1
Basic Optics

Krishna Thyagarajan and Ajoy Ghatak

1.1
Introduction

This chapter on optics provides the reader with the basic understanding of light rays and light waves, image formation and aberrations, interference and diffraction effects, and resolution limits that one encounters because of diffraction. Laser sources are one of the primary sources used in various applications such as interferometry, thermography, photoelasticity, and so on, and Section 1.10 provides the basics of lasers with their special characteristics. Also included is a short section on optical fibers since optical fibers are used in various applications such as holography, and so on. The treatment given here is condensed and short; more detailed analyses of optical phenomena can be found in many detailed texts on optics [1–6].

1.2
Light as an Electromagnetic Wave

Light is a transverse electromagnetic wave and is characterized by electric and magnetic fields which satisfy Maxwell's equations [1, 2]. Using these equations in free space, Maxwell showed that each of the Cartesian components of the electric and magnetic field satisfies the following equation:

$$\nabla^2 \Psi = \varepsilon_0 \mu_0 \frac{\partial^2 \Psi}{\partial t^2} \tag{1.1}$$

where ε_0 and μ_0 represent the dielectric permittivity and magnetic permeability of free space. After deriving the wave equation, Maxwell could predict the existence of electromagnetic waves whose velocity (in free space) is given by

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \tag{1.2}$$

Since

$$\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{C^2\,N^{-1}\,m^{-2}} \quad \text{and} \quad \mu_0 = 4\pi \times 10^{-7}\,\mathrm{Ns^2\,C^{-2}} \tag{1.3}$$

Optical Methods for Solid Mechanics: A Full-Field Approach, First Edition. Edited by Pramod Rastogi and Erwin Hack.

we obtain the velocity of light waves in free space

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} = 2.99794 \times 10^8 \text{ ms}^{-1} \tag{1.4}$$

In a linear, homogeneous and isotropic medium, the velocity of light is given by

$$v = \frac{1}{\sqrt{\varepsilon \mu}} \tag{1.5}$$

where ε and μ represent the dielectric permittivity and magnetic permeability of the medium. The refractive index n of the medium is given by the ratio of the velocity of light in free space to that in the medium

$$n = \sqrt{\frac{\varepsilon \mu}{\varepsilon_0 \mu_0}} \tag{1.6}$$

In most optical media, the magnetic permeability is very close to μ_0 and hence we can write Eq. (1.6) as

$$n \approx \sqrt{\frac{\varepsilon}{\varepsilon_0}} = \sqrt{K} \tag{1.7}$$

where K represents the relative permittivity of the medium, also referred to as the *dielectric constant*.

The most basic light wave is a plane wave described by the following electric and magnetic field variations:

$$\begin{aligned} \boldsymbol{E} &= \hat{\boldsymbol{x}} E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \\ \boldsymbol{H} &= \hat{\boldsymbol{y}} H_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \end{aligned} \tag{1.8}$$

Here, we have assumed the direction of propagation to be along the $+z$-direction and the electric field to be oriented along the x-direction. Note that the electric and magnetic fields are oscillating in phase.

The wave given by Eq. (1.8) represents a plane wave since the surface of constant phase is a plane perpendicular to the z-axis. It is a monochromatic wave since it is described by a single angular frequency ω. It has its electric field along the x-direction and hence is a linearly polarized wave, in this case an x-polarized wave. The electric field amplitude is described by E_0. The amplitudes of the electric and magnetic fields are related through the following equation:

$$H_0 = \frac{k}{\omega \mu} E_0 = \sqrt{\frac{\varepsilon}{\mu}} E_0 \tag{1.9}$$

and ω and k are related through

$$k^2 = \varepsilon \mu \omega^2 = \frac{\omega^2}{v^2} \tag{1.10}$$

The propagation constant of the wave represented by k is related to the wavelength through the following equation:

$$k = \frac{2\pi}{\lambda_0} n \tag{1.11}$$

where λ_0 is the free space wavelength and n represents the refractive index of the medium in which the plane wave is propagating.

The intensity or irradiance of a light wave is the amount of energy crossing a unit area perpendicular to the propagation direction per unit time and is given by the time average of the Poynting vector **S**, which is defined by

$$\langle \boldsymbol{S} \rangle = \frac{1}{2} \operatorname{Re}\left(\boldsymbol{E} \times \boldsymbol{H}^*\right) \tag{1.12}$$

where the * in the superscript represents complex conjugate and angular brackets represent time average. Thus for the wave described by Eq. (1.8), the intensity is given by

$$\langle \boldsymbol{S} \rangle = \frac{k}{2\omega\mu} E_0^2 \hat{z} \tag{1.13}$$

Thus, the intensity of the wave is proportional to the square of the electric field amplitude. If E_0 is a complex quantity, then in Eq. (1.13) E_0^2 gets replaced by $|E_0|^2$. Equation (1.13) can be written in terms of refractive index of the medium by using Eq. (1.11) as

$$I = \frac{n}{2c\mu_0} |E_0|^2 \tag{1.14}$$

where we have assumed the magnetic permeability of the medium to be μ_0.

The following represents an x-polarized wave propagating along the $-z$-direction:

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 e^{\mathrm{i}(\omega t + kz)} \tag{1.15}$$

A y-polarized light wave propagating along the z-direction is described by the following expression:

$$\boldsymbol{E} = \hat{\boldsymbol{y}} E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \tag{1.16}$$

and is orthogonal to the x-polarized wave. The x- and y-polarized waves form two independent polarization states and any plane wave propagating along the z-direction can be expressed as a linear combination of the two components with different amplitudes and phases.

Thus the following two combinations represent right circularly polarized (RCP) and left circularly polarized (LCP) waves, respectively:

$$\boldsymbol{E} = (\hat{\boldsymbol{x}} - \mathrm{i}\hat{\boldsymbol{y}}) E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \tag{1.17}$$

and

$$\boldsymbol{E} = (\hat{\boldsymbol{x}} + \mathrm{i}\hat{\boldsymbol{y}}) E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \tag{1.18}$$

They are represented as superpositions of the x- and y-polarized waves with equal amplitudes and phase differences of $+\pi/2$ or $-\pi/2$. If the amplitudes of the two components are unequal then they would represent elliptically polarized waves, which are the most general polarization states.

The general expression for the electric field of a plane wave propagating along the z-direction is given by

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_1 \mathbf{e}^{\mathrm{i}(\omega t - kz)} + \hat{\boldsymbol{y}} E_2 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \tag{1.19}$$

where E_1 and E_2 are complex quantities and represent the components of the electric field vector along the x- and y-directions, respectively. Linear, circular, and elliptical polarization states are special cases of the expression given in Eq. (1.19).

In the above discussion, we have written a circularly polarized wave as a linear combination of two orthogonally linearly polarized components. In fact, instead of choosing two orthogonally linearly polarized waves as a basis, we can as well choose any pair of orthogonal polarization states as basis states. Thus if we choose right circular and left circular polarization states as the basis states, then we can write a linearly polarized light wave as a superposition of a right and a left circularly polarized wave as follows:

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} = \frac{1}{2}\left(\hat{\boldsymbol{x}} - \mathrm{i}\hat{\boldsymbol{y}}\right) E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} + \frac{1}{2}\left(\hat{\boldsymbol{x}} + \mathrm{i}\hat{\boldsymbol{y}}\right) E_0 \mathbf{e}^{\mathrm{i}(\omega t - kz)} \tag{1.20}$$

The first term on the rightmost equation represents an RCP wave and the second term represents an LCP wave.

The choice of the basis set depends on the problem. Thus, for anisotropic media (which is discussed in Section 1.9), it is appropriate to choose the basis set as linearly polarized waves, since the eigen modes are linearly polarized components. On the other hand, in the case of Faraday effect, it is appropriate to choose circularly polarized wave for the basis set, because when a magnetic field is applied to a medium along the propagation of a linearly polarized light wave, the plane of polarization rotates as the wave propagates.

In Section 1.10, we shall discuss Jones vector representation for the description of polarized light and the effect of various polarization components on the state of polarization of the light wave.

A plane wave propagating along a general direction can be written as

$$\boldsymbol{E} = \hat{\boldsymbol{n}} E_0 \mathbf{e}^{\mathrm{i}(\omega t - \boldsymbol{k}.\boldsymbol{r})} \tag{1.21}$$

where $\boldsymbol{k}$ represents the direction of propagation of the wave and its magnitude is given by Eq. (1.11). The state of polarization of the wave is contained in the unit vector n. Since light waves are transverse waves we have

$$\hat{\boldsymbol{n}}.\hat{\boldsymbol{k}} = 0 \tag{1.22}$$

that is, the propagation vector is perpendicular to the electric field vector of the wave. The vectors $\boldsymbol{E}$, $\boldsymbol{H}$, and $\boldsymbol{k}$ form a right-handed coordinate system (Figure 1.1). The vector $\boldsymbol{k}$ gives the direction of propagation of the wavefronts.

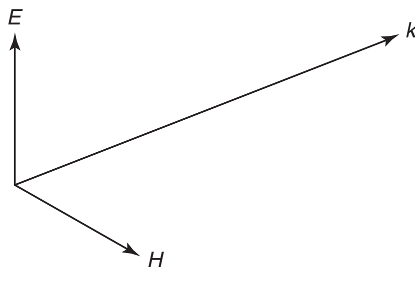

Figure 1.1 A plane wave propagating along a direction specified by k. The electric and magnetic fields associated with the wave are at right angles to the direction of propagation.

Tutorial Exercise 1.1

Consider the superposition of an x-polarized and a y-polarized wave with unequal amplitudes E_1 and E_2 but with the same phase. Write the resulting wave, discuss its nature, and obtain the intensity of this wave.

Solution:

Combining Eqs. (1.1) and (1.4) with amplitudes E_1 and E_2 instead of E_0, we have

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_1 \mathbf{e}^{\mathrm{i}(\omega t - kz)} + \hat{\boldsymbol{y}} E_2 \mathbf{e}^{\mathrm{i}(\omega t - kz)} = (\hat{\boldsymbol{x}} E_1 + \hat{\boldsymbol{y}} E_2)\mathbf{e}^{\mathrm{i}(\omega t - kz)}$$

The above equation represents another linearly polarized wave with its electric field vector $\boldsymbol{E}$ making an angle of $\tan^{-1}(E_2/E_1)$ with the x-axis.Its intensity is given by

$$\begin{aligned} I &= \frac{n}{2c\mu_0}|E|^2 = \frac{n}{2c\mu_0}(\hat{\boldsymbol{x}} E_1 + \hat{\boldsymbol{y}} E_2)\mathbf{e}^{\mathrm{i}(\omega t - kz)} \cdot (\hat{\boldsymbol{x}} E_1^* + \hat{\boldsymbol{y}} E_2^*)\mathbf{e}^{-\mathrm{i}(\omega t - kz)} \\ &= \frac{n}{2c\mu_0}\left(|E_1|^2 + |E_2|^2\right) = I_1 + I_2 \end{aligned}$$

Tutorial Exercise 1.2

Consider a laser emitting a power of 1 mW and having a beam diameter of 2 mm. Calculate the intensity of the laser beam and its field amplitude in air.

