object. Since astronomical objects are spherical in shape, the inversion of the image is not an issue. Moreover, different arrangement can be used for capturing the images on a screen and for record purpose. For terrestrial usage, an additional converging lens or Porro prisms which are total internal reflecting are used to invert the image so that an upright image with respect to the object is formed.

Using three converging lenses as shown in figure 3.7, the image is erect and magnified. The objective lens has a long focal length f_o, the erecting lens f and the eyepiece lens f_e both have short focal lengths. The image I_1 formed by a distance object is real, inverted, and is formed at the principal focus of the objective. The erecting lens is placed at twice its focal length from I_1. So, forms an image I_2, the same size and inverted with respect to I_1, and at twice the focal length but with respect to the object, the image is erect. The eyepiece is place such that its principle focus coincide at twice the focal length of the erecting lens, such that the final image is formed at infinity, magnified, virtual, but erect with respect to the original object

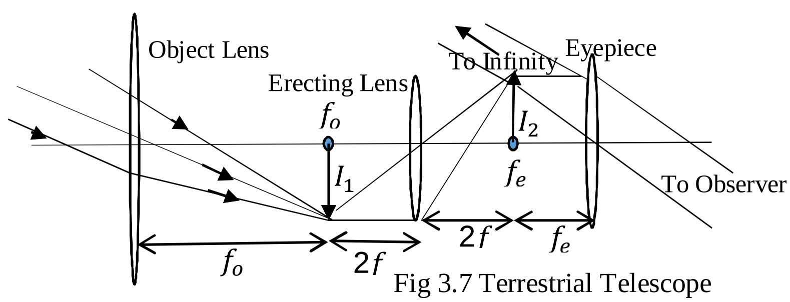


Fig 3.7 Terrestrial Telescope

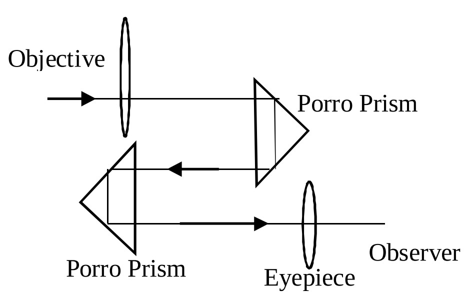


Fig 3.8 Prism Binocular

Though the image is erect, but the disadvantages of including the erecting lens are (i) the reduction of the light intensity getting to the eyepiece and the increase in the length of the telescope with a minimum length of $f_o + 4f + f_e$. This disadvantages can be overcome using Porro prism to invert the image, since they are totally internal reflecting, theoretically, light intensity is not diminished and the additional $4f$ can be easily reduced to only the size of the prism.

The prism telescope consist of a converging objective lens of long focal length f_o, two total reflecting prisms with angles $45°, 90°, 45°$, and a converging eyepiece lens of short focal length f_e. one of the prism corrects the image for vertical inversion, the second prism correct the image for horizontal inversion. The eyepiece then magnifies the image. The final image is erect and magnified.

3.10 THE PROJECTOR

The lens camera produces a small image on a film, the projector is used for the production of large images on a white screen. The essential features are

i. A very powerful small source of light.
ii. A condenser made up of two plano-concave lenses, which collects the light, beams it towards the slides, and illuminates it strongly.
iii. A projection lens near the slide (series of images) so that the object distance for the slide is less than $2f$ but more than f, the focal length of the projection lens, leading to the formation of an inverted real image of the slide, which is produced on a distant screen.

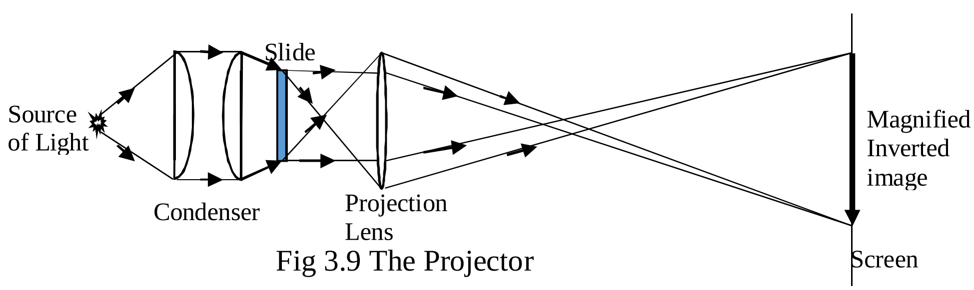


Fig 3.9 The Projector

3.11 OCULARS AND EYEPIECES

Due to light aberration while using single lens for the eyepiece of most optical instruments, it is necessary to design special lens combination suitable for each instrument and to eliminate most of the aberrations (discussed in Chapter Four). Such eyepieces are called oculars. One of the most important considerations in the designs of oculars is the correction for lateral chromatic aberration. For this reason, the basic structure of most oculars involves two lenses of the same glass separated by a distance equal to half the sum of their focal lengths.

The two most popular oculars are the Huygens eyepiece and Ramsden eyepiece. In both these system, the lens nearest the eye is called the eye lens EL and the lens nearest the objective is called the field lens FL. Other designs include the Kellner achromatized Ramsden ocular (with almost zero chromatic aberration, and high reduction in spherical aberration), orthoscopic ocular (with wide field and free from distortion), symmetrical ocular (with wider field especially used in telescopic gun sight). Various types of oculars are shown in figure 3.10, the distance between EL and FL is not drawn to scale.

3.12 STOPS IN OPTICAL SYSTEMS

Stops are used in optical systems to control

(i) the amount of light entering an optical system, and hence the brightness of the image

(ii) the depth of the focus and the depth of the field - Depth of focus is defined as the greatest distance the film of a camera can be moved without spoiling the definition of the image and depth of field is the distance between the positions of the object at which it is sufficiently in focus.

(iii) the extent of the field of view – The field of view of an instrument is the angle subtended at the eye by the longest object, the whole of which is just seen

(iv) The aberration of defects of image.

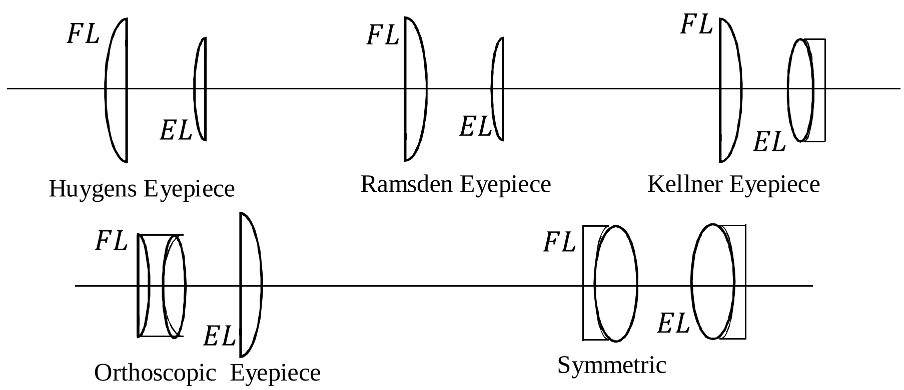


Fig 3.10 Various types of Oculars

Example 3.1

The far point and the near point of a short-sighted person are $200\ cm$ and $15\ cm$ respectively from the eye. (i) What type lens placed $2.0\ cm$ from the eye would enable him see an object at infinity? (ii) Where would the near point be when he uses spectacles of the focal length calculated?

Solution

(i) The lens must form a virtual image at a distance of 200 cm from the eye i.e. $200.0 - 2.0\ cm = 198.0\ cm$ from the lens to see an object at infinity, i.e. $u = \infty; v = -198\ cm$

Using $\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \rightarrow \frac{1}{\infty} - \frac{1}{198} \rightarrow f = -198\ cm$, a concave lens.

$$p = \frac{1}{1.98} = -0.505\ diopters$$

(ii) The original near point is $15\ cm$ from the eye or $15.0 - 2.0\ cm = 13\ cm$ from the lens,

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \rightarrow \frac{1}{-198} = \frac{1}{u} - \frac{1}{13} \rightarrow u = 13.9\ cm$$

The distance from the near point to the eye will be $13.9 + 2.0 = 15\ 9\ cm$

Example 3.2

The near point of a man is $75\ cm$ from the eye. What type of lens should he use in order to be able to read a book held at $25\ cm$ from the eyes?

Solution

$u = 25\ cm, v = -75\ cm$ (Virtual image),

$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \rightarrow \frac{1}{f} = \frac{1}{25} - \frac{1}{75} \rightarrow f = 37.5\ cm; p = \frac{1}{37.5} = 2.67\ diopters$

Example 3.3

A man's near point is $50\ cm$ from his eye, and his far point is at $3\ m$. find the focal lengths of two sets of spectacle lenses he needs to enable him see distant objects, and to read prints at $25\ cm$.

Solution

To correct for short-sightedness $u = \infty; v = -3\ m, f = -3\ m, p = -0.33\ diopters$

To correct for far-sightedness, $u = 25\ cm, v = -50\ cm, f = 50\ cm, p = 2\ diopters$

Exercise Three

1. Explain the following terms (i) Chromatic Aberration (ii) Spherical Aberration. Discuss any noticeable changes in the two effects if any for a lens system when (i) monochromatic light is used (ii) the lens aperture is reduced.
2. A boy's near point is 20 cm from his eye, and his far point is 2 m away from the eye. Find the type and focal length of the spectacle he needs to see distant object clearly and the range of vision when he is wearing them.
3. What is the closest distance of distinct vision for a man whose near point is $50\ cm$ from his eye when wearing spectacles of power $1\ diopters$
4. What do you understand by astigmatism? How can it be corrected?
5. A lady finds that she has to hold a book at arm's length in order to read clearly although she can see distant objects clearly. What type of eye defect is she suffering from and how can it be corrected?

CHAPTER FOUR
DEFECTS IN OPTICAL INSTRUMENTS

4.1 INTRODUCTION

The image formed by a simple lens, particularly if the lens is powerful and the image is highly magnified, shows very noticeable defects; it is colored at the edges, not all of it is in focus, and it is distorted. If two such lenses are arranged to form a model telescope or microscope, the defects appear far worse. Such a single-lens objective is unsatisfactory for most precision and astronomical purposes as the images produced suffer from defects or aberrations of different kinds. Considerable effort has been applied to the design of objectives to remove or reduce the aberrations and refractor telescopes show a variety of construction depending on their intended function and the way particular aberrations have been compensated. These aberrations are wavelength dependent (chromatic) or the imperfection in the spherical curvature of the instruments (Spherical). Most mirrors show only spherical aberration while lenses show both spherical and chromatic aberration.

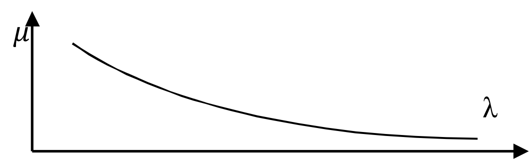


Fig 4.1 Variation of Reflective index with wavelength

Aberrations fall into two classes: monochromatic and chromatic. Monochromatic aberrations are caused by the geometry of the lens or mirror and occur both when light is reflected and when it is refracted. They appear even when using monochromatic light, hence the name. Chromatic aberrations are caused by dispersion, the variation of a lens's refractive index with wavelength. They do not appear when monochromatic light is used. Single lens objectives in optical instruments may suffer from the following defects: (i) Chromatic Aberration (ii) Spherical Aberration (iii) Coma (iv) Astigmatism (v) Curvature of Field (vi) Distortion of Field.

4.2 CHROMATIC ABERRATION

A closer look at the lens-maker's formula reveals why the image produced by a single lens suffers from chromatic aberration. For a given lens of known shape, the lens-maker's formula may be re-written so that the focal length may be expressed by

$$\frac{1}{f} = (\mu - 1)\left(\frac{1}{r_1} + \frac{1}{r_2}\right) = K(\mu - 1) \qquad 4.1$$

where $K = \left(\frac{1}{r_1} + \frac{1}{r_2}\right)$ and for a positive lens, K is also positive. The term $(\mu - 1)$ is known as the refractive power of the material of the lens. The refractive index and, hence, refractive power of all common materials used for lenses exhibits dispersion i.e. its value is wavelength-dependent. A typical dispersion curve is illustrated in figure 4.1

and it depicts the fact that the refractive index progressively decreases as the wavelength of the light increases. Thus, the focal length of a single lens depends on the wavelength of light that is used and, for a simple positive lens, the focal length increases as the wavelength increases. If, therefore, a single positive lens is illuminated by a parallel beam of white light, a spread of images is produced along the optic axis of the lens. This is illustrated in figure 4.2.

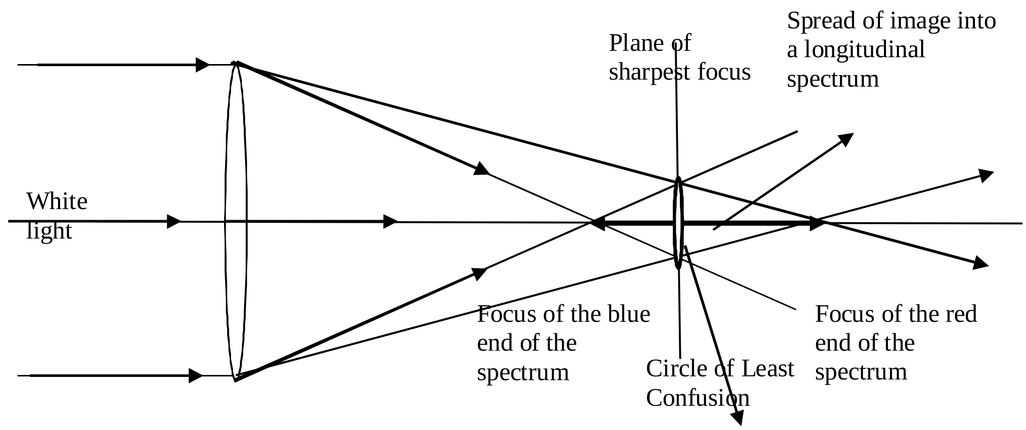


Fig 4.2 A ray diagram illustrating the effect of chromatic aberration resulting from the use of a single positive lens.

At no point along the optic axis is there a position where a point image can be seen: at the position where an image is formed for the extreme blue end of the spectrum, this image is surrounded by a red halo; similarly at the position where an image is formed for the extreme red end of the spectrum, this image is surrounded by a blue halo. Between these extreme positions, there is a plane which contains the smallest possible image which can be obtained. The image in this plane of sharpest focus is not a point image. It is known as the circle of least confusion and its size can be determined from the physical properties of the lens.

The spread of the image along the optic axis is known as longitudinal chromatic aberration and the spread of the image in the plane of sharpest focus is known as lateral chromatic aberration. It will be remembered that it was first supposed that chromatic aberration could not be removed from lens systems but, eventually, Dolland proposed a method for achromatism. It is possible by using in combination with a positive lens, a negative lens of different material — and hence negative refractive power — to cancel the dispersion without complete cancellation of the refractive power. This is easily demonstrated as follows.

If two lenses of focal lengths f_1 andf_2 are fixed together by optical cement, the focal length of the combination, f, is given by

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \rightarrow f = \frac{f_1 f_2}{f_1 + f_2} \qquad 4.2$$

By expressing the individual focal lengths in terms of the shape of each lens and their powers, equation (4.2) can be rewritten as

$$p_i = p_{1i} + p_{2i} = K_1(\mu_{1i} - 1) + K_2(\mu_{2i} - 1) \qquad 4.3$$

The subscript i indicate that particular color. Thus, for any two colors say red and blue, we can write

$$p_r = K_1(\mu_{1r} - 1) + K_2(\mu_{2r} - 1);\ p_b = K_1(\mu_{1b} - 1) + K_2(\mu_{2b} - 1) \qquad 4.4$$

Now the aim of the combination is to provide a system whose focal length is identical for the blue and red ends of the spectrum, and achromatic, thus $p_r = p_b$, thus achromatic condition is achieved when $K_1(\mu_{1r} - 1) + K_2(\mu_{2r} - 1) = K_1(\mu_{1b} - 1) + K_2(\mu_{2b} - 1)$ 4.5

Multiplying and cancelling out, we have

$$K_1\mu_{1r} + K_2\mu_{2r} = K_1\mu_{1b} + K_2\mu_{2b} \rightarrow -K_1(\mu_{1b} - \mu_{1r}) = K_2(\mu_{2b} - \mu_{2r})$$

$$\rightarrow \frac{K_1}{K_2} = -\frac{(\mu_{2b}-\mu_{2r})}{(\mu_{1b}-\mu_{1r})} \qquad 4.6$$

Since both the numerator and the denominator on the right hand side of equation (4.6), have positive values, the minus sign shows that one of the K must be negative and the other positive. Hence, achromatization can only be achieved by combining a positive and a negative lens.

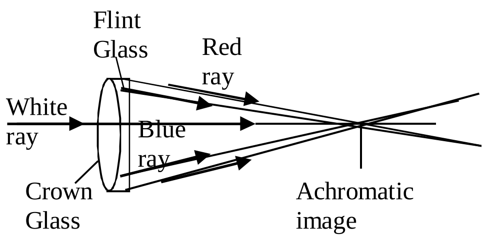


Fig 4.3a An achromatic doublet, made by cementing a positive crown glass lens to a negative flint glass lens, depicts how light rays of different colors are brought to the same focus.

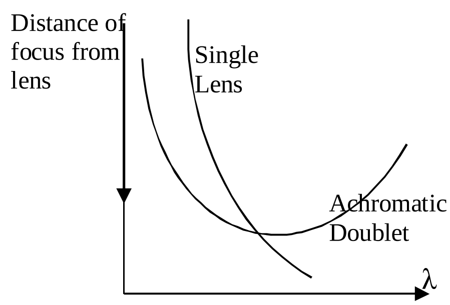


Fig 4.3b The variation of the focal length with wavelength of (a) a single positive lens and (b) a corrected achromatic doublet.

The choice of the types of glass available for the two lenses imposes a condition on the ratio K_1/K_2. According to the required focal length of the combination, K_1 and K_2 may then be determined individually: the particular values of the radii of each surface are usually decided from conditions set for reducing other aberrations.

An achromatic doublet or achromat usually consists of a positive lens made of a glass with a medium refractive index and dispersion followed by a negative lens made of glass with a higher refractive index and dispersion. A typical doublet is depicted in figure 4.3a. The same figure illustrates the way in which the first lens introduces separation of the rays corresponding to the blue and red wavelengths present in the incident light and the second lens produces a closing of these rays so that they will cross at the same point along the optic axis. A simple achromatic doublet does not completely remove all the chromatic aberration. There still remains a small spread of focus along the optic axis of the lens but this has been reduced considerably over the single lens. At each position within this reduced spread, light rays corresponding to a pair of wavelengths are brought to the same focus.

Figure 4.3b illustrates how the focal length of an achromatic lens might vary with wavelength. The same fig also illustrates how the focal length of a single lens would vary over the same spectral range. There are, in fact, many different designs of achromatic objective. The wavelengths for which the best achromatism occurs can be chosen according the purpose of the telescope. Hence, some objectives are termed to be visual or photographic. Not all objectives are constructed in the form of a cemented doublet; there are some variations which have an air-gap between the two elements. One special objective, known as a photo visual, is constructed of three elements. The achromatism produced by this system is so effective that it can be used either visually or with photographic plates, as its name implies. For any given line, the power equation becomes for two lenses combined

$$p_1 = K_1(\mu_{1i} - 1); p_2 = K_2(\mu_{2i} - 1) \rightarrow \frac{K_1}{K_2} = \frac{p_1(\mu_{2i} - 1)}{p_2(\mu_{1i} - 1)} \qquad 4.7$$

Combining equation (4.7 and 4.6), we have

$$\frac{p_1(\mu_{2i} - 1)}{p_2(\mu_{1i} - 1)} = -\frac{(\mu_{2b} - \mu_{2r})}{(\mu_{1b} - \mu_{1r})} \qquad 4.8$$

Solving for p gives $\frac{p_1}{p_2} = -\frac{(\mu_{1i} - 1)/(\mu_{1b} - \mu_{1r})}{(\mu_{2i} - 1)/(\mu_{2b} - \mu_{2r})} = \frac{\gamma_1}{\gamma_2}$,

where $\gamma_1 = \frac{(\mu_{1i} - 1)}{(\mu_{1b} - \mu_{1r})}; \gamma_2 = \frac{(\mu_{2i} - 1)}{(\mu_{2b} - \mu_{2r})}$

is called the dispersion constant. The use of the above formulas in calculating the radii for a desired achromatic lens involves the following

i. A specified focal length f_i and power p_i, where i is most likely the yellow light.
ii. Selecting the crown and flint glass.
iii. Estimating the dispersive constant.
iv. The values of K_1 and K_2 are calculated, from which the radii are calculated before polishing.

4.3 SPHERICAL ABERRATION

The simple theory for lens design is based on the effects of refraction produced by spherical surfaces and this is particularly convenient as a spherical surface is one of the easiest to obtain on an optical grinding and polishing machine. However, simple lens theory only takes into account paraxial rays, i.e. rays that are very close to the optic axis, allowing sin θ to be written as θ. For a point source at infinity, lying on the axis of a single lens, the image produced by the lens does not retain a point-like appearance and is spread out into a disc. This effect results from the rays which cannot be considered as paraxial. The position of the focus for any incident ray depends on its distance from the optic axis. The defect of the image is known as spherical aberration and its effect is illustrated in figure 4.4

If the spread in focus is denoted by Δf, the severity of any spherical aberration may be expressed by assessing the value of the ratio $\Delta f / f$. As in the case of chromatic aberration, there is one plane through the spread of focus which contains the smallest image, again known as the circle of least confusion. Also the spread of an image along the optic axis is known as longitudinal spherical aberration and the spread of an image in the plane containing the circle of least confusion is known as lateral spherical aberration. The size of any image can be predicted by the physical properties of the lens or by performing a ray-tracing analysis

The amount of spherical aberration depends on the shape of a lens. It is, therefore, convenient to define what is known as the shape factor of a lens. By denoting the radii of the two lens surfaces as R_1 and R_2, the shape factor, q, is expressed as

$$q = \frac{r_2 + r_1}{r_2 - r_1} \qquad 4.9$$

Typical shapes of lenses have been drawn in figure 4.5a with a range of shape factors running between $q < -1$ and $q > +1$. By examination of lenses over the complete range of shape factors, it is found that spherical aberration has minimal effect when q is close to $+0.7$ – it never goes to zero. Spherical aberration can be overcome completely by figuring a lens so that the curvature of the faces is not constant. This process is known as aspherizing and is sometimes used in producing telescope objectives. However, this process can be costly and it is not the only way to remove spherical aberration

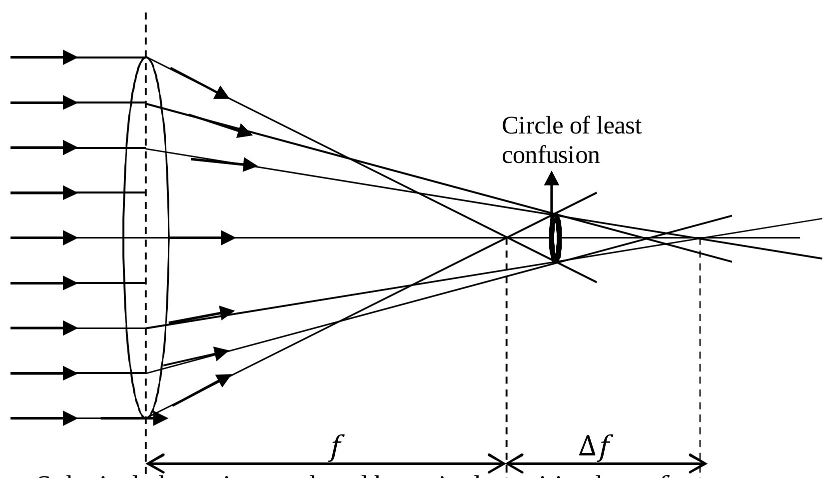


Fig 4.4 Spherical aberration produced by a single positive lens; f represents the focal length measured to the circle of least confusion and Δf (exaggerated for clarity) represents the spread of the image.

For a negative lens, the spherical aberration is also negative, i.e. the numerical value of the focal length increases as incident rays which are more distant from the optic axis are considered. A combination of a positive and a negative lens can be designed to provide a system which is free from spherical aberration. This is particularly convenient as we have already seen that chromatic aberration can be reduced by a two-lens system. Thus, achromatic lenses are designed to have minimal spherical aberration.

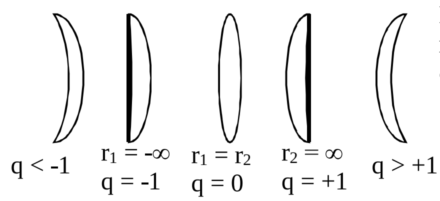


Fig 4.5a. A range of lenses with different shape factors.

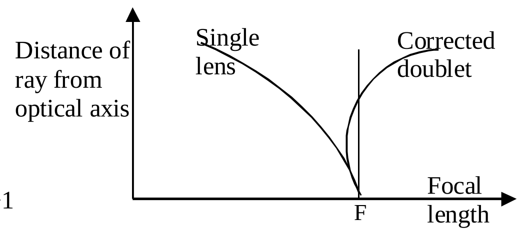


Fig 4.5b Longitudinal spherical aberration for a single lens and a corrected doublet.

The residual effect of spherical aberration for a corrected objective is illustrated in fig 4.5b. In the same figure, the effect of spherical aberration of a single lens is drawn for contrast. If effects of chromatic and spherical aberration have not been removed completely, they may be noticeable when the primary image is viewed with an eyepiece. However, an eyepiece normally limits the field of view to a small angle and these may be the only defects of the image that will be detected. By increasing the field of view, either by using a special wide-angle eyepiece or by placing a two dimensional imaging detector in the focal plane of the telescope, other types of aberration may become apparent. Such aberrations that might be detected in this way are: coma, astigmatism, field-curvature and field-distortion. Any of these aberrations may be present in images which result from incident rays which arrive at an angle to the optic axis.

4.4 COMA

The effect of **coma** derives its name from the comet-like appearance that an image can have when a point object is off the axis of a lens. Its cause can be considered by treating a lens as being made of a series of annuli. Each annular zone gives rise to an annular image in the focal plane. The total aberrated image results from the combination of each component image. When coma is present, any annulus of the lens produces an annular image; the total aberrated image can be thought of as being made up of a series of such annular images, the sizes increasing as the outer zones of the lens are considered. For any single lens, the size of the comatic image depends on the shape factor. In contrast with spherical aberration, it is possible for an image to be free of coma for a particular value of the shape factor. For objects at infinity, the coma-free condition is obtained when the shape factor is close to the value given by

$$q = \frac{(\mu-1)(2\mu+1)}{\mu+1} \qquad 4.10$$

and for a lens made of glass with refractive index $\mu = 1.5$, $q = +0.8$. This value is close to the value which results in a lens having minimum spherical aberration. When an achromatic doublet is designed, it is possible to correct coma and spherical aberration at the same time. A telescope objective which is free from these defects is known as an aplanatic lens.

4.5 ASTIGMATISM

A lens system which has been corrected for both spherical aberration and coma may not be free from aberration completely, especially when images are formed of objects which are at a considerable distance from the optic axis. Such images may suffer from astigmatism. The effect of astigmatism is illustrated in figure 4.6.

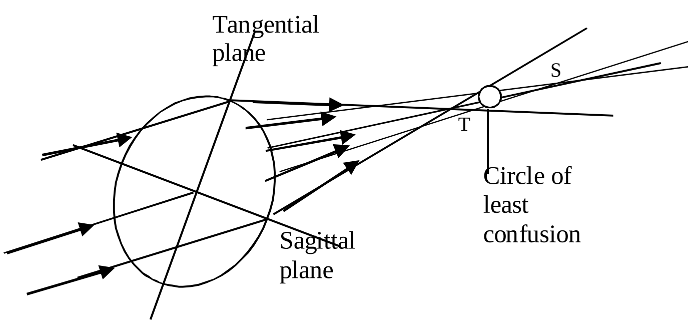


Fig 4.6 Points *T* and *S* represent the foci of rays passing through the tangential and sagittal planes of a lens; the spread of the astigmatic image lies between the points *T* and *S*.

The position of the focus depends on which section of the lens is used to form the image. The spread in the position of the image lies between two points which correspond to the image position for rays in the plane formed by the point object and the optic axis (i.e. the tangential plane) and to the image position for rays in the plane at right angles to this (i.e. the sagittal plane). Between these points there is a position where the smallest image can be found and again this image is known as the circle of least confusion.

By examining the images produced by objects with a range of distances from the optic axis, it is possible to record the surfaces on which the focus, T, of the rays of the tangential plane and the focus, S, of the rays of the sagittal plane lie. They are found to approximate to paraboloids of revolution. A single achromatic combination is likely to exhibit appreciable astigmatism. However, the combination of two achromats at the correct spacing can provide a paraboloidal surface containing both the T and S images. This surface is known as the Petzval surface, after the investigator of this property.

4.6 CURVATURE OF FIELD

A system which is designed to remove completely astigmatism suffers from curvature of field, as the points of sharp focus lie on a curve rather than on a plane. Flattening of the field can be achieved at the expense of having incomplete removal of astigmatism and usually some compromise is made. The judicial placement of stops in the combination of lenses can also be used to reduce the effect of field curvature.

4.7 DISTORTION OF FIELD

Objectives that have been designed to remove aberrations which would show up by the blurring of images of point objects may still suffer from distortion of field. The effect occurs when the correspondence between the distance of an object from the optic axis and the distance of its image from the optic axis in the focal plane is not constant over the field of view, i.e. the plate scale varies over the focal plane.

Barrel distortion is said to occur when the correspondence of the image position decreases with distance from the optic axis and pincushion distortion is said to occur when the correspondence increases with distance from the optic axis. In the design of compound objectives, the effect of field distortion can be reduced by the judicious placing of stops.

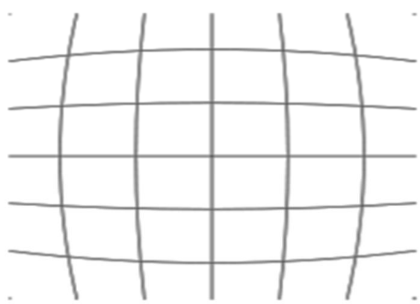


Fig. 4.7a: Barrel distortion

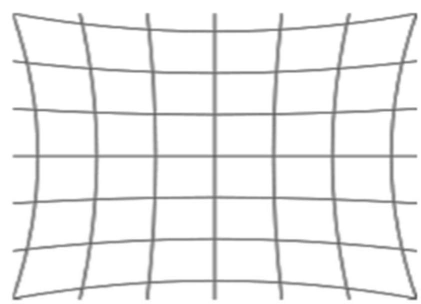


Fig. 4. 7b: Pincushion distortion

Distortion can be thought of as stretching the image non-uniformly, or, equivalently, as a variation in magnification across the field. While "distortion" can include arbitrary deformation of an image, the most pronounced modes of distortion produced by conventional imaging optics is "barrel distortion", in which the center of the image is magnified more than the perimeter (figure 4.7a). The reverse, in which the perimeter is magnified more than the center, is known as "pincushion distortion" (figure 4.7b). This effect is called lens distortion or image distortion. Usually, there are algorithms/adaptive optics to correct these distortions.

In pinhole projection, the magnification of an object is inversely proportional to its distance to the camera along the optical axis so that a camera pointing directly at a flat surface reproduces that flat surface. Such systems free of distortion are called orthoscopic or rectilinear system.

4.8 MATHEMATICAL FORMULATION OF DEFECTS IN OPTICAL INSTRUMENTS

The ray tracing equation use in paraxial ray optics were actually corrected to the first order in the inclination angles of the rays and the normal to the reflecting or refracting surfaces (where $\tan\theta \sim \sin\theta \sim \theta$). When higher-order approximations are used for the trigonometric functions of the angle, departures from the predictions of the paraxial optics will be found. No longer will it be true that all the rays leaving a point object will exactly meet to form a point image after being refracted or reflected, or that the magnification of a given transverse plane is constant. Furthermore, the properties of the system (refracting/reflecting) may be wavelength dependent. These deviations may be considered the mathematical representation of the aberration and optical defects noticed in optical instruments.

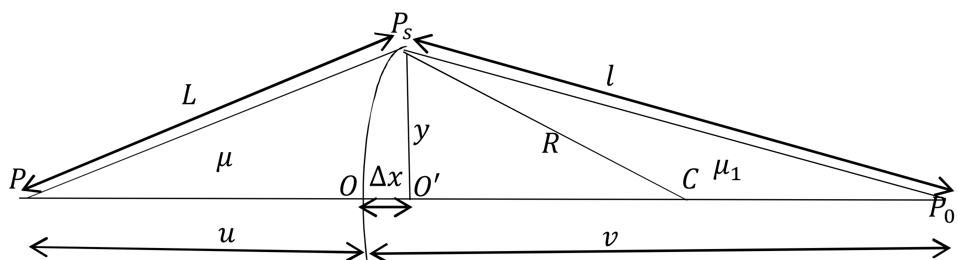


Fig 4.8 Geometry For Understanding Defects in Optical Instruments

We consider first the aberrations of a single spherical refracting surface between media of different refractive indices μ and μ_1. In paraxial ray approximation, $\Delta x = 0$, but in practice, $\Delta x \neq 0$, though may be considered small for thin lenses (where $\Delta x \ll \ll R, u, v$ and $y \ll R, u, u$). So that the optical path is given by vector addition (see figure 4.8)

$$\overline{PP_S} = L = \sqrt{u^2 + 2u\Delta x + y^2} \qquad 4.11a$$

And

$$\overline{P_SP_0} = l = \sqrt{v^2 - 2v\Delta x + y^2} \qquad 4.11b$$

To eliminate y, from equations A and B, we make use of triangle P_SCO', so we can write (C is the center of curvature, thus $R = O'C + \Delta x$)

$$R^2 = y^2 + (R - \Delta x)^2 = y^2 + R^2 - 2R\Delta x + (\Delta x)^2 \qquad 4.12$$

$\rightarrow \Delta x \cong \frac{y^2}{2R}$. Here we have neglected the term $(\Delta x)^2, \Delta x \lll R$. Substituting this in equations (1) we have

$L \cong \sqrt{u^2 + 2u\frac{y^2}{2R} + y^2} = u\sqrt{1 + \frac{y^2}{u}\left(\frac{1}{u} + \frac{1}{R}\right)}$ 4.13a

And

$l \cong \sqrt{v^2 - 2v\frac{y^2}{2R} + y^2} = v\sqrt{1 + \frac{y^2}{v}\left(\frac{1}{v} - \frac{1}{R}\right)}$ 4.13b

Making use of Taylor series expansion which gives

$\sqrt{1 + \epsilon} = 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{8} + \cdots$ 4.14

where we have neglected higher-order terms if $\epsilon \ll 1$. Applying this to equations (4.13), we have

$L \cong u\left[1 + \frac{y^2}{2u}\left(\frac{1}{u} + \frac{1}{R}\right) - \frac{y^4}{8u^2}\left(\frac{1}{u} + \frac{1}{R}\right)^2 + \cdots\right]$ 4.15a

And

$l \cong v\left[1 + \frac{y^2}{2v}\left(\frac{1}{v} - \frac{1}{R}\right) - \frac{y^4}{8v^2}\left(\frac{1}{v} - \frac{1}{R}\right)^2 + \cdots\right]$ 4.15b

Equation (4.15) is now a function of y. The optical path difference δ which will also be a function of y is given by (where $\overline{PO} = u$, and $\overline{OP_0} = v$)

$\delta(y) = \mu L + \mu_1 l - \mu\overline{PO} - \mu_1\overline{OP_0} = \mu L + \mu_1 l - \mu u + \mu_1 v = \mu(L - u) + \mu_1(l - v)$
4.16

Using equation (4.15), the optical path difference δ, becomes on simplification

$\delta(y) = \frac{y^2}{2}\left[\frac{\mu}{u} + \frac{\mu_1}{v} - \left(\frac{\mu - \mu_1}{R}\right)\right] - \frac{y^4}{8}\left\{\left[\mu\left(\frac{1}{u} + \frac{1}{R}\right)\right]^2\frac{1}{\mu u} + \left[\mu_1\left(\frac{1}{v} + \frac{1}{R}\right)\right]^2\frac{1}{\mu_1 v}\right\} + \cdots$
4.17

From equation (4.17), it is clear that there are the terms different from paraxial ray formulation, using the paraxial ray formula $\frac{\mu}{u} + \frac{\mu_1}{v} = \left(\frac{\mu - \mu_1}{R}\right)$, equation (4.17) reduces to

$\delta(y) = -\frac{y^4}{4}\left\{\left[\mu\left(\frac{1}{u} + \frac{1}{R}\right)\right]^2\frac{1}{\mu u} + \left[\mu_1\left(\frac{1}{v} + \frac{1}{R}\right)\right]^2\frac{1}{\mu_1 v}\right\} + \cdots = -\frac{y^4}{8}\sigma + \cdots$
4.18

where

$\sigma = \frac{1}{2}\left\{\left[\mu\left(\frac{1}{u} + \frac{1}{R}\right)\right]^2\frac{1}{\mu u} + \left[\mu_1\left(\frac{1}{v} + \frac{1}{R}\right)\right]^2\frac{1}{\mu_1 v}\right\}$
4.19

which is the lowest nonzero contribution to the path difference (δ). A full analysis of this introduces other nonzero contributions to the variation in the path length which depends basically on the radius of curvature, the nature of the media, the distance light travelled in the various media. These contributions consist of the defects notices in optical instruments.

Exercise Four

1. Give an account of the following with reference to the formation of images by lenses: (a) Spherical Aberration (b) Astigmatism (c) Curvature of Field (d) Distortion (e) Coma.
2. Assuming that the focal length of a thin double-convex lens (bi-convex) for marginal rays is less than for rays near the axis, explain why, for a given object distance, there is a particular shape for a thin converging lens of a given focal power which produces minimum spherical aberration. Why does this shape depend on the object distance?
3. Discuss the conditions under which a thin lens will form a sharp image of an object.
4. What do you understand by aberration? Discuss ways in which it can be reduced.
5. A beam of light parallel to the axis is incident on a simple lens. What effect on the position of the focus has (a) the color of the light (b) the distance of the parallel beam from the axis

CHAPTER FIVE
WAVE OPTICS

5.1 INTRODUCTION

Light shows some definite properties, which can only be explained if light is regarded as a wave. These properties include – interference, diffraction and polarization. It was Huygens in 1678 that first suggested the wave nature of light. In 1801, Thomas Young provided the first clear demonstration that light is a wave since they interfere under appropriate conditions. Several years later, Augustine Fresnel (1788 – 1829), performed a number of experiments dealing with interference and diffraction of light, providing further evidences to the wave theory of light. Maxwell's theory of light as electromagnetic wave was confirmed experimentally in 1887 by Hertz, proving beyond doubt that light can be considered as a wave.

5.2 WAVE THEORY OF LIGHT

All light waves are classified as transverse wave, which are those in which the wave's plane of vibration is perpendicular to its plane of propagation. When a vibrating source sends out transverse wave through a homogenous medium, they can be represented by a sinusoidal varying function.

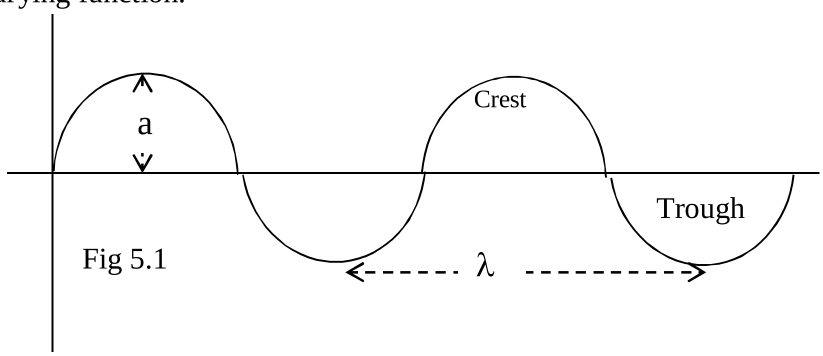


Fig 5.1

The distance between two similar points of any two consecutive waveforms is called the wavelength λ. One wavelength is equal to the distance between two wave crests or two wave troughs (fig 5.1). The maximum displacement of the wave about its mean position is called the amplitude. The frequency F defines the number of wave fronts passing at any giving point per second and is specified in unit of Hertz (Hz). The mathematical relation between the speed of wave v, its frequency ν and wavelength λ is $v = F\lambda$.

We define a wave front as the surface where points on the surface oscillate in phase with one another. In homogenous medium, the direction of propagation of wave is always normal to the wave front, and we can identify two types of wave front – spherical wave front and plane wave front. The simplest mathematical function that can be used to represent a transverse wave is a sinusoidal function. Consider transverse waves in which the motions of all parts are perpendicular to the direction of propagation, the displacement y of any point on the wave is given as

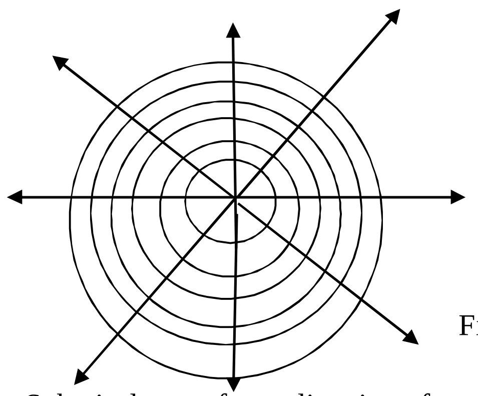
$$y = a \sin \frac{2\pi x}{\lambda} \qquad 5.1$$

If the wave moves to the right with a velocity v, then at any time t, the displacement in y is given as

$$y = a \sin \frac{2\pi}{\lambda} (x - vt) \qquad 5.2$$

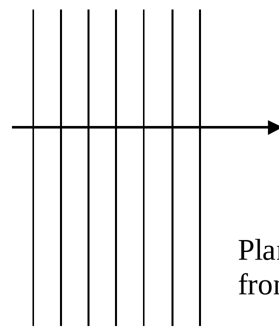
where a is the amplitude (maximum positive or negative displacement from the mean position). Using the $v - \lambda - F$ relation and defining the period T as the time taken for one complete revolution, we can write other forms of equation 5.2 as

$$y = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right) = a sin \left\{\frac{2\pi}{T}\left(t - \frac{x}{v}\right)\right\} = a \sin \left\{2\pi F \left(t - \frac{x}{v}\right)\right\} \qquad 5.3$$



Spherical wave front, direction of propagation is normal to the wave

Fig 5.2



Direction of Propagation

Planar wave front

5.2.1 Phase Angles

In wave motions, the instantaneous displacements and directions of propagation of two wave fronts are described by specifying the position and the angle with reference to an initial time. For instance at time $t = 0$, a wave propagating at an earlier time or later time can be written as

$y = a \sin (\omega t - \theta)$,

where θ (in radians) measured anti-clockwise from the positive $x - axis$ specifies the initial position at time $t = 0$ is called the phase angle, and ω is the angular frequency define as $\omega = 2\pi f$. A useful and concise way of expressing the equation of a wave oscillating in time and space is in terms of the angular frequency and wave number k, define by $k = 2\pi/\lambda$ as

$$y = a \sin (kx \pm \omega t + \theta) + a \cos (kx \pm \omega t + \theta) = exp\ i(kx \pm \omega t + \theta) \qquad 5.4$$

5.2.2 Phase Velocity and Wave Velocity

From the definition of phase angle, we can define the progression of a wave by saying that it constitutes the progression of a condition of constant phase. This condition might be, for instance, the crest of the wave (where the phase is such as to yield the maximum upward displacement); and the speed with which the crest moves along is called the phase velocity. The phase velocity may be equivalent to the wave velocity, which is evaluated by considering the rate of change of the x coordinate (considering our 1-dimensional wave), under the condition that the phase remains constant. With that condition, from equation (5.4) we can write $kx \pm \omega t \ = \ C$, and the wave velocity is

$$v \ = \frac{dx}{dt} \ = \ \frac{\omega}{k} \qquad 5.5$$

The ω/k is the phase velocity, and for any kind of wave, it depends on the physical properties of the medium in which the wave propagates and on the frequency of the wave itself.

5.2.3 Huygens's Principle

Huygens's Principle is a geometrical construction for determining the position of a wave later, if its position at an earlier time is known. Huygens's Principle is a statement that all points on a given wave front are taken as point sources, for the production of secondary waves called wavelets, which propagate through a medium with a speed characteristics of waves in that medium. This implies that every point of a wave front may be considered the source of secondary wavelets that spread out in all direction with a speed equal to the speed of propagation of the wave. The wave front at a later time can be found by constructing a surface tangent to the secondary wavelets.

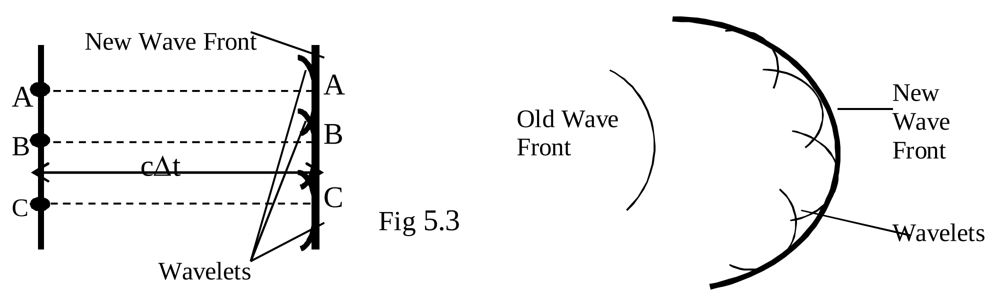


Fig 5.3

We can apply Huygens's principle to the laws of reflection and refraction. Consider figure 6.4, the AA^1 represents a wave front of the incident light. As ray 3 travels from A^1 to C, ray 1 reflects from A and produces a spherical wavelet of radius AD; because the two wavelets having radii A^1C and AD are in the same medium, they have the same speed v, therefore, line $A^1C \ = \ AD$.

From Huygens's principle, the reflected wave front can be represented by line CD; and from triangles AA^1C and ADC being right-angled triangle and congruent (they have common hypotenuse) we have $AA^1 = AD$, this implies that, the sine of angle A^1AC (θ_1) in ΔAA^1C is equal to sine of angle ACD (θ) in ΔADC,

$\Rightarrow sin\,\theta_1 = sin\,\theta$; $\Rightarrow \theta_1 = \theta$ 5.6

Since θ_1 is the incident angle and θ the reflected angle, equation 5.6 proves the law of reflection.

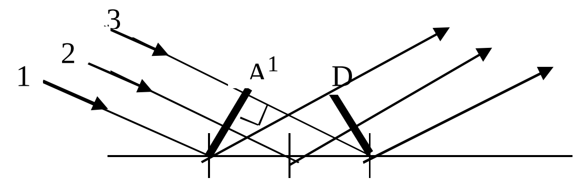


Fig 5.4

To derive Snell's law, we note that the in the time interval Δt, for which a ray moved from say A to B in the second medium is the same time interval for a second ray to moves from say A^1 to C. Considering the point A as the origin, the radius of the outgoing spherical wavelets centered at A is equal to $v_2\Delta t$ in the second medium, and the in the first medium, the distance $A^1C = v_1\Delta t$ (v is the velocity of the wave in each media). Using geometric arguments, we note that angle $A^1AC = \theta_1$ and that angle ACB equals θ_2. Defining the sine of the two angles, we have

$$\sin\theta_1 = \frac{v_1\Delta t}{AC}; \sin\theta_2 = \frac{v_2\Delta t}{AC} \Rightarrow AC = \frac{v_1\Delta t}{sin\,\theta_1} = \frac{v_2\Delta t}{sin\,\theta_2}$$

$$\Rightarrow \frac{sin\,\theta_1}{sin\theta_2} = \frac{v_1}{v_2} \Rightarrow \mu_1 \sin\theta_1 = \mu_2 \sin\theta_2 \qquad 5.7$$

Since from the definition of refractive index μ, $v_1 = \frac{c}{\mu_1}$ and $v_2 = \frac{c}{\mu_2}$

(c is the speed of light in vacuum). Since $\mu > 1$.

For other media except free space, $v < c$ always

5.3 INTERFERENCE OF LIGHT

When two waves from two sources over laps, the resultant wave is the vector sum of the waves from each source at that point. The resultant wave has amplitude that may be different in magnitude and direction from the two initial waves at that point. This effect is called interference. The interference produced by the superposition of two or more waves may be constructive interference (in this case, the amplitude of the resultant wave is greater than that of individual waves); or destructive interference (the resultant amplitude is less than that of individual waves). Light waves interfere, and the interference arises when the electromagnetic fields that constitute the individual waves combine to produce a new field.

Observing interference of light waves require certain conditions, this arises due of the nature of light and its extremely small wavelength - $400 - 700$ Å. This can be better appreciated by considering light from two bulbs; with such sources, interference effects are not observed, because the light waves from one bulb are emitted independently of those from the other bulb, thus, the emission from the two bulbs do not maintain a constant phase relationship with each other over time. Moreover, light waves from an ordinary source of light such as light bulbs and other natural sources (such light are termed incoherent) undergo random changes about every $10^{-8}\,s$, therefore, when interference occurs, the eye cannot follow such rapid changes, thus no interference effects are observed. The conditions necessary for observable interference to occur for light waves include (i) The light waves must come from two coherent sources – they must maintain a constant phase with respect to each other. (ii) The sources should be monochromatic – a single wavelength of light. (iii) The amplitude of the waves must be the same or nearly the same. (iv) The wavelength of light is about $400 - 700$ Å, thus the separation between the two sources must be of that range. (v) The path difference of the two light waves that interfere must be small.

For stable interference pattern, the coherence condition also implies that the light have the same frequency. If light waves of significantly different frequencies are used, the resulting interference pattern will be a rapidly varying time dependent pattern which may not be observable.

5.3.1 Young's Double Slit Experiment

Thomas Young experiment in 1801 to demonstrate the wave nature of light made use of the apparatus, schematically shown (figure 5.5). The apparatus consist of a monochromatic source of light M that is very bright, placed behind a narrow slit C, which acts as a point source of light, a double slit A, B with a separation of less than $0.5\,mm$ to act as two coherent sources of light and a screen. Light incident on the first barrier, M emerges through the second barrier as two coherent sources of wave; they interfere and produce on the screen, a visible pattern of bright and dark parallel band called interference fringe.

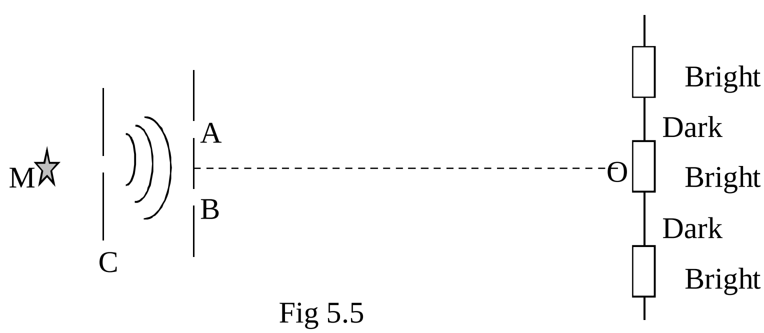


Fig 5.5

The interference fringes obtained from Young's double slit experiment can be explained using Huygens wave theory of light. The single slit sends out spherical wave fronts, since A and B are placed equi-distant from C, the same wave front arrives there at the same time. Then, points of the same wave fronts at A and B emit wavelets in phase with each other (coherent wavelets). These secondary wavelets interfere on the screen producing the fringe patterns we observe. This method of obtaining two coherent sources of wave is known as division of wave front.

The superposition of the waves from A and B produces interference pattern at the screen. For constructive interference to occur, the path difference between the waves must be an integral number of wavelengths (this ensures that two troughs or crests from the two sources arrive at the point of interference at the same time),

$\Rightarrow \Delta d = n\lambda$ 5.9

where d is the path length, λ is the wavelength, and $n = 0, \pm 1, \pm 2 \ldots$ is called the order number. When $n = 0$ (zeroth-order), $\Delta d = 0$, and a central maximum is obtained (located at O). The first bright fringe occurs when $n = \pm 1$ (first-order), this implies that, the nth bright fringe from the central maximum corresponds to path difference of $\Delta d = n\lambda$. A dark fringe is obtained where destructive interference occurred, and this happens at the point when the path difference equals half-integer wavelength (this ensures that a crest and a trough from the two sources arrive at the same time at the point of interference),

$\Rightarrow \Delta d = (n + ½)\lambda$ 5.10

5.3.2 Mathematical Analysis of the Interference Pattern

We can describe Young's double slit experiment quantitatively with the help of figure 5.6. The viewing screen is located a distance L from the slits, S_1 and S_2 are separated by a distance d. To reach any arbitrary point on the screen say, P, the wave from S_2 would have travel a further distance δ, which is the path difference

$\Rightarrow \delta = R_2 - R_1, but\ R_2 - R_1 = d \sin\theta \Rightarrow \delta = R_2 - R_1 = d \sin\theta$ 5.11

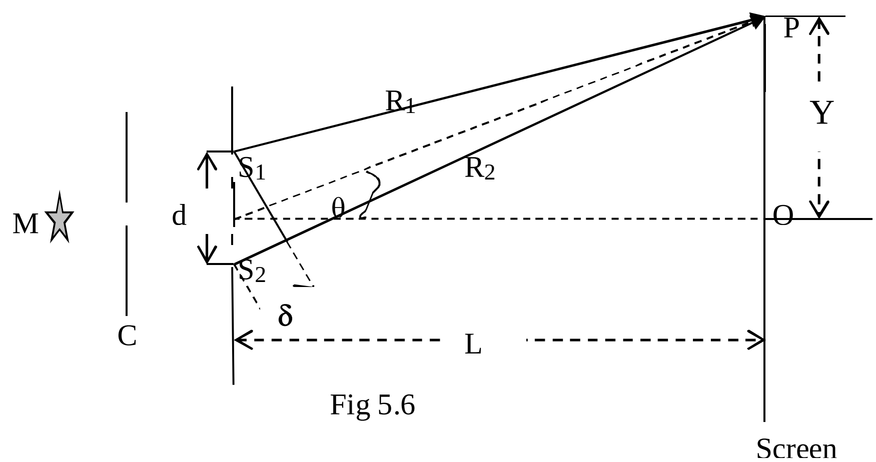


Fig 5.6

The value of δ, determines whether the two waves are in phase. If δ is either zero or some integer multiple of the wavelength, the two waves arriving at P are in phase and constructive interference occurs

$\Rightarrow \delta = n\lambda = d \sin\theta$ 5.10

For destructive interference, the value of δ is an odd multiple of half-integer wavelength

$\Rightarrow \delta = d \sin\theta = (n + ½)\lambda$ 5.11

To obtain an expression for the positions of the dark and bright fringes measured vertically along the screen, we assume that $L \gg d \gg \lambda$, then, geometric arguments give $Y_n = L \tan\theta_n$. Assuming $L \gg d \gg \lambda, \Rightarrow \tan\theta = \sin\theta = \theta$ (for small angle approximation)

$\Rightarrow Y_n = L \sin\theta_n$ and $\sin\theta_n = Y_n / L$.

From equations (5.10) and (5.11) $\sin\theta$ is given by

For destructive interference $\sin\theta_n = \frac{n+\frac{1}{2}\lambda}{d}$ 5.12

For constructive interference $\sin\theta_n = \frac{n\lambda}{d}$ 5.13

This implies that for destructive interference (dark fringes) the distance separating two consecutive dark fringes is

$Y_n = \frac{n+\frac{1}{2}\lambda}{d} L$ 5.14

And for constructive interference (bright fringes), we have

$Y_n = \frac{n\lambda}{d} L$ 5.15

From equations (5.14) and (5.15), we note that the distance between adjacent bright bands/dark bands in the pattern is inversely proportional to the distance d, between the slit, the closer the distance d between that slits, the more the patterns spread out. Young's double slit experiment provides us with a method for measuring the wavelength of light. Since L, d, and Y_n can be fixed, the wavelength λ of light used can easily be calculated from equations (5.14) and (5.15).

5.3.3 Factors That Affect The Interference Fringe Pattern

The variation of the following factors listed below affects the observable interference pattern as indicated

i. Separation between the single slit and the double slit – The fringe pattern becomes bright when the separation decreases, since the intensity of light reaching the double slit has increased, but the fringe separation remains the same.
ii. Slit separation d – Decreasing the slit separation d increases the separation between fringe pattern and vice-versa.
iii. Using white light instead of monochromatic light causes the central maximum to appear white, while other bright fringes appear multi-colored.
iv. If the experiment is performed in a medium other air, the fringe separations are narrower by a factor equal to the refractive index of the medium.
v. If one of the slit is covered by a transparent film of refractive index μ, the central maximum will shift towards the slit covered with the transparent film, while an nth order dark fringe will appear at the former central maximum. This shift depends on μ and the thickness of the film.

vi. If Polaroid of the same orientation in polarization plane covers the two slits, interference pattern will be clearly seen, but if the planes of polarization are perpendicular to each other, no interference pattern will be seen.

Examples 5.1

In a two-slit interference experiment, the slits are $0.20\,mm$ apart, the screen at a distance of $1.0\,m$ from the slits. The third bright fringe is $7.50\,mm$ displaced from the central maximum. Calculate the wavelength of light used.

Solution

$d = 0.20\,mm$; $Y_3 = 7.50\,mm$; $n = 3$; $L = 1.0\,m$

From $Y_n = \frac{n\lambda}{d}L$ we can write $\lambda = \frac{dY_3}{3L} = \frac{7.5\,mm x 0.2\,mm}{3x1\,m} = 500\,nm$

Examples 5.2

A radio station operating at a frequency of $1.5\,MHz$ has two identical dipole antennae spaced $400\,m$ apart, oscillating in phase. At a distance much greater that $400\,m$, in what direction is the intensity maximum/minimum in the resulting radiation pattern?

Solution

Since both antennae are emitting electromagnetic waves, the radiation pattern at any point will be a superposition of the waves from the two antennae. For maximum intensities, we recall the equation

$$d\sin\theta_n = n\lambda \Rightarrow \sin\theta_n = \frac{n\lambda}{d},$$

where $d = 400m$; but $\lambda = \frac{c}{f} = \frac{3x10^8}{1.5x10^6} = 200\,m$;

$\Rightarrow \sin\theta_n = n\,x\frac{200}{400} = \frac{n}{2}, n = 0, \pm 1, \pm 2.$

For $n = 0$, $\theta = 0$; $n = \pm 1$, $\theta = \pm 30^o$; $n = \pm 2$, $\theta = \pm 90^o$.

For minima intensities, we use $\sin\theta_n = \frac{n+\frac{1}{2}\lambda}{d} \Rightarrow \sin\theta_n = \frac{n+\frac{1}{2}}{2}$

$\Rightarrow n$ can only be $-2, -1, 0, and\ 1$ with the corresponding angles $\theta = \pm 48.5^o., \pm 14.5^o$

Examples 5.3

A viewing screen is separated from a double slit source by $1.2\,m$. The distance between the two slits is $0.03\,mm$; the second order bright fringe is $4.50\,cm$ from the centerline. Determine the wavelength of the light used and the distance between adjacent fringes.

Solution

$Y_n = \frac{n\lambda L}{d} \Rightarrow \lambda = \frac{0.03mm\ x\ 4.50cm}{2x1.2m} = 560nm$

$Y_{n+1} - Y_n = \frac{\lambda L(n+1)}{d} - \frac{n\lambda L}{d} = \frac{\lambda L}{d} = \frac{(1.2m\ x\ 560nm\ )}{0.03mm} = 2.2\,x10^{-2}m.$

Examples 5.4

A light source emits visible light of two wavelengths $\lambda = 430\,nm\ and\ \lambda' = 510\,nm$. The source is used in a double-slit experiment in which $L = 1.5\,m, and\ d = 0.025\,mm$. Find the distance between the third order fringes.

Solution

For the third order bright fringe, $Y_n = \frac{n\lambda L}{d}$; substituting the values of the parameters given, we have

$Y_3 = 7.74\, x10^{-2}\, m, and\, Y'_3 = 9.18\, x10^{-2} m$,

$\Rightarrow \Delta Y = Y_3 - Y'_3 = 1.4\, x10^{-2}\, m$

Examples 5.5

In a Young double-slit experiment, if the wavelength $\lambda = 640\, nm$, and $d = 0.2\ mm$, initially the distance between screen and double slit is $1.00\, m$. What is the angular separation between two neighboring bright fringes formed on the screen? If the screen is shifted further away from the double slit, what changes will occur in the fringe pattern?

Solution

Recall the equation $Y_n = \frac{n\lambda L}{d}$; and from figure 5.6 and small angle approximation, $\theta_n = tan\, \theta_n = \frac{Y_n}{L}$, which for first order bright fringe becomes after simplification

$$\theta = \frac{\lambda}{d} = \frac{640\ nm}{0.20\ mm} = 3.20\, x10^{-3} rad$$

If L is increased, $\theta = \frac{\lambda}{d}$, remains constant, but $Y_n = \frac{n\lambda L}{d}$ increases as L increases; also, the intensity of the fringe pattern reduces.

5.4 INTENSITY DISTRIBUTION OF DOUBLE-SLIT INTERFERENCE PATTERN

To determine the intensity at the bright regions, we considered the two slits to represent coherent sources of light of sinusoidal waves, such that the two waves from the slits have same angular frequency ω, and a constant phase angle φ. From Maxwell's theory of light, we recall that the total magnitude of the electric field at any point P is a superposition of the two waves at that point (figure 5.7). If $E_1 = E_0 sin\, \omega t, and\, E_2 = E_0 sin\, (\omega t + \varphi)$; $\varphi = 0$ at the slit, but at P, φ depends on the path difference δ, given by $\delta = dsin\, \theta = R_2 - R_1$.