Solution:

Intensity is power per unit area; thus

$$I = \frac{10^{-3}\ \mathrm{W}}{\pi \times \left(10^{-3}\ \mathrm{m}\right)^2} = \frac{10^3}{\pi}\frac{\mathrm{W}}{\mathrm{m}^2}$$

From Eq. (1.15), the field amplitude is given by $|E_0| = \sqrt{(2c\mu_0 I)}$. This corresponds to an electric field amplitude of approximately 500 V m^{-1}.

Absorbing media (such as metals) can be described by a complex refractive index:

$$n = n_\mathrm{r} - \mathrm{i}n_\mathrm{i} \tag{1.23}$$

where n_r and n_i represent the real and imaginary parts of the refractive index. The propagation constant also becomes complex and is given by

$$k = k_\mathrm{r} - \mathrm{i}k_\mathrm{i} \tag{1.24}$$

In such media, a plane wave propagating along the z-direction has the following variation of electric field

$$\boldsymbol{E} = \hat{\boldsymbol{x}} E_0 \mathbf{e}^{\mathrm{i}(\omega t - \{k_\mathrm{r} - \mathrm{i}k_\mathrm{i}\}z)} = \hat{\boldsymbol{x}} E_0 \mathbf{e}^{-k_\mathrm{i} z}\mathbf{e}^{\mathrm{i}(\omega t - k_\mathrm{r} z)} \tag{1.25}$$

which shows that the wave attenuates exponentially as it propagates. The attenuation constant is given by k_i.

If we consider a point source of light, then the waves originating from the source would be spherical waves and the electric field of a spherical wave would be described by

$$E = \frac{E_0}{r}\mathbf{e}^{\mathrm{i}(\omega t - kr)} \tag{1.26}$$

Here r is the distance from the point source and is the radial coordinate of the spherical polar coordinate system. The amplitude of the electric field decreases as $1/r$ so that the intensity, which is proportional to the square of the amplitude, decreases as $1/r^2$ in order to satisfy energy conservation. The surfaces of constant phase are given by $r =$ constant and thus represent spheres. Thus the wavefronts are spherical. The wave given by Eq. (1.26) represents a diverging spherical wave. A converging spherical wave would be given by

$$E = \frac{E_0}{r}\mathbf{e}^{\mathrm{i}(\omega t + kr)} \tag{1.27}$$

Light emanating from incoherent sources such as incandescent lamps or sodium lamp are randomly polarized. In many experiments, it is desired to have linearly polarized waves and this can be achieved by passing the randomly polarized light through an optical element called a *polarizer*. The polarizer could be an element that absorbs light polarized along one orientation while passing that along the perpendicular orientation. A Polaroid sheet is such an element and consists of long-chain polymer molecules that contain atoms (such as iodine) that provide high conductivity along the length of the chain. These long-chain molecules are aligned so that they are almost parallel to each other. When a light beam is incident on such a Polaroid, the molecules (aligned parallel to each other) absorb the component of electric field that is parallel to the direction of alignment because of the high conductivity provided by the iodine atoms; the component perpendicular to it passes through. Thus, linearly polarized light waves are produced.

When a randomly polarized laser light is incident on a Polaroid, then the Polaroid transmits only half of the incident light intensity (assuming that there are no other losses and that the Polaroid passes entirely the component parallel to its pass axis). On the other hand, if an x-polarized beam is passed through a Polaroid whose pass axis makes an angle θ with the x-axis, then the intensity of the emerging beam is given by

$$I = I_0 \cos^2 \theta \tag{1.28}$$

where I_0 represents the intensity of the emergent beam when the pass axis of the polarizer is also along the x-axis (i.e., when $\theta = 0$). Equation (1.28) represents Malus' law.

Linearly polarized light can also be produced by the simple process of reflection. If a randomly polarized plane wave is incident at an interface separating media of refractive indices n_1 and n_2 at an angle of incidence (θ) such that

$$\theta_1 = \theta_{\mathrm{B}} = \tan^{-1}\left(\frac{n_2}{n_1}\right) \tag{1.29}$$

then the reflected beam will be linearly polarized with its electric vector perpendicular to the plane of incidence (Tutorial Exercise 1.4).The above equation is known as *Brewster's law* and the angle θ_B is known as the *polarizing angle* (or *Brewster angle*).

Using anisotropic media, it is possible to change the state of polarization of an input light into another desired state of polarization. Anisotropic media are briefly discussed in Section 1.9.

1.2.1 Reflection and Refraction of Light Waves at a Dielectric Interface

Electric and magnetic fields need to satisfy certain boundary conditions at an interface separating two media. When a light wave is incident on an interface separating two dielectrics, in general, it will generate a reflected wave and a transmitted wave (Figure 1.2). The angles of reflection and transmission as well as the amplitudes of the reflected and transmitted waves can be deduced by applying the boundary conditions at the interface.

It follows that the angle of reflection is equal to the angle of incidence, while the angle of refraction and the angle of incidence are related through Snell's law:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{1.30}$$

Thus, as the wave propagates from a rarer medium to a denser medium ($n_1 < n_2$), the wave bends toward the normal of the interface as it gets refracted. On the other hand, when the wave propagates from a denser medium to a rarer medium ($n_1 > n_2$), the wave bends away from the normal. In fact for a certain angle of incidence, the angle of refraction would become equal to 90° and this angle is referred to as the *critical angle*. The critical angle θ_c is given by

$$\sin\theta_c = \frac{n_2}{n_1} \tag{1.31}$$

For an interface between glass of refractive index 1.5 and air, the critical angle is approximately given by 41.8°. Hence for angles of incidence greater than the critical angle, there would be no refracted wave, and such a phenomenon is referred to as *total internal reflection*.

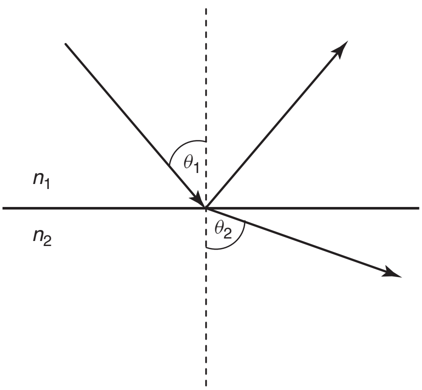

Figure 1.2 A plane wave incident on a dielectric interface generates a reflected and a transmitted wave.

For a light wave polarized in the plane of incidence, the amplitude reflection coefficient, which is the ratio of the amplitude of the electric field of the reflected wave and the amplitude of the electric field of the incident wave, is given by [1]

$$r_p = \frac{n_1 \cos\theta_2 - n_2 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1} \tag{1.32}$$

where n_1 and n_2 represent the refractive indices of the two media and θ_1 and θ_2 represent the angles of incidence and refraction (Figure 1.2).

Similarly, for a light wave polarized perpendicular to the plane of incidence, the amplitude reflection coefficient is given by

$$r_s = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \tag{1.33}$$

The corresponding energy reflection coefficients are given by $R_p = |r_p|^2$ and $R_s = |r_s|^2$, respectively.

Tutorial Exercise 1.3

A light wave is normally incident on an air–glass interface; the refractive index of glass is 1.5. Calculate the amplitude reflection coefficient and the corresponding energy reflection coefficient.

Solution:

For normal incidence, the incident angle is $\theta_1 = 0$; hence the refraction angle, from Eq. (1.30), is also 0°. Using Eq. (1.32) then yields an amplitude reflection coefficient of -0.2 and the energy reflection coefficient is 0.04. The negative sign in the amplitude reflection coefficient signifies a phase change of π on reflection. If the wave is incident on the glass–air interface, then the amplitude reflection coefficient would be positive and there is no phase change on reflection.

Tutorial Exercise 1.4

The reflection coefficient for parallel polarization, r_p, in Eq. (1.32) can become zero. Calculate the corresponding angle. What is the reflection coefficient, r_s for this case? What will happen, if the incoming light is randomly polarized?

Solution:

From Eq. (1.32), it follows that the reflection coefficient will become zero if

$$n_1 \cos\theta_2 = n_2 \cos\theta_1$$

Using Snell's law (Eq. (1.30)), we can simplify the above equation to obtain the condition $\theta_1 + \theta_2 = \pi/2$; this gives us the following angle of incidence:

$$\theta_B = \tan^{-1}\left(\frac{n_2}{n_1}\right)$$

which is referred to as the *Brewster angle*. Inserting $\cos\theta_2 = n_2/n_1 \cos\theta_1$ into Eq. (1.33) yields

$$r_s = \frac{n_1^2 - n_2^2}{n_1^2 + n_2^2}$$

When unpolarized light is incident at the Brewster angle, the reflected light will show s-polarization only, because $r_p = 0$. Hence, unpolarized light will be polarized under reflection at the Brewster angle.

When the light wave undergoes total internal reflection, the angle of refraction satisfies the following inequality derived from Snell's law

$$\sin\theta_2 = \frac{n_1}{n_2}\sin\theta_1 > 1 \tag{1.34}$$

Since $\sin\theta_2 > 1$, $\cos\theta_2$ becomes purely imaginary. Thus the amplitude reflection coefficient (for example, for the case when the incident light is polarized perpendicular to the plane of incidence) becomes

$$r_s = \frac{n_1\cos\theta_1 + \mathrm{i}\alpha}{n_1\cos\theta_1 - \mathrm{i}\alpha} \tag{1.35}$$

where

$$\alpha = \sqrt{n_1^2\sin^2\theta_1 - n_2^2} \tag{1.36}$$

is a real quantity. Thus, r_s will be a complex quantity with unit magnitude and can be written as

$$r_s = \mathbf{e}^{-\mathrm{i}\Phi} \tag{1.37}$$

where

$$\Phi = -2\tan^{-1}\left(\frac{\sqrt{n_1^2\sin^2\theta_1 - n_2^2}}{n_1\cos\theta_1}\right) \tag{1.38}$$

Thus, under total internal reflection, the energy reflection coefficient R_s is unity, that is, the intensity of the reflected wave and the incident wave are equal. The reflected wave undergoes a phase shift on reflection and this phase shift is a function of the angle of incidence of the wave. It can be shown that under total internal reflection, there is still a wave in the rarer medium; the amplitude of this wave decreases exponentially as we move away from the interface and it is a wave that is propagating parallel to the interface. This wave is referred to as an *evanescent wave*.