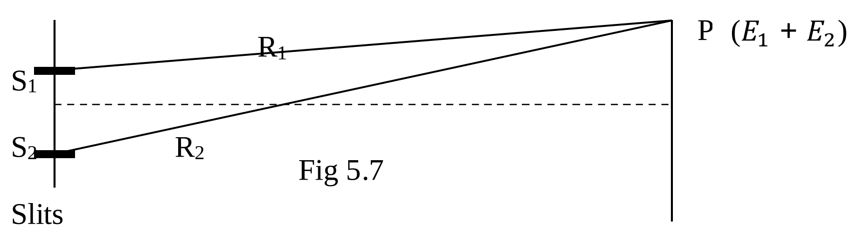


Fig 5.7

But a path difference of one wavelength, corresponds to a phase difference of 2π radian (constructive interference), thus, their ratio is given by

$$\frac{\delta}{\lambda} = \frac{\varphi}{2\pi} \Rightarrow \varphi = \frac{2\pi\delta}{\lambda} = \frac{2\pi dsin\theta}{\lambda} \qquad 5.16$$

Using the superposition principle, the resultant electric field vector E_p at the point P is

$$E_p = E_1 + E_2 = E_0\{\sin \omega t + \sin(\omega t + \varphi)\} \quad 5.17$$

Using trigonometric identity, equation 6.17 can be written as

$$E_p = 2E_0 \sin\frac{(\omega t + \varphi + \omega t)}{2} \cos\frac{(\omega t + \varphi - \omega t)}{2} = 2E_0 \sin\left(\omega t + \frac{\varphi}{2}\right) \cos\frac{\varphi}{2} \quad 5.18$$

The implication of equation (5.18) is that the electric field at the point P has the same frequency ω as the electric field at the slit, but the amplitude is multiplied by a factor of $2\cos(\frac{\varphi}{2})$. To check the consistency of equation (5.18), we evaluate it at various values of φ. When $\varphi = 0, 2\pi, 4\pi$, and even integer multiples of π, then the magnitude of the electric field at P is $E_p = 2E_0$, this shows that the amplitude of the electric field at point P has increased, which indicates constructive interference. If $\varphi = \pi, 3\pi, 5\pi$, and odd integer multiples of π, then the magnitude of the electric field at P is zero; this is consistent with destructive interference.

To obtain an expression for the light intensity I at P, we note that the intensity of a light wave at any point is proportional to the square of the resultant electric field at that point

$$I = E_p^2 = 4E_0 \sin^2\left(\omega t + \frac{\varphi}{2}\right)\cos^2\frac{\varphi}{2} \quad 5.19$$

Most light detecting instruments measure time-average, thus integrating the sine component of equation (5.19) over the full-cycle with period T, we have

$$I^2 = I_0^2 \frac{1}{T}\int_0^T \sin^2\left(\omega t + \frac{\varphi}{2}\right) \mathrm{dt} \quad 5.20$$

Using $\sin^2\theta = \frac{1-\cos 2\theta}{2}$, equation (5.20) can be written as

$$\sin^2\left(\omega t + \frac{\varphi}{2}\right) = \frac{1}{T}\int_0^T \left\{\frac{1}{2} - \frac{\cos^2\left(\omega t + \frac{\varphi}{2}\right)}{2}\right\} \mathrm{dt} = \frac{1}{2} \quad 5.21$$

Thus, we can write that the average intensity at the point P is

$$I = I_0 \cos^2\frac{\varphi}{2} \quad 5.22$$

where I_0 is the maximum intensity at the screen. Substituting the value of φ given in equation (5.16), equation (5.22) can be written as

$$I = I_0 \cos^2\left\{\frac{\pi d \sin\theta}{\lambda}\right\} = I_0 \cos^2\left\{\frac{Y d\pi}{\lambda L}\right\} \quad 5.23$$

We have used the fact that for small angle approximation, $\theta = \sin\theta = \frac{Y}{L}$. The plot of light intensity versus $d\sin\theta$, shows that the interference pattern consists of equally spaced bright fringes of equal width. The analysis of the intensity distribution can also be achieved using phasor addition of waves (figure 5.8).

Let $E_1 = E_0 \sin\omega t, E_2 = E_0 \sin(\omega t + \varphi)$ and $\theta = \pi - \varphi$.

Using cosine rule for the addition of vectors,

$$E_r^2 = E_0^2 + E_0^2 - 2E_0^2 \cos(\pi - \varphi) = 2E_0^2(1 + \cos\varphi)$$

5.24

Where we note that $(1 + \cos\varphi) = 2\cos\frac{\varphi}{2}$ from trigonometry. Simplifying, we get the magnitude at any point P as

$$E_r = 2E_0 \cos\frac{\varphi}{2} \qquad 5.25$$

Thus, at any point in time, the electric vector E_0 at any point P is

$$E_p = E_r \sin\left(\omega t + \frac{\varphi}{2}\right) = 2E_0 \cos\frac{\varphi}{2} \sin\left(\omega t + \frac{\varphi}{2}\right) \qquad 5.26$$

An alternative route to the calculating the intensity at any point P is by considering the phase difference and the path difference. Consider when the two waves are in phase, $\varphi = 0^o$ or 360^o, when out of phase $\varphi = 180^o$. Thus, the ratio of the phase difference φ to 2π should be equal to the ratio of the path difference.

$$\Rightarrow \frac{\varphi}{2\pi} = \frac{d \sin\theta}{\lambda}; \Rightarrow \varphi = \frac{2\pi d\sin\theta}{\lambda} = k\, d\sin\theta \qquad 5.27$$

Where $k = \frac{2\pi}{\lambda}$ is called the wave number (unit m^{-1}).

For maximum intensity, $\cos\varphi = \pm 1, \Rightarrow I = I_0 \cos^2(½ k d\sin\theta)$

$$\Rightarrow ½ k d\sin\theta_n = \frac{1}{2}\frac{2\pi}{\lambda} d\sin\theta_n = n\pi \Rightarrow d\sin\theta_n = n\lambda \qquad 5.28$$

While for intensity minima,

$$\cos\varphi = 0 \Rightarrow \sin\theta_n = n(1 + ½) \Rightarrow d\sin\theta_n = n\lambda(1 + ½) \qquad 5.29$$

Equation (5.28) and (5.29), were obtained initially by geometric consideration.

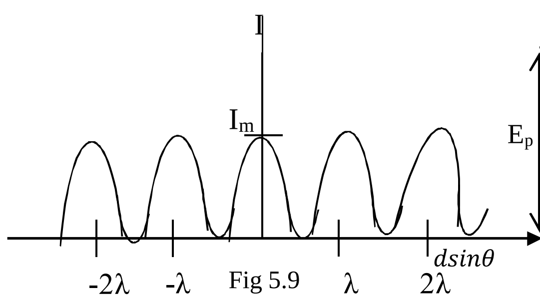


Fig 5.9

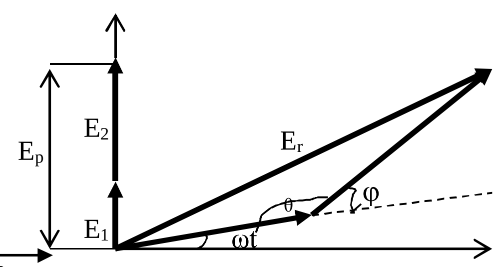


Fig 5.8

Examples 5.6

If two similar antennae oscillating at $60\,MHz$ are separated by 10m, have intensity $I_0 = 0.020\,Wm^{-2}$, at a distance of $700\,m$ in a positive x-direction, what is the intensity in the direction 40°? In what direction is the intensity $\frac{I_0}{2}$? In what direction is the intensity zero?

Solution $\lambda = \frac{c}{f} = \frac{3x10^8}{60x10^6} = 5.0\,m$

$$I = I_0 \cos^2\left(\frac{\pi d\sin\theta}{\lambda}\right) = 0.020\cos^2\left(\frac{\pi * 10\sin 40}{5}\right) = 0.164\,Wm^{-2}$$

This is significantly different (about 82%) from the value when $\theta = 0$.

For $I = \frac{I_0}{2}, \Rightarrow \cos^2\frac{\varphi}{2} = \frac{1}{2}$ or $\cos\frac{\varphi}{2} = \frac{1}{\sqrt{2}} \Rightarrow \varphi = \pm\frac{\pi}{2}$

$\Rightarrow 2\pi \sin\theta = \pm\frac{5}{40} \Rightarrow \sin\theta = \pm \Rightarrow \theta = \pm 7.18^o$

For $I = 0$, then $cos^2\left(\frac{\pi d sin\theta}{\lambda}\right) = 0 \rightarrow sin\theta = \pm\frac{5}{40} \rightarrow \theta = \pm 14.48^o$

and this is only possible when $cos(2\pi sin\theta)$= 0 or $2\pi sin\theta = \frac{\pi}{2}$ and $\pm\frac{3\pi}{2}$

$\Rightarrow for\frac{\pi}{2}, \ \theta = 14.48^o$ and

$for\frac{3\pi}{2}, \ \theta = 48.59^o$; $while\ for -\frac{3\pi}{2}, \theta = -48.59^o$ respectively.

Examples 5.7

In a Young double-slit experiment, the slit separation is $0.05\,cm$, and the distance between the double slit and the screen is $200\,cm$, when blue light is used, the distance of the first order bright fringe from the central bright fringe is $0.13\,cm$. Calculate the wavelength of the blue wavelength and the distance of the fourth order dark fringe from the center maximum.

Solution

$d = 0.05\,cm, L = 200\,cm, Y_1 = 0.13\,cm$, and when can write

$\lambda = \frac{Yd}{nL} \Rightarrow \lambda = \frac{0.05x0.13}{200}cm = 325\,nm$;

and for the fourth dark fringe,

$n = 4, Y = \left(n + \frac{1}{2}\right)\frac{\lambda L}{d} = \frac{4.5\ x\ 325\ nm\ x\ 200\ cm}{0.05\ cm} = 5.85\,x10^{-3}m$

5.5 THREE-SLIT INTERFERENCE PATTERN

Using phasor diagrams, we can also analyze the interference pattern caused by three equally spaced slits. Let the electric field component at any point P of the resultant wave, representing the light waves from the three slits be given by

$E_1 = E_0\,sin\omega t,\ E_2 = E_0 sin\,(wt + \varphi), and\ E_3 = E_0\,sin\,(\omega t + 2\varphi)$,

where φ is the phase difference (figure 5.11).

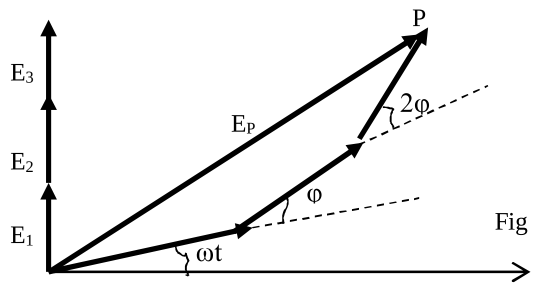


Fig 5.11

Using phasor diagrams, the values of E_p (for constructive interference) for various values of φ are shown in figure 5.12. The resultant magnitude of the electric field at the point P has a maximum value of $3\,E_0$ when $\varphi = 0, \pm 2\pi, \pm 4\pi, ...$ These points are called the primary maxima, also we find secondary maxima of amplitude of value E_0 occurring when $\varphi = \pm\pi, \pm 3\pi$ etc. For these points, the wave from one slit exactly cancels out that from another slit, with the implication that only light from the third slit contributes to the resultant. Total destructive interference occurs whenever the three phasors forms a close triangle (figure 5.12c), and this happens for $\varphi = \pm\pi/3, \pm 2\pi/3, \pm 4\pi/3$ etc.

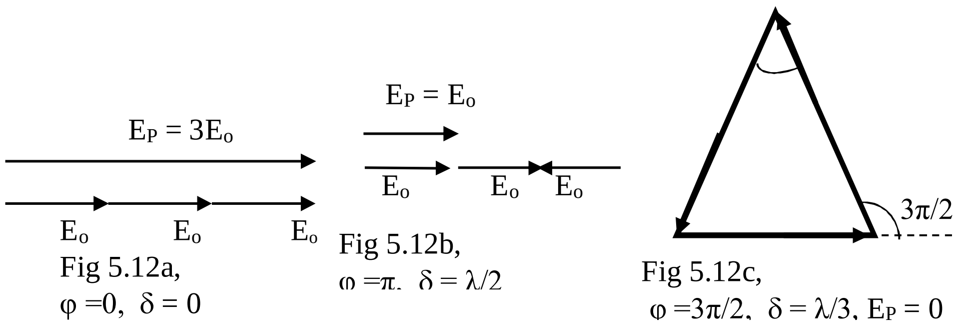


Fig 5.12a, φ =0, δ = 0

Fig 5.12b, φ =π, δ = λ/2

Fig 5.12c, φ =3π/2, δ = λ/3, EP = 0

The measurement of radiation intensity for three-slits interfernce fringe pattern indicates that the primary maxima are nine times more intense than the secondary maxima. this is due to the fact that intensity varies as E^2, where E is the magnitude of the electric field where the intensity is being measured. Thus, for N-slits, the intensity of the primary maxima is N^2 times greater than that due to a single slit, and as the number of the slits increases, the primary maxima intensity increases but narrows in width, while the secondary maxima decreases in intensity but its number between primary maxima increases (figure 5.13). In fact, the number of secondary maxima between adjacent primary maxima is N – 2, where N is the number of slits (this is employed in diffraction grating).

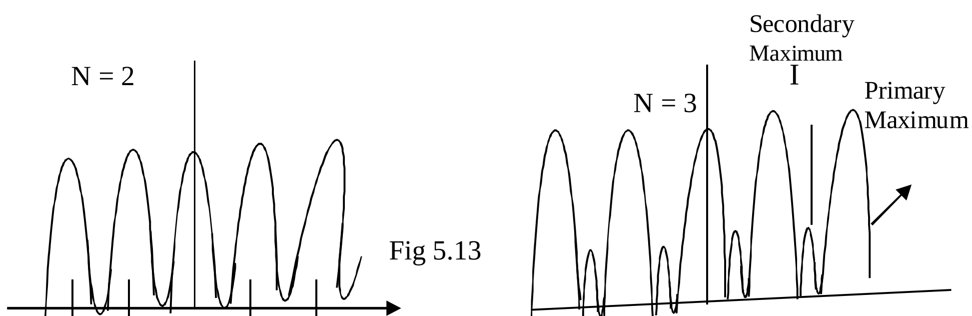


Fig 5.13

5.6 FRESNEL BIPRISM

After the double-slit experiment was performed by Young, due to the fact that corpuscular theory of light was more establish then, objections where raised that the fringe patterns he observed were probably due to some complicated modification of the light by the edges of the slits and not due to true interfernce. Not long after, Fresnel performed several experiments in which the interference of light was proved beyond doubt, two of the optical instruments he used are the Fresnel Biprism and Fresnel double mirror. The Fresnel biprism is an optical arrangement that consists of two thin prisms joined at their bases, and can produce interference fringes. A single spherical wave front impinges on both prisms; the top portion of the wave front is refracted downwards, while the lower segment is refracted upwards. When the two refracted waves superpose, interference occurs and this can be viewed on a screen.

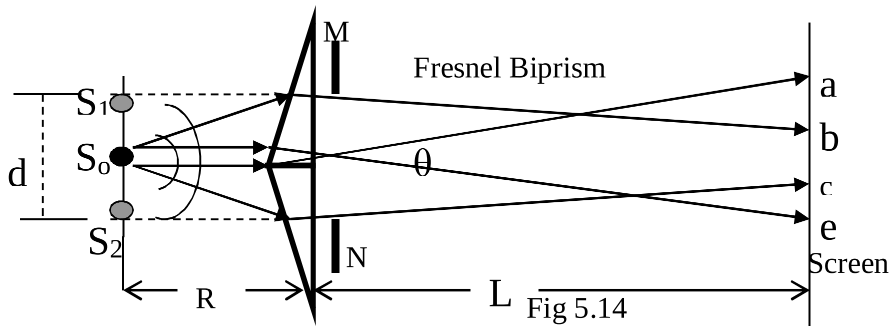


Fig 5.14

The schematic diagram of Fresnel biprism is shown in figure 5.14, the thin double prism refracts the light from the slit sources S into two overlapping beams *ac* and *be*. With the screen *MN* placed as shown, interference fringes are observed only in the region *bc*. When the screen *MN* is removed from the light path, the two beams will overlap the whole region *ae*. Just as in Young's double-slit experiment, the wavelength of light can be determined from measurements of the interference fringes produced by the biprism. Let L be screen distance, and R source distance from the biprism respectively; d the distance between the virtual sources S_1 and S_2; and ΔY the separation between adjacent fringe pattern. Then the wavelength of light used can be shown to be given by $\lambda = \frac{d\,\Delta Y}{R + L}$

5.28

In order to find d, one can measure the their angular separation, and assume to a good approximation that $d = R\theta$

5.7 FRESNEL DOUBLE MIRROR

Fresnel double mirror consists of two plane front-silvered mirrors inclined to each other at a very small angle (figure 5.15). One portion of the wave front coming from the slit S is reflected from the first mirror, while another portion of the wave front is reflected from the second mirror. An interference field exists in space in the region where the two reflected waves superpose. The analysis of the interference pattern can be better appreciated if we consider the images of S_1 and S_2 of the slit S in the two mirrors to be two coherent sources separated by a distance d. It follows from the laws of reflection, $SA = S_1A, SB = S_2B$, so that $SA + AP = R_1$ and $SB + BP = R_2$, then, the optical path-length difference δ between the two rays is simply $\delta = R_1 - R_2$. As the case with Young's double-slit interference, the bright maxima occur when the $n\lambda = R_1 - R_2$, and the separation of the fringes is given by

$$\Delta Y = \frac{\lambda L}{d} \qquad 5.29$$

Equation 5.29 can be used to calculate the λ of the light used

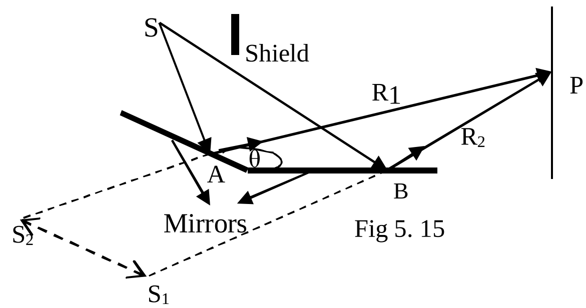


Fig 5. 15

5.8 LLOYD'S MIRROR

Another simple way of producing interference pattern with a single mirror is by the use of Lloyd's mirror. Here, a light source S is placed at a point close to a mirror and a viewing screen is position at a distance $L >> d$ (figure 5.16), and at right angle to the mirror. Light waves can reach the point P on the screen, either by direct path SP or by the path involving reflection from the mirror S_1P. The reflected ray can be treated as a ray originating from the virtual source S_1. This arrangement is analogous to double-slit and at the screen, interference fringes are observed. However, the position of the dark and the bright fringes are reversed relative to the pattern created by double-slit experiment. This is because the coherent source at S and S_1 differ in phase by $\varphi = \pi$, which is a phase change introduced by reflection. This must be the reason, since if path difference alone were responsible for the interference pattern, then we expect to see bright fringe at center point (where $\delta = 0$); but since a dark fringe is observed at the point P^1, a phase change is introduce by reflection. Thus, an electromagnetic wave undergoes a phase change of $\varphi = \pi$ upon reflection from a medium that has a higher refractive index than the one in which it was propagating.

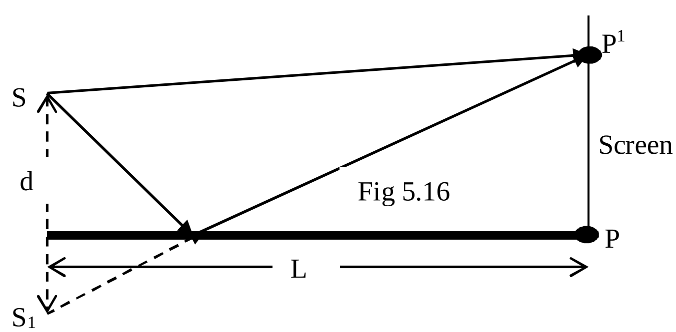


Fig 5.16

5.9 INTERFERENCE IN THIN FILMS

Interference effects are commonly observed in thin films such as layer of oil in water, thin surface of soap bubble etc. The various colors observed when white light is incident on such films results from the interference of light waves reflected from the two surfaces of the film. Consider a film of uniform thickness t and refractive index μ, to determine whether the reflected rays interfere constructively or destructively, we note the following

(1) A wave traveling from a medium of refractive index μ_1 towards a medium of refractive index μ_2, undergoes a phase change $\varphi = \pi$ upon reflection provided that $\mu_2 > \mu_1$ and undergoes no phase change $(\varphi = 0)$ if $\mu_2 < \mu_1$

(2) We define the wavelength, $\lambda' = \lambda/\mu$ to be the wavelength of light in a medium of refractive index μ, if the wavelength of the light in free space is λ.

In figure 5.17, ray 1 is π out of phase with ray 2, this is equivalent to a path difference of $\delta = \frac{\lambda'}{2}$ but note that ray 2 traveled a further distance of $2t$ before the two waves can recombine at a point say P above the thin film. If $2t = \frac{\lambda'}{2}$, then ray 1 and 2 recombine in phase and the result is constructive interference. Thus, for thin films, the condition for constructive interference is

$$2t = (n + ½)\,\lambda' \qquad 5.30$$

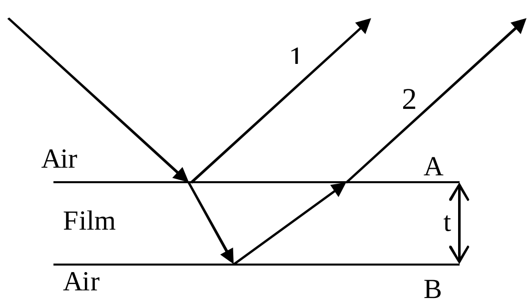
where $n = 0, 1, 2, \ldots$ Equation (5.30) takes into account two important factors

(1) The difference in path length due to thickness of the thin film.

(2) The $\varphi = \pi$ phase change upon reflection

Recall that $\lambda' = \lambda/\mu$, thus, equation (5.30) can be written as

$$2\mu t = (n + ½)\,\lambda \qquad 5.31$$



Ray 1 undergoes a phase change φ = π, but ray 2 propagating from a medium with if $\boldsymbol{\mu_2 > \mu_1}$, undergo… hange (φ = 0)

Fig 5.17

When the extra distance traveled by the by ray 2 corresponds to an integer multiple of λ', the two waves combine out of phase, and the result is destructive interference. Thus, the general condition necessary for destructive interference in thin films can be expressed as

$$2\mu t = n\,\lambda \qquad 5.32$$

The above conditions for constructive and destructive interference in thin films are only valid when the medium above the top surface of the film is the same as the medium below the bottom surface, so that even if the medium has a greater or lesser index of refraction, the reflected rays from the two surfaces are still out of phase by $\varphi = \pi$. If the film is placed between two media, one with $\mu_1 < \mu_{film}$, and the other with $\mu_2 > \mu_{film}$ then the conditions for constructive and destructive interference are reversed. In this case, either there is a phase change of $\varphi = \pi$ for both ray1 reflected from surface A and ray 2 reflected from surface B (figure 5.17), or there is no phase change for either ray; hence, the net change in phase due to reflection is zero. In such situation, the interference patterns observed are due to differences in path length.

5.10 NEWTON'S RING

When a plano-convex lens is placed on top of a glass surface, an interference of light waves occurs. In this arrangement, the air film between the glass surfaces varies in thickness from zero at the point of contact to t at the point P (fig 5.18). If the radius of curvature R of the lens is much greater than the distance r, and if the system is viewed from above using a monochromatic light of wavelength λ, a pattern of circular bright and dark fringe is observed.

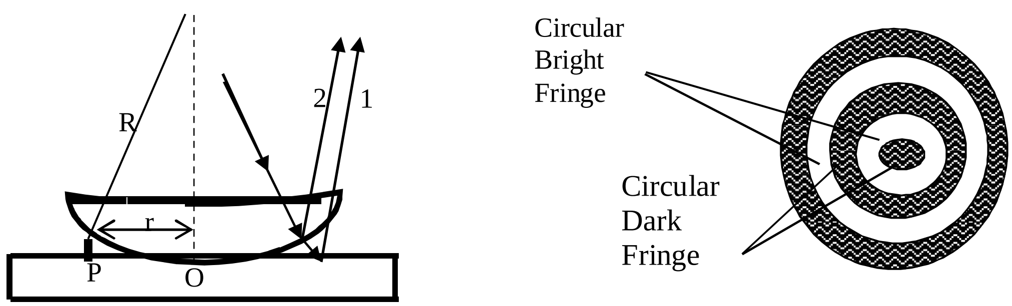


Fig 5.18. The combination of rays reflected from the flat plate and the curved lens surfaces gives rise to an interference pattern shown and called Newton's ring

These circular fringes are called Newton's ring. The interference effect is due to the combination of ray 1 reflected from the plate, with ray 2 reflected from the curved surface of the lens. Ray 1 undergoes a phase change of $\varphi = \pi$ upon reflection since $\mu_{glass} > \mu_{air}$ (total internal reflection); whereas ray 2 undergoes no phase change (because it is reflected from a medium of lower refractive index), hence condition for constructive interference is the same as equation (5.31), while for destructive interference, equation (5.32) holds. Using the geometry of the figure, we can obtain an expression for the radii of the bright and dark bands in terms of the radius of curvature R and wavelength λ, considering the right angled triangle (figure 5.18),

$R^2 = r^2 + (R - t)^2 \Rightarrow r^2 = 2Rt - t^2, \Rightarrow r_n^2 = 2Rt_n$

since $R \gg t$ and for destructive interference (nth dark fringe), $t_n = \frac{n\lambda}{2\mu}$

$$\Rightarrow r_n^2 = \frac{n\lambda R}{\mu}; \Rightarrow r_n = \sqrt{n\lambda R} \qquad 5.33$$

Where we have used the fact that $\mu = 1$ for air.

For bright fringes, $t = \frac{\left(n+\frac{1}{2}\right)\lambda}{2\mu} \Rightarrow r_n = \sqrt{\{(n + ½)\,\lambda R\}}$ 5.34

Example 5.8

Calculate the minimum thickness of a soap bubble film with μ = 1.33 that results in constructive interference in reflected light, if the film is illuminated with light of wavelength in free space $\lambda = 600\,nm$.

Solution

The minimum thickness corresponds to $n = 0, \Rightarrow 2\mu t = (n + ½)\lambda, \Rightarrow t = \frac{\lambda}{4\mu} = 112.78\,nm$

Other film thickness that will produce constructive interference correspond to $n = 1, 2, \ldots$

$$\Rightarrow t_n = (2n + 1)x\,112.78\,nm.$$

One major application of interference by thin films is in non-reflective coatings in camera, vehicle rear light, and solar cell panels (solar cell panels are often coated with transparent thin films of SiO with $\mu = 1.45$ to minimize reflective losses from the surface of the Si solar cell, where a reduction in loss of about 90% is achievable.)

Example 5.9

A possible way for making an airplane invisible to radar is to coat the plane with an anti-reflective polymer. If radar waves have a wavelength of $3.00\,cm$, and the refractive index of the polymer is $\mu = 1.50$. How thick would you make the coating?

Solution

For the plane to be invisible, we need the condition for destructive interference.

$$\Rightarrow t = \frac{n\lambda}{2\mu} = \frac{3.00}{2\,x\,1.50} = 1.00\,cm\,for\,n = 1.$$

5.11 DIELECTRIC FILMS

Dielectric film can be described as thin film (mostly non-metallic) with thickness of the order of about $10^{-3}\ m \sim 10^{-6}\ m$. Example of dielectric film include, soap films, thin reflective coating, etc. Let's consider in details the nature of interference fringes produce by thin film of thickness t and refractive index μ, when light is incident on it. Suppose that the thin film is non-absorbing and that the reflection coefficients at the interfaces are low so that only the first two reflected beam E_1 and E_2 (having been reflected once) need be considered. Let S be a monochromatic point source, and the film serving as an amplitude splitting device, so that E_1 and E_2 may be considered as arising from two virtual sources lying behind the film (figure 5.19). The reflected rays are parallel on leaving the film and can be brought together at a point through a convex lens (observation through a telescope achieves this).

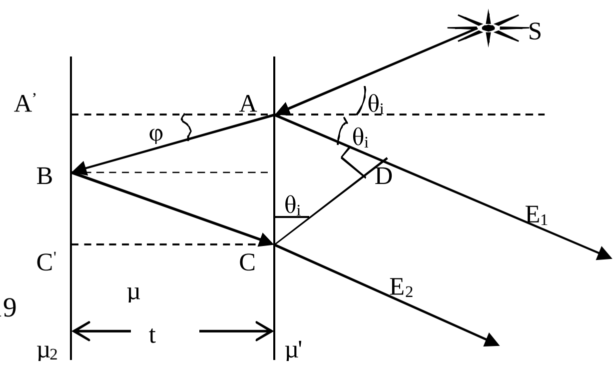


Fig 5.19

The path difference for the two beams E_1 and E_2 is given by

$$\delta = \mu\{AB + BC\} - \mu' AD \qquad 5.35$$

But $\frac{A'B}{t} = tan\varphi \Rightarrow AB = BC = \frac{t}{cos\varphi}$ 5.36a

Also $AD = AC\, sin\theta_i$.from Snell's law,

$$\mu\, sin\varphi = \mu' sin\theta_i \Rightarrow sin\theta_i = \frac{\mu sin\varphi}{\mu'} \Rightarrow AD = AC\frac{\mu sin\varphi}{\mu'} \qquad 5.36b$$

Note that $A'B = BC' \Rightarrow AC = 2\,A'B$;

but $A'B = t\, tan\varphi \Rightarrow AC = 2t\, tan\varphi \Rightarrow AD = 2t\, \frac{\mu sin\varphi}{\mu'} tan\varphi$ 5.36c

Substituting equations (5.36a), (5.36b) and (5.36c) into equation (5.35), we have

$$\delta = \frac{2\mu t}{cos\varphi} - 2\mu` t\, tan\,\varphi\, \frac{\mu sin\theta}{\mu'} = \frac{2\mu t}{cos\varphi}\{1 - sin^2\varphi\} = 2\mu t cos\varphi \qquad 5.37$$

The corresponding phase difference associated with the path length difference is just the product of the wave number *k*, and the path difference δ i.e. $\psi = k\delta$. If the film is immersed in a single medium, $\mu_2 = \mu'$; $if\ \mu' < \mu$, (recall the soap film in air) or $\mu > \mu'$ (recall Newton's ring apparatus) or an air film between two sheets of glass; in all cases there will be an additional phase shift arising from reflection themselves, and thus will experience a relative phase shift of π,

$$\Rightarrow \psi = k\delta \pm \pi = \left(\frac{2\pi}{\lambda} * 2\mu t cos\,\varphi\right) \pm \pi = \frac{4\pi\mu t\, cos\varphi}{\lambda} \pm \pi \qquad 5.38$$

But $cos\varphi = \{1 - sin^2\varphi\}^{1/2}$; from equation (5.36b), $sin\varphi = \frac{\mu' sin\theta_i}{\mu}$, we now substitute these expressions into equation (5.38) to get

$$\psi = \left\{\frac{4\pi\mu t}{\lambda}\sqrt{1 - \frac{\mu'^2}{\mu^2} sin^2\theta_i}\right\} \pm \pi = \left\{\frac{4\pi t}{\lambda}\sqrt{\mu^2 - \mu'^2 sin^2\theta_i}\right\} \pm \pi \qquad 5.39$$

Equation (5.39) is the general phase difference, when light is reflected from thin film. In reflected light, an interference maximum occurs at the screen when $\psi = 2n\pi$ or even integer multiples of π, then from equation (5.38) we have

$$\psi = \frac{4\pi\mu t cos\varphi}{\lambda} \pm \pi = 2n\pi \Rightarrow t\cos\varphi = \frac{(2n+1)\lambda}{4\mu} = \frac{(2n+1)\lambda'}{4} \qquad 5.40$$

Where we have used only the negative sign, and $\lambda' = \frac{\lambda}{4}$, for interference minima,

$$\psi = (2n+1)\pi, \Rightarrow t\cos\varphi = \frac{n\lambda'}{2} \qquad 5.41$$

Notice that equation (5.40) and (5.41) are the same as equations (5.31) and (5.32).