1.3 Rays of Light

Like any wave, light waves also undergo diffraction as they propagate (Section 1.8). However, we can neglect diffraction effects whenever the dimensions of the object with which light interacts are very large compared to the wavelength or when we do not look closely at points such as the focus of a lens or a caustic. In such a case, we can describe light propagation in terms of light rays. Light rays are directed lines perpendicular to the wavefront and represent the direction of propagation of energy. Thus the light rays corresponding to a plane wave would be just parallel straight arrows; for a spherical wave it would correspond to arrows emerging from the point source and for a more complex wavefront, light rays would be represented by arrows perpendicular to the wavefront at every point. The field of optics dealing with rays is referred to as *geometrical optics* since simple geometry can be used to construct the position of images and their magnification formed by optical instruments. In Section 1.4, we will discuss image formation by optical systems using the concept of light rays.

However, they cannot be used to estimate, for example, the ultimate resolution of the instruments since this is determined by diffraction effects. In Section 1.8, we shall discuss the diffraction phenomenon and how it ultimately limits the resolution of optical instruments such as microscopes, cameras, and so on.

Box 1.1: Rays in an Inhomogeneous Medium

The path of rays in a medium with a general refractive index variation given by $n(x, y, z)$ is described by the following ray equation [1]:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left(n\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}s}\right) = \nabla n$$

where ds is the arc length along the ray and is given by

$$\mathrm{d}s = dz\sqrt{1 + \left(\frac{\mathrm{d}x}{\mathrm{d}z}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}z}\right)^2}$$

For a given refractive index distribution, the solution of the ray equation will give us the path of rays in that medium.

Example: Consider a medium with a parabolic index variation given by

$$n^2(x) = n_1^2\left(1 - 2\Delta\left(\frac{x}{a}\right)^2\right)$$

Substituting the value of $n^2(x)$ in the ray equation, we obtain

$$\frac{\mathrm{d}^2x}{\mathrm{d}z^2} + \Gamma^2 x = 0$$

where

$$\Gamma = \frac{n_1\sqrt{2\Delta}}{\tilde{\beta}a}$$

The solution of the ray equation gives us the ray paths as

$$x(z) = A\sin\Gamma z + B\cos\Gamma z$$

showing that the rays in such a medium follow sinusoidal paths. The constants A and B are determined by the initial launching conditions on the ray.

In homogeneous media, the refractive index n is constant and light rays travel along straight lines. However, in graded index media, in which n depends on the spatial coordinates, light rays propagate along curved paths (Box 1.1). For example, in a medium with a refractive index varying with only x, we can assume the plane of propagation of the ray to be the x–z plane and the ray equation becomes

$$\frac{d^2x}{dz^2} = \frac{1}{2\tilde{\beta}^2}\frac{dn^2(x)}{dx} \tag{1.39}$$

where

$$\tilde{\beta} = n(x)\cos\theta(x) \tag{1.40}$$

with $\theta(x)$ representing the x-dependent angle made by the ray with the z-axis. The quantity $\tilde{\beta}$ is a constant of motion for a given ray, and as the ray propagates this quantity remains constant.

Tutorial Exercise 1.5

Consider a medium with the following refractive index variation:

$$n^2(x) = n_1^2(1 + \alpha x)$$

over the region $0 < x < x_0$. This represents a linear variation of refractive index. Calculate the ray path.

Solution:

In such a medium, the ray equation can be integrated easily and the ray path in the region $0 < x < x_0$ is given by

$$x(z) = \left(\frac{\alpha n_1^2}{4\tilde{\beta}^2}\right) z^2 + C_1 z + C_2$$

where C_1 and C_2 are constants determined by initial launch conditions of the ray. The ray paths are thus parabolic.

The ray analysis given above is used in obtaining ray paths and imaging characteristics of optical systems containing homogeneous lenses or graded index (GRIN) lenses, in understanding light propagation through multimode optical fibers, and in many other applications.

1.4
Imaging through Optical Systems

Since the wavelength of the light is negligible in comparison to the dimensions of optical devices such as lenses and mirrors, we can approximate the propagation of light through geometrical optics.

The propagation of rays through such systems is primarily based on the laws of refraction at the boundary between media of different refractive indices and reflection at mirror surfaces. In the paraxial approximation, we assume that the angle made by the rays with the axis of the optical system is small and also that they propagate close to the axis. Using such an approximation, we find that optical systems can form perfect images and that we can obtain the basic properties such as position of the images, their magnifications, and so on, using ray optics. In this section, we shall use the paraxial approximation to study formation of images by lenses and then in Section 1.5, we shall discuss the various aberrations suffered by the images. More details can be found in Ref. [7].

When an optical system consists of many components then in the paraxial approximation, it is easy to formulate the properties of the optical system in terms of matrices which describe the changes in the height and the angle made by the ray with the axis of the system as the ray undergoes refraction at different interfaces or propagates through different thicknesses of the media. These matrices are obtained by applying Snell's law at the interfaces and also using the fact that rays propagate along straight lines in homogeneous media.

We represent a ray by a column matrix as follows:

$$\begin{pmatrix} x \\ \alpha \end{pmatrix} \tag{1.41}$$

where x represents the height or radial distance of the ray from the axis and α represents the angle made by the ray with the axis. Since we are using the paraxial approximation, we assume $\sin\alpha \approx \tan\alpha \approx \alpha$. We assume that rays propagate from left to right and that convex surfaces have positive radii of curvature while concave surfaces have negative radii of curvatures. We also define that rays pointing upwards have a positive value of α while rays pointing downwards have a negative value of α.

Figure 1.3(a) shows a ray refracting at a spherical interface of radius of curvature R between media of refractive indices n_1 and n_2 and Fig. 1.3(b) propagation through a homogeneous medium of thickness d. The effect of refraction at the interface

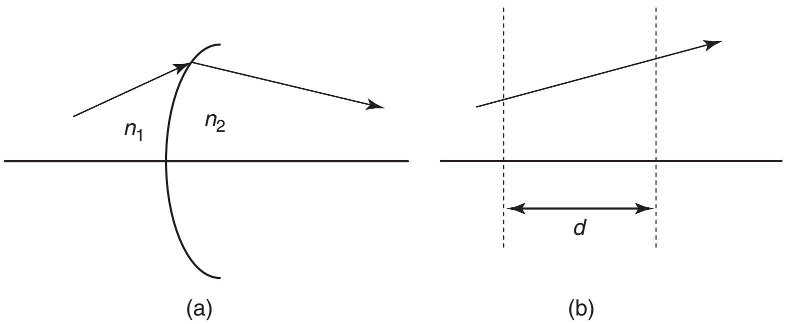

Figure 1.3 (a) Shows a ray refracting at an interface between two media of refractive indices n_1 and n_2. (b) A ray propagating through homogeneous medium.

and propagation through the medium are represented by square matrices.

$$\text{Refraction:}\begin{pmatrix} 1 & 0 \\ \frac{(n_1-n_2)}{n_2 R} & \frac{n_1}{n_2} \end{pmatrix}$$

$$\text{Propagation:}\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$$

1.4.1
Thin Lens

A thin lens consists of two refracting surfaces and since it is thin we neglect the effect of the propagation of the ray between the two interfaces within the lens. The effect of the two interfaces is a product of the matrices corresponding to refraction at the first interface of radius of curvature R_1 between media of refractive indices n_1 and n_2 and the matrix corresponding to refraction at the second interface of radius of curvature R_2 between media of refractive indices n_2 and n_1. Thus the effect of a thin lens is given by matrix corresponding to a thin lens:

$$\begin{pmatrix} 1 & 0 \\ \frac{(n_2-n_1)}{n_1 R_2} & \frac{n_2}{n_1} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \frac{(n_1-n_2)}{n_2 R_1} & \frac{n_1}{n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \tag{1.42}$$

where

$$\frac{1}{f} = \frac{(n_2-n_1)}{n_1}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{1.43}$$

f is the focal length of the lens.

1.4.2
Thick Lens

In case we cannot neglect the finite thickness of the lens, then we also need to consider the effect of propagation of the ray between the two refracting surfaces. Thus if the thickness of the lens is d (the distance between the points of intersection of the lens surfaces with the axis), then the effect of a thick lens is given by

$$\begin{pmatrix} 1 & 0 \\ \frac{(n_2-n_1)}{n_1 R_2} & \frac{n_2}{n_1} \end{pmatrix}\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \frac{(n_1-n_2)}{n_2 R_1} & \frac{n_1}{n_2} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{1.44}$$

where

$$A = 1 - \frac{(n_2-n_1)}{n_2 R_1} d \tag{1.45}$$

$$B = \frac{n_1}{n_2} d \tag{1.46}$$

$$C = \frac{(n_2-n_1)}{n_1}\left(\frac{1}{R_1} - \frac{1}{R_2}\right) - \frac{(n_2-n_1)^2}{n_1 n_2 R_1 R_2} d \tag{1.47}$$

$$D = 1 + \frac{(n_2 - n_1)}{n_2 R_2} d \tag{1.48}$$

Note that for thin lenses, we can assume $d = 0$ and we get back the matrix for a thin lens, as shown in Eq. (1.42).

We can use the formalism given above to study a combination of lenses. For example, if we have two lenses of focal lengths f_1 and f_2 separated by a distance d, then the overall matrix representing the combination would be given by the product of the matrices corresponding to the first lens, propagation through free space, and the second lens:

$$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f_2} & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_1} & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{d}{f_1} & d \\ -\frac{1}{f_1} - \frac{1}{f_2} + \frac{d}{f_1 f_2} & 1 - \frac{d}{f_2} \end{pmatrix}$$

The first element in the second row contains the focal length of the lens combination and hence the combination is equivalent to a lens of focal length

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2} \tag{1.49}$$

1.4.3
Principal Points of a Lens

Let us consider a pair of planes enclosing an optical system, which could be a thick lens, a combination of thin or thick lenses, and so on (Figure 1.4). Knowing the components and their spacing, we can use the above formalism to get the matrix connecting the rays between the two planes. Let us assume a ray starts from a height x_o making an angle of θ_o from the input plane and let us assume that the coordinates of the ray in the final plane are x_f and θ_f. Hence we have

$$\begin{pmatrix} x_f \\ \theta_f \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x_o \\ \theta_o \end{pmatrix} \tag{1.50}$$

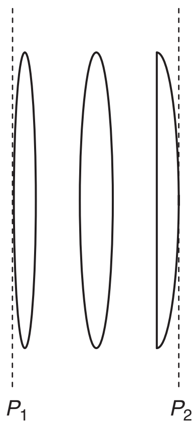

Figure 1.4 An optical system formed by three lenses. The system matrix describes the matrix of propagation of the ray from plane P_1 to plane P_2.