5.12 INTERFEROMETERS

Interferometers are optical instruments that produce interference by splitting a light beam into two parts and recombining them again to produce interference pattern. In general, there are two basic types of interferometers – wave-splitting interferometers, which include Fresnel double mirror, Fresnel biprism, Young double-slit apparatus, Lloyd's mirror; and amplitude splitting interferometers, which include Michelson interferometer, Mach-Zehnder interferometer, and Sagnac and Pohl interferometers.

5.12. 1. Michelson Interferometer

The schematic diagram of Michelson interferometer is shown in figure 5.20. This interferometer can be used to measure lengths with great precision. A beam of light from a monochromatic source is split into two rays by the mirror M (the beam split, since it is a half-silvered mirror and transmit half of the light incident on it and reflects the other half), which is inclined at $45°$ to the incident beam. The reflected beam propagates vertically upwards to M_1, while the transmitted beam propagates horizontally to M_2. Hence, the two beams of light travel separate distances L_1 *and* L_2. After reflecting from M_1 *and* M_2, the two beams eventually recombine at *M* to produce interference pattern, which can be viewed through a telescope. The glass plate P, equal in thickness to mirror *M*, is placed in the path of the horizontal beam to ensure that the two returning beams travel the same thickness of glass.

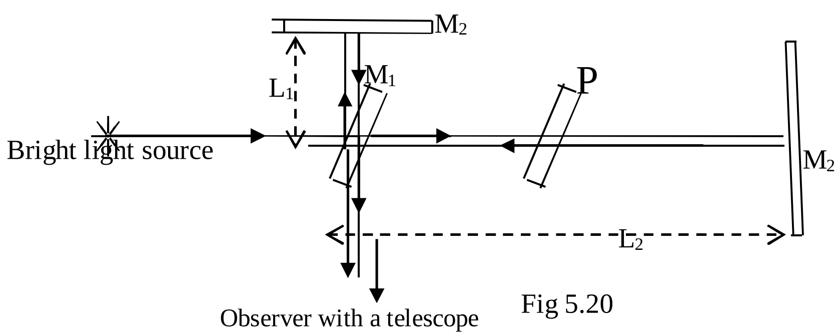


Fig 5.20

The interference condition for the two beams is determined by their path length differences and for such amplitude splitting interferometers, the path difference is given by $\delta = 2d\sin\theta = n\lambda$ and destructive interference occurs at the center (this time circular fringe). M_1 is adjustable, and as it is adjusted, the fringe pattern shifts. For example, if a dark fringe appears in the field of view (corresponding to destructive interference), and M_1 is moved a distance $\lambda/4$ towards M, the path difference changes by $\lambda/2$, and bright fringe appears on the field of view. Thus, the fringe shifts by one-half fringe each time M_1 moves a distance of $\lambda/4$. The wavelength of light used can be determined by counting the number of fringe shifts for a given displacement of M_1.

5.13. MULTIPLE BEAM INTERFERENCE

We have examined situations in which two coherent beams have combined to produce interference fringes; however, other circumstances exist under which a much larger number of mutually coherent waves are made to interfere. For instance, if the amplitude-reflection coefficients of the medium are not small, higher order reflected waves becomes quite significant, an example is a glass plate slightly silvered on both side have reflection coefficient nearly 1, and will generate a large number of multiple internally reflected rays.

Some of the most beautiful effects of interference result from the multiple reflection of light between the two surfaces of a film of transparent material. For instance, the brilliant colors in peacock's feathers are due to interference of reflected light from the multiplayer structure of the feathers, the colors of butterflies, multi-colors displays by thin oil film floating in water are all as a result of interference from multiple reflections. To understand this class of interference, we begin by considering idealized case of reflection and refraction from the boundary separating the different optical media. Let us denote a ray of light in free space incident on a plane surface of a transparent medium by *A*, if *T* and *R* are the transmission and reflection coefficient respectively, then the reflected and refracted rays are indicated by *AR* and *AT* respectively or simply in terms of ratio by R and T respectively (we will use this to simplify our notation).

We examined the case of non-absorbing medium; our interest from the standpoint of physical optics is to evaluate the change of phase that will result when waves are reflected from different boundaries. This change of phase will differ as whether the waves are propagating from regions of higher refractive indices to regions of lower refractive indices or vice-versa. Let's represent a ray reflected once by *R*, twice by *RR*, etc.; and a ray of light transmitted/refracted once by *T*, twice by *TT*; and we prime the quantities to differentiate between quantities by different media (figure 5.21)

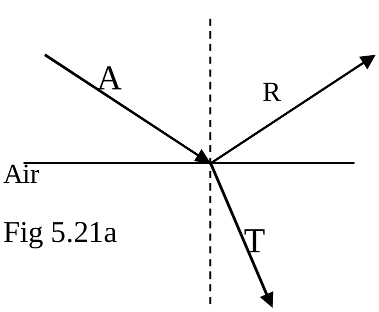


Fig 5.21a

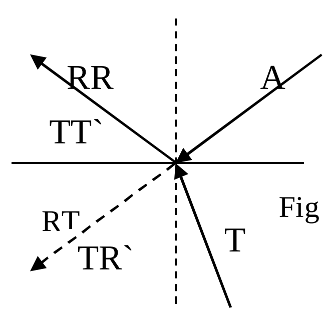


Fig 5.21b

Using the notation established, the amplitude of the incident ray is *A*, the amplitude of reflected ray *R*, and the amplitude of the refracted ray is *T* (figure 5.21a). Applying the principle of reversibility, the wave of amplitude R gives a reflected wave of amplitude *RR*, and refracted wave of amplitude *RT*. If we denote the fractions of the amplitude of reflected and refracted ray traveling from below $R`$ and $T`$, then this contributes amplitudes *TT'* and $TR`$ to the waves (figure 5.21b). From conservation principle, the resultant effect must consist of a wave in air of amplitude *A*, thus we have

$$TT' + RR = A \quad and \quad RT + TR' = 0 \qquad 5.42$$

Which implies (we have assumed A=1, as we work in ratio)

$$TT` = 1 - R^2 \; and \; R' = -R \qquad 5.43$$

These equations we first obtained by Stokes and are called Stokes' relation. These relations show that the reflectance (fraction of the intensity reflected) is the same for a wave incident from either side of the boundary, since the negative sign disappears upon squaring the amplitudes. The difference in sign of the amplitudes (equation 5.43) indicates a difference in phase of π between the two cases, since a reversal of sign means a displacement in the opposite sense. If there is no phase change on reflection from above, there must be phase change of π on reflection from below and vice-versa.

5.13.1 Reflection From A Plane-Parallel Film

Let a ray of light from a source S be incident on the surface of a plane film (non-absorbing) at point Q, part of this ray will be reflected as ray 1 and part refracted in the direction QF. Upon arrival at F, part will reflected to B and part refracted towards H. At B the ray will again be divided (figure 5.22), and a continuation of this process yields two sets of parallel rays on either sides of the film. In each of these sets, the intensity decreases from one ray to the next. If the set of parallel-reflected rays is focused at a point, each ray would have traveled a different distance, and the phase relations may be such as to produce either constructive or destructive interference, such interference produces the various colors of thin films when viewed by the eye.

In order to find the phase difference between these rays, we must first evaluate the difference in optical path transverse by a pair of successive rays, say ray 1 and 2. This has been established in equation (5.37) as $\delta = 2\mu t cos\, \varphi$. If this path difference is a whole number of wavelengths, we might expect ray 1 and 2 to arrive at the point of superposition say P in phase to produce an interference maximum, however, we must take account of the fact that ray 1 undergoes a phase change of π at reflection, while ray 2 does not. Thus, the relation $2\mu t cos\, \varphi = n\lambda$ $(n = 0, 1, 2, \ldots$ $and\ \lambda$ wavelength of light used) becomes a condition for destructive interference as far as ray 1 and 2 are concerned. Next we examine the phases of the remaining rays 3, 4, 5, …. Since the geometry is the same, the path difference between rays 2 and 3 will also be given by $\delta = 2\mu t cos\, \varphi$, but here there are only internal reflection involved, so that if $2\mu t cos\, \varphi = n\lambda$ is fulfilled, ray 3 will be in phase with ray 2 and interference maximum will occur. The same holds for all succeeding pairs, and we conclude that under these conditions rays 1 and 2 will be out of phase (interference minima) with rays 2, 3, 4, 5,…, will be in phase with each other.

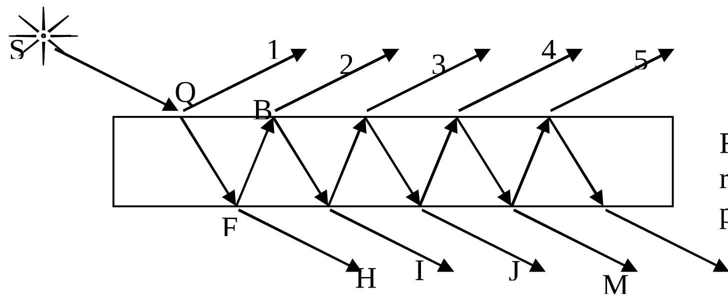


Fig 5.22. Multiple reflections in a plane parallel film

On the other hand, if the condition is such that $2\mu t cos\,\varphi = (n + ½)\lambda$, rays 1 and 2 will be in phase but 3, 5, 7, … will be out of phase with 2, 4, 6, …. Since ray 2 is more intense than ray 3, 4 more than 5 etc, these pairs will not cancel out each other, and the stronger series combine with ray 1, the strongest of all, these will produce intensity maxima. For the intensity minima, ray 2 is out of phase with ray 1, but ray 1 has considerable greater amplitude than ray 2, so that these two will not completely cancel out. We now show that the addition of rays 3, 4, 5, …, which are all in phase with ray 2 will give a net amplitude just sufficient to produce complete darkness at the minima. Using our usual notations (this time retaining the A as the initial amplitude of the incident light ray and AR as the amplitude of the once reflected light and AT as the amplitude of once transmitted/refracted light), and in accordance with equation (5.43), we have taken the fraction reflected internally and externally to be the same. Adding the amplitudes of all the rays except the first on the upper side (ray 2) of the film, we obtain the resultant amplitude A_R as

$$A_R = ATRT` + ATR^3T` + ATR^5T` + ATR^7T` + \ldots$$
$$= ATRT`\{1 + R^2 + R^4 + R^6 + R^8 + \ldots\} \qquad 5.44$$

Since R is necessarily less than 1, the geometrical series in parentheses has a finite sum given by

$$\{1 + R^2 + R^4 + R^6 + R^8 + + \ldots\} = \frac{1}{1-R^2} \Rightarrow A_R = \frac{ATRT\prime}{1-R^2} \qquad 5.45$$

but $TT` = 1 - R^2$ (equation 5.43), thus we obtain finally $A_R = AR$, this is just the amplitude of the first reflected ray, thus we conclude that under the condition $2\mu t cos\,\varphi = n\lambda$, there will be complete destructive interference.

5.13.2 Interference in the Transmitted Light

The rays emerging from the lower side of the film (figure 5.22) i.e. is rays H, I, J, M, and N, can also be brought to a point and cause to interfere. However, there are no phase changes at reflection for any of these rays, and the relation $2\mu t cos\,\varphi = n\lambda$, now becomes the condition for interference maxima, while the relation

$$2\mu t cos\varphi = (n + ½)\lambda \qquad 5.46$$

becomes the condition for interference minima. For interference maxima, the rays H, I, J, …, are all in phase, while for interference minima, the rays I, M …, are out of phase with H, J, …. Since the reflection coefficient (reflectance) might be small, the amplitude of ray H is much greater than ray 1 etc., thus, there is no complete annulment and the interference minima are not completely dark.

5.13.3 The Fabry-Perot Interferometer

The Fabry-Perot interferometer is a multiple beam interferometer that utilizes the fringes produced in the transmitted light after multiple reflections in the air gap between two plates thinly silvered on the inner surfaces. Fabry-Perot interferometer is an important optical instrument that can be used as a spectroscopic device of extremely high resolving power, and can serve as a basic laser resonant cavity. The device consists of two plane-parallel-highly reflecting surfaces (partially silvered mirror) separated by a distance t. The enclosed air gap generally varies from several mm to several cm (when the apparatus is used as an interferometer), but when serving as a laser resonant cavity, the air gap increases much more.

A broad source of bright monochromatic light emits light rays; only one ray (S_1) on the source is traced through the interferometer. Entering by the way of the partially silvered plate, it is multiple reflected with the gap. The transmitted rays are collected by a lens and brought to a focus at *P* on a screen. Any other ray emitted from a different point S_2 parallel to the S_1 and in the same plane of incidence, will form a spot at the same point *P* on the screen. The multiple light waves generated in the cavity, arriving at *P* from either S_1 *or* S_2 are coherent and in phase amongst themselves, but the rays arising from S_1 are completely incoherent with respect to those from S_2, so there is no sustained mutual interference; but the ray from S_1 incident at an angle θ produces a series of parallel rays at the same angle, which are brought together at the point *P* on the screen. The condition for reinforcement of the transmitted rays is $2t\mu\, cos\theta = n\lambda$. This condition will be fulfilled by all point on a circle through *P* (forming a circular bright fringe) centered on the middle of the screen. When the angle θ is decreased, the cosine will increase until another maximum is reached for which *n* is greater by integer values, so that we have for the maxima a series of concentric rings on the screen.

5.14 INTERFERENCE FILTERS

Thin films, both single and multi-layers have found wide variety of optical usage and these include

(1) Coating to eliminate unwanted refections off a diversity of surface.
(2) Usage as multi-layer, non-absorbing beam splitter and dichroic mirror (color selective beam splitter)

(3) Coating of solar cells, astronomers kit (helmet, visor) with heat control covering.

(4) Application in multi-layer and narrow band-pass filters that can transmit only over a specific spectral range. For example, in visible light, thin film play an important role in splitting up the image in color television cameras.

(5) In infrared region, thin films are used as filters in missile guidance systems

(6) Thin films are also used in CO_2 lasers and remote-sensing satellite sensors.

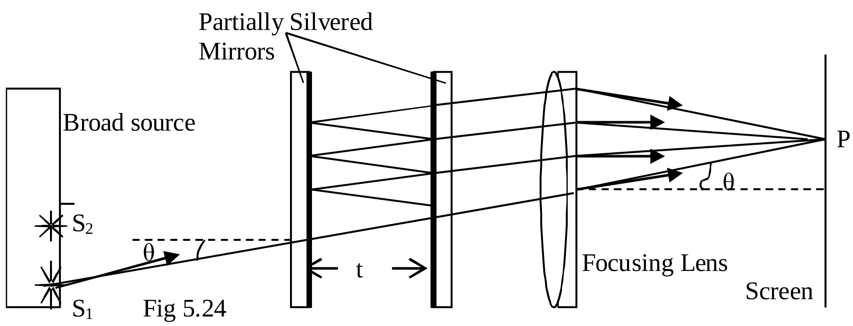


Fig 5.24

Exercises Five

1. Astronomers observed a 60MHz radio source both directly and by reflection from the sea, if the receiving dish is 20.0m above sea level, what is the angle of the radio source above the horizon at first order maximum.

2. Astronomers observe the chromosphere of the Sun with a filter that passes the red hydrogen spectral line of wavelength $656\,nm$. The filter consists of a transparent dielectric of thickness t held between two partially aluminized glass plates. Find the minimum value of t that produce maximum transmission of the light wave, if the dielectric has an index of refraction of 1.38

3. Find the amplitude and phase constant of the sum of the two waves represented by $E_1 = 120\,sin\,(15x - 4.5t)$ and $E_2 = 120\,sin\,(15x - 4.5t + 70)$ by using both trigonometric identities and phasor diagram representation.

4. Find the amplitude and the phase constant of the sum of the three waves represented by $E_1 = 120\,sin\,(15x - 4.5t + 70)$, $E_2 = 155\,sin\,(15x - 4.5t - 80)$, and $E_3 = 170\,sin\,(15x - 4.5t + 160)$.

CHAPTER SIX
DIFFRACTION

6.1 INTRODUCTION

Diffraction can be define as the bending or spreading of a wave when it passes through an opening of the same order of magnitude as its wavelength. In general, diffraction occurs when waves pass through a small opening (the same order in magnitude as the wavelength), around obstacles or past edges. The diffraction of light can be viewed as the failure of light to propagate in straight lines. Diffraction phenomena are commonly conveniently divided into two classes –

- Fraunhofer diffraction in which the source of light and the screen on which the pattern is observed are assumed to be at infinite distance from the aperture causing the diffraction;
- Fresnel diffraction in which either the screen or the source, or both are at finite distance from the aperture.

Fraunhofer diffraction is much simpler to analyze mathematically, and is readily observed in practice by rendering the light from a source parallel with a lens and focusing it on a screen with another lens placed behind the aperture, this arrangement effectively removes the source and the screen to infinity. Fresnel diffraction does not require lenses, but wave fronts are not planar but divergent, and theoretical treatment is more complex.

6.2 FRAUNHOFER DIFFRACTION

Fraunhofer diffraction is the diffraction of light waves produced by a narrow slit when plane light waves are incident on the slit, and the light waves emerging from the slit incident on the observing screen are plane waves, i.e. all the light rays passing through the narrow slit are assumed approximately parallel to another. Experimentally, this can be achieved either by placing the screen far from the slit (diffraction grating), or by using a converging lens to focus the rays once they pass through the slit.

6.3 DIFFRACTION FROM A SINGLE NARROW SLIT

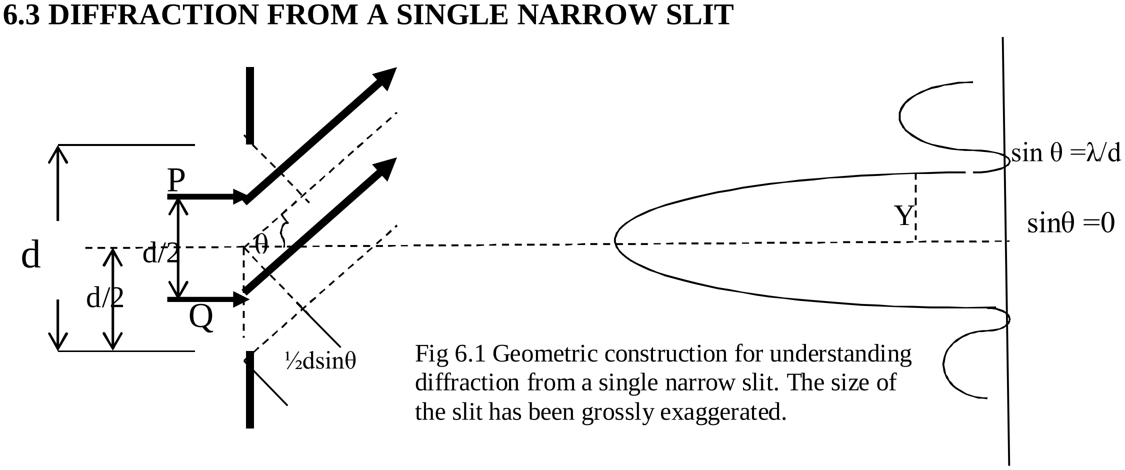


Fig 6.1 Geometric construction for understanding diffraction from a single narrow slit. The size of the slit has been grossly exaggerated.

Fraunhofer diffraction from a narrow slit can be explained using Huygens's wave principle. According to Huygens's principle, each portion of the slit acts as a source of light wave, hence light waves from one portion of the slit can interfere with light waves from another portion, and the resultant light intensity on a viewing screen depends path difference of these waves determine by the direction of the angle θ (figure 6.1).

To understand the diffraction pattern, it is better to divide the slit into two halves (remember that all the rays are in phase on leaving the slit). Consider the rays P and Q (figure 6.1), as the two rays travels towards the screen, ray Q travels farther than ray P by an amount equal to the path difference $\delta = \frac{1}{2} d \sin\theta$; where d is the width of the slit. If the path difference δ is exactly half a wavelength, (corresponding to phase difference $\varphi = \pi$), then the two wave interfere destructively and a dark fringe is obtained. The general condition for destructive interference will be

$$\frac{d}{2} \sin \theta_n = \frac{n\lambda}{2} \Rightarrow \sin \theta_n = \frac{n\lambda}{d} \qquad 6.1$$

Where $n = \pm 1, \pm 2, \pm 3, ...$ This equation gives the value of θ for which the diffraction pattern has zero intensity, but it never reveals anything about the variation in light intensity along the screen.

Example 6.1

Light of wavelength $580\ nm$ is incident on a slit of width $0.30\ mm$. The viewing screen is $2.00\ m$ from the slit. Find the first dark fringes and the width of the central bright fringe.

Solution

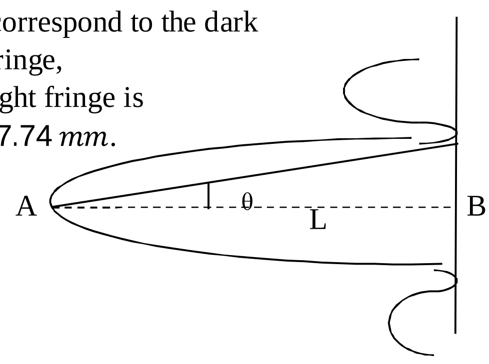
The first two dark fringes that flanks the central bright fringe correspond to $n = \pm 1$;

$hence\ \sin \theta = \pm \frac{\lambda}{d} = \frac{580nm}{0.030mm} = \pm 1.93\ x10^{-3} rad.$

From $\Delta ABC, \theta = \sin\theta = \tan\theta = \frac{Y}{L} \Rightarrow Y_1 = L \sin\theta = \pm \frac{L\lambda}{L} = \pm 3.87\ x10^{-3}\ m.$

The positive and the negative signs correspond to the dark fringes on either side of the central fringe, hence the width W of the central bright fringe is

$$W = 2\ x\ 3.87\ x10^{-3} m = 7.74\ mm.$$



Example 6.2

In a diffraction pattern of a single slit, the separation between the first minimum on one side and the first minimum on the other side is $5.2\ mm$. The distance of the screen from the slit is $80.00\ cm$ and the wavelength of the light used is $546\ nm$. What is the width of the slit?

Solution

$Y = ½ x\ 5.2\ mm = 2.6\ mm$; $if\ \theta\ is\ small, \theta = sin\ \theta = \frac{\lambda}{d}, but\ tan\ \theta = \frac{Y}{L}$

$\Rightarrow \frac{\lambda}{d} = \frac{Y}{L} \Rightarrow d = \frac{\lambda L}{d} = \frac{546\ nm * 80\ cm}{2.6\ mm} = 1.68\ x 10^{-4}\ m$

Example 6.3

What is the angular separation between the 1st minimum and the central maximum for a diffraction pattern of a single slit, if the slit width is equal to (a) one wavelength, (b) five wavelengths (c) ten wavelengths?

Solution

$sin\ \theta = \frac{\lambda}{d} \Rightarrow d = \lambda;\ \theta = sin^{-1} 1 = 90^o;\ d = 5\lambda, \Rightarrow \theta = sin^{-1} \frac{1}{5} = 11.54^o;$
$d = 10\lambda \Rightarrow \theta = sin^{-1} \frac{1}{10} = 5.74^o$

6.4 INTENSITY OF SINGLE SLIT DIFFRACTION PATTERN

Consider a slit divided into large number of zones of width Δy (figure 6.2, only four zones where shown and the slit width highly exaggerated), each zone acts as a coherent source of light. At the point P, each zone contibutes an increment electric field ΔE to the resultant resultant electric field magnitude. The light intensity at P is proportional to the square of the magnitude of the electric field vector. Let us assume that the incremental electric field ΔE between adjacent zones are out of phase with one another by an amount $\Delta\beta$ (which is the difference between each phase) and this phase difference is related to the path difference by

$$\delta = \Delta\beta = \frac{2\pi}{\lambda}\Delta y\ sin\ \theta \qquad 6.2$$

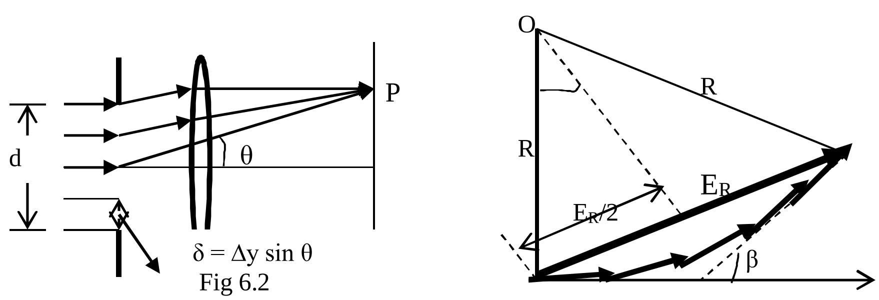


Fig 6.2

To find the magnitude of the total electric field at the P on the screen at any angle θ, we sum all the incremental magniude ΔE due to each zone, and for small values of θ, we can assume that all ΔE have the same value. If $\theta = 0$, all the phasors are aligned, and the total elctric field at the center of the screen is $E_R = E_0 = N\Delta E$, where N is the number of zones. At some other angles, $E_R < E_0$, and the total phase difference between the top zone and the bottom zone of the slit is

$$\beta = N\Delta\beta = \frac{2\pi}{\lambda} N\Delta y\ sin\ \theta = \frac{2\pi}{\lambda} d\ sin\ \theta \qquad 6.3$$

where $d = N\Delta y$. When $\beta = 2\pi$, equation 7.3 reduces to $sin\,\theta = \frac{\lambda}{d}$, which is the relation for first minimum in diffraction pattern (equation 6.1). As θ increases, the chain of the phasor tightens and β increases beyond 2π. The second maximum will occur when $\beta = 3\pi$, while the second minimum occurs at $\beta = 4\pi$ which satisfy the condition $\theta = \frac{2\lambda}{d}$.

The total electric field E_R and the light intensity I at any point P on the screen can be obtained by considering the limiting case in which $\Delta y \rightarrow 0$ to become dy as $N \rightarrow \infty$. At some angle θ, the resultant elctric field E_R on the screen is equal to the chord length (figure 6.2b), and from the ΔABC, we see that sin $\beta/2 = ½E_R/R$, but $E_0 = R\beta$ where β is measured in radian. Combining these relations we have

$$E_R = 2Rsin\frac{\beta}{2} = \frac{2E_0}{\beta}sin\frac{\beta}{2} = E_0\left\{\frac{\sin\beta/2}{\beta/2}\right\} \qquad 6.4$$

But $I \propto E_0^2 \Rightarrow I = I_0\frac{\sin^2(\beta/2)}{(\beta/2)^2} = I_0\left\{\frac{\sin^2\left(\frac{\pi d}{\lambda}\right)sin\theta}{\left(\frac{\pi d}{\lambda}\right)^2\sin^2\theta}\right\}$ 6.5

where $\beta = \frac{2\pi}{\lambda}\,d\,sin\,\theta$; and the condition for minimum intensity is when $\frac{2\pi}{\lambda}\,dsin\,\theta = n\pi \Rightarrow sin\,\theta_n = \frac{n\lambda}{d}$, which is again in agreement with our earlier result.

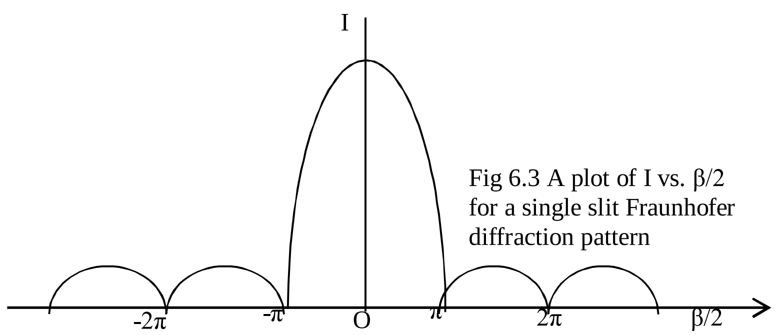


Fig 6.3 A plot of I vs. β/2 for a single slit Fraunhofer diffraction pattern

Example 6.4

Find the ratio of the intensities of the secondary maxima to the intensity of the central maxima for a single slit Fraunhofer diffraction pattern.

Solution

To a good approximation, secondary maxima occur at $\beta/2 = 3\pi/2, 5\pi/2$ etc.

Using $I' = I\frac{\left(\sin^2\frac{\beta}{2}\right)}{\left(\frac{\beta}{2}\right)^2} \Rightarrow = \frac{I'}{I} = \frac{sin^2\left(\frac{3\pi}{2}\right)}{\left(\frac{3\pi}{2}\right)^2} = 0.045$

Which about 4.5% of maximum intensity (where I\` is the first maximum and I is the central maximum).

For the second maximum I'' we have $\frac{I''}{I} = \frac{sin^2\left(\frac{5\pi}{2}\right)}{\left(\frac{5\pi}{2}\right)^2}, = \frac{4}{25\pi^2} = 0.016 \Rightarrow 1.6\%\ of\ I.$

6.5 MATHEMATICAL FORMULATION OF SINGLE-SLIT DIFFRACTION PATTERN

We have used mainly geometric argument to evaluate the intensity of a single-slit diffraction pattern, we now treat the same using mathematical formulation. Figure 6.4 represents a section of a slit of width d, illuminated by parallel rays of ligth; we consider an infintesmal wave front of width x (we note that E is the electric field component of the electromagnetic wave and ∂d the infinitesimal element of the wave front) that according to Huygens principle will emit secondary wavelets. We assume that the parts of the secondary wavelets which travel normal to the plane will be focused at P_O on the screen, while those secondary wavelets which travel at an given angle, say θ will be focused at P. Considering first the wavelet emitted by the element ∂d situated at the point O center of the slit, taken as the origin, its amplitude at any point will be directly proportional to the length ∂d andinversely proportional to its x distance from the slit. At P on the screen, it will produce an infinitesimal wavefront and which for a spherical wave, may be expressed as (recall that

$$E = E_0 \sin(\omega t - kx) \rightarrow \partial E_0 = \frac{E_0 \partial d}{x} \sin(\omega t - kx) \qquad 6.6$$

As the position of ∂d is varied from the origin O, the displacement it produces on the screen will vary in phase because of the different path length to P, when it is at a distance say d below the origin, the contribution will be

$$\partial E_d = \frac{E_0 \partial d}{x} \sin\{\omega t - k(x + \delta)\} = \frac{E_0 \partial d}{x} \sin(\omega t - kx - kd \sin\theta) \qquad 6.7$$

where $\delta = kd \sin\theta$.We now wish to sum the effects of all the elements from one edge of the slit to the other, this is done by intergrating equation 6.7 from $-d\ to\ +d$. The simplest way is to integrate the contributions from pairs of infitesimal elements symmetrical placed at $-d$ and $+d$, each contribution being

$$dE = \partial E_{-d} + \partial E_{+d} = \frac{E_0 \partial d}{x} \sin\{\omega t - k\,x - \delta\} + \frac{E_0 \partial d}{x} \sin\{\omega t - k\,x + \delta\}$$

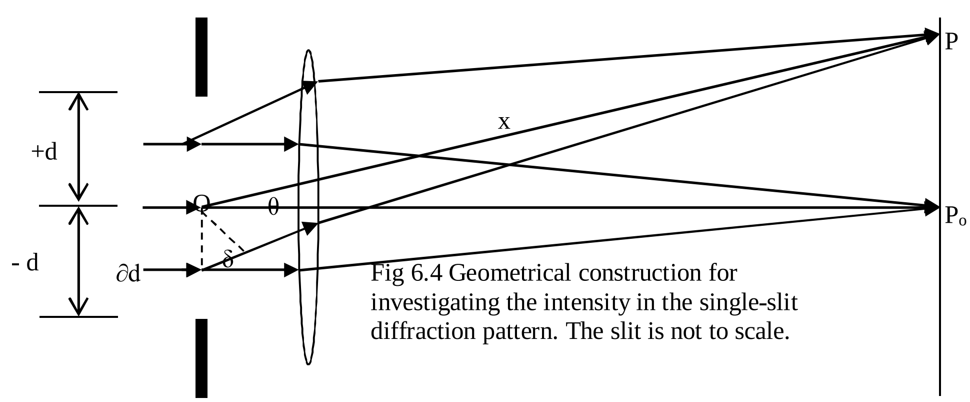


Fig 6.4 Geometrical construction for investigating the intensity in the single-slit diffraction pattern. The slit is not to scale.

Substituting the value of δ in the above equation we have

$$dE = \frac{E_0 \partial d}{x} \sin(\omega t - kx - kd \sin\theta) + \frac{E_0 \partial d}{x} \sin(\omega t - kx + kd \sin\theta)$$

Using trignometric identity, $sin\,\theta_1 + sin\,\theta_2 = 2\,cos\,½(\theta_1 - \theta_2)sin½(\theta_1 + \theta_2)$, we have

$$dE = \frac{E_0 \partial d}{x}\{2cos\,(kd\,sin\,\theta)\,sin\,(\omega t + kx)\} \qquad 6.8$$

If the width of the slit is b, we now integrate equation (6.8) from $b = 0\,to\,d$, taking x as constant as it does not affect the amplitude to obtain

$$E = \frac{2E_0}{x}\,sin\,(\omega t + kx)\int_0^d cos(kdsin\theta)\,\partial d = \frac{E_0 d}{x}\,\frac{sin(kdsin\theta)}{kdsin\theta}\,sin\,(\omega t + kx) \qquad 6.9$$

Equation (6.9) which is the resultant at any point P (determined by θ) on the screen and is a harmonic vibration which can be written as

$$E = E_R\,\frac{sin\beta}{\beta}\,sin\,(\omega t + kx) \qquad 6.10$$

with amplitude given by $\frac{E_R\,sin\,\beta}{\beta}$ and $\beta = kdsin\theta$; $E_R = \frac{E_0 d}{x}$; $k = \frac{2\pi}{\lambda}$, and recalling that intensity $I \propto E^2$, we have

$$I = E_R^2\,\frac{sin^2\beta}{\beta^2} = I_0\,\frac{sin^2\left(\frac{\pi b}{\lambda}sin\theta\right)}{\left(\frac{\pi b}{\lambda}\right)^2 sin^2\theta} \qquad 6.11$$

Equation (6.11) is similar to equation (6.5) derived from purely geometrical consideration. The quantity β is convenient, which signifies one-half the phase difference between the contributions coming from opposite edges of the slit. If the light wave is incident on the slit at an angle φ instead of being perpendicular, it can be shown that the general expression for β will be

$$\beta = \pi kb\,(sin\,\theta + sin\,\varphi) = \frac{\pi b}{\lambda}\,(sin\,\theta + sin\,\varphi) \qquad 6.12$$

6.6 INTENSITY OF TWO-SLIT DIFFRACTION PATTERN

When two slits are present, we must consider not only the diffraction due to the individual slits, but also the interference of the waves coming from different slits. To determine the effects of both interference and diffraction on the intensity at any point P on the screen, we simply combine the equations for both diffraction intensity (equation 6.11) and interference intensity (equation 5.23) to get

$$I = I_0\,cos^2\left(\frac{\pi dsin\theta}{\lambda}\right) * \left\{\frac{sin^2\left(\frac{\pi b}{\lambda}sin\theta\right)}{\left(\frac{\pi b}{\lambda}\right)^2 sin^2\theta}\right\} = I_{max}\,cos^2\left(\frac{\pi dsin\theta}{\lambda}\right)\left\{\frac{sin^2\left(\frac{\pi b}{\lambda}sin\theta\right)}{\left(\frac{\pi b}{\lambda}\right)^2 sin^2\theta}\right\} \qquad 6.13$$

Equation (6.13) represents the diffraction pattern acting as an envelope for a two slit interference pattern. Recall that the condition for interference maxima is given by $d\,sin\,\theta_n = n\lambda$, where d is the distance between the two slits, while the diffraction minima occur when $d'sin\,\theta = \lambda$, where d' is the diameter of the slit. Dividing one by the other allows us to determine which interference maximum coincides with the first diffraction minimum, i.e.

$$\frac{dsin\theta}{d'\,sin\theta} = \frac{n\lambda}{\lambda} = \frac{d}{d'}\,n \qquad 6.14$$

For instance, let $d = 18.0\,\mu m$, $d' = 6.0\,\mu m$, $\Rightarrow n = 3$. Thus, the 3rd interference maxima is aligned with the first diffraction minimum and cannot be seen.

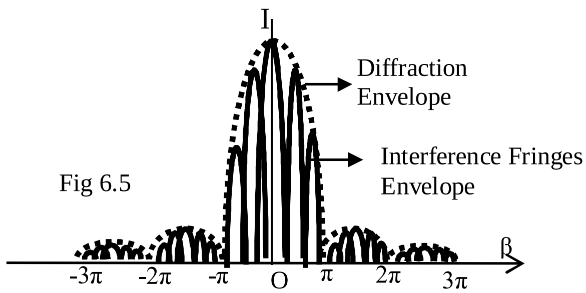


Fig 6.5

A fundamental way to derive the intensity distribution of a double-slit diffraction pattern is to follow the same procedure used for single slit, changing the limits of integration to include the portions of the wave front transmitted by the double slit (the result we will obtain can be generalize for N slits). Consider two equal slits of width d (figure 6.6), separated by an opaque space of width D, the origin may be taken to be at the center of D, then, the integration is extended from $\frac{D}{2} - \frac{d}{2}\ \ to \frac{D}{2}\ + \frac{d}{2}$, this gives

$$E\ = \frac{2E_0}{xk\ sin\theta}\ \ \{sin\ [\ ½k\ (D\ +\ d)sin\ \theta] - \ sin\ \ [½k\ (D\ + d)sin\ \theta]\}sin\ (\omega t - \ kx)$$

The quantity in the curly bracket is of the form $sin\ (A\ +\ B) - \ sin\ (A - \ B)$, and when it is expanded, we obtain

$$E\ = \frac{2dE_0}{x}\frac{sin\beta}{\beta}\ cos\ \gamma\ sin\ (\omega t - \ kx) \qquad 6.15$$

where $\beta\ = \frac{kd}{2} sin\ \theta;\ k\ \ = \frac{2\pi}{\lambda}; d'\ =\ d\ +\ D;\ \gamma\ \ = \frac{k(d+D)}{2} sin\ \theta\ \ = \frac{kd'\,sin\theta}{2}\ \ = \frac{\pi d'\,sin\theta}{\lambda}$.
The intensity as before is proportional to the square of the amplitude of equation (6.15) which is given as

$$I\ =\ 4E_R^2 \frac{sin^2\beta}{\beta^2}\ cos^2\gamma \qquad 6.16$$

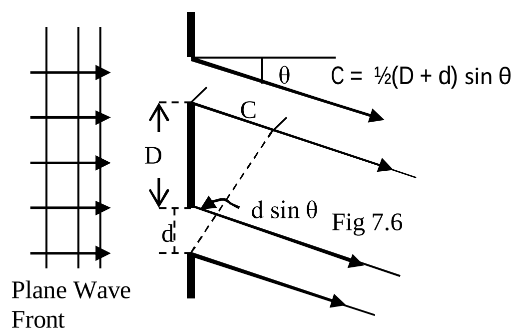


Fig 7.6

The factor $\frac{sin^2\beta}{\beta}$ in equation (6.16) is the single-slit diffraction pattern and the second factor $cos^2\gamma$ is the characteristic of the interference pattern produced by two beams of equal intensity and phase difference φ, where the resultant intensity was found to be proportional to $cos^2\frac{\varphi}{2}$ (equation 5.22); so that the expression of equation (6.16) corresponds to equation (6.13) if we put $\gamma\ \ = \frac{\varphi}{2}$, and the resultant intensity will be zero when either of the two factors is zero. As shown in figure 6.5. For the first factor, this will occurs when $\beta = \pi, 2\pi, 3\pi, \ldots$, and for the second factor when $\gamma = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$. The difference in path length δ from the two edges of a given slit to the screen is $d \sin\theta$, while the corresponding phase difference φ is $\varphi = \frac{2\pi}{\lambda} d \sin\theta = 2\beta$; while the path difference given by $\delta' = ½(D + d)\sin\theta$ is for any two corresponding points in the two slits with the phase difference $\varphi' = \frac{\pi}{\lambda}(D + d)\, sin\,\theta = 2\gamma$. Thus, in terms of dimension of the slits

$$\frac{\varphi'}{2\beta} = \frac{\gamma}{\beta} = \frac{D+d}{2d} \qquad 6.17$$

6.7 DISTINCTION BETWEEN INTERFERENCE AND DIFFRACTION

One is justified in explaining the above result (equation 6.16) by saying that the light waves from the two slits undergo interference to produce the fringes of the type obtained with two beams, but the intensities of these fringes are limited by the amount of light arriving at the given point on the screen by virtue of the diffraction occurring at each slit. In the result (equation 6.16), the relative intensities in the resultant pattern are just those obtained by multiplying the intensity function for diffraction from a single slit of separation $½(D + d)$, by the intensity function for the diffraction from a single slit of width d. Thus, the result may be regarded as due to the joint action of interference between the rays coming from corresponding points in the two slit and of diffraction (as we assumed to arrive at equation 6.13), which determines the amount of light emerging from either slit at any given angle. However, diffraction is merely the result of the interference of all the secondary wavelets originating from the different elements of the wave front. Hence, it is proper to say that the whole pattern is an interference pattern just as we can also called diffraction pattern. However, it might be proper to reserve the term interference for those cases in which a modification of amplitude is produced by the superposition of a finite number of beams, and diffraction for those in which the amplitude is determined by an integration over an infinitesimal elements of the wave front (as we assumed to derive equation 6.11).

6.8 DIFFRACTION GRATING

The diffraction grating consists of a large number of equally spaced parallel slits and serves as a useful device for analyzing light sources. Diffraction grating may be a transmission grating or reflecting grating. Transmission gratings are made by cutting parallel lines on a glass/plastic plates with a precision machine and the space between the lines are transparent and act as separate slits (obscure glass louver is a transmission grating). Reflection gratings are made by cutting parallel lines on the surfaces of reflective materials. The reflection of light from the lines cut into material is diffused, while the reflection of light from the spaces between the lines is specular. Thus, the spaces between the cuttings act as a parallel source of reflected light Disc plates –

(DVD, CD, VCD) are reflection grating. Diffraction gratings have many lines very close to each other and that means very small slit spacing. For instance, a diffraction grating of 5000 lines per cm has a slit size $d = 2\,\mu m$.

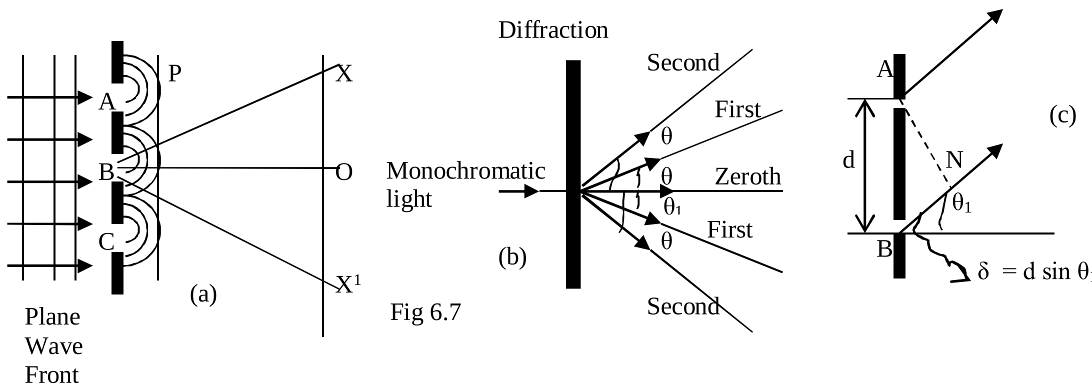


Fig 6.7

The effect produced by diffraction grating can be explained using Huygens principle. Since the slits in the grating are very small, they act as sources of secondary wavelets that are diffracted and interfere with one another producing the final pattern on the screen. They wavelets from each slit are in phase as they leave the slits, however, for some arbitrary angle θ measured from the horizontal, the waves travel different path length before reaching any given point on the screen. The fringe pattern produce at the screen is determines by the cumulative difference in path length. Figure 6.7 shows plane waves incident normally on a diffraction grating, the secondary wavelets from the diffraction grating forms secondary wave front such as *P* moving in the direction *BO*, to form a bright fringe. This is the Zeroth order diffraction and light of very high intensity is detected. Another set of waves may move in the direction CX^1 *or AX*, to form the first order bright fringe.

The relationship between θ_1, the angle the first order diffraction (which occurs when the path length difference is equal to $\delta = \lambda$, where λ is the wavelength of light used) makes with the zero order, and the separation between the slits d is given by figure 6.7c

$$sin\theta_1 = \frac{BN}{AB} = \frac{\lambda}{d} \quad \Rightarrow \quad d\, sin\theta_1 = \lambda\,; \qquad 6.18$$

In general, for the nth order bright diffraction fringe, it can be shown that

$$d\, sin\theta_n = n\,\lambda \qquad 6.19$$

where $d = \frac{1}{N}$, *N* being the number of line per unit length. If white light is used instead of monochromatic light, the zero order is white. For the order bright orders, white light will be separated into its color spectra. This is because the various components of white light have different wavelengths and for the same order, red light will be diffracted differently from blue light. Thus, a spectrum of white light is formed at other bright diffraction order. An example of this occurs when looking at a CD – plate. The width of the spectrum increases as the diffraction order increases, but the intensity decreases.

To derive the intensity distribution function of an equally spaced N-slit diffraction grating, we employ the same procedure used for double slit, only which this times to simplify the integration, we apply the method of adding complex amplitudes. We denote the amplitude contributed by each individual slit as E_0; the phase will change by equal amount φ from one slit to the next, so the resultant complex amplitude is the sum of the series

$$E = E_0\,(1 + \exp(i\varphi) + \exp(2i\varphi) + \exp(3i\,\varphi) + \ldots + exp\{i\,(N-1)\varphi\}$$

$$= E_0\,\frac{1-exp(iN\varphi)}{1-exp(i\varphi)}$$

To find the intensity, the above expression must be multiply by its complex conjugate to get

$$E^2 = E_0^2\,\frac{1-exp(iN\varphi)}{1-exp(i\varphi)}\;x\;\frac{1-exp(iN\varphi)}{1-exp(i\varphi)} = E_0^2\,\frac{1-\cos N\varphi}{1-\cos\varphi} \qquad 6.20$$

Using trignometric relation, $1 - cos\,\theta = 2\,sin2\,½\theta$, equation 7.20 can be written as

$$E^2 = E_0^2\,\frac{sin^2\left(\frac{1}{2}N\varphi\right)}{sin^2\left(\frac{1}{2}\varphi\right)} = E_0^2\,\frac{sin^2 N\gamma}{sin^2\gamma} \qquad 6.21$$

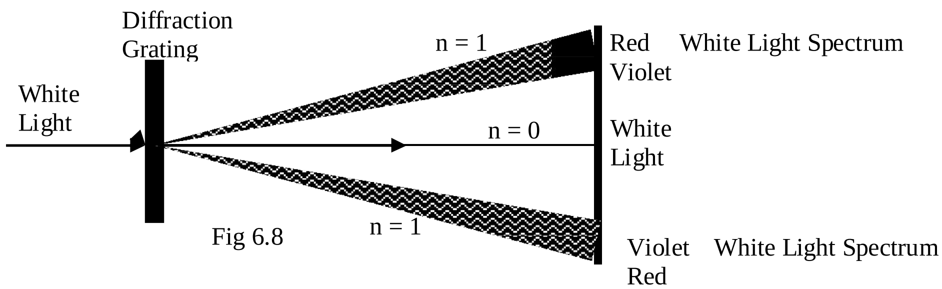


Fig 6.8

Where as in the double slit $\gamma = \frac{\varphi}{2}$. The factor E_0 represents the intensity distribution of a single slit, and after inserting its value into equation 6.21, we obtain for the intensity in the Fraunhofer diffraction pattern of an ideal grating as

$$I = E_R^2\,\frac{sin^2\beta}{\beta^2}\,\frac{sin^2 N\gamma}{sin^2\gamma} \qquad 6.22$$

The new factor $sin^2\,N\gamma / sin^2\,\gamma$ may be said to represent the interference term for N slits. It possesses maximum value equal to N^2 for $\gamma = 0, \pi, 2\pi, \ldots$.Although the quotient becomes indeterminate at these values, this result can be obtained by noting that

$$\lim_{\gamma\to n\pi}\frac{sinN\gamma}{\gamma} = \frac{N\cos N\gamma}{\cos\gamma} = \pm N \qquad 6.23$$

These maxima correspond in position to those of the double slit, however they are more intense in the ratio of the square of the number of slits; as an example, for an N-slit diffraction grating, principle maximum should be $N^2/4$ times more intense than that of a double slit.

To find the minima of the diffraction intensity function, which is determined principally by the new factor in equation (6.22) given by $sin^2 N\gamma / sin^2 \gamma$, we note that the numerator becomes zero more often than the denominator, and this occurs at the values of $N\gamma = 0, \pi, 2\pi, \ldots p\pi$ or in general integral values of π. In the special cases when $p = 0, N, 2N, \ldots, \gamma$ will be $0, \pi, 2\pi, \ldots$; so for these values the denominator will also vanish, and we have the principal maxima. Hence the condition necessary for a minimum is $\gamma = \frac{p\pi}{N}$, excluding the values of $p = nN$, where n is the order. These values of γ correspond to path differences

$$\delta = d \sin\theta = \frac{\lambda}{N}, \frac{2\lambda}{N}, \frac{3\lambda}{N}, \ldots \frac{(N-1)\lambda}{N}, \frac{(N+1)\lambda}{N} \ldots \qquad 6.24$$

While omitting the values $0, \frac{N\lambda}{N}, \frac{2N\lambda}{N}, \frac{3N\lambda}{N}$, for which $\delta = dsin\theta = n\lambda$ represent the principal maxima.

Example 6.4

Light from a source is incident normally on a diffraction grating which has 4000 lines per cm. If the light consists of two lines of wavelrngth between $656\,nm$ and $410\,nm$ respectively, determine the angularseparation between the two lines in the second order spectrum produced by the grating.

Solution

$d\,sin\theta_n = n\lambda, for\ n = 2$, and $d = (400{,}000)^{-1}\,m^{-1} = 2.5\,x10^{-6}\,m^{-1}$

For $\lambda = 656\,nm, sin\,\theta_2 = \frac{2x656\,nm}{2.5\,\mu m} = 0.525 \Rightarrow \quad \theta_2 = 31.65^o$

$For\ \lambda = 410\,nm, \theta_2 = 19.15^o \quad \Rightarrow$ angular separation $= 31.65^o - 19.15^o = 12.50^o$

Example 6.5

How many lines can be observe using a diffraction grating of length $2.0\,cm$ having 10^4 lines when illuminated normally with light of wavelength $495\,nm$.

Solution

$n \leq \frac{dsin\theta_n}{\lambda}$, but θ_n must begreater or less than π rad $(\theta_n \leq \pi rad)$

$\Rightarrow sin\theta_n \leq 1\ and\ d = \frac{2\,cm}{10^4} = 2\,x10^{-6}m; n \leq \frac{d}{\lambda} = \frac{2\,x10^{-6}\,m}{495\,x\,10^{-9}\,m} = 4.04$

$\Rightarrow$ Only up to the fourth order brigth diffraction fringe will be observed

Example 6.6

White light from a source passes througha filter which transmits only wavelength of $400\,nm$ to $600\,nm$. When the filtered light falls on diffraction grating, light of wavelength $400\,nm$ in one order of spectrum is diffracted at the same angle 30° as the $600\,nm$ ligth in the adjacent order. Find the spacing between the lines in the grating.

Solution

$d\,sin\theta_n = n\lambda \Rightarrow dsin\,30^o = 600n = (n + 1)400n \Rightarrow \frac{n}{n+1} = \frac{400}{600}$

$\Rightarrow n = 2 \Rightarrow d = \frac{2\,x\,600\,nm}{sin30} = 2.4\,\mu m$

Example 6.7

A diffraction grating is ruled with $40\ lines\ per\ mm$. A monochromatic parallel light of wavelength $5.9\ nm$ is incident normally on the grating. The diffraction is focused onto a screen by a converging lens of focal length $100\ cm$. Calculate the linear separation between the zero order and first order bright fringe

Solution

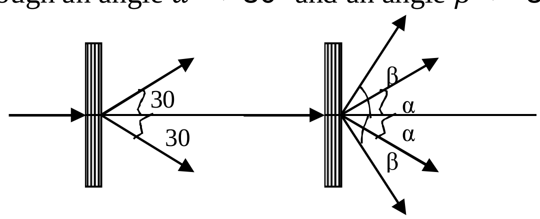
$d = 2.5\,x10^{-5}\ m, for\ n = 1,\ L = f = 100\ cm$ (f is the focal length of the lens)

$sin\,\theta = \frac{\lambda}{d} = 0.0236\ rad$, since if θ is small $sin\,\theta = \theta(rad)$

$\Rightarrow \theta = \frac{Y}{L}\ or\ Y = \theta L = 0.0236\ m = 2.36\ cm$

Example 6.8

Fig a shows the action ofv a diffraction grating PQ on a beam of monochromatic light which is incident normally on the grating in air. Fig b shows the arrangement when the grating is immersed in water of refractive index $\mu = 1.33$. The beam of light is diffracted through an angle $\alpha < 30°$ and an angle $\beta > 30°$ to the normal.



a

b

Explain the change produced by water. If the grating has $8.0\ x10^{5}$ lines per m, calculate the wavelength of light in air, and when immersed in water. Calculate the angle α and β.

Solution

Let λ be wavelength in air, then wavelength in water $\lambda = \frac{\lambda}{\mu} = \frac{\lambda}{1.33}$

1st order in air, $d\ sin\ 30^{o} = \lambda$,

and for 1st order in water, $d \sin\alpha = \frac{\lambda}{1.33}$

comparing the two $\Rightarrow \frac{dsin30}{dsin\alpha} = \frac{\lambda}{\frac{\lambda}{1.33}} = 22.08^{o} < 30^{o}$

In water, if n is the maximum order of diffraction, $d\ sin\ \theta_n = \frac{n\lambda}{1.33}$

But $sin\ \theta_n \leq 1 \Rightarrow n \leq \frac{1.33d}{\lambda}$;

but we have established that $d \sin 30^{o} = \lambda \rightarrow \frac{\lambda}{d} = \sin 30^{o} = \frac{1}{2}$

$\Rightarrow n \leq 1.33x\,2 = 2.66\ i.e.\ n = 2$ at most.

Thus the highest order diffraction obtainable in water is 2.

Similarly, $d\ sin\ \beta = \frac{2\lambda}{1.33} \Rightarrow sin\ \beta = \frac{2}{1.33}\ x\frac{\lambda}{d} \Rightarrow sin\ \beta = \frac{2sin30}{1.33}\ or\ \beta = 48.75^{o} > 30^{o}$

In air, $dsin30 = \lambda$ and $d = \frac{1}{8.0\,x10^{5}\,m^{-1}} = 1.25\,x10^{-6}m \Rightarrow \lambda = 625\,nm$,

while in water $\lambda` = \frac{\lambda}{1.33} = 470\,nm$

Example 6.10

When the spectrum ligth which consist of red and violet are observed using diffraction grating, it is found that the fourth line from the ceter excluding the zero order line is a mixture of red and violet. Explain this phenomenon. If the diffraction grating has $500 lines\ per\ mm$, and the angle for the combine line is 43.6°, find the wavelength of the red and violet components. What is the color of the fifth line and what is its angle of diffraction?

Solution

(a) The wavelength of red light is longer than the wavelength of violet light, thus, from $d\ sin\ \theta_n = n\lambda$, we conclude that for the same order red light is diffracted more than violet, and as the order increases, lower order red will begin to overlap higher order violet.

(b) Since the combined line is the fourth line, this correspond to $n = 2$ for red and $n = 3$ for violet, and $d = 2\,x\,10^{-6}m$; $\theta = 43.6^{o} \Rightarrow \lambda_r = \frac{dsin\theta}{2} = 6.9\,x\,10^{-7}m$; $\lambda_v = \frac{dsin\theta}{3} = 4.6\,x\,10^{-7}m$

If we assume that the fifth line is red, then, $n = 3$ from $d\,sin\,\theta = n\lambda$ we have $sin\,\theta = n\lambda/3 = 1.035$, but $sin\,\theta$ cannot be more than 1, thus the fifth line must be violet and $n = 4$, substituting the values we have

$sin\,\theta = \frac{n\lambda}{4} = 0.92 \Rightarrow \theta = 66.9^{o}$

Example 6.11

Parallel rays of ligth is incident normally on a grating having N lines per unit length. The spectrum consists of two close lines of wavelengths $\lambda\ and\ \lambda + \delta\lambda$. If θ is the mean angle of deviation, show that δθ, the angular separation between the two emergent beams of light is given by $\delta\theta = \frac{\delta\lambda tan\theta}{\lambda}$. Calculate $\delta\theta$ if $\lambda = 589.0\,nm$; $\lambda + \delta\lambda = 589.6nm$, $n = 2$ and the number of lines per unit length is $4.0\,x\,10^{5}m^{-1}$

Solution

From $d\ sin\theta = n\lambda$ (a)

We differentiate equation (a) to get $d\ cos\theta\ \delta\theta = n\ \delta\lambda$ (b)

Divide (a) by (b) to get $\frac{sin\theta}{cos\theta\ \delta\theta} = \frac{\lambda}{\delta\lambda}$ (c)

$\Rightarrow \quad \delta\theta = \frac{\delta\lambda}{\lambda} tan\,\theta$ (d)

Equation (d) is the required equation.

From $d\,sin\,\theta = n\lambda$, we get that $\theta = sin^{-1}\frac{n\lambda}{d} = \frac{2\,x\,589.6nm}{2.5\mu m} = 28.11^{o}$

And $\delta\theta = \frac{\delta\lambda}{\lambda} tan\,\theta = \frac{tan28.11\,x\,0.6nm}{589nm} = 5.44\,x10^{-4}rad = 1.71^{o}\,x10^{-3}$.

Example 6.12

Suggest an upper limit to the useful number of lines per cm of a grating for use with light of $\lambda = 500\, nm$ at normal incidence. A grating of $2.50\, cm$ long has $4.6\, x 10^4\, lines$. A parallel beam of white light with wavelength between $380\, nm$ to $700\, nm$ illuminates it normally. Describe with the aid of a diagram what happens to the beam and calculate the range of wavelengths in the diffracted beam, which emerge from the grating. A beam of electrons accelerated from rest through a potential difference of $500\, V$ is incident on such a grating. Calculate the angle, which the first and the tenth order diffracted beams make with the axis of the system. Hence, discuss whether diffraction effects can be observed in practice using optical grating with electrons at normal incidence.

Solution

a. A grating must at least produce a first order diffraction i.e. for $n = 1, \theta < \pi/2$.

Thus, from $d \sin\theta = n\lambda, \Rightarrow \sin\theta < \frac{\lambda}{d}$ or $d < \frac{\lambda}{\sin 90} = 500\, nm$.

$N = \frac{1}{d} \Rightarrow N < \frac{1}{500\, nm} = 2\, x\, 10^6 lines\ per\ m\ or\ N < 2\, x\, 104\ lines/cm.$

b. When white light is incident on the grating a white light is obtained for the zero order, and

$n = 1$, and $d = \frac{2.5\, x 10^{-2}}{4.6 x 10^4} = 5.43\, x 10^{-7} m;\ \lambda = 380\, nm$

Then $\theta_1 = \sin^{-1}(\lambda/d) = 44.36^o$ and $\theta_2 = \sin^{-1}(2\lambda/d) = \sin^{-1}(1.25)$, this is not possible, thus the second order diffraction do not appear. The minimum angle of diffraction is 90° and this occurs for maximum wavelength

$\lambda_{max} = d \sin\theta = 5.43\, x 10^{-7} m * \sin 90 = 543\, nm.$

c. The wavelength of electron is $\lambda = h/p$ where $p = mv$, h is Planck's constant, m mass of the electron, v its velocity, p the linear momentum and the kinetic energy of an electron of electronic charge e and accelerating potential V is

$K.E = \frac{mv^2}{2} = eV \Rightarrow p = mv = \sqrt{2meV};\ and\ \lambda = \frac{h}{\sqrt{2meV}} = 5.49\, x 10^{-11}\, m$

where $V = 500\, V;\ m = 9.11\, x 10^{-31} kg, h = 6.63 \frac{x 10^{-34} J}{s}$, and $e = 1.6\, x 10^{-19} C$

When $n = 1, \theta_1\ (rad) = \frac{\lambda}{d} = \frac{5.49\, x 10^{-11} m}{5.43\, x 10^{-7} m} = 1.01\, x\, 10^{-4}\, rad;$

$n = 10, \theta_2 (rad) = 10 \frac{\lambda}{d} = 1.01\, x\, 10^{-3}\ rad$

The angular separation between neighboring order of diffraction is

$\Delta Y = 1.01\, x \frac{10^{-3} rad}{10} = 1.01\, x\, 10^{-4}\, rad$

This angle is very small, and thus the various order of diffraction would not be resolved using optical grating for electron beam.

6.9 RESOLUTION OF A SINGLE SLIT AND CIRCULAR APERTURE

The ability of optical systems to distinguish between closely spaced objects is limited because of the wave nature of light. Consider two non-coherent point sources S1 and S2 (may be star light), if no diffraction occurred two distinct bright spots (images) would be observed on the viewing screen. However, because of diffraction, each source is imaged as a bright central region flanked by weaker bright and dark fringes. What is observed on the screen is the sum of the two diffraction patterns, one from S_1 and the other from S_2 (figure 6.9)

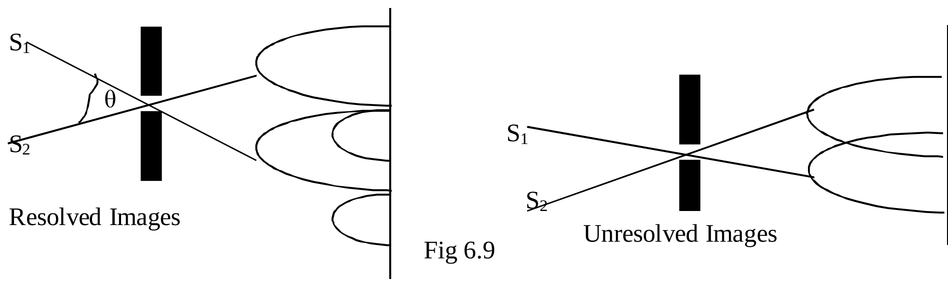


Fig 6.9

If the two sources are far apart to keep their central maxima from overlapping, their central image are said to be resolved. In determining whether two image are resolved a limiting condition called Rayleigh's criterion is often used. The criterion may be stated as follows – two images are said to be resolved when the central diffraction maximum of one image falls on the first minimum of the other image. Rayleigh's criterion enables us to determine the minimum angular separation θ_{min} subtended by the sources at the aperture of the optical instrument for which the images are just resolved. In diffraction, the first minimum in single-slit diffraction pattern occurs at an angle given by sin θ =λ/d, but since in practice $d \gg \lambda$, we assume that $sin\,\theta = \theta\ (radian)$, then the limiting angle of resolution for a slit of width d is

$$\theta_{min} = \frac{\lambda}{d} \qquad 6.25$$

Many optical systems use circular apertures rather than slit. The diffraction pattern of circular aperture consists of a central circular bright disk surrounded by progressively dark and fainter rings. It can be shown that the limiting angle of resolution of circular aperture is

$$\theta_{min} = \frac{1.22\lambda}{D} \qquad 6.26$$

where D is the diameter of the aperture. This condition is never achieved in practice due to material and engineering design factors.