This matrix equation is equivalent to the two vector component equations:

$$x_f = Ax_o + B\theta_o$$
$$\theta_f = Cx_o + D\theta_o \tag{1.51}$$

If we choose an input plane such that θ_f is independent of θ_o, then this would imply the condition

$$D = 0 \tag{1.52}$$

This input plane must be the front focal plane because all rays from a point on the front focal plane entering the optical system under different angles emerge parallel from the final plane. Similarly, if x_f is independent of x_o, then this implies that rays coming into the optical system at a particular angle converge to one point x_i and the corresponding plane should be the back focal plane. Hence for the back focal plane

$$A = 0 \tag{1.53}$$

If the two planes are such that $B = 0$, then this implies that the output plane is the image plane and A represents the magnification. Similarly if $C = 0$, then a parallel bundle of rays will emerge as a parallel beam and D represents the angular magnification.

Tutorial Exercise 1.6

As an example, we consider a combination of two thin lenses shown in Figure 1.5. The first lens is assumed to be a converging lens of focal length 20 cm and the second a diverging lens of focal length 8 cm. We assume that they are separated by a distance of 13 cm. Calculate the ABCD matrix connecting the plane S_1 and S_2 and the focal length of this lens system:

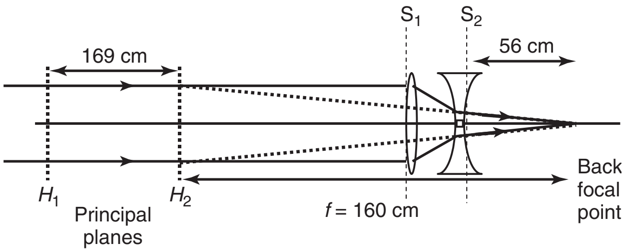

Figure 1.5 An optical system consisting of a double convex and a double concave lens. H_1 and H_2 are the principal planes.

Solution:

Inserting the values (in centimeters) into the corresponding matrices we obtain

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{1}{8} & 1 \end{pmatrix} \begin{pmatrix} 1 & 13 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{20} & 1 \end{pmatrix} = \begin{pmatrix} 0.35 & 13 \\ -0.00625 & 2.625 \end{pmatrix}$$

The focal length of the combination is given by $1/0.00625 = 160$ cm.

In the example discussed above, we obtained the focal length of the lens combination. However, we need to know from where this distance needs to be measured. For a single thin lens, we measure distances from the center of the lens; however, for combination of lenses, we need to determine the planes from where we must measure the distances. To understand this let us consider a plane at a distance u in front of the first lens and a plane at a distance v from the second lens. The matrix connecting the rays at these two planes is

$$\begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A + Cv & B - Au + v(D - Cu) \\ C & D - Cu \end{pmatrix} \tag{1.54}$$

The image plane is determined by the condition

$$B - Au + v(D - Cu) = 0$$

which can be simplified to

$$-\frac{1}{(u - u_\mathrm{p})} + \frac{1}{(v - v_\mathrm{p})} = \frac{1}{f} \tag{1.55}$$

where

$$u_\mathrm{p} = \frac{(D - 1)}{C};\ v_\mathrm{p} = \frac{(1 - A)}{C};\ f = -\frac{1}{C} \tag{1.56}$$

Thus imaging by the lens combination can be described by the same formula as for a simple lens provided we measure all distances from appropriate planes. The object distance is measured from a point with a coordinate u_p with respect to the front plane S_1 of the lens combination and the image distance is measured from the point with the coordinate v_p with respect to the back plane S_2 of the lens combination. Positive values of these quantities imply that they are on the right of the corresponding planes and negative values imply that they are on the left of the corresponding planes. The planes perpendicular to the axis and passing through these points are referred to as the *principal planes* of the imaging system. The back focal length of the system is f and measured from the second principal plane. Hence, it is at a distance $-A/C$ from the back plane S_2. Similarly, the front focal distance measured from the plane S_1 is $-D/C$.

Tutorial Exercise 1.7

Obtain the positions of the front and back principal planes of Tutorial Exercise 1.6.

Solution:

Substituting the values of the various quantities, we obtain the back focal distance from the second lens as $-0.35/(-0.000625) = 56\,\mathrm{cm}$ and the front focal distance from the first lens as $2.625/0.00625 = 420\,\mathrm{cm}$. For the given system $u_\mathrm{p} = 260$ cm and $v_\mathrm{p} = -104$ cm. The first and second principal planes are situated as shown in Figure 1.3.

The advantage of using matrices to describe the optical system is that it is easily amenable to programing on a computer and complex systems can be analyzed easily. Of course, the analysis is based on paraxial approximation. To obtain the quality of the image in terms of aberrations, and so on, one would need to carry out more precise ray tracing through the optical system.

1.5 Aberrations of Optical Systems

The paraxial analysis used in Section 1.4 assumes that the rays do not make large angles with the axis and also lie close to the axis. Principally in this analysis $\sin\theta$ is replaced by θ, where θ is the angle made by the ray with the axis of the optical system. In such a situation, perfect images can be formed by optical systems. In actual practice, not all rays forming images are paraxial and this leads to imperfect images or aberrated images. Thus, the approximation of replacing $\sin\theta$ by θ fails and higher order terms in the expansion have to be considered. The first additional term that would appear would be of third power in θ and hence the aberrations are termed as *third-order aberrations*. These aberrations do not depend on the wavelength of the light and are termed as *monochromatic aberrations*. When the illumination is not monochromatic, we have additional contribution coming from chromatic aberration because of the dispersion of the optical material used in the lenses.

1.5.1 Monochromatic Aberrations

There are five primary monochromatic aberrations: spherical aberration, coma, astigmatism, curvature of field, and distortion.

1.5.2 Spherical Aberration

According to paraxial optics, all rays parallel to the axis of a converging system focus at one point behind the lens. However, if we trace rays through the system then we find that rays farther from the axis intersect the axis at a different point compared to rays closer to the axis. This is termed as *longitudinal spherical aberration*. Thus, if we place a plane corresponding to the paraxial focal plane, then all rays would not converge at one point and rays farther from the axis will intersect the plane at different heights. The transverse distance measured on the focal plane is termed *transverse spherical aberration*. This implies that if we begin with a very small aperture in front of the lens, then the image would almost be perfect. However, as we increase the size of the aperture, the size of the focused spot on the focal plane will increase, leading to a drop in quality of the image.

It is possible to minimize spherical aberration of a single converging lens by ensuring that both the surfaces contribute equally to the focusing of the incident light. By using lens combinations, it is possible to eliminate spherical aberration by using aspherical lenses or combination of lenses. Thus a plano-convex lens with the spherical surface facing the incident light would have lower aberration than the same lens with the plane surface facing the incident light. In fact, for far off objects the plano-convex lens is close to having the smallest spherical aberration and hence is often used in optical systems.

For points lying on the axis of the optical system, the aberration of the image is only due to spherical aberration.

1.5.3 Coma

For off-axis points, each circular zone of the lens has a different magnification and forms a circular image and the circular images together form a coma-type image shape. By adjusting the radii of the surfaces of the lens, keeping the focal length the same, it is possible to minimize coma similar to spherical aberration. An optical system free of spherical aberration and coma is referred to as an *aplanatic lens*.

1.5.4 Astigmatism and Curvature of Field

Consider rays emerging from an off-axis point and passing through an optical system that is free of spherical aberration and coma. In this case, rays in the tangential plane and the sagittal planes come to focus at single points but they do not coincide. Thus, on the planes where these individually focus the images are in the form of lines brought about by the other set of rays. This aberration is termed as *astigmatism*. At a point intermediate between these two line images, the image will become circular and this is termed as the *circle of least confusion*.

Now for different off-axis points, the surface where the tangential rays form the image is not perpendicular to the axis but is in the form of a paraboloidal surface. Similarly, the sagittal rays form an image on a different paraboloidal surface. Elimination of astigmatism would imply that these two surfaces coincide but they would still not be a plane perpendicular to the axis. This effect is called *curvature* of field.

1.5.5 Distortion

Even if all the above aberrations are eliminated, the lateral magnification could be different for points at different distances from the axis. If the magnification increases with the distance then it leads to pincushion distortion while if it decreases with distance from the axis, then it is referred to as *barrel distortion*. In either of the two cases, the image will be sharp but will be distorted.

Aberration minimization is very important in optical system design. Nowadays, lens design programs are available that can trace different sets of rays through the system without making any approximations and by plotting points where they intersect an image plane, one can form a spot diagram. By changing the parameters of the optical system, one can see changes in the spot diagram, and some optimization routine is used to finally get an optimized optical system.

1.6 Interference of Light

Light waves follow the superposition principle and hence when two or more light waves superpose at any point in space then the total electric field is a superposition of the electric fields of the two waves at that point and depending on their phase difference, they may interfere constructively or destructively. This phenomenon of interference leads to very interesting effects and has wide applications in nondestructive testing, vibration analysis, holography, and so on.

In order that the interfering waves form an observable interference pattern, they must originate from coherent sources. Section 1.7 discusses the concept of coherence.

1.6.1 Young's Double-Slit Arrangement

Figure 1.6 shows Young's double-slit experiment in which light emerging from a slit S with infinitesimal width illuminates a pair of infinitesimal slits S_1 and S_2 separated by distance d from each other. On the other side of the double slit, there is a screen placed at a distance D from the slits. If we consider a point P at a distance x from the axis of the setup, then assuming $D \gg d$, the path difference between the waves arriving at the point P from S_1 and S_2 would be approximately given by

$$\Delta = d\frac{x}{D} \tag{1.57}$$

When the path difference is an even multiple of $\lambda_0/2$, then the waves from S_1 and S_2 interfere constructively, leading to a maximum in intensity. On the other hand, if the path difference is an odd multiple of $\lambda_0/2$, then the waves will interfere destructively, leading to a minimum in intensity. Thus the positions of constructive interference in the screen are given by

$$x_{\max} = m\frac{\lambda_0 D}{d}; \quad m = 0, \pm 1, \pm 2, \ldots \tag{1.58}$$

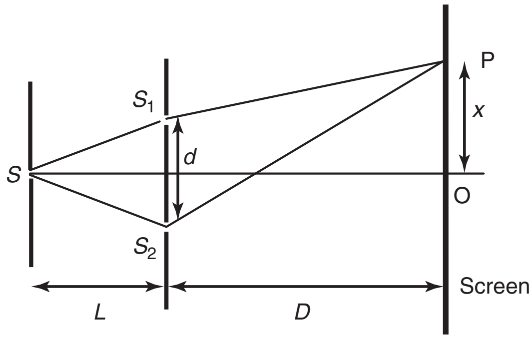

Figure 1.6 Young's double-slit experiment. Light from a source S illuminates the pair of slits S_1 and S_2. Light waves emanating from S_1 and S_2 interfere on the screen to produce an interference pattern.

The separation β between two adjacent maxima, which is referred to as the *fringe width* is

$$\beta = \frac{\lambda_0 D}{d} \tag{1.59}$$

The intensity pattern on the screen along the x-direction would be given by

$$I = 4I_0 \cos^2\left(\frac{\pi x d}{\lambda_0 D}\right) \tag{1.60}$$

where I_0 is the intensity produced by the source S_1 or S_2 in the absence of the other source.

Box 1.2: Antireflective Coating and Fabry-Perot Filters

If we consider a thin film of refractive index n_f and thickness d coated on a medium of refractive index n_s placed in air, then light waves at a wavelength λ_0 incident normally on the film will undergo reflection at both the upper and lower interfaces. If the reflectivities at the interfaces are not large, then we can neglect multiple reflections of the light waves. The path difference between the two reflected waves (one reflected from the upper surface and one from the lower surface) would be

$$\Delta = 2n_f d$$

Since in this case each of the two reflections occurs from a denser medium, it undergoes a phase shift of π on reflection; hence the only phase difference between the two waves is because of the extra path length that the reflected wave from the lower surface takes. If we assume that $1 < n_f < n_s$, then we would have constructive interference when

$$\Delta = m\lambda_0; \quad m = 1, 2, 3, \ldots$$

and for

$$\Delta = \left(m + \frac{1}{2}\right)\lambda_0; \quad m = 1, 2, 3, \ldots$$

we would have destructive interference. If the refractive index of the film is the geometric mean of the refractive indices of the two media surrounding the film [2] then the amplitudes of the two interfering beams are equal and we would have complete destructive interference. This is the principle behind antireflection coatings. The required minimum film thickness is

$$d = \frac{\lambda_0}{4n_f}$$

Example: Let us calculate the refractive index and thickness of a film to reduce the reflection at a wavelength of 600 nm from a medium of refractive index 2 placed in air. The required refractive index is $\sqrt{2} \approx 1.414$ and thickness is 0.106 μm.

If the reflectivities of the two surfaces become large, then we cannot neglect the presence of multiple reflections. If we assume that the reflectivities of the two surfaces are equal and is represented by R, then the overall intensity transmittance of the film would be given by

$$T = \frac{1}{1 + F\sin^2\left(\frac{\delta}{2}\right)}$$

where

$$F = \frac{4R}{(1-R)^2}$$

is called the *coefficient of finesse* and δ represents the phase difference accumulated during one back and forth propagation of the wave through the film. For normal incidence, it is just $(2\pi/\lambda_0)2n_f d$. If the angle made by the waves inside the film is θ_f, then

$$\delta = \frac{2\pi}{\lambda_0} 2n_f d \cos\theta_f$$

Note that when δ is an integral multiple of 2π, then all the incident light gets transmitted. If R is close to unity, then the transmittance drops very quickly as δ changes. Figure 1.7 shows a typical transmission spectrum of a film with $R = 0.9$. The changes in δ could be brought about by changes in the wavelength of the incident radiation, the thickness or the refractive index of the medium between the two highly reflecting surfaces, or by changes in the angle of illumination. Thus in transmission, this produces very sharp interference effects. This interference phenomenon is referred to as *multiple beam interference* and the Fabry–Perot interferometer and the Fabry–Perot etalon are based on this principle. By scanning the distance between the two highly reflecting surfaces, it is possible to measure very precisely the spectrum of the incident radiation and this is widely used in spectroscopy. Fabry–Perot etalons are also widely used inside laser cavities for selecting a single frequency of oscillation.

The resolving power of a Fabry–Perot interferometer is given by

$$R = \frac{\lambda_0}{\Delta\lambda} = \frac{\pi d\sqrt{F}}{\lambda}$$

where it is assumed that the Fabry–Perot interferometer operates at normal incidence.

In almost all interferometers light from the given source is split into two parts by for example using beam splitters, and made to interfere after propagating through two different paths by optical components such as mirrors. This ensures that the interfering beams are coherent, leading to formation of good contrast interference. Creating any difference in the propagation paths of the two interfering beams changes the interference condition, leading to changes in intensity. This is used in instrumentation for measuring spectra of light sources, for very accurate displacement measurement, for surface evaluation, and for many other applications.

Figure 1.8 shows a Michelson interferometer arrangement. S represents an extended near monochromatic source, G represents a 50% beam splitter, and M_1 and M_2 are two plane mirrors. The mirror M_2 is fixed while the mirror M_1 can be moved either toward or away from G. Light from the source S is incident on

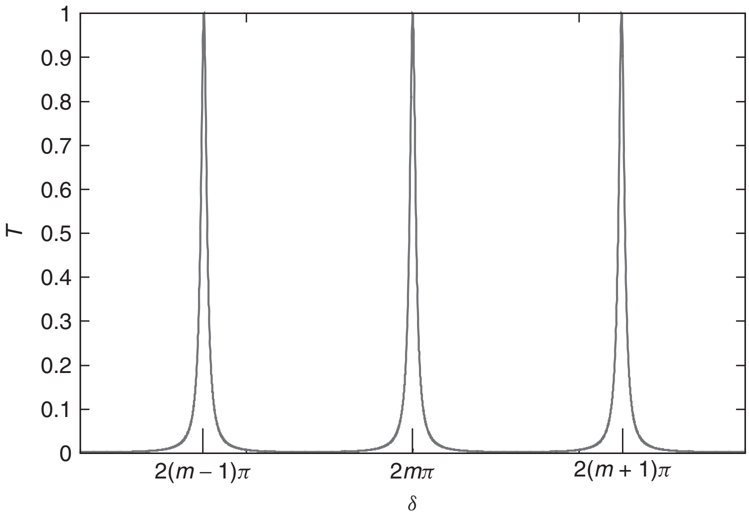

Figure 1.7 The transmittance of a Fabry–Perot interferometer. The maxima of transmission correspond to integral multiples of 2π. The higher the reflectivity of the mirrors the sharper would be the peaks.

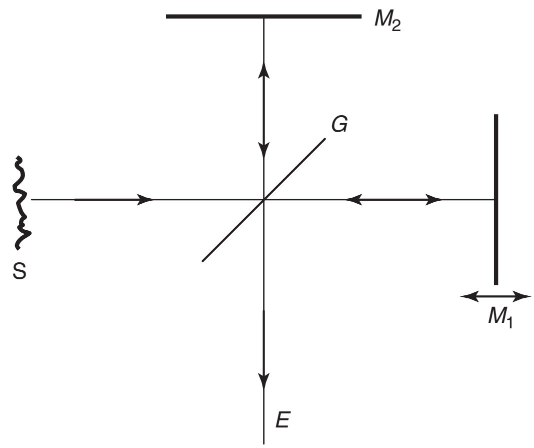

Figure 1.8 Michelson interferometer arrangement.

G and is divided into two equal amplitude portions; one part travels toward M_1 and is reflected back to the same beam splitter and the other part is reflected back from M_2 to the beam splitter. At the beam splitter, both beams partially undergo reflection and transmission and interfere and produce interference fringes that are visible from the direction *E*. If the two mirrors are perpendicular to each other and the beam splitter is at 45° to the incident beam, then the system is equivalent to interference fringes formed by a parallel plate illuminated by a near monochromatic extended source, and we obtain circular fringes of equal inclination. If the mirror M_1 is moved toward or away from the beam splitter, depending on its distance from the beam splitter vis-à-vis mirror M_2, the circular fringes either contract toward the center or expand away from the center. Each fringe collapsing at the center of the pattern corresponds to a movement of $\lambda_0/2$ of the mirror M_1. Thus, movements can be measured very precisely using this interferometer. When illuminated by broadband sources, the fringes will appear only when the path lengths to both mirrors are almost equal. (In this case, one has to use an additional glass plate identical to the beam splitter to ensure that the optical path lengths of the two arms are the same for all wavelengths of the incident light beam.) A measurement of variation of contrast with the displacement of mirror M_1 can be used to measure the spectrum of the source and this is referred to as *Fourier transform spectrometry*.

Instead of incoherent sources, lasers are often used with the interferometers and in this case even large path differences achieve good contrast interference. Laser interferometers are used for high precision measurements, measuring and controlling displacements from a few nanometers to about 100 m, used for measuring angles very precisely, to measure flatness of surfaces, velocity of moving objects, vibrations of objects, and so on. It is also possible to use interference among two lasers with slightly different frequencies leading to beating of the two laser beams. This is termed *heterodyne interferometry*.

In holography (Chapter 6) interference between a reference wave and the wave scattered from the object is recorded. This is used in many applications such as for measuring tiny displacements of objects in real time or to identify defects using interference between the object wave emanating from the object under two different strain conditions. Vibration interferometry can be used to identify the state of vibration of the object using this principle. In holographic interferometry, the object need not be specularly reflecting, and thus this is a very powerful technique for nondestructive testing of various objects.

When objects with a rough surface are viewed under laser illumination, one sees a granular pattern, which is referred to as the *speckle pattern*. In this, the light waves get scattered from different points and propagate in different directions. Thus the light reaching any point on a screen consists of these various scattered waves. Owing to the nature of the surface the phases of the various waves reaching the given point on the screen may lie anywhere between 0 and 2π. With an illumination from a coherent source such as a laser, when scattered waves with these random phases are added, the resultant could lie anywhere between a maximum and a minimum value. At a nearby point, the waves may add to generate a completely different intensity value. In such a circumstance what we observe on the screen is a speckle pattern. The mean speckle diameter is approximately given by

$$s \approx 1.22\frac{\lambda_0 L}{d} \tag{1.61}$$

where λ_0 is the wavelength of illumination, L is the distance between the screen and the rough surface, and d is the diameter of the region of illumination of the object. If instead of allowing the light to fall on a screen, an imaging system is used to form an image, then again we see a speckle pattern due to interference effects. In this case, the mean speckle diameter s is approximately given by the following relation:

$$s \approx 1.22\lambda_0 F(1 + M) \tag{1.62}$$

where F is the F-number of the lens (focal length divided by diameter of the lens) and M is the magnification.