6.10 RESOLVING POWER OF OPTICAL INSTRUMENTS

The resolving power of an optical instrument is the ability of the instrument to differentiate clearly the images of two close objects. The resolving power of an optical instrument is measured by the smallest angle subtended at the optical instrument when the images are just precisely separated; the smaller the angle subtended, the better the resolving power of the instrument.

6.10.1 Limiting Resolution Of A Microscope

Light of wavelength $\lambda = 589\ nm$ is used to view an object under a microscope, if the aperture of the objective has a diameter of $0.90\ cm$, what is the limiting angle of resolution? If it where possibleto use visible light of any wavelength, what would be the maximum limit of resolution for this microscope? Suppose that water of refractive index $\mu = 1.33$ fills the space between the object and objective, what effect does this have on the resolving power when light of $\lambda = 589\ nm$ is used?

Solution

(a) $\theta_{min} = \frac{1.22\lambda}{D} = \frac{1.22*589\ nm}{0.9\ cm} = 7.98\ x10^{-5} rad$

(b) Violet being the shortest visible light spectrum has a wavelength $\lambda = 400\ nm$, thus

$$\theta_{min} = \frac{1.22\lambda}{D} = \frac{1.22*\ 400\ nm}{0.9\ cm} = 5.42\ x10^{-5} rad$$

(c) The wavelength of the light will change by $\lambda_w = \frac{\lambda}{\mu} = \frac{589\ nm}{\lambda} = 443\ nm$

this results in the reduction in the limiting angle of resolution.

6.10.2 Limiting Resolution of a Telescope

As an example, let evaluate the resolution of a typical telescope. The Hale telescope at Mount Palomar has a diameter of $500\ cm$, what is its limiting angle of resolution?

Solution

To solve this problem, we assume a mean wavelength for white light, but if the observation was carried out using a filter, the wavelength of the filtered light is used. In general, the wavelength of visible light ranges from $400\ nm - 700\ nm$, thus, a simple average will be $550\ nm$

$$\theta_{min} = \frac{1.22\lambda}{D} = \frac{1.22*\ 550\ nm}{500\ cm} = 1.34\ x10^{-7}\ rad = 0.02\ arcsec.$$

The arcsec is a unit used by astronomers to measure the angular dimension of astronomical objects. It is define in terms of parallax a star will subtend on Earth due to its apparent motion. Any two stars that subtend an angle greater than or equal to this value can never be resolved. Though, this value would have given astronomers a good resolution of astronomical objects, atmospheric conditions impose a severe limitation to obtaining such resolution in practice. In practice, seeing condition of the atmosphere never allows resolution of less than $0.1\ arcsec$ to be obtained, this is one of the reasons for the launch of Hubble Space Telescope (HST).

For radio waves the resolution is worse, due to longer wavelength. For instance, the large radio telescope at Arecibo Puerto Rico has a diameter of $305\ m$ and is designed to detect radio waves of $\lambda = 0.75\ m$, the minimum angle of resolution is

~ $3\ x10^{-7}\ rad = 10.3\ arcmin = 619\ arcsec$. This is a poor resolution in astronomical measurement. To achieve better resolution interferometers are employed.

6.10.3 Resolution of the Eye

Let us estimate the limiting angle of resolution of the human eye, assuming that its resolution is only limited by diffraction. The average wavelength of white light is $550\ nm$, average size of the pupil is $2\ mm$, $\theta_{min} = 3.36\ x10^{-4}\ rad$. Thus, the minimum linear separation between two objects that an average eye can resolve completely must be $d = L\theta_{min}$ (where L is the near sight and equals $25\ cm$) $\Rightarrow d = 0.25m\ x\ 3.36\ x10^{-4}\ rad = 85\ x10^{-6}\ m = 85\ \mu m$.

Example 6.13

Suppose the pupil of an average eye is dilated to $5.0\ mm$, and that two point sources $3.0\ m$ away are being viewed, how far apart must the sources be if the eye is to resolve them?

Solution

$d = L\theta_{min} = \frac{0.25\ m\ x\ 1.22\ x550\ x10^{-9}}{5\ x\ 10^{-3}m} = 33.55\ x10^{-6}\ m.$

Example 6.14

The headlight of a car are separated by a distance of $1.40\ m$, if the wavelength of light from the lamp is $500\ nm$, and the diameter of the pupil is $5.0\ mm$. What is the minimum distance of the car from the eye when the two lamps are resolved?

Solution

$\theta_{min} = \frac{1.22\lambda}{D} = \frac{d}{L} \Rightarrow L = \frac{Dd}{1.22\lambda} = 11.67\ km$

where $d = 1.40\ m$, $D = 5.0\ mm$, $\lambda = 500nm$.

Example 6.15

A distance of $0.1\ arcsec$ separates two stars, if the stars are observed in the blue light of $\lambda = 450\ nm$, calculate the diameter of the telescope that will just barely resolve them.

Solution

$\theta_{min} = \frac{1.22\lambda}{D} \Rightarrow D = \frac{1.22\lambda}{\theta_{min}} = 1.13m$,

where $0.1arcsec = 4.85\ x10^{-7}\ rad$

6.11 RESOLVING POWER OF THE DIFFRACTION GRATING

The diffraction grating is the most useful instrument for measuring wavelengths accurately. With large number of slits, the maxima are extremely narrow, and its resolving power R for nearly two equal wavelengths of light λ_1, and λ_2 is given as

$$R = \frac{\lambda}{\Delta\lambda} \qquad 6.27$$

is correspondingly high, $and\ \lambda = ½(\lambda_1 + \lambda_2), and\ \Delta\lambda = (\lambda_1 - \lambda_2)$. To evaluate the resolving power of any diffraction grating, we apply Rayleigh criterion, which require that the images formed by the two wavelengths that are barely resolved must be separated by the angle $\Delta\theta$ (which is the angular half-width of the principal maximum) given by

$$\Delta\theta = \frac{\lambda}{D\ cos\theta} = \frac{\lambda}{Nd\ cos\theta} \qquad 6.28$$

$D\ (width\ of\ the\ emergent\ beam) = Nd$, and d dimension of the slit. Thus, the light of wavelength $\lambda + \Delta\lambda$ must form its principle maximum of order n at the same angle as that for the first minimum of wavelength λ in that order. Hence equating the path difference to obtain resolution we have $nN\lambda + \lambda = nN\lambda + nN\Delta\lambda$, which give as the resolution of nth order of diffraction grating of N-slits as

$$R = \frac{\lambda}{\Delta\lambda} = nN \qquad 6.29$$

That the resolving power is proportional to the order n is to be understood from the fact that the width of a principal maximum depends on the width D of the emergent beam. From equation (6.29), we note that the resolving power in a given order is proportional to the total number of slits N but is independent of the spacing between the slits. However, at any given angle of incidence i and diffraction angle θ, the resolution power is independent of N but on W the width of the grating as

$$R = \frac{\lambda}{\Delta\lambda} = W\frac{\sin i + \sin\theta}{\lambda} \qquad 6.30$$

Example 6.16

Two strong components of atomic spectrum of Na have $\lambda = 589.00\ nm$ and $589.59\ nm$. What must be the resolving power of a grating if these components are to be distinguished? To resolve this line in the second order spectrum, how many lines of the grating must be illuminated?

Solution

$$R = \frac{\lambda}{\Delta\lambda} = \frac{\frac{1}{2}(569.00+589.59)}{589.00-589.59} = 999;\ N = \frac{R}{n} = \frac{999}{2} = 500\ lines.$$

6.12 DIFFRACTION BY A CIRCULAR APERTURE

Consider a circular hole of radius R in the xy plane. Coherent light enters the hole from the direction of the negative z axis (see figure 6.11). We consider light rays leaving the hole parallel to the xz plane forming an angle θ with the z axis. The light waves interfere on a screen far away. The phase difference between a wave through a point (x, y) and a wave going through the center of the hole can be calculated from the different path length

$$s = x\ sin\ \theta\ .\delta = \frac{2\pi s}{\lambda} = \frac{2\pi x}{\lambda}sin\theta \equiv kx \qquad 6.31$$

Thus, the phase difference δ depends on the x coordinate only. The sum of the amplitudes of the waves from a small surface element is proportional to the area of the element $dxdy$. Let the amplitude coming through the centre of the hole be

$$d\boldsymbol{a_0} = dx\, dy\boldsymbol{i}. \quad 6.32$$

The amplitude coming from the point (x, y) is then

$$d\boldsymbol{a} = dx\, dy\, (\cos\delta\, \boldsymbol{i} + \sin\delta\, \boldsymbol{j})\,. \quad 6.33$$

We sum up the amplitudes coming from different points of the hole:

$$a = \int da \int_{-x}^{+x}\int_{-y}^{+y}(coskxi + sinkxj)\, dydx = 2\int_{-x}^{+x}\sqrt{R^2 - x^2}\,(coskx\boldsymbol{i} + sinkx\boldsymbol{j})d \quad 6.34$$

where $x = R$ and $y = \sqrt{R^2 - x^2}$ and since sine is an odd function $(sin(-kx) = -sin(kx))$, we get zero when we integrate the sine term. Cosine is an even function, and so

$$a = 2\int_0^R \sqrt{R^2 - x^2}\, coskxi\, dx \quad 6.35$$

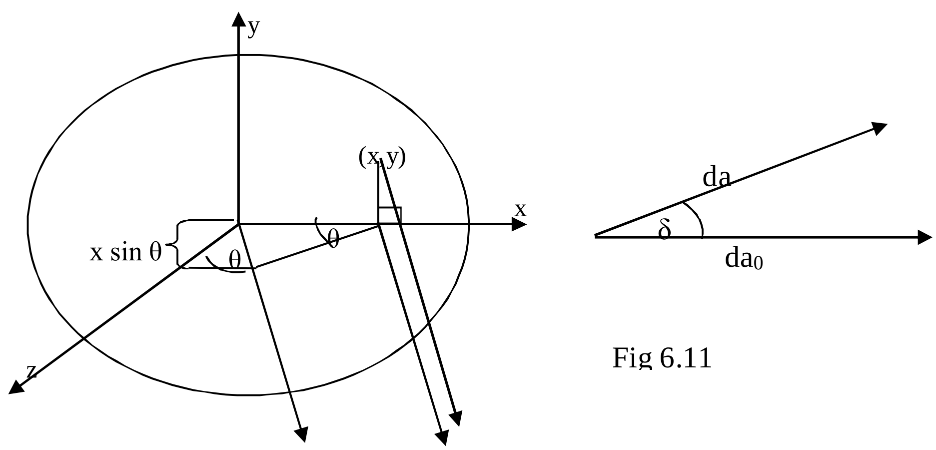


Fig 6.11

We substitute $x = Rt$ and define $p = kR = \frac{2\pi}{\lambda} sin\theta$, thus getting

$$a = \int_0^1 (1 - t^2)\cos pt\, dt \quad 6.36$$

The zero points of the intensity observed on the screen are obtained from the zero points of the amplitude,

$$J(p) = \int_0^1 (1 - t^2)\cos pt\, dt \quad 6.37$$

Inspecting the function $J(p)$, we see that the first zero is at $p = 3.8317$, or $\frac{2\pi}{\lambda} R\sin\theta = 3.8317$. The radius of the diffraction disc in angular units can be estimated from the condition

$$\sin\theta = 3.8317\frac{2\pi}{\lambda} \approx \frac{1.22\lambda}{D}\ , \quad 6.38$$

where $D = 2R$ is the diameter of the hole. In mirror telescopes diffraction is caused also by the support structure of the secondary mirror. If the aperture is more complex and only elementary mathematics is used calculations may become rather cumbersome. However, it can be shown that the diffraction pattern can be obtained as the Fourier transform of the aperture.

Exercise Six

1. Parallel light of wavelength 6563 Å is incident on a slit $0.3850\ mm$ wide. A lens with a focal length of $50.0\ cm$ is located just behind the slit bringing the diffraction pattern to form a focus on a white screen. Find the distance from the center of the principle maximum to (a) the first minimum (b) the fifth minimium.
2. Plane wave of blue light, with wavelength $\lambda = 4340$ Å, fall on a single slit, then pass through a lens with a focal length of $85.0\ cm$. If the central band of the diffraction pattern on the screen has a width of $2.450\ mm$, find the width of the single slit.
3. The objective of a telescope has a diameter of $12.0\ cm$. At what distance would two small green oblects $30.0\ cm$ apart be hardly resolved by the telescope, assuming the resolutio to be limited by the objective only? Assume $\lambda = 5400\ nm$.
4. Two slits of a double slit each hae a width of $0.140\ mm$ and a distance between the centers of $0.840\ mm$. (a) What orders are missing, (b) What is the approximate intensity of orders $n = 0\ to\ n = 6$?
5. Derive the expression for the number of interference maxima occuring under the central diffraction maximum of the double-slit pattern in terms of the separation between the slits and the slit width

CHAPTER SEVEN
POLARIZATION

7.1 INTRODUCTION

Polarization is the restriction of the vibration in a wave so that the vibrations occur in a define plane. Polarization is a property of transverse waves only. An ordinary light beam consists of a large number of waves emitted by the atoms of the light source. Each atom produces a wave having some particular orientation of the electric field vector E, corresponding to the direction of the atomic vibration. The direction of polarization of each individual wave is taken to be the direction in which the electric field is vibrating. For an N-number of atoms in a subtance radiating light waves, where N is very large and all possible energy levels filled, all directions of vibration from such a wave source are possible. Hence, the resultant electromagnetic wave is a superposition of waves vibrating in many different directions. The result is an un-polarized wave.

A plane-polarized wave is a wave in which the resultant electric field vector E vibrates in the same direction at all times at a particular point. Such a wave can be obtained from un-polarized wave by removing all waves from the beam except those whose electric field vectors oscillate along the chosen plane. An optical instrument called the Polaroid is often employed in achieving plane polarization (or linear polarized wave as they are often called)

7.2 MATHEMATICAL REPRESENTATION OF POLARIZATION OF LIGHT

Let us represent two orthogonal optical disturbances in their electric vector by E_x and E_y, if we assume that the waves are propagating in z-direction (for simplification), waves theory of light allows us to represent the waves mathematically as

$$E_x(z,t) = iE_0x\, cos\,(kz - \omega t) \qquad 7.1$$

$$E_y(z,t) = jE_0y\, cos\,(kz - \omega t + \varphi) \qquad 7.2$$

Where E_0 is the amplitude of the vibrating electric vector, $k = 2\pi/\lambda$ is the wave number, φ the phase difference, and $i\ and\ j$ are the unit vectors in the directions of x and y respectively. If the two waves are added, the resultant is given by

$$E(z,t) = E_y(z,t) + E_x(z,t) \qquad 7.3$$

Analysis of equation (7.3) indicates that the resultant wave $E(z,t)$ depends on the values of φ, E_{0y} and E_{0x}. Let us consider these cases, we will start with when both waves are in phase and $\varphi = 0, 2\pi, 4\pi$ ..., equation (7.3) becomes

$E(z,t) = \left(iE_{0x} + jE_{0y}\right)\cos(kz - \omega t)$; If $E_{0x} = E_{0y} = E_0$

$$\Rightarrow \quad E_0(i + j)cos\,(kz - \omega t) \qquad 7.4$$

Equation (7.4) represents a linearly polarized wave, lying in the axis defines by the resultant of $i + j$ vector but propagating in the z-direction. When the phase difference $\varphi = \pm n\pi$ (where n = 1, 3, 5,), equation 7.4 now becomes

$$E\,(z,t) = (iE_{0x} - jE_{0y})cos\,(kz - \omega t). \qquad 7.5$$

This again is a plane-polarized wave, except that the plane of polarization has been rotated by π. Now we consider a case when both waves have equal amplitude but differ in phase by $\pi/2$, the resultant is given as

$$E\,(z,t) = iE_{0x}\cos(kz-\omega t) + jE_{0y}\cos(kz-\omega t + ½\pi)\,E(z,t)$$

$$= E_x\,\{icos\,(kz-\omega t) \pm jsin\,(kz-\omega t)\} \qquad 7.6$$

Equation (7.6) represents a wave equation that has constant amplitude but traces a circular path due to the changing direction of the changing direction of the two superposed waves. When the phase difference is negative half-integer values of π ($\varphi = -½\pi, -3\pi/2, -5\pi/2 \ldots,$) the resultant wave propagates towards the right as it traces a circular path in clockwise direction; thus, we said that the wave is right circularly polarized. On the other hand, if $\varphi = +½\pi, +3\pi/2, +5\pi/2, \ldots$ the resultant wave travels to the left as the resultant electric vector rotates in the anti-clockwise direction, and the wave is said to be left circularly polarized. We note that we can also generate a linearly polarized wave by superimposing left-circularly and right-circularly polarized waves of equal amplitude as

$$E\,(z,t) = E_0\{icos(kz-\omega t) + jsin(kz-\omega t) + icos(kz-\omega t) - jsin(kz-\omega t)\}$$

$$= 2iE_0\,cos(kz-\omega t) \qquad 7.7$$

When $E_{0x} \neq E_{0y}$ and φ is different from integer or half-integer multiples of π, the resultant wave is elliptically polarized and it traces out a helix as it propagates. Common methods used in producing and demonstrating the polarization of light includes the following – dichroic (selective absorption), reflection, scattering, and birefringence (double refraction).

7.3 POLARIZATION BY SELECTIVE ABSORPTION – DICHROISM

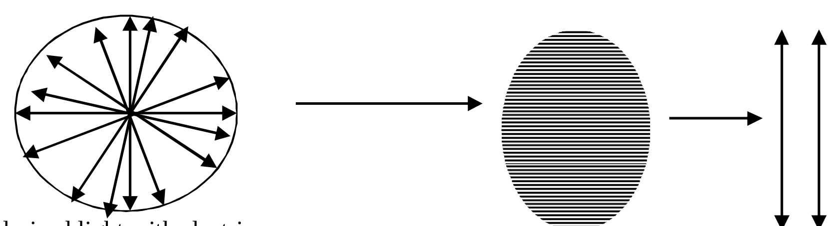


Fig 7.1

The most common technique for producing polarized light is to use a material that transmit waves whose electric fields vector vibrate in a plane parallel to a certain direction, but absorb waves whose electric field vector vibrates in any other direction, such materials are called Polaroid. Polaroid polarizes light through selective absorption by orientated molecules (the commonest Polaroid are fabricated in thin sheet of long-chain hydrocarbon that are stretched during manufacturing so that the long-chain molecules align). The molecules readily absorb light whose electric field vector is parallel to their length, but allow light whose electric field vector is perpendicular to their length.

The direction perpendicular to the molecular chain is called the transmission axis. In an ideal polarizer, all light with electric field vector parallel to the transmission axis are transmitted, and all light with the electric field vector perpendicular to the transmission axis are absorbed.

Since the polarizer transmits all the components of the electric field vector parallel to the transmission axis, while it absorbs all the light perpendicular to it, we can represent the electric vector amplitude of transmitted light by

$$E = E_0 \cos\theta \qquad 7.8$$

where E_0 is the amplitude of un-polarized light, θ is the angle between the transmission axis and emergent light. The intensity of light wave varies as the square of the electric field magnitude; thus, the intensity of the transmitted light is given by

$$I = I_0 \cos^2\theta \qquad 7.9$$

Equation 8.9 is known as Malus's law, and we deduce from it that the intensity of the transmitted light is maximum when $\theta = 0 \; or \; \pi$, but minimum (zero) when $\theta = \pm \frac{1}{2}\pi$. Recall that the average value of $\cos^2\theta$ is ½, thus for an ideal Polaroid, the maximum intensity of the transmitted light is one-half the intensity of un-polarized light.

7.4 POLARIZATION BY REFLECTION

When an un-polarized light beam is reflected from a surface, the reflected light may be completely polarized, partially polarized or un-polarized depending on the angle of incidence. If the angle of incidence is zero, the reflected light is un-polarized; but for any other angle of incidence, the reflected light is polarized to some degree, and for a particular angle of incidence called the Brewster angle (polarizing angle), the reflected light is completely polarized.

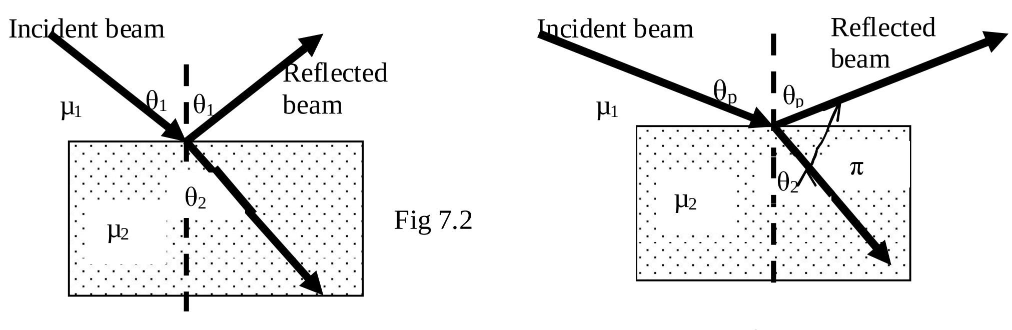


Fig 7.2

When an un-polarized light beam is incident on a surface, each individual electric field vector can be resolved into two components – one parallel to the surface and the other one perpendicular to both the first component and the direction of propagation. If we describe polarization in terms of the electric field components, it is found that the parallel component is strongly reflected more than the perpendicular component, (which is more readily transmitted/absorbed more than the parallel component). This results in partially polarized-reflected beam; by the same argument, the transmitted beam is partially polarized.

If we vary the angle of incidence θ_1, until the angle between the reflected ray and the refracted ray is ½π (figure 7.2), at this particular incidence angle θ_p we can say that the reflected ray is completely polarized, since its electric field vector is now parallel to the surface, while the refracted ray is still partially polarized. To obtain an expression relating the polarization angle θ_p to the refractive index of the medium, we note from figure 7.2 that

$\theta_p + \theta_2 + ½\pi = \pi, and \Rightarrow \theta_2 = ½\pi - \theta_p.$

By Snell's law, $\mu_2 \sin\theta_2 = \mu_1 \sin\theta_p$, and noting that $\mu_1 = 1$ (for air), we have

$$\mu_2 = \frac{\sin\theta_p}{\sin\theta_2} = \frac{\sin\theta_p}{\sin(½\pi - \theta_p)} = \frac{\sin\theta_p}{\cos\theta_p} = \tan\theta_p \qquad 7.10$$

Equation 8.10 is called the Brewster's law, and we can drop the subscript and for any given medium in air, the Brewster angle is calculated from $\theta_p = \tan^{-1}\mu$. Polarization by reflection is a common phenomenon – sunlight reflected from water, glass, and snow is partially polarized. Sunglasses are made from polarizing material to reduce the glare of reflected light; the transmission axes of the glass are oriented in such a way to strongly absorb the unwanted component of the reflected light (usually the horizontal component).

7.4 POLARIZATION BY DOUBLE REFRACTION – BIREFRINGENCE

In certain crystalline materials such as calcite ($CaCO_3$), quartz, etc., the speed of light is not the same in all direction; such materials are characterized by two indices of refraction unlike most amorphous solids like glass, in which the speed of light in them is the same in all direction. Substances with two indices of refraction are referred to as birefringent materials.

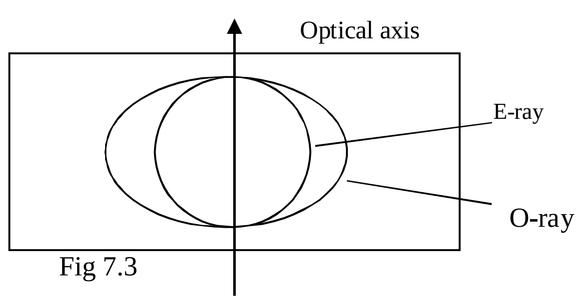


Fig 7.3

Upon entering a birefringent crystal or solution, un-polarized light slit into two plane-polarized mutually perpendicular rays that travels with different velocities corresponding to the two angles of refraction. One ray is called the ordinary ray *O*, characterized by an index of refraction μ_0 that is the same in all directions; this implies that a point source of light inside the crystal will spread out from the source as spherical waves. The second plane-polarized ray is called the extra-ordinary ray E, and travels with different speeds in different directions, and hence is characterized by an index of refraction μ_e that varies with direction of propagation. The extra-ordinary component of a point source of light will spread out in form of an ellipse inside the crystal (figure 7.3). In one particular direction called the optical axis of substances that show birefringence, the $E - ray$ and the $O - ray$ have the same speed ($\mu_0 \;=\; \mu_e$).

Viewing an object through birefringent materials such as calcite will show two images. These two images correspond to one formed by the ordinary ray and one formed by the extra-ordinary ray. If the two images are viewed through a sheet of rotating polarizing glass, they alternately appear and disappear because the ordinary and the extra-ordinary rays are plane-polarized along mutually perpendicular directions.

7.5 POLARIZATION BY SCATTERING

Scattering may be treated, as the resultant effect that results when light is incident on a medium. Light incident on a medium is reflected, refracted, absorbed and transmitted, this is because the incident light can be absorbed by the electrons in the material and re-radiate part of the light in diverse direction, such absorption and re-radiation of light by electrons is called scattering. The scattering of sunlight by the gas molecules of the atmosphere causes the sunlight reaching us to be partially polarized. Figure 7.4 illustrates how sunlight reaching us is scattered. When an un-polarized beam of sunlight traveling in the horizontal direction strikes an air molecule, it sets the electrons of the molecules into vibrations.

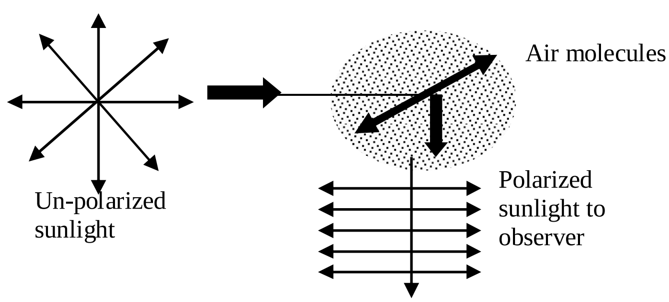


Fig 7.4. The scattering of light of un-polarized sunlight by the air molecules. The scattered light traveling perpendicular to the incident light is plane-polarized because the vertical vibrations of the charge particles in the air molecules send no light in this direction.

These vibrating electrons act like the vibrating charges in antennae. The horizontal component of the electric field vector in the incident wave results in a horizontal component of vibration of the electrons; while the vertical component of the electric field vector results in a vertical component of vibration of electron. An observer looking straight up (perpendicular to the original direction of propagation of the sunlight), will not observe the radiations from the vertical oscillation of charges. The observer sees

only the light that is completely polarized in the horizontal direction. If the observer looks in any other direction, the light is partially polarized in the horizontal direction.
Some phenomena involving the scattering of light in the atmosphere can be understood as follows. When light of various wavelengths λ is incident on gas molecules of diameter d, where $d << \lambda$, the relative intensity of the scattered light varies as λ^{-4}, this is known as the inverse fourth-power law for scattering. The condition that $d << \lambda$ is satisfied by the atmospheric gases like oxygen, nitrogen, carbon dioxide (whose diameters are about $0.2\,nm$ compared with white light with λ ranging from $400\,nm - 700\,nm$. Hence, short wavelengths of sunlight are scattered more efficiently than long wavelengths. This scattering of the sunlight by atmospheric gases and particle is responsible for the blue sky and the red sunset.

7.5 THE BLUE SKY

From the inverse fourth-power law of scattering, the ratio of the intensity I_s/I_l of scattered shorter wavelength λ_s to longer wavelength λ_l is $(\lambda_l/\lambda_s)^4$, thus, the intensity of a scattered shorter wavelength is to some factor raised to the fourth power higher than the intensity of the longer wavelength when scattered by the same particle. The sunlight spectrum, consist of light with short bluer wavelengths and long redder wavelengths. The scattering of sunlight by atmospheric gases and particles is such that the intensity of the short bluer wavelengths are several factors much more than the long redder wavelength, thus, the blue color of the sky. If there are no atmospheric gases and particles to scatter sunlight, the sky will be dark, and one would only see sunlight when in direct path of the Sun's ray.

7.5 THE RED SUNSET

During the day, when the Sun is up in the horizon, the scattered sunlight causes the sky to appear blue, and observing Sun rays which have traveled through some layers of atmospheric components we will see all the spectrum of sunlight in equal proportion, thus the white sunlight. When the Sun is setting down the horizon, the rays of light from Sun reaching an observer would have passed through thicker layers of the atmosphere and these atmospheric component will scatter the sunlight, and the bluer light being scattered more efficiently, would have been refracted and scattered away leaving only the redder longer wavelength for the observer to see.

7.6 OPTICAL ACTIVITY

This describes the continuous rotation of the plane of vibration of light as it propagates along the optical axis of the medium. Many important applications of polarized light involve materials that display optical activity. A substance is said to be optically active if it rotates the plane of polarization of any light transmitted through it. The angle through which the light is rotated depends on the path length through the medium and the density of the medium or concentration if the material is in solution. Sugar dextrose solution, proteins, stressed plastic, sodium chlorate, organic solutions like turpentine, sugar crystals, quartz crystals, tartaric acid are some examples of optically active substances.

It is molecular asymmetry determines whether a material is optically active, for examples proteins and enzymes are optically active because of their spiral shape. In general, any medium that causes the electric field vector of an incident linear plane polarized wave to rotate is optically active. Optically active materials are classified depending on the direction through which the rotate the linear plane electric vector field incident on them. If the rotation is clockwise, the substance is referred to as dextrorotatory (d-rotatory) or right-handed; the substance is referred to as levo-rotatory (l-rotatory) or left-handed, if the rotation is anti-clockwise.