Tutorial Exercise 1.8

Obtain the speckle size for the following situations:

1) Reflections from a laser spot of $d = 2$ cm and a wavelength of $\lambda = 500$ nm on a screen in a distance of $L = 1$ m.
2) Let us assume that we observe from a distance of 100 cm with the naked eye a rough surface illuminated by a laser with a wavelength of 600 nm. Assume a pupil diameter of 4 mm and an eye length of 24 mm.

Solution:

1) Substituting the values into Eq. (1.61) we obtain $s \sim 30$ μm.

2) From the pupil diameter and the length of the eye we obtain an *F*-number of 6. Using Eq. (1.61), we obtain for the approximate size of the speckle as formed on the retina $s \sim 4.4\ \mu\text{m}$.

Speckle contrast measurement has proved to be a powerful tool for the non-destructive testing of small surface roughness within the light wavelength. More details of speckle techniques are given in Chapter 6.

1.7 Coherence

Light sources are never perfectly monochromatic; they emit over a range of wavelengths. If the spectral bandwidth of emission is very small the source is termed *quasi-monochromatic*. When the source emits more than one wavelength, then in an interference setup each wavelength would form its own interference pattern and what one observes is a superposition of the interference patterns from different wavelengths. Thus, if we consider a source emitting two wavelengths (λ_0 and $\lambda_0 - \Delta\lambda$), then when we start from zero path difference between the two interfering waves, the maxima and minima of the two wavelengths would coincide and we will observe very good contrast fringes. Since the fringe width depends on the wavelength, with increasing path difference between the two waves, the maxima and minima of each wavelength would occur at different positions and thus the contrast would begin to fall. In the case of two wavelengths, the contrast will become zero when the maxima of one wavelength fall on the minima of the other wavelength and vice versa. This will happen for a path difference l_c given by

$$l_c = m\lambda_0 = \left(m + \frac{1}{2}\right)(\lambda_0 - \Delta\lambda) \tag{1.63}$$

Eliminating m from the two equations, we obtain

$$l_c = \frac{\lambda_0^2}{2\Delta\lambda} \tag{1.64}$$

For a source emitting a continuous range of wavelengths from λ_0 to $(\lambda_0 - \Delta\lambda)$, the expression for coherence length becomes

$$l_c = \frac{\lambda_0^2}{\Delta\lambda} = \frac{c}{\Delta\nu} \tag{1.65}$$

where $\Delta\nu$ represents the spectral width in frequency.

Coherence length is a very important property of a source as it defines the maximum path difference permitted between the interfering waves so that the interference pattern formed has good contrast. For good contrast fringes, the path difference must be much smaller than the coherence length and as the path difference approaches the coherence length and exceeds it, the contrast in the fringes would decrease steadily, and finally no interference pattern will be visible.

Tutorial Exercise 1.9

Consider two lasers operating at 800 nm with spectral widths of (i) 1 nm and (ii) 0.001 nm and calculate the coherence lengths. Which one would you prefer to measure displacements of a few centimeters?

Solution:

Using Eq. (1.65) the coherence length of the two lasers would be (i) 0.64 mm and (ii) 64 cm, respectively. The latter laser has a much larger coherence length and setting up interference experiments with that laser would be much easier. Using the first laser, one would have to ensure that the maximum path difference is much less than 0.64 mm for good contrast interference.

A laser beam has a well-defined phase front and all points across the wavefront are coherent with respect to each other. Thus if we illuminate a double slit with the laser beam, then the waves emerging from the two slits will exhibit a good contrast interference pattern. When extended incoherent sources such as sodium lamps are used, then the different points across the source are not coherent with respect to each other. In order to form good contrast interference, we would also need to ensure spatial coherence of the source.

Consider Young's double-slit experiment (Figure 1.6); assume we had two sources, one on the axis, S, as shown in the figure, and another source S' displaced from the axis by a distance l. Then at point O, the source S will produce a maximum of intensity, but the intensity produced by S' would depend on l. In fact, if L represents the distance from the plane of the source S to the double-slit arrangement, and if l is such that the path difference $(S'S_2 - S'S_1) = \lambda_0/2$, then the source S' would produce a minimum of intensity at O and the contrast in the interference pattern at O will be nearly zero. This will happen when the following condition is satisfied (assuming $L \gg l, d$):

$$S'S_2 - S'S_1 = d\frac{l}{L} = \frac{\lambda_0}{2} \tag{1.66}$$

If we represent by $\theta(\sim l/L)$ the angle subtended by the pair of sources S and S' at the plane of the slits, then Eq. (1.66) gives

$$d = l_\mathrm{w} = \frac{\lambda_0}{2\theta} \tag{1.67}$$

where l_w is referred to as the *lateral coherence width*. The above discussion pertains to a pair of sources; if the source is an extended source and subtends an angle θ at the plane of the slits, then the lateral coherence width is given by

$$l_\mathrm{w} = \frac{\lambda_0}{\theta} \tag{1.68}$$

Tutorial Exercise 1.10

Consider a double-slit experiment to be conducted using a sodium lamp ($\lambda_0 = 589$ nm) with a pinhole having a diameter of 2 mm placed in front of it.

The distance of the slits is assumed to be 0.3 mm. What is the minimum distance to the lamp needed to obtain good contrast fringes?

Solution:

The lateral coherence width of the source at a distance of 1 m would be $l_w \sim 0.3$ mm. Thus for forming good contrast interference pattern, the distance to the lamp must be larger than 1 m. The lateral coherence width can be increased either by decreasing the size of the source or by increasing the distance between the source and the double slit.

1.8 Diffraction of Light

Plane waves described by Eq. (1.9) have infinite extent in the transverse direction; for such a wave, the transverse amplitude distribution remains the same as the wave propagates; only the phase changes because of propagation. However, any wave that has an amplitude or phase depending on the transverse coordinate will undergo changes in the transverse field distribution. This phenomenon is referred to as *diffraction*.

If $f(x,y)$ is the transverse complex amplitude distribution of a wave on a plane $z = 0$, then the amplitude distribution on any plane z is given by the following equation [1]:

$$f(x, y, z) = \frac{\mathrm{i}}{\lambda z}\mathbf{e}^{-\mathrm{i}kz} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta) \exp\left[-\mathrm{i}\frac{k}{2z}\left\{(x-\xi)^2 + (y-\eta)^2\right\}\right] \mathrm{d}\xi\,\mathrm{d}\eta \tag{1.69}$$

Thus the transverse amplitude distribution changes, in general, as the field propagates and this is referred to as *Fresnel diffraction*. Note that Fresnel diffraction is a two-dimensional convolution operation as given below:

$$f(x, y, z) = \frac{\mathrm{i}}{\lambda z}\mathbf{e}^{-\mathrm{i}kz} f(x, y) \otimes \exp\left[-\mathrm{i}\frac{k}{2z}\left(x^2 + y^2\right)\right] \tag{1.70}$$

If the distance of the observation plane z satisfies the following condition

$$z \gg \frac{a^2}{\lambda} \tag{1.71}$$

where a is the transverse dimension of the field pattern on the plane $z = 0$, then the Fresnel diffraction pattern reduces to the Fraunhofer diffraction pattern. In this case, we can approximate the field distribution also referred to as *far field distribution* by the following equation:

$$f(x, y, z) = C \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)\mathbf{e}^{2\pi\mathrm{i}(u\xi + v\eta)} \mathrm{d}\xi\,\mathrm{d}\eta \tag{1.72}$$

where

$$u = \frac{x}{\lambda z}; \quad v = \frac{y}{\lambda z} \tag{1.73}$$

represent spatial frequencies along the coordinate directions x and y respectively, and C is a constant. It can be seen that Eq. (1.72) is nothing but a two-dimensional Fourier transform. Thus the Fraunhofer diffraction is the Fourier transform of the field distribution on the initial plane.

When we observe the field distribution on the back focal plane of a lens, we obtain the Fraunhofer diffraction pattern. This can be seen from the fact that a parallel beam of light falling on the lens focuses on a point on the back focal plane. Thus, each point on the back focal plane corresponds to a certain direction of propagation in space and the pattern on the back focal plane represents the Fraunhofer diffraction pattern.

If we consider a plane wave falling on a converging lens, then the wave focuses to the focal point of the lens. In this case, the transverse dimension of the lens acts as an aperture and thus on the back focal plane of the lens we would observe the Fraunhofer diffraction pattern corresponding to a circular aperture of dimension equal to that of the lens aperture. Thus, although geometrical optics predicts focusing on a point image (in the absence of aberrations), the focused image will have a finite size because of diffraction effects.

The Fraunhofer diffraction through a circular aperture is evaluated in many texts and is given by Ghatak [2]

$$I = I_0 \left[\frac{2J_1(\beta)}{\beta} \right]^2 \tag{1.74}$$

where I_0 is the intensity on the axis of the lens and

$$\beta = ka \ \sin\theta \tag{1.75}$$

where θ is defined in Figure 1.9 and J_1 is the Bessel function of first kind of order 1. The distribution given by Eq. (1.74) is referred to as the *Airy pattern* (Figure 1.9b). The focal pattern consists of a bright central spot surrounded by rings of smaller and smaller intensities. The radius of the first dark ring corresponds to the first zero of J_1 and occurs at $\beta \sim 3.83$, which corresponds to

$$\sin\theta \approx \frac{1.22\lambda}{2a} \tag{1.76}$$

and is referred to as the *radius* of the Airy pattern.

For small θ, that is, paraxial condition, we can approximate $\sin\theta$ by θ; in such a case, the transverse dimension of the Airy spot is approximately given by

$$\Delta r \approx \frac{1.22\lambda f}{2a} = 1.22\lambda f^{\#} \tag{1.77}$$

where f is the focal length of the lens and $f^{\#}$, the ratio of focal length to the diameter of the lens is called the *f-number* of the lens. The smaller the f-number, the smaller is the spot size and the sharper is the image. For a given focal length, the larger the diameter of the lens, the smaller is the focused spot. The quantity $1.22\lambda/2a$

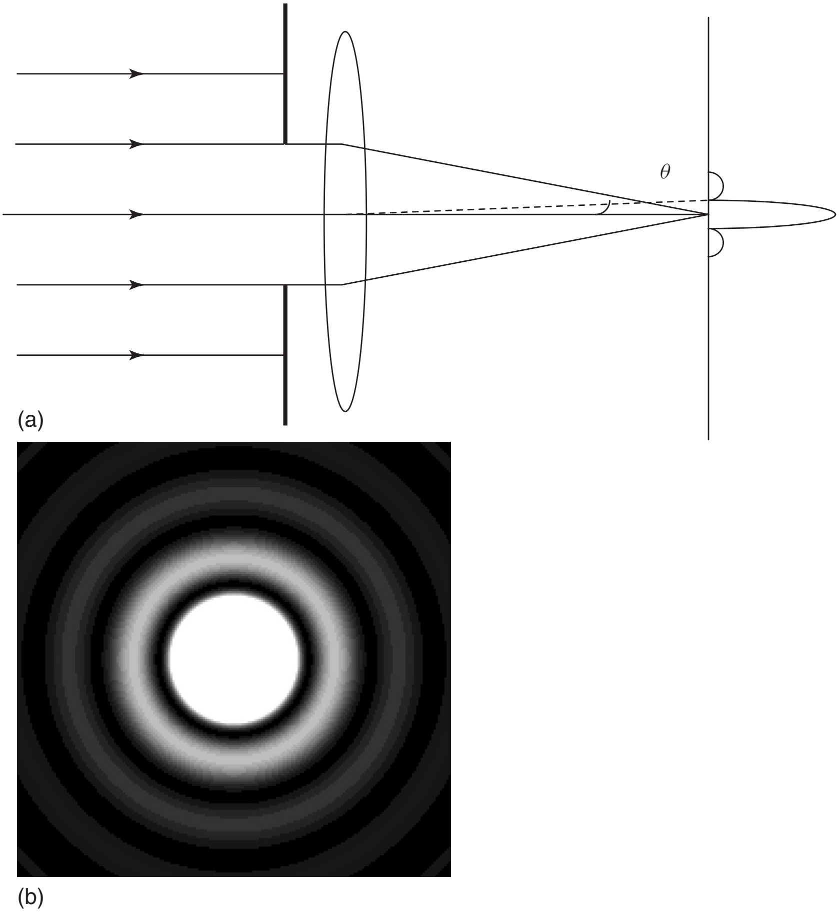

Figure 1.9 (a) A plane wave falling on a circular aperture and forming an Airy pattern on the focal plane of the lens. (b) Airy pattern that will be observed on the back focal plane of the lens.

corresponds to the angle made by the first dark ring of the Airy spot with the center of the lens.

Box 1.3: Diffraction Gratings

A diffraction grating is a very important optical component that is used in many instruments such as spectrometers for dispersing different wavelengths present in the illumination. It consists of a number of equally spaced identical long narrow slits placed parallel to each other. Each of the slits produces its own diffraction and the interference among the diffracted light waves from the various apertures produces the diffraction pattern of the grating given by

$$I(\theta) = I_0 \left(\frac{\sin\beta}{\beta}\right)^2 \left(\frac{\sin N\gamma}{\sin\gamma}\right)^2$$

where

$$\beta = \frac{\pi b \sin\theta}{\lambda_0}$$

$$\gamma = \frac{\pi d \sin\theta}{\lambda_0}$$

Here b represents the width of each slit and d is the spacing between the slits. The first term within the brackets represents the diffraction by individual slits and the second term represents the interference between the various diffraction patterns.

Whenever $\gamma = m\pi$, $m = 0, 1, 2, \ldots$, the second bracketed term in the first equation becomes N^2 and these correspond to the principal maxima of the diffraction grating. This gives us the angles of the principal maxima as

$$d \sin\theta = m\lambda_0$$

where m represents the order of the diffraction. Since the angles of the principal maxima depend on the wavelength, the diffraction grating disperses the incident wavelengths along different directions, thus forming a spectrum.

The resolving power of the diffraction grating is given by

$$\mathrm{RP} = \frac{\lambda_0}{\Delta\lambda} = mN$$

where N is the total number of slits in the grating that is illuminated by the incident light and $\Delta\lambda$ is the minimum resolvable wavelength difference. If we consider a diffraction grating with 5000 lines per centimeter and illuminate 2 cm width of the grating, then in the first order the resolving power would be 10 000. Thus at a wavelength of 500 nm, the grating can resolve two lines spaced by 0.05 nm.

1.8.1
Resolution of Optical Instruments

According to ray optics, an aberration-less optical system should form point images of point objects and thus have infinite resolution. Since light has a finite wavelength, diffraction effects limit the size of the focused spot for a point object and thus diffraction would decide the resolving power of optical instruments.

When a telescope is used to image far-off objects such as stars, then each star produces an Airy pattern and if the Airy patterns of two stars are too close to each other then it would not be possible for us to resolve the two stars. Thus diffraction effects will ultimately limit the resolving ability of telescopes. If the aperture of the objective of the telescope is d and the wavelength is λ_0, then the angular resolution of the telescope is

$$\Delta\theta_{\mathrm{min}} = \frac{1.22\lambda_0}{d} \tag{1.78}$$

The larger the diameter of the objective lens, the better is the angular resolution. Terrestrial-based telescopes do not actually operate at this limit because of atmospheric turbulence and hence practical values may be higher than the value predicted by Eq. (1.78). The diameter of the objective of the Keck telescope is 10 m. With this size of the objective, the expected resolution is about 0.01 arc sec.

In the case of a microscope in which the object is illuminated by incoherent illumination, the spatial resolution is given by

$$\Delta x = \frac{0.61\lambda_0}{\mathrm{NA}} \tag{1.79}$$

where NA is the numerical aperture of the microscope objective and is given by $n \sin\theta_{\mathrm{m}}$ where n is the refractive index in which the object is placed (for example, in oil immersion objectives, n can be larger than 1) and θ_{m} is the limiting angular aperture. Typically a 40× objective with an NA of 0.65 will have a resolution of about 0.42 μm at 550 nm wavelength. As demand for increased resolution of microscopes increases, many techniques to overcome this limitation are being developed and microscopes with resolutions of better than 100 nm are now available [8].

Tutorial Exercise 1.11

Calculate the resolution for the following optical systems for a wavelength of 550 nm, which corresponds to the center of the visible spectrum.

1) An *f*/16 lens, that is, a lens with an *F*-number of 16 is used to focus a laser beam. What is the diameter of the focused spot? How does this diameter change when the pupil of the lens is changed to *f*/1.4?
2) The average pupil size of the human eye is about 4 mm and the focal length is about 17 mm. What is the diameter of a far away object such as a star on the retina?
3) Compare the resolution of the 200 in. Mt Palomar telescope with the human eye. What is its spatial resolution when observing the moon surface?
4) Let us consider a microscope with an NA of 0.8. What is the minimum spatial resolution?

Solution:

1) Using Eq. (1.79), the diameter of the focused image would be about 21 μm, but only 1.9 μm for the *f*/1.4 lens. The *f*/1.4 lens has obviously better resolution than the *f*/16 lens. In camera lenses you would see the *f*-number written on the lens and that gives you an indication of its resolution capabilities.

2) The diameter of the first dark ring of the Airy pattern would be about 3 μm.
3) For the telescope with an objective diameter of 200 in., the minimum angular resolution is given by 0.13×10^{-6} rad. This telescope can resolve spatial details of 50 m on the moon (assuming the distance between earth and moon to be 384 000 km).
4) The minimum spatial resolution is given by 0.41 μm.

Box 1.4: Diffraction of a Gaussian Beam

The beam of a laser has a Gaussian distribution in the transverse plane. It is described by the following equation:

$$f(x, y) = A \exp\left[-\frac{(x^2 + y^2)}{w_0^2}\right]$$

where w_0 represents the spot size of the Gaussian beam. We can use Eq. (1.69) to study the propagation of such a Gaussian beam. Substituting in Eq. (1.69) and integrating, we obtain the field distribution on any plane z as

$$u(x, y, z) \approx \frac{a}{(1 - i\gamma)} \exp\left[-\frac{x^2 + y^2}{w^2(z)}\right] \mathbf{e}^{-i\Phi}$$

where

$$\gamma = \frac{\lambda z}{\pi w_0^2},$$

$$w(z) = w_0\sqrt{1 + \gamma^2},$$

$$\Phi = kz + \frac{k}{2R(z)}\left(x^2 + y^2\right)$$

$$R(z) \equiv z\left(1 + \frac{1}{\gamma^2}\right)$$

Thus, the intensity distribution $|u|^2$ at any z is given by

$$I(x, y, z) = \frac{I_0}{1 + \gamma^2} \exp\left[-\frac{2\left(x^2 + y^2\right)}{w^2(z)}\right]$$

which again represents a Gaussian distribution. Thus, as a Gaussian beam propagates in a homogeneous medium, the transverse distribution remains Gaussian with its spot size changing with z as given by $w(z)$.

On the plane $z = 0$, the phase is independent of the transverse coordinates x and y and hence on this plane the Gaussian beam has a plane wavefront. As the beam propagates, the phase distribution changes and is given by the function Φ. Now, we note that the transverse phase distribution at any value of z is given by

$$\exp\left[-\mathrm{i}\frac{k}{2R(z)}(x^2+y^2)\right]$$

The above phase distribution corresponds to the paraxial approximation of the phase distribution of a spherical wave and thus as the Gaussian beam propagates, its phase front becomes spherical with a radius of curvature given by $R(z)$. The radius of curvature of the phase front is infinite at $z = 0$ representing a plane phase front and as z tends to infinity it again tends to infinity.

The equations representing the variation of the spot size and radius of curvature of the phase front of the Gaussian beam are valid for all values of z (positive or negative) and z is measured from the plane where the beam has a plane phase front. This plane is called the *waist* of the Gaussian beam, and as can be seen from $w(z)$ the beam has the minimum value of spot size at the waist.

For values of z satisfying

$$z \gg \pi w_0^2/\lambda$$

that is, in the far field, the spot size increases linearly with z:

$$w(z) \approx \frac{\lambda z}{\pi w_0}$$

which is similar to the case of a circular aperture except for a different factor. In fact, for a beam having a transverse dimension w, the angle of diffraction is approximately given by λ/w.

1.9 Anisotropic Media

The electric displacement vector **D** inside a material and the applied electric field $\boldsymbol{E}$ are related through the following equation:

$$\boldsymbol{D} = \varepsilon \boldsymbol{E} \tag{1.80}$$

where ε is the electric permittivity of the medium. In isotropic media, the displacement and the electric field are parallel to each other and ε is a scalar quantity. In anisotropic media, the two vectors are not parallel to each other and ε is a tensor and we write Eq. (1.80) in the form

$$\boldsymbol{D} = \overline{\varepsilon} \boldsymbol{E} \tag{1.81}$$

where we have put a bar on ε to indicate that it is not a scalar.