To understand optical activity, consider the right clockwise circularly polarized light to be the R-state, while the left anti-clockwise polarized light to be the L-state. We have shown that a linear polarized wave can be obtained by superposing the R-state light wave and L-state light wave. Since all optically active media show circular birefringence, we let the refractive index for L-state to b$\mu_L$and the refractive index for R-state to be μ_R. When a linear polarized light waves transverse an optically active medium, the plane waves break into two circularly polarized wave of opposite direction due to different refractive indices for the R-state and the L-state. The result is that after propagating through a given path length through the optically active medium, the R-state and the L-state will be out of phase, and the initial plane-polarized wave would appear rotated. Mathematically, both E_R (right circularly polarized electric field vector) and E_L (left circularly polarized electric field vector) can be written as

$$E_R = \frac{E_0}{2}\{ icos\,(k_R z - \omega t) + jsin\,(k_R z - \omega t) \qquad 7.11a$$

$$E_L = \frac{E_0}{2}\{ icos\,(k_L z - \omega t) - jsin\,(k_L z - \omega t) \qquad 7.11b$$

Where $i\ and\ j$ are the unit vectors in x and y directions respectively. Since ω is constant, $\Rightarrow k_R = \mu_R k_0$ and $k_L = \mu_L k_0$. E_0 is the amplitude of the plane-polarized wave while k_0 is the wave number. Superposing equations (7.11a) and (7.11b), the resultant is

$$E = E_0 cos\,\{k_R + k_L) - \omega t\}[icos\,(k_R - k_L) + jsin\,(k_R - k_L)\,] \qquad 7.12$$

For a plane-polarized wave in the x-direction, $z = 0$ and equation 7.12 reduces to

$$E = iE_0 cos\,wt \qquad 7.13$$

If $\mu_L > \mu_R$ $(k_R > k_L)$ the resultant electric vector E will rotate clockwise, but if $\mu_L < \mu_R$ $(k_R < k_R)$, E will rotate anti-clockwise. The angle β through which E rotates (the phase difference between E_R and E_L) is given by

$$\beta = -(k_R - k_L)\frac{z}{2} \qquad 7.14$$

if β is positive, we have clockwise rotation, but if β is negative, we have anti-clockwise rotation. For a medium with thickness d (path length), the angle through which the plane of vibration will rotate is given by

$$\beta = \frac{\pi d}{\lambda}(\mu_L - \mu_R) \qquad 7.15$$

The specific rotary power ρ is define as

$$\rho = \frac{\beta}{d} = (\mu_L - \mu_R) \qquad 7.16$$

Polarization by optical active substances have been employed in determining their concentrations (an example is in the determination of concentration of sugar solution); optical stress analysis (in designing structures ranging from bridges to even small tools) – here models are made with plastics that are optically active, and analyzed under different load conditions, this helps to determine regions of potential weakness and failure when under stress. The LCD (Liquid Crystal Display) found in most display terminal have their optical activity changed by the application of electric potential across different parts of the display unit, leading to the display of specific figure as per input signals.

7.7 OPTICAL PROPERTIES OF GEMSTONES

Throughout history, gemstones have held a fascination for humanity – diamonds, emeralds, rubies, sapphires. Numerous studies have been carried out to understand these gemstones and why they glitter and to reproduce them in the laboratory. Many gemstones have been synthesized in the lab with similar chemical and physical properties as the natural ones, even with a more perfect crystal structure. The principal attraction of a well-cut gems is the size, freedom from flaws, their fire and luster.

The first important gemstones to be synthesized in the laboratory belong to the corundum family, which is a hexagonal crystal form of alumina (Al_2O_3). Those of very high purity which are transparent are called white sapphire. If a few percent of chromium oxide (Cr_2O_3) is added to the growing crystal, we obtain a ruby (red or pink in color). Other metallic oxide (iron, titanium) added produces different colors. Other complex gemstones like the star rubies, star sapphires, emeralds, pure white diamonds, pale blue and pale yellow diamonds have all been synthesized in the laboratory.

It is now understood that these gemstones have very high refractive indices ranging from about 2.410 for ordinary diamond to 3.330 for rutile (a highly reflective gemstone). They sparkle by total internal reflection, polarization of light by reflection, and transmission of directed light and dispersion of light. All these phenomenon combines to give gemstones their unusual color, brilliance, luster and fire. The distinct color, sparkle and fire from gemstones (natural and artificial) also depend on the percentage impurity and the time during the crystallization period it was added, and most importantly how it was cut and its shape.

7.8 PHASE-CONTRAST MICROSCOPE

The eye readily detects differences in amplitude by intensity changes, but it is not able to see changes in phase directly. Thus, as long as the objects on a microscope slide are colored, opaque, or absorbing, they can be seen in the image; but if they are transparent, and differ only slightly from their surroundings in refractive index or in thickness, they will ordinary not be visible. It is nevertheless possible to convert these phase changes produced by such objects into amplitude changes in the final images. This is the principle behind the phase-contrast microscope first devised by F. Zernike in 1935, he was awarded a Nobel Prize in physics in 1953 for his discovery of the phase-contrast principle.

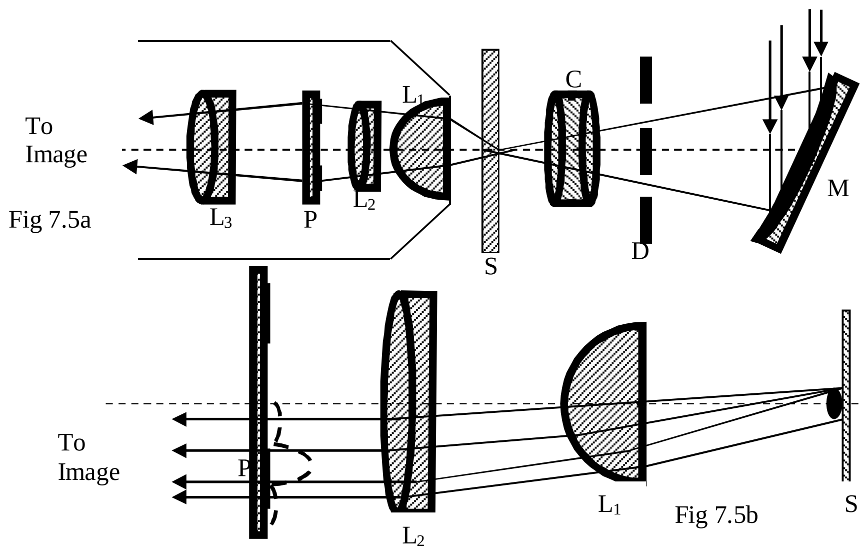


Fig 8.5 Optical components of phase contrast microscope. Jenkins and White (1985)

Fig 7.6 Vector diagram for the waves at the transform plate of the objective lens in a phase-contrast microscope.
(a) Relative phases of the waves arriving at the phase plate;
(b) for bright contrast illumination
(c) dark-contrast illumination

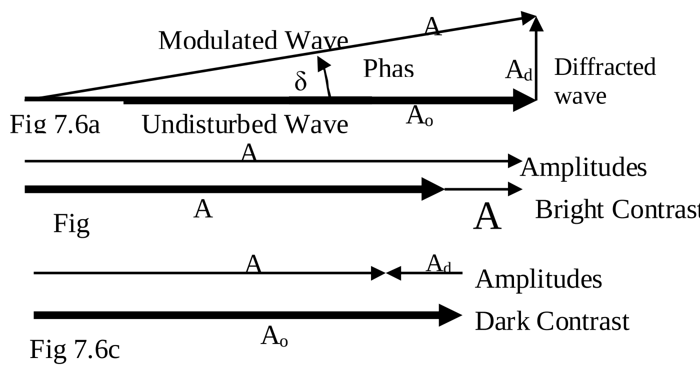


Figure 7.5 shows two essential additions to an ordinary microscope – the phase plate P and an annular diaphragm D. The annular diaphragm is placed in front focal plane of the sub stage condenser C, and an image of light is focused upon D by the concave mirror M. A hollow cone of parallel light therefore illuminates the object on the slide S. If there were no diffraction by objects on the slide, this light would be focused again by the first two lenses of the objective O to form an image of D on the phase plate P. The phase plate consists of a glass plate upon which is evaporated an

annular layer of transparent material to such a thickness that it increases the optical path by such a size to match the image of D (practical thickness is of the order of one-quarter the wavelength of green light).

Let us assume that a small transparent object on the slide retards the phase of the light passing through it by a small angle δ, relative to the phase of the undisturbed light transmitted by the unobstructed parts of the slide (figure 7.6a). A small phase shift of this sort produces a modulated wave, which is the sum of the undisturbed wave and a new diffracted wave retarded in phase by approximately $\pi/2$ (Jenkins and White, 1985). This retarded wave is typically characterized by a varying spatial structure and will form a relatively broad and complex diffraction pattern at the transform plane P. Most of the light in this diffracted wave will therefore miss the annular ring, while the undisturbed wave, which is not diffracted, will pass through the thicker annular layer, where it undergoes a phase retardation of $\pi/2$ with respect to the diffracted light. Thus, the phase plate brings the two waves into phase, with the resulting increase in the intensity at the corresponding point of the final image (figure 7.6c), the diffracting object is then rendered visible by what is called negative or bright contrast. For positive or dark contrast, the annular phase plate is made thinner so that direct light is advanced in phase with respect to the diffracted light. The interference at the image is then, destructive and the object is dark. For better result, a metal film is deposited on the annular portion of the phase plate to make it absorbing, since otherwise the undisturbed light is too strong relative to the diffracted light and destructive interference is not sufficiently complete to reveal the object through dark contrast. It is thus apparent that by introducing phase changes in the transform plane, an object which influences the transmitted beam only through changing its optical path can be made visible, provided that such an object produces diffraction pattern.

Exercise Seven

1. Light is reflected from the smooth surface of water at the polarizing angle. Assume $\mu = 1.333$, find (i) the angle of incidence (ii) the angle of refraction (c) Describe what would be seen if the reflected light were viewed through a calcite crystal which is rotated about the direction of the reflected beam.
2. The effective intensity of a source of light is controlled by the use of a polarizer and analyzer by changing the angle θ between their principal sections. To what accuracy in degrees must θ be known to obtain an accuracy of $2\,\%$ in the intensity of the transimmited light at a setting which reduces the maximum intensity to $10\,\%$.
3. A beam of white is partially polarized by passing it through a single glass plate at the polarizing angle. Assuming $15\,\%$ reflection of the s vibration at each surface, find the degree of polarization (i) if multiple reflections within the plate are neglected (ii) if internal reflection are taken into account (iii) find the degree of polarization if there are 12 plates.
4. An ordinary beam of light is sent through three dichoric polarizer, the second of which is oriented at $25°$ with the first and the third at $50°$ with the first in the

same direction. What intensity gets through the system, relative to that of the incident un-polarized light (i) neglecting the light reflected from the six surfaces (ii) assuming 4.0 % of the light reflected at each surface?

5. Calculate the degree of polarization of the light due to Rayleigh scattering at $70°$ with the direction of the primary beam. Calculate the intensity of the light relative to that scattered straight backwards.

CHAPTER EIGHT
QUANTUM OPTICS

8.1 INTRODUCTION

Quantum optics is a field of research that uses semi-classical and quantum-mechanical physics to investigate phenomena involving light and its interactions with matter at submicroscopic levels. Light propagating in a vacuum has its energy and momentum quantized according to an integer number of particles known as photons

$$E = hf \qquad 8.1$$

where E is the energy of the photon, $h = 6.63x10^{-34}\ J.s$ is Planck constant and f is the frequency of the photon. Quantum optics studies the nature and effects of light as quantized photons. The first major development leading to that understanding was the correct modeling of the blackbody radiation spectrum by Max Planck in 1899 under the hypothesis of light being emitted in discrete units of energy. The photoelectric effect was further evidence of this quantization as explained by Einstein in a 1905 paper. Niels Bohr showed that the hypothesis of optical radiation being quantized corresponded to his theory of the quantized energy levels of atoms, and the spectrum of discharge emission from hydrogen in particular. The understanding of the interaction between light and matter following these developments was crucial for the development of quantum mechanics as a whole, even though, the subfields of quantum mechanics dealing with matter-light interaction were principally regarded as research into matter rather than into light. Laser science (i.e., research into principles, design and application of these quantum optical devices) became an important field, and the quantum mechanics underlying these optical devices was studied now with more emphasis on the properties of light and the name quantum optics became customary.

As quantum optics needed good theoretical foundations, following the work of Dirac in quantum field theory, George Sudarshan, Roy J. Glauber, and Leonard Mandel applied quantum theory to the electromagnetic field in the 1950s and 1960s to gain a more detailed understanding of photodetection and the statistics of light. This led to the introduction of the coherent state as a concept which addressed variations between laser light, thermal light, exotic squeezed states, etc. as it became understood that light cannot be fully described just referring to the electromagnetic fields describing the waves in the classical picture. In 1977, a group of scientists demonstrated a single atom emitting one photon at a time, further compelling evidence that light consists of photons. Previously unknown quantum states of light with characteristics unlike classical states, such as squeezed light were subsequently discovered.

Today's fields of interest among quantum optics researchers include shorter (attosecond) light pulses, use of quantum optics for quantum information, manipulation of single atoms, Bose–Einstein condensates, their application, and how to manipulate them (a sub-field often called atom optics), coherent perfect absorbers, and much more. Topics classified under the term of quantum optics, especially as applied to engineering and technological innovation, often go under the modern term photonics.

8.2 CONCEPTS OF QUANTUM OPTICS

The nineteen century physicist understood electromagnetic field as a disturbance of the all-pervading ether, thus if two charges interacted, it was because the ether in which they are imbedded was distorted by their presence and the resulting strain transmitted from one to the other. Maxwell's field equations described this measurable disturbances of the medium without discussing the ether itself. Light was simply regarded as a wave train consisting of oscillatory mechanical stresses within the ether. The relativity theory of Einstein and the experiment of Michelson-Morley put aside such hypothesis. In the twenty century, light was regarded as electromagnetic field waves that display particle like properties, which can be emitted and absorbed in lumps of matter called photons, and by implication can be created and destroyed. Thus, the electromagnetic field can transport observable physical characteristic such as energy, charge, mass, momentum, but advances through space as wave. Within the context of quantum field theory, particles are regarded as localized wave packets described mathematically by wave functions that are ascribed certain well-behaved mathematical properties. Thus, interactions in quantum field theory arises from creation and annihilation of particles with photons being the means of 'communication' in electromagnetic interactions; in gravitational interaction, gravitons is the exchange particle; in sound, phonons is the exchange particles, etc.

According to quantum theory, light may be considered not only as an electromagnetic wave but also as a "stream" of particles called photons which travel with velocity c, the vacuum speed of light. These particles should not be considered to be classical billiard balls, but as quantum mechanical particles described by a wave function spread over a finite region. Each particle carries one quantum of energy, equal to hf, where h is Planck's constant and f is the frequency of the light. The use of statistical quantum mechanics is fundamental to the concepts of quantum optics: Light is described in terms of field operators for creation and annihilation of photons—i.e. in the language of quantum electrodynamics which is the statistical quantum treatment of photon particles of light with the mean values as the observed. Atoms are considered as quantum mechanical oscillators with a discrete energy spectrum, with the transitions between the energy levels being driven by the absorption or emission of light according to Einstein's theory. For solid state matter, one uses the energy band models of solid state physics.

8.3 THE QUANTUM THEORY OF ATOM

Historically, the atomic and molecular structure of nearly all known chemical elements was established through quantum theory and the various relationships between the various frequencies of light they emit. According to Bohr, the hydrogen atom is composed of a single electron of mass m_e and charge $-e$, rotating n a circular orbit around a positively charged more massive nucleus with mass m and charge $+Ze$, where

Z is the atomic number. This circular motion is governed by the classical laws of electrodynamics, given by

$$\frac{mv}{r} = \frac{Ze^2}{4\pi\varepsilon_0 r^2} \qquad 8.2$$

where r is the radius of the orbit, v velocity of the electron in the orbit and ε_0 the permittivity of free space. Bohr assumed that the orbits are quantized and electrons will stay in only orbits that meet the quantized condition given by

$$m_e vr = \frac{nh}{2\pi} = n\hbar, \qquad n = 1,2,3 \ldots \qquad 8.3$$

n is called the principal quantum number. Solving equations (8.2) and (8.3) gives the allowed orbital radius as

$$r = kn^2 = 0.53n^2 \text{ Å} \qquad k = \frac{\hbar^2}{Z4\pi\varepsilon_0 m_e e^2} \qquad 8.4$$

and velocity given by

$$v = \frac{Z4\pi\varepsilon_0 me^2}{n\hbar} = \frac{2.19x10^6}{n} \; ms^{-1} \qquad 8.5$$

Equations (8.4) and (8.5) implies that not all obits are allowed, and that electrons in each allowed orbit must have allowed energies to exist in that orbit. The energies of the electrons in their allowed orbit is given by

$$E = -\frac{8\pi^2\varepsilon_0^2 Z^2 m_e e^4}{\hbar^2 n^2} = -\frac{2.18x10^{-18}}{n^2} \; J = -\frac{13.6}{n^2} \; eV \qquad 8.6$$

Bohr, assumed that for emission to occur, an electron must jump from higher allowed orbit to lower allowed orbit and the emitted photon energy will be equal to the difference between the energies of the two orbits. For such photons, their frequency will be

$$f = \frac{\Delta E}{h} = \frac{E_i - E_f}{h} = 3.29x10^{15}\left(\frac{1}{n_i^2} - \frac{1}{n_f^2}\right) \; Hz \qquad 8.7$$

This is the origin of various emission spectral lines (Lyman series, Balmer series, Paschen series, Brackett series, Pfund series) observed in hydrogen atom. Similar, the absorption spectral lines occur when an electron absorbs an energy equivalent to the difference between two energy levels and move to a high orbit from a lower orbit.

Though Bohr Theory could explain the spectra lines from hydrogen, it could not properly account for the spectral lines observed with other more complex atoms. To account for more complex atoms, physicists developed quantized elliptical orbits and four quantum numbers with each set of quantum numbers corresponding to a given energy level and the Pauli Exclusion Principle, which states that no two quantum particles in a system will have the four quantum numbers exactly the same. These quantum numbers are principal quantum number$n$, orbital angular momentum quantum number l_n, orbital magnetic quantum number m_l, and the spin quantum number s.

To account for the continuous spectral observed from dense matter, since the exclusion principle forbids quantum particles in a system to have exactly the same quantum numbers, thus, to occupy the same energy level, the energy levels of each particle in the same state splits (for 1 mole of a substance – $6.02x10^{23}$), thus we have allowed energy bands leading to continuous emission spectral.

8.4 RESONANCE RADIATION

Resonance occur when a system A is set into vibration by another system B vibrating at the natural frequency of system A. The atomic process of resonance can be demonstrated by light from sodium lamp in passing through sodium flame (say from a nearby Bunsen Burner). An excited atom in the sodium lamp emit photons of wavelength $\lambda = 5890$ Å, by transition between the two excited (3^2P) states to the (3^2S) ground state. This photon, coming close to a normal sodium atom in the flame will be absorbed and will raise the single valence electron to the corresponding 3^2P level. The second atom will in turn emit the same frequency again, to be absorbed by another atom in the flame, or to escape from the flame in some random direction, but since reemission will be in random direction and seldom in the original direction from the lamp, a shadow will be cast.

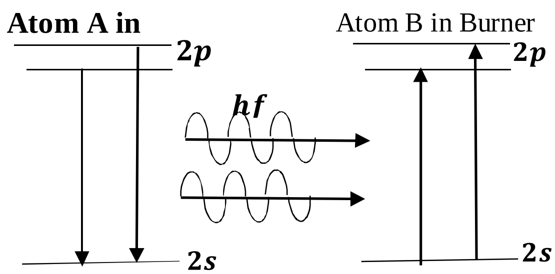


Fig 8.1 Energy Level diagram illustrating light emission and resonance

If the sodium lamp is replaced by a source of white light from a hot solid, those frequencies corresponding to the resonance lines 5890 Å, 5896 Å and all the entire principal series of sodium will be absorbed by the flame and will be seen in a spectrograph as dark lines.

In gases like those found in Bunsen burner or electric discharge tubes, their atoms have their valence electrons in the ground state, when by collision with another particle or atom, the valence electrons are raise to an excited state, they remain there for approximately $1.6\ x10^{-8}\ s$, before jumping to the lower state with the emission of photon. Transitions to lower states are governed by established selection rules, such that not all transitions are allowed. For all atoms with one valence electron, the selection rules can be given by

$$\Delta n = 0, \pm 1, \pm 2, \pm 3, \ldots, \qquad \Delta l = \pm 1 \text{ only} \qquad 8.8$$

For atoms with two electrons in the valence orbit, transitions may occur as one electron jumps from orbit to orbit or two electrons may jump simultaneously, with the emission of a single radiated frequency. Selection rules for two electron systems in general may be written as

$\Delta n = 0, \pm 1, \pm 2, \pm 3, \ldots,$ $\Delta l_1 = \pm l$ and $\Delta l_2 = 0, \pm 2$ only 8.9

If a single electron jumps, the l value of one changes by 1 and the other remains unchanged. If two electrons jump simultaneously, the l value of one changes by 1 and the other by 0 or 2. There are no restrictions on the number of total quantum number n of either electron. For calcium, the two-electron transition between $3d$ to $4p$ and $4s$ to $3d$ gives rise to three groups of lines called the multiplets which constitute some of the strongest lines in the visible spectrum for calcium atom.

In some atoms, it is not possible for an electron to get back to the ground state with the emission of light, from some energy states (forbidden transition –due to selection rules). However, it can return to the ground state by collision with another atom, transferring some of its excitation energy to the colliding atom. Such collisions are called collision of the second kind. Atoms in such state are said to be in metastable state. The existence of metastable states and the transfer of energy from one atom in a metastable state to another by collision are of importance to LASER (Light Amplification by Stimulated Emission Radiation) and MASER (Microwave Amplification by Stimulated Emission Radiation). In metastable state, the average lifetime of an electron is several order of magnitude greater than the $10^{-8}\ s$, which is the mean lifetime of electron in other excited state.

Nearly all atoms in solids, liquids or gases at near absolute zero temperature are in their ground state. As the temperature is raised by some form of energy input, more and more electrons are raised into excited state. The population of electrons in the higher energy level increases at the expense of those at the ground state. At sufficiently high constant temperature, a steady state will be reached and just as many electrons will be jumping into any level as will be the electrons jumping out of that level. If a metastable state exists, the situation is different. As atoms are excited to higher level, more and more of them may get trapped in the metastable state and relatively few of them get out except through mechanical collisions with other atoms. At this point, though a steady state may be reached, but the average populations of atoms in the metastable states may be several order of magnitude greater than that of any other state, except the ground state. If they exceed those of the ground state, it is called population inversion. By using light of appropriate frequency, population inversion can be created, this process is called optical pumping.

8.5 LASER

A laser is a device that produces an intense, concentrated, and highly parallel beam of coherent light. So coherent and parallel is the laser beam that a beam of visible laser light $10\ cm$ in diameter at the moon surface $384,000\ km$ away may not be more than $5\ km$ at the earth. The efficiency of a laser is unusually high: over 30%. Lasers were invented in 1960, since then, they have become ubiquitous, finding utility in thousands of highly varied applications in every section of modern society, including consumer electronics, information technology (fiber optic communication using laser technology is of great importance in internet usage), science, medicine, industry, entertainment, law enforcement and military warfare. The first use of lasers in the daily lives of the general population was the supermarket barcode scanner, introduced in 1974. The laserdisc player and compact disc player was the first laser-equipped device to become common, followed shortly by laser printers. Some other uses are:

1. Medicine: Bloodless surgery, laser healing, surgical treatment, kidney stone treatment, eye treatment, dentistry
2. Industry: Cutting, welding, material heat treatment, marking parts, non-contact measurement of parts
3. Military: Marking targets, guiding munitions, missile defence, electro-optical countermeasure, alternative to radar, blinding troops.
4. Law enforcement: used for latent fingerprint detection in the forensic identification field.
5. Research: Spectroscopy, laser ablation, laser annealing, laser scattering, laser interferometry, laser capture microdissection, fluorescence microscopy
6. Product development/commercial: laser printers, optical discs (e.g. CDs and the like), barcode scanners, thermometers, laser pointers, holograms, bubblegrams.
7. Laser lighting displays: Laser light shows.
8. Cosmetic skin treatments: acne treatment, cellulite, and hair removal.

In laser technology, the principles involve include – Metastable State, Optical Pumping, Fluorescence, Population Inversion, Resonance, Stimulated Emission, Coherence, Polarization, Cavity Oscillation and Interferometry.

Consider a gas enclosed in a vessel containing free atoms having a number of energy levels, which include the metastable state. By optical pumping or any other process, many atoms can be raised, through resonance, from the ground state to excited state. As the electrons jump back, many may be trapped in the metastable state, and if the pumping light is intense enough, population inversion state will be achieved. Fluorescence occur when an electron from the metastable state jumps to lower state with the emission of photon. By resonance, the photon emitted can stimulate other nearby atoms to radiate photons of similar frequency, with similar frequency, direction and polarization as the primary photon (this is called spatial coherence). This process is also possible in the reverse direction, where the emitted primary photon causes can cause transition from the lower state to excited state, and by suitable process, a chain reaction can be produced, resulting in the emission of high intense coherent radiation. In order to produce a laser beam, one must collimate the stimulated emission, and this is done by a properly designed oscillation cavity in which the waves are used over and over again, applying the principle of Fabry-Perot interferometer.

There are various types of laser, classified based on method of lasing or the lasing medium, these include

1. Gas Lasers – Gas lasers using many different gases have been built and used for many purposes. The He-Ne is able to operate at a number of different wavelengths, these relatively low cost but highly coherent lasers are extremely common in optical research and educational laboratories. Commercial CO_2 lasers can emit many hundreds of watts in a single spatial mode with the emission in the thermal infrared at 10.6 μm; such lasers are regularly used in industry for cutting and welding. Argon-ion lasers can operate at a number of lasing transitions between 351 and 528.7 nm. Depending on the optical design one or more of these transitions can be lasing simultaneously; the most commonly used lines are 458 nm, 488 nm and 514.5 nm. Metal ion lasers are gas lasers that generate deep ultraviolet wavelengths, He-Ag at 224 nm and Ne-Cu at 248 nm are two examples. In most low-pressure gas lasers, the gain media of these lasers have quite narrow oscillation line-widths, making them candidates for use in fluorescence suppressed spectroscopy.

2. Chemical Lasers – Chemical lasers are powered by a chemical reaction permitting a large amount of energy to be released quickly. Such very high power lasers are especially of interest to the military, however continuous wave chemical lasers at very high power levels, fed by streams of gasses, have been developed and have some industrial applications. As examples, in the hydrogen fluoride laser at 2700–2900 nm and the deuterium fluoride laser at 3800 nm.

3. Excimer Lasers – Excimer lasers are a special sort of gas laser powered by an electric discharge in which the lasing medium is an excimer. These are molecules which can only exist with one atom in an excited electronic state. Once the molecule transfers its excitation energy to a photon, therefore, its atoms are no longer bound to each other and the molecule disintegrates. This drastically reduces the population of the lower energy state thus greatly facilitating a population inversion. Excimers currently used are all noble gas compounds; noble gasses are chemically inert and can only form compounds while in an excited state. Excimer lasers typically operate at ultraviolet wavelengths with major applications including semiconductor photolithography and eye surgery. Commonly used excimer molecules include ArF (emission at 193 nm), KrCl (222 nm), KrF (248 nm), XeCl (308 nm), and XeF (351 nm).

4. Solid-State Lasers – Solid-state lasers use a crystalline or glass rod which is "doped" with ions that provide the required energy states. For example, the first working laser was a ruby laser, made from ruby (chromium-doped corundum). The population inversion is actually maintained in the dopant. These materials are pumped optically using a shorter wavelength than the lasing wavelength, often from a flashtube or from another laser. The usage of the term "solid-state" in laser physics is narrower than in typical use. Semiconductor lasers (laser diodes) are typically *not* referred to as solid-state lasers.

Neodymium is a common dopant in various solid-state laser crystals, these lasers can produce high powers in the infrared spectrum at 1064 nm. They are used for cutting, welding and marking of metals and other materials, and also in spectroscopy and for pumping dye lasers. These lasers are also commonly frequency doubled, tripled or quadrupled to produce 532 nm (green, visible), 355 nm and 266 nm (UV) beams, respectively.

Thermal limitations in solid-state lasers arise from unconverted pump power that heats the medium. This heat, when coupled with a high thermo-optic coefficient can cause thermal lensing and reduce the quantum efficiency of these class of solid-state lasers. Diode-pumped thin disc laser overcome these issues by having a gain medium that is much thinner than the diameter of the pump beam. This allows for a more uniform temperature in the material. Thin disc lasers have been shown to produce beams of up to one kilowatt.

5. Fiber Lasers – Solid-state lasers or laser amplifiers where the light is guided due to the total internal reflection in a single mode optical fiber are instead called fiber lasers. Guiding of light allows extremely long gain regions providing good cooling conditions; fibers have high surface area to volume ratio which allows efficient cooling. In addition, the fiber's wave-guiding properties tend to reduce thermal distortion of the beam. Erbium and ytterbium ions are common active species in such lasers.

Quite often, the fiber laser is designed as a double-clad fiber. This type of fiber consists of a fiber core, an inner cladding and an outer cladding. The index of the three concentric layers is chosen so that the fiber core acts as a single-mode fiber for the laser emission while the outer cladding acts as a highly multimode core for the pump laser. This lets the pump propagate a large amount of power into and through the active inner core region, while still having a high numerical aperture to have easy launching conditions.

Fiber lasers have a fundamental limit in that the intensity of the light in the fiber cannot be so high that optical nonlinearities induced by the local electric field strength can become dominant and prevent laser operation and/or lead to the material destruction of the fiber. This effect is called photo-darkening. In bulk laser materials, the cooling is not so efficient, and it is difficult to separate the effects of photo-darkening from the thermal effects.

6. Semiconductor Lasers – Semiconductor lasers are diodes which are electrically pumped. Recombination of electrons and holes created by the applied current introduces optical gain. Reflection from the ends of the crystal form an optical resonator, although the resonator can be external to the semiconductor in some designs.

Commercial laser diodes emit at wavelengths from 375 nm to 3500 nm. Low to medium power laser diodes are used in laser pointers, laser printers and CD/DVD players. Laser diodes are also frequently used to optically pump other lasers with high efficiency. The highest power industrial laser diodes, with power up to 10 kW are used in industry for cutting and welding. External-cavity semiconductor lasers have a semiconductor active medium in a larger cavity. These devices can generate high power outputs with good beam quality, wavelength-tunable narrow-line-width radiation, or ultra-short laser pulses.

Vertical Cavity Surface-Emitting Lasers (VCSELs) are semiconductor lasers whose emission direction is perpendicular to the surface of the wafer. These devices typically have a more circular output beam than conventional laser diodes. The development of a silicon laser is important in the field of optical computing. Silicon is the material of choice for integrated circuit, and so electronic and silicon photonic components could be fabricated on the same chip. Unfortunately, silicon is a difficult lasing material to deal with, since it has certain properties which block lasing. However, recently scientist have produced silicon lasers through methods such as fabricating the lasing material from silicon and other semiconductor materials, such as indium(III) phosphide or gallium(III) arsenide, materials which allow coherent light to be produced from silicon. These are called hybrid silicon laser.

7. **Dye lasers –** Dye Lasers use an organic dye as the gain medium. The wide gain spectrum of available dyes, or mixtures of dyes, allows these lasers to be highly tunable, or to produce very short-duration pulses. Although these tunable lasers are mainly known in their liquid form, researchers have also demonstrated narrow-linewidth tunable emission in dispersive oscillator configurations incorporating solid-state dye gain media. In their most prevalent form these solid state dye lasers use dye-doped polymers as laser media.

8. **Free-Electron Lasers –** Free-Electron Lasers (FELs), generate coherent, high power radiation that is widely tunable, currently ranging in wavelength from microwaves through infrared to the visible spectrum, to soft X-rays. They have the widest frequency range of any laser type. While FEL beams share the same optical traits as other lasers, such as coherent radiation, FEL operation is quite different. Unlike gas, liquid, or solid-state lasers, which rely on bound atomic or molecular states, FELs use a relativistic electron beam as the lasing medium, hence the term free-electron.

9. **Exotic media –** Living cells have been used to produce laser light. The cells were genetically engineered to produce Green Fluorescent Protein (GFP). The GFP is used as the laser's "gain medium", where light amplification takes place. The cells were then placed between two tiny mirrors, just 20 millionths of a meter across, which acted as the "laser cavity" in which light could bounce many times through the cell. Upon bathing the cell with blue light, it could be seen to emit directed and intense green laser light. Other current active research area is the space-based X-ray lasers pumped by a nuclear explosion have also been proposed as antimissile weapons.

8.6 HOLOGRAPHY

Holography is the science and practice of making holograms. Normally, a hologram is a photographic recording of a light field, rather than of an image formed by a lens, and it is used to display a fully 3D image of the holographed subject, which is seen without the aid of special glasses. The hologram itself is not an image and it is usually unintelligible when viewed by diffuse ambient light. It is an encoding of the light field as an interference pattern of seemingly random variations in the opacity, density, or surface profile of the photographic medium. When suitably lit, the interference pattern diffracts the light into a reproduction of the original light field and the objects that were in it appear to still be there, exhibiting visual depth cues such as parallax and perspective that change realistically with any change in the relative position of the observer.

In its pure form, holography requires the use of laser light for illuminating the subject and for viewing the finished hologram. In a side-by-side comparison under optimal conditions, a holographic image is visually indistinguishable from the actual subject, if the hologram and the subject are lit just as they were at the time of recording. A microscopic level of detail throughout the recorded volume of space can be reproduced. In common practice, however, major image quality compromises are made to eliminate the need for laser illumination when viewing the hologram, and sometimes, to the extent possible, also when making it. Holograms can now also be entirely computer-generated and show objects or scenes that never existed.

Holography should not be confused with lenticular and other earlier autostereoscopic 3D display technologies, which can produce superficially similar results but are based on conventional lens imaging. Stage illusions such as Pepper's Ghost and other unusual, baffling, or seemingly magical images are also often carelessly called holograms.

8.6.1 Overview And History

The physicist Dennis Gabor was awarded the Nobel Prize in Physics in 1971 for his invention and development of the holographic method. The development of the laser enabled the first practical optical holograms that recorded 3D objects to be made in 1962 by Yuri Denisyuk in the Soviet Union and by Emmett Leith and Juris Upatnieks at USA. Early holograms used silver halide photographic emulsions as the recording medium. They were not very efficient as the produced grating absorbed much of the incident light. Various methods of converting the variation in transmission to a variation in refractive index (known as "bleaching") were developed which enabled much more efficient holograms to be produced.

Most holograms produced are of static objects but systems for displaying changing scenes on a holographic volumetric display are now being developed. Holograms can also be used to store, retrieve, and process information optically. In its early days, holography required high-power expensive lasers, but nowadays, mass-produced low-cost semi-conductor or diode lasers can be used to make holograms. Today, holograms with x-rays are generated by using synchrotrons or x-ray free-electron laser as radiation sources and pixelated detectors such as CCDs as recording medium. The reconstruction is then retrieved via computation. Due to the shorter wavelength of x-rays compared to visible light, this approach allows to image objects with higher spatial resolution.

8.6.2 How Holography Works

Holography is a technique that enables a light field, which is generally the product of a light source scattered off objects, to be recorded and later reconstructed when the original light field is no longer present, due to the absence of the original objects. Holography can be thought of as somewhat similar to sound recording, whereby a sound field created by vibrating matter like musical instruments or vocal cords, is encoded in such a way that it can be reproduced later, without the presence of the original vibrating matter.

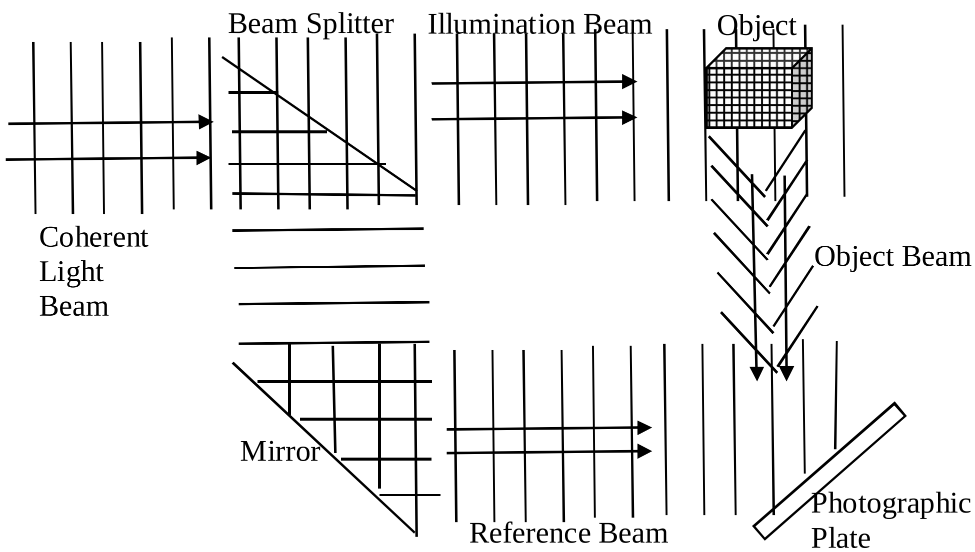


Fig 8.2 Recording a Hologram

Hologram is recorded using a flash of light that illuminates a scene and then imprints on a recording medium, much in the way a photograph is recorded. In addition, however, part of the light beam must be shone directly onto the recording medium (figure 8.2) - this second light beam is known as the reference beam. A hologram requires a laser as the sole light source. To prevent external light from interfering, holograms are usually taken in darkness, or in low level light of a different color from the laser light used in making the hologram. Holography requires a specific exposure time, which can be controlled using a shutter, or by electronically timing the laser. A more flexible arrangement for recording a hologram requires the laser beam to be aimed through a series of elements that change it in different ways. The first element is a beam splitter that divides the beam into two identical beams, each aimed in different directions:

- One beam (known as the *illumination* or *object beam*) is spread using lenses and directed onto the scene using mirrors. Some of the light scattered (reflected) from the scene then falls onto the recording medium.

- The second beam (known as the *reference beam*) is also spread through the use of lenses, but is directed so that it doesn't come in contact with the scene, and instead travels directly onto the recording medium.

Several different materials can be used as the recording medium. One of the most common is a film very similar to photographic film, but with a much higher concentration of light-reactive grains, making it capable of the much higher resolution that holograms require. A layer of this recording medium (e.g., silver halide) is attached to a transparent substrate, which is commonly glass, but may also be plastic.

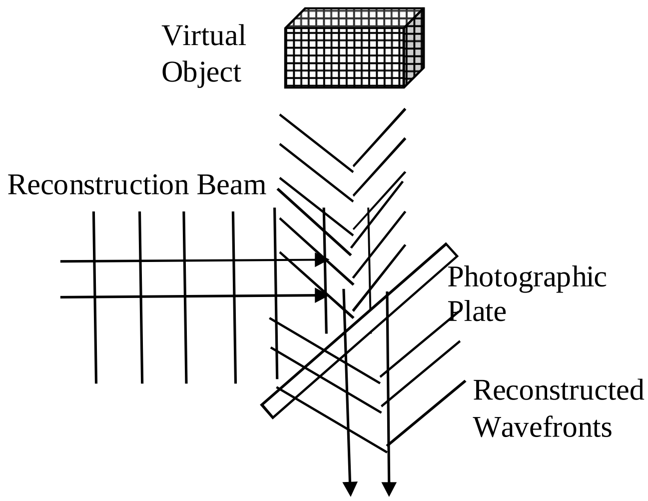


Fig 8.3 Viewing a Hologram

When the two laser beams reach the recording medium, their light waves interfere with each other. It is this interference pattern that is imprinted on the recording medium. The pattern itself is seemingly random, as it represents the way in which the scene's light *interfered* with the original light source — but not the original light source itself. The interference pattern can be considered an encoded version of the scene, requiring a particular key — the original light source — in order to view its contents. This missing key is provided later by shining a laser, identical to the one used to record the hologram, onto the developed film. When this beam illuminates the hologram, it is diffracted by the hologram's surface pattern. This produces a light field identical to the one originally produced by the scene and scattered onto the hologram (Figure 8.3).

8.6.3 Holography Versus Photography

Holography may be better understood via an examination of its differences from ordinary photography:

- A hologram represents a recording of information regarding the light that came from the original scene as scattered in a range of directions rather than from only one direction, as in a photograph. This allows the scene to be viewed from a range of different angles, as if it were still present.

- A photograph can be recorded using normal light sources (sunlight or electric lighting) whereas a laser is required to record a hologram.
- A lens is required in photography to record the image, whereas in holography, the light from the object is scattered directly onto the recording medium.
- A holographic recording requires a second light beam (the reference beam) to be directed onto the recording medium.
- A photograph can be viewed in a wide range of lighting conditions, whereas holograms can only be viewed with very specific forms of illumination.
- When a photograph is cut in half, each piece shows half of the scene. When a hologram is cut in half, the whole scene can still be seen in each piece. This is because, whereas each point in a photograph only represents light scattered from a single point in the scene, *each point* on a holographic recording includes information about light scattered from *every point* in the scene. It can be thought of as viewing a street outside a house through a $1.2\ m\ by\ 1.2\ m$ window, then through a $60\ cm\ by\ 60\ cm$ window. One can see all of the same things through the smaller window (by moving the head to change the viewing angle), but the viewer can see more *at once* through the $1.2\ m$ window.
- A photograph is a two-dimensional representation that can only reproduce a rudimentary three-dimensional effect, whereas the reproduced viewing range of a hologram adds many more depth perception cues that were present in the original scene. These cues are recognized by the human brain and translated into the same perception of a three-dimensional image as when the original scene might have been viewed.
- A photograph clearly maps out the light field of the original scene. The developed hologram's surface consists of a very fine, seemingly random pattern, which appears to bear no relationship to the scene it recorded.

8.6.4 Physics of Holography

A simple hologram can be made by superimposing two plane waves from the same light source on a holographic recording medium. The two waves interfere giving a straight line fringe pattern whose intensity varies sinusoidally across the medium. The spacing of the fringe pattern is determined by the angle between the two waves, and of the wavelength of the light. The recorded light pattern is a diffraction pattern. When it is illuminated by only one of the waves used to create it, it can be shown that one of the diffracted waves emerges at the same angle as that at which the second wave was originally incident so that the second wave has been 'reconstructed'. Thus, the recorded light pattern is a holographic recording as defined above. If the recording medium is illuminated with a point source and a normally incident plane wave, the resulting pattern is a sinusoidal zone plate which acts as a negative Fresnel lens whose focal length is equal to the separation of the point source and the recording plane.

To record a hologram of a complex object, a laser beam is first split into two separate beams of light. One beam illuminates the object, which then scatters light onto the recording medium. According to diffraction theory, each point in the object acts as a point source of light so the recording medium can be considered to be illuminated by a set of point sources located at varying distances from the medium. The second (reference) beam illuminates the recording medium directly. Each point source wave interferes with the reference beam, giving rise to its own sinusoidal zone plate in the recording medium. The resulting pattern is the sum of all these 'zone plates' which combine to produce a random (speckle) pattern as in the photograph above. When the hologram is illuminated by the original reference beam, each of the individual zone plates reconstructs the object wave which produced it, and these individual wavefronts add together to reconstruct the whole of the object beam. The viewer perceives a wavefront that is identical to the wavefront scattered from the object onto the recording medium, so that it appears to him or her that the object is still in place even if it has been removed. To make a hologram, the following are required:

- A suitable object or set of objects
- A suitable laser beam
- Part of the laser beam to be directed so that it illuminates the object (the object beam) and another part so that it illuminates the recording medium directly (the reference beam), enabling the reference beam and the light which is scattered from the object onto the recording medium to form an interference pattern
- A recording medium which converts this interference pattern into an optical element which modifies either the amplitude or the phase of an incident light beam according to the intensity of the interference pattern.
- An environment which provides sufficient mechanical and thermal stability that the interference pattern is stable during the time in which the interference pattern is recorded.

8.6.5 Hologram Classifications

In the classification of holograms, three important properties are considered – Amplitude and phase of the wave, thickness of the recoding medium and the direction of incident and reference beams with respect to each other.

Amplitude And Phase Modulation Holograms – An amplitude modulation hologram is one where the amplitude of light diffracted by the hologram is proportional to the intensity of the recorded light. A straightforward example of this is photographic emulsion on a transparent substrate. The emulsion is exposed to the interference pattern, and is subsequently developed giving a transmittance which varies with the intensity of the pattern - the more light that fell on the plate at a given point, the darker the developed plate at that point. A phase hologram is made by changing either the thickness or the refractive index of the material in proportion to the intensity of the holographic interference pattern. This is a phase grating and it can be shown that when such a plate is illuminated by the original reference beam, it reconstructs the original object wavefront. The efficiency (i.e., the fraction of the illuminated beam which is converted to reconstructed object beam) is greater for phase than for amplitude modulated holograms.

Thin Holograms and Thick (Volume) Holograms – A thin hologram is one where the thickness of the recording medium is much less than the spacing of the interference fringes which make up the holographic recording. A thick or volume hologram is one where the thickness of the recording medium is greater than the spacing of the interference pattern. The recorded hologram is now a three dimensional structure, and it can be shown that incident light is diffracted by the grating only at a particular angle, known as the Bragg angle. If the hologram is illuminated with a light source incident at the original reference beam angle but a broad spectrum of wavelengths; reconstruction occurs only at the wavelength of the original laser used. If the angle of illumination is changed, reconstruction will occur at a different wavelength and the color of the re-constructed scene changes. A volume hologram effectively acts as a color filter.

Transmission and Reflection Holograms – A transmission hologram is one where the object and reference beams are incident on the recording medium from the same side. In practice, several more mirrors may be used to direct the beams in the required directions. Normally, transmission holograms can only be reconstructed using a laser or a quasi-monochromatic source, but a particular type of transmission hologram, known as a rainbow hologram, can be viewed with white light. In a reflection hologram, the object and reference beams are incident on the plate from opposite sides of the plate. The reconstructed object is then viewed from the same side of the plate as that at which the re-constructing beam is incident. Only volume holograms can be used to make reflection holograms, as only a very low intensity diffracted beam would be reflected by a thin hologram.

Dynamic Holography – In static holography, recording, developing and reconstructing occur sequentially, and a permanent hologram is produced. There also exist holographic materials that do not need the developing process and can record a hologram in a very short time. This allows one to use holography to perform some simple operations in an all-optical way. Examples of applications of such real-time holograms include phase conjugate mirrors ("time-reversal" of light), optical cache memories, image processing (pattern recognition of time-varying images), and optical computing. The amount of processed information can be very high (terabits/s), since the operation is performed in parallel on a whole image. This compensates for the fact that the recording time, which is in the order of a microsecond, is still very long compared to the processing time of an electronic computer. The optical processing performed by a dynamic hologram is also much less flexible than electronic processing.

On one side, one has to perform the operation always on the whole image, and on the other side, the operation a hologram can perform is basically either a multiplication or a phase conjugation. In optics, addition and Fourier Transforms are already easily performed in linear materials, this enables some applications, such as a device that compares images in an optical way. Common materials used for dynamic holography include photorefractive crystals, semiconductor materials, atomic vapors and gases, plasmas and even liquids.

8.6.6 Applications

Hologram has found many applications in art designs, data storage, security (holograms are very difficult to forge), in sensors and detection, film making and data transmission. One particularly promising application is the optical phase conjugation. It allows the removal of the wavefront distortions a light beam receives when passing through an aberrating medium, by sending it back through the same aberrating medium with a conjugated phase. This is useful, for example, in free-space optical communications to compensate for atmospheric turbulence (the phenomenon that gives rise to the twinkling of starlight). Holographic scanners are in use in post offices, larger shipping firms, and automated conveyor systems to determine the three-dimensional size of a package. They are often used in tandem with weight-checkers to allow automated pre-packing of given volumes, such as a truck or pallet for bulk shipment of goods. Holograms produced in elastomers can be used as stress-strain reporters due to its elasticity and compressibility, the pressure and force applied are correlated to the reflected wavelength, therefore its color. More technical and technological application include Holographic Interferometry and Holographic microscopy.

8.6.7 Non-optical holography

In principle, it is possible to make a hologram for any wave. Electron holography is the application of holography techniques to electron waves rather than light waves. Electron holography was invented by Dennis Gabor to improve the resolution and avoid the aberrations of the transmission electron microscope. Today it is commonly used to study electric and magnetic fields in thin films, as magnetic and electric fields can shift the phase of the interfering wave passing through the sample.

Acoustic holography is a method used to estimate the sound field near a source by measuring acoustic parameters away from the source via an array of pressure and/or particle velocity transducers. Measuring techniques included within acoustic holography are becoming increasingly popular in various fields, most notably those of transportation, vehicle and aircraft design. The general idea of acoustic holography has led to different versions such as near-field acoustic holography and statistically optimal near-field acoustic holography.

Atomic holography has evolved out of the development of the basic elements of atom optics. With the Fresnel diffraction lens and atomic mirrors, atomic holography follows a natural step in the development of the physics (and applications) of atomic beams. Recent developments including atomic mirrors and especially ridged mirrors have provided the tools necessary for the creation of atomic holograms, although such holograms have not yet been commercialized.

8.7 MAGNETO-OPTICS AND ELECTRO-OPTICS

In wave optics, electromagnetic theory is capable of explaining the main the main features of propagation of light through space and through matter. In further support of the photon and wave theory of light, there is a group of phenomena that demonstrates the interaction between light and matter, when light is subjected to strong external magnetic or electric field. These phenomena are classed under

1. Magneto-Optics – Effects due to application of external magnetic field – Zeeman Effect, Inverse Zeeman Effect, Voigt Effect, Cotton-Mouton Effect, Faraday Effect, Kerr Magneto-optics Effect.
2. Electro-Optics - Effects due to application of external electric field – Stark Effect, Inverse Stark Effect, Electric Double Refraction, Kerr Electro-Optics Effect.

A magneto-optic effect is any one of a number of phenomena in which an electromagnetic wave propagates through a medium that has been altered by the presence of a quasi-static magnetic field. In such a material, which is also called gyromagnetic, left- and right-rotating elliptical polarizations can propagate at different speeds, leading to a number of phenomena. In particular, in a magneto-optic material the presence of a magnetic field (either externally applied or because the material itself is ferromagnetic) can cause a change in the permittivity ε of the material. Thus, the permittivity of the medium ε becomes anisotropic, depending on the frequency ω of incident light, with the principle axes becoming complex, these correspond to elliptically-polarized light where left- and right-rotating polarizations can travel at different speeds.

8.7.1 ZEEMAN EFFECT AND INVERSE ZEEMAN EFFECT

Zeeman Effect is the splitting/broadening of atomic spectral lines in the presence of strong external magnetic field. In the year 1896, Zeeman discovered that when a sodium flame is placed between the poles of a powerful magnetic field, the two sodium yellow lines are considerably broadened. Lorentz explained the observation by assuming that electrons which are responsible for the emitted light being charged particle have their motion modified by the magnitude and the direction of the applied magnetic field. Lorentz predicted that each spectrum line when produced in such a field should split into two components when viewed parallel to the field and into three components when viewed perpendicular to the field. He further predicted that in the longitudinal direction, the lines should be circularly polarized while in the transverse direction, plane polarized. These predictions were later verified experimentally.

If the electron is moving in a circular orbit, the plane of which is normal to the magnetic field direction, the electron may be speeded up or slowed down by an amount proportional to the magnetic induction. A semi-classical treatment shows that if ω_0 represents the orbital frequency of the electron in a field-free space, the frequency in the presence of a field will be given by $\omega_0 \pm \Delta\omega$, where

$$\pm\Delta\omega = \pm\frac{eB}{2m} \qquad 8.10$$

e is the electronic charge, m_e is the electronic mass, and B is the magnetic field. Equation (8.3) indicates a symmetry in the number of lines around the fundamental frequency. However, in very strong magnetic fields, this symmetry is broken and complete quantum field theory is needed to understand the observed complex spectra.

The Zeeman Effect obtained in absorption is called the inverse Zeeman Effect. This phenomenon is observed by sending white light through an absorbing vapor subjected to a uniform magnetic field. The Zeeman components of any spectrum lines obtained in absorption along field direction are not completely absorbed, and the light that does get through is found to be circularly polarized in directions opposite to those of the corresponding components obtained in emission.

Zeeman Effect have observed in solar spectra, indicating the existence of strong magnetic fields in sunspots. Such fields can be quite high, on the order of 0.1 tesla or higher. Today, the Zeeman Effect is used to produce magnetograms showing the variation of magnetic field on the sun. Zeeman Effect is also utilized in many laser cooling applications such as a magneto-optical trap.

8.7.2 QUADRATIC MAGNETIC ROTATION EFFECT (QMR EFFECT)

Quadratic Magnetic Rotation Effect (QMR Effect) is a type of magneto-optic effect, discovered in the 1980s by Soviet physicists. QMR, like the Faraday Effect, establishes a relationship between the magnetic field and rotation of polarization of the plane of linearly polarized light, but in contrast to Faraday Effect, QMR states the quadratic proportionality between the angle of the rotation of the plane of polarization and the strength of the magnetic field. Mostly QMR can be observed in the transverse geometry when the direction of the magnetic field strength is perpendicular to the direction of light propagation. The first evidence of QMR effect was obtained in the antiferromagnetic crystal of cobalt fluoride in 1985. Generally, considerations of the symmetry of the media, light and axial vector of the magnetic field forbid QMR in non-magnetic or magnetically disordered media.

8.7.3 FARADAY EFFECT

The Faraday Effect or Faraday Rotation is an interaction between light and a magnetic field in a medium. The Faraday Effect causes a rotation of the plane of polarization which is linearly proportional to the component of the magnetic field in the direction of propagation. Discovered by Michael Faraday in 1845, the Faraday Effect was the first experimental evidence that light and electromagnetism are related. This effect occurs in most optically transparent dielectric materials (including liquids) under the influence of magnetic fields.

The Faraday Effect is caused by left and right circularly polarized waves propagating at slightly different speeds, a property known as circular birefringence. Since a linear polarization can be decomposed into the superposition of two equal-amplitude circularly polarized components of opposite handedness and different phase, the effect of a relative phase shift, induced by the Faraday Effect, is to rotate the orientation of a wave's linear polarization. The Faraday Effect has a few applications in measuring instruments. For instance, the Faraday Effect has been used to measure optical rotatory power and for remote sensing of magnetic fields. The Faraday Effect is used in spintronics research to study the polarization of electron spins in semiconductors. Faraday rotators can be used for amplitude modulation of light, and are the basis of optical isolators and optical circulators required in optical telecommunications and other laser applications.

In circularly polarized light, the direction of the electric field rotates at the frequency of the light, either clockwise or anticlockwise. In a material, this electric field causes a force on the charged particles comprising the material (because of their light mass the electrons are most heavily affected). The affected motion will be circular, and circularly moving charges will create their own (magnetic) field in addition to the external magnetic field. There will thus be two different cases, the created field will be parallel to the external field for one (circular) polarization, and in the opposing direction for the other polarization direction - thus the net magnetic field is enhanced in one direction and diminished in the opposite direction. This changes the dynamics of the interaction for each beam and one of the beams will be slowed down more than the other, causing a phase difference between the left- and right-polarized beam. When you add the two beams after this phase shift, the result is again a linearly polarized beam, but with a rotation in the polarization direction. The direction of polarization rotation depends on the properties of the material through which the light is shone.

Faraday Rotation In The Interstellar Medium

The Faraday Effect is imposed on light over the course of its propagation from its origin to the Earth, through the interstellar medium. The effect is caused by free electrons and can be characterized as a difference in the refractive index seen by the two circularly polarized propagation modes. Hence, in contrast to the Faraday Effect in solids or liquids, interstellar Faraday rotation measure (β) has a simple dependence on the wavelength of light (λ), given by

$$\beta = RM\lambda^2 \qquad 8.11$$

where the overall strength of the effect is characterized by RM, the rotation measure. This in turn depends on the axial component of the interstellar magnetic field $B_{\parallel}$ and the number density of electrons n_e, both of which vary along the propagation path. The rotation measure RM is given by

$$RM = \frac{e^3}{8\pi^2 \varepsilon_0 c^3 m_e^2} \int_0^l n_e(l)\, B_{\parallel}(l) dl \qquad 8.12$$

where $n_e(l)$ is the density of electrons at each point s along the path, $B_{\parallel}(l)$ is the component of the interstellar magnetic field in the direction of propagation at each point s along the path, e is the electronic charge and ε_0 is the permittivity of free space.

Faraday rotation is an important tool in astronomy for the measurement of magnetic fields, which can be estimated from rotation measures given a knowledge of the electron number density. In the case of radio pulsars, the dispersion caused by these electrons results in a time delay between pulses received at different wavelengths, which can be measured in terms of the electron column density, or dispersion measure. A measurement of both the dispersion measure and the rotation measure therefore yields the weighted mean of the magnetic field along the line of sight. In particular, Faraday rotation measurements of polarized radio signals from extragalactic radio sources occulted by the solar corona can be used to estimate both the electron density distribution and the direction and strength of the magnetic field in the coronal plasma.

Faraday Rotation In The Ionosphere

Radio waves passing through the Earth's ionosphere are subject to the Faraday Effect. The ionosphere consists of a plasma containing free electrons which contribute to Faraday rotation. Due to the fact that the positive ions are relatively massive, they have little influence. In conjunction with the earth's magnetic field, rotation of the polarization of radio waves thus occurs. Since the density of electrons in the ionosphere varies greatly on a daily basis, as well as over the sunspot cycle, the magnitude of the effect varies. However the effect is always proportional to the square of the wavelength, so even at the Ultra-High Frequency (UHF) television frequency of 500 MHz (λ = 60 cm), there can be more than a complete rotation of the axis of polarization. A consequence is that although most radio transmitting antennas are either vertically or horizontally polarized, the polarization of a medium or short wave signal after reflection by the ionosphere is rather unpredictable. However the Faraday Effect due to free electrons diminishes rapidly at higher frequencies (shorter wavelengths) so that at microwave frequencies, used by satellite communications, the effect is minimized.

8.7.4 THE VOIGT EFFECT

The Voigt Effect is one of a class of effects, resulting in what is called magnetic birefringence or magnetic double refraction. It is a magneto-optical phenomenon with a similar origin to the Faraday Effect. In the Faraday Effect, the polarization of light can be rotated when passed through a transparent medium to which an external magnetic field is applied. The Voigt effect is similar, but while the Faraday Effect is linear in the applied field, the Voigt effect is quadratic in the presence of strong magnetic field. This quadratic scaling stems from an arrangement whereby the external magnetic field is applied at right angles to the direction of propagation. In this case, all the effects that are proportional to the magnetic field vanish. The Voigt effect was discovered in 1902 by Woldemar Voigt. The term Voigt effect is usually reserved for the observation of the aforementioned polarization shift when a vapor plays the role of the transparent medium. When a liquid plays this role, the effect is much stronger (i.e. the proportionality to the square of the magnetic field is greater), and is known as the Cotton-Mouton Effect. Voigt Effect is utilized in Voigt Filters, a type of atomic line filter, where the Voigt effect makes a vapor cell act as a half-wave plate.

8.7.5 COTTON–MOUTON EFFECT

In physical optics, the Cotton–Mouton Effect refers to birefringence in a liquid in the presence of a constant transverse magnetic field. It is a similar but stronger effect than the Voigt Effect. It was discovered in 1907 by Aime Cotton and Henri Mouton. This double refraction is attributed to a lining up of the magnetically and optically anisotropic molecules in the presence of the applied magnetic field direction. This lining up would result whether the magnetic dipole moments of the molecules were induced or permanent. The effect is proportional to the square of the applied field strength, depends on temperature and decreases rapidly with increase in temperature.

8.7.6 KERR MAGNETO-OPTIC EFFECT

In 1888, John Kerr a Scottish physicist discovered that when a plane-polarized light is reflected at normal incidence from the polished pole of electromagnet, it becomes elliptically polarized, with the major axis of the ellipse rotated with respect to the incident light wave vibrations. Thus, the applied magnetic field has given rise to a vibration component called Kerr Component which is perpendicular to the incident light vibrations.

8.7.7 STARK EFFECT AND INVERSE STARK EFFECT.

The Stark Effect is the splitting of spectra lines in the presence of strong electric field. It is analogous to the Zeeman Effect except that the Stark effect arises due to the application of strong electric field while Zeeman Effect arises due to the application on magnetic field. When viewed perpendicular to the electric field, some of the components are observed to be plane-polarized with the electric vector parallel to the field, and others plane-polarized with the electric vector perpendicular to the field This is the transverse Stark Effect. When viewed parallel to the electric field vector, only the normal components will appear, but as ordinary unpolarized light. This is the longitudinal Stark Effect. The Stark Effect with the lines appearing in absorption is called the Inverse Stark Effect, and the effect is proportional to the square of the electric field strength.

8.7.8 ELECTRIC DOUBLE REFRACTION

Electric double refraction is related to the Stark effect and is analogous to magnetic double refraction (Voigt Effect). Double refraction is attributed to the very small difference in frequency of the absorption line for light polarized parallel and perpendicular to the electric lines of force.

8.7.9 KERR ELECTRO-OPTIC EFFECT

This is the effect observed when a medium through which light is passing through is subjected to strong electric field, the medium becomes double refracting. This effect is observed in all state of matter, thus, it cannot be attributed to strain in the matter. This effect is proportional to the length of the medium l, the square of the strength of the electric field E, the wavelength of the light λ and is temperature dependent. Kerr Electro-Optic Effect is different from Electric Double refraction which may arise due to strain induced in the medium by the presence of the applied external field, but rather from induced anisotropy of the molecule and the lining up of the molecules in the direction of the field. This alignment causes the medium to as a whole to be optically anisotropic.

8.7.10 POCKELS ELECTRO-OPTIC EFFECT

In a variety of uniaxial crystals, the induced birefringence varies linearly with the applied electric field. With present technology, variety of electro-optic crystals (Pockels cell) have been developed and are applied in light modulator or shutter. Pockels cell together with Kerr cell (magneto-optic crystals) are used in wide range of electro-optics and magneto-optics systems, including switches to produce ultra-short pulses in lasers, laser beam communication systems, etc.

Exercise Eight

1. Calculate the orbital frequency of the electrons in (i) the first (ii) the second and (iii) the third, Bohr circular orbits. To what wavelength in angstroms would such frequencies belong?
2. Calculate the diameters of (i) the tenth (ii) the twenty-fifth and (iii) the hundredth circular orbits of hydrogen atom according to Bohr's theory.
3. What would be the approximate quantum number n for an orbit of hydrogen to be $1.00\ mm$ in diameter?

References

Jenkins, F.A. and White, H.E., 1985, Fundamentals of optics, 4th Edition, McGraw-Hill Publishers, Singapore, pp745

Young, H.D., and Freedman, R.A., 1995, University Physics, Extended Version 9th Edition, Addison Wesley Longman Publishers, London, pp1463

Yong, P.L., Anyakoha, M.W., and Okeke, P.N, 2002, University Physics II, Africana-FEP Publishers, Onitsha, pp641,

Hecht, E., and Zajac A., 1974, Optics, Addison Wesley Publishers, Philippines, pp565

www.wikipedia.com

Index