In the principal axis system of the medium, $\overline{\varepsilon}$ can be represented by a diagonal matrix:

$$\overline{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix} \tag{1.82}$$

And the three diagonal terms give the principal dielectric permittivities of the medium.

- For isotropic media

$$\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{zz} = \varepsilon \tag{1.83}$$

- For uniaxial media

$$\varepsilon_{xx} = \varepsilon_{yy} \neq \varepsilon_{zz} \tag{1.84}$$

- and for biaxial media

$$\varepsilon_{xx} \neq \varepsilon_{yy} \neq \varepsilon_{zz} \tag{1.85}$$

The principal dielectric constants and the principal refractive indices of the anisotropic medium are and defined through the following equations:

$$K_{ij} = \frac{\varepsilon_{ij}}{\varepsilon_0}; \quad n_{ij}^2 = \sqrt{K_{ij}} \tag{1.86}$$

Since in the principal axis system ε is diagonal, the principal refractive indices are also sometimes referred to as n_x, n_y, and n_z. For isotropic media $n_x = n_y = n_z$; for uniaxial media $n_x = n_y \neq n_z$, while for biaxial media all the three principal refractive indices are different.

In isotropic media, the speed of wave propagation is independent of the state of polarization of the light beam and is also independent of the direction of propagation. On the other hand, in anisotropic media, along any given direction of propagation there are two linearly polarized eigenmodes which propagate, in general, with different phase velocities. Thus any incident light beam can be broken into the two eigenmodes of propagation and since their velocities are in general different, the phase difference between the two components changes as the light beam propagates. This results in a change of the state of polarization of the light beam as it propagates along the medium.

In uniaxial media, there is one direction of propagation along which the two velocities are equal and this direction is referred to as the *optic axis*. There is only one such direction in uniaxial media and hence the name. In biaxial media, there are two such optic axes.

When a plane wave propagates in a uniaxial medium, one of the polarization components travels always at the same velocity, which is independent of the direction of propagation. This wave is referred to as the *ordinary wave* and its velocity is given by c/n_o. On the other hand, the orthogonal polarization, which is referred to as the *extraordinary wave*, has a velocity of propagation that depends on the direction of propagation with respect to the optic axis. If the direction of the propagation vector of the extraordinary wave makes an angle of ψ with respect to the optic axis, then the velocity of the extraordinary wave is given by $c/n_e(\psi)$ where

$$\frac{1}{n_e^2(\psi)} = \frac{\cos^2\psi}{n_o^2} + \frac{\sin^2\psi}{n_e^2} \tag{1.87}$$

The velocity of the extraordinary wave varies between c/n_o and c/n_e.

For the o-wave, the displacement vector ***D*** and the electric field vector ***E*** are perpendicular to the plane containing the propagation vector ***k*** and the optic axis. For the e-wave, the displacement vector ***D*** lies in the plane containing the propagation vector ***k*** and the optic axis and is normal to ***k***.

If we consider a plane wave propagating perpendicular to the optic axis of a uniaxial crystal, then the two eigen polarizations have refractive indices n_o and n_e. If the free space wavelength of the wave is λ_0, then in propagating through a distance l, the phase change due to propagation would be

$$\Delta\phi = \frac{2\pi}{\lambda_0}\left(|n_o - n_e|\right) l \tag{1.88}$$

If a linearly polarized wave is incident on such a medium with the polarization direction making an angle θ with the optic axis and propagating perpendicular to the optic axis, then since in the linearly polarized wave the two components are in phase, as the wave propagates through the uniaxial medium, it will develop a phase difference given by Eq. (1.88). If the phase difference is π, then the polarization state would be still linear but now oriented at an angle $-\theta$ with the optic axis. Such a device is called a half wave plate (HWP) and the required thickness is

$$l_h = \frac{\lambda_0}{2|n_o - n_e|} \tag{1.89}$$

Using a HWP it is possible to change the orientation of the linearly polarized wave by appropriately orienting the optic axis of the HWP.

If the thickness of the uniaxial medium is half of that given by Eq. (1.89), then the phase difference introduced is $\pi/2$ and such a plate is called a quarter wave plate (QWP). If the angle θ is chosen to be $\pi/4$, then the output of a QWP will have two equal components of linear polarizations but with a phase difference of $\pi/2$; such a wave would correspond to a circularly polarized wave. Thus a QWP can be used to convert a linearly polarized wave into a circularly polarized wave and vice versa. If the angle θ is not $\pi/4$, then the output would be an elliptically polarized wave.

HWP and QWPs are very useful components in experiments involving polarization (Chapter 9).

1.10
Jones Calculus

Jones calculus is a very convenient method to determine the changes in the state of polarization of a light wave as it traverses various polarization components. In this calculus, the state of polarization is represented as a (2×1) column vector and the polarization components are represented by (2×2) matrices. The propagation through each polarization component is represented by multiplying the column vector by the Jones matrix corresponding to the polarization component.

We first discuss the Jones vector representation of the state of polarization of a plane wave. Equation (1.19) gives a general expression for an elliptically polarized plane light wave propagating along the z-direction. The complex amplitudes of the electric field vector can be expressed as the components of the Jones vector $\boldsymbol{J}$ where

$$\boldsymbol{J} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{1.90}$$

If we write

$$|x\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{1.91}$$

and

$$|y\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.92}$$

then the Jones vector can also be written as

$$\boldsymbol{J} = E_x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + E_y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = E_x |x\rangle + E_y |y\rangle \tag{1.93}$$

The vectors $|x\rangle$ and $|y\rangle$ represent normalized Jones vectors for an x-polarized and a y-polarized wave, respectively. This is an equivalent representation of the fact that a general polarized wave can be represented as a superposition of a linearly polarized wave along x and a linearly polarized wave along y with appropriate amplitude and phase.

The normalized Jones vector representing a linearly polarized wave making an angle ϕ with the x-axis is given by

$$|\phi\rangle = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} = \cos\phi\, |x\rangle + \sin\phi\, |y\rangle \tag{1.94}$$

The electric field vector of an RCP wave propagating in the z-direction is given in Eq. (1.17); hence, the normalized Jones vector of an RCP is given by (neglecting a common phase factor)

$$|\mathrm{RCP}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} \tag{1.95}$$

Similarly, the normalized Jones vector representing the LCP wave is given by

$$|\mathrm{LCP}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ +\mathrm{i} \end{pmatrix} \tag{1.96}$$

The normalized Jones vector for a right elliptically polarized wave would be

$$|\mathrm{REP}\rangle = \frac{1}{\sqrt{\left(E_1^2 + E_2^2\right)}} \begin{pmatrix} E_1 \\ -\mathrm{i}E_2 \end{pmatrix} \tag{1.97}$$

Thus, we have represented a general polarized plane wave by a (2×1) column vector.

Optical elements such as HWP, QWP, polarizer, and so on, act on the polarization state of the light beam and thus modify the Jones vector of the light beam propagating through them. The input to these elements would be described by a

Jones vector of the input light and the output would correspond to the Jones vector of the output light wave. Thus these elements convert a (2 × 1) column vector to another (2 × 1) column vector and hence can be represented by a (2 × 2) matrix.

The Jones matrix for a linear polarizer is given by

$$J_{\text{LP}}(\alpha) = \begin{pmatrix} \cos^2\alpha & \sin\alpha\cos\alpha \\ \cos\alpha\sin\alpha & \sin^2\alpha \end{pmatrix} \tag{1.98}$$

where α represents the angle made by the polarizer pass axis with the x-axis. This can be easily seen as follows: if the input Jones vector is given by

$$|\text{in}\rangle = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \tag{1.99}$$

then the Jones vector of the output wave would be given by

$$|\text{out}\rangle = \begin{pmatrix} \cos^2\alpha & \sin\alpha\cos\alpha \\ \sin\alpha\cos\alpha & \sin^2\alpha \end{pmatrix}\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} = \begin{pmatrix} E_0\cos\alpha \\ E_0\sin\alpha \end{pmatrix} \tag{1.100}$$

with

$$E_0 = E_1\cos\alpha + E_2\sin\alpha \tag{1.101}$$

As expected Eq. (1.100) represents a linearly polarized wave. Thus no matter what the input polarization may be, the output will always be linearly polarized.

For a QWP with its fast axis along the y-direction, the Jones matrix is given by

$$J_{\text{QWP}} = \begin{pmatrix} 1 & 0 \\ 0 & \text{i} \end{pmatrix} \tag{1.102}$$

When a linearly polarized wave (making an angle 45° with the x-axis) is incident on such a QWP, the output polarization is given by

$$J_{\text{QWP}}\left|45^\circ\right\rangle = \begin{pmatrix} 1 & 0 \\ 0 & \text{i} \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \text{i} \end{pmatrix} \tag{1.103}$$

which represents an LCP (Figure 1.10). For a QWP with its fast axis along the x-direction, the Jones matrix is given by

$$J_{\text{QWP}} = \begin{pmatrix} \text{i} & 0 \\ 0 & 1 \end{pmatrix} \tag{1.104}$$

For a HWP with its fast axis along the y-direction, the Jones matrix is given by

$$J_{\text{HWP}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.105}$$

When a light beam linearly polarized at 45° to the x-axis is incident on a HWP, then the output is given by

$$J_{\text{HWP}}\left|45^\circ\right\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{1.106}$$

which is again a linearly polarized beam, polarized at −45° to the x-axis.

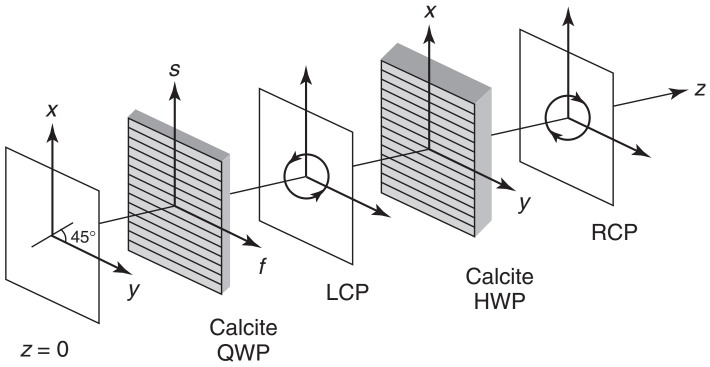

Figure 1.10 A linearly polarized beam making an angle 45° with the x-axis gets converted to an LCP after propagating through a calcite QWP; further, an LCP gets converted to a RCP after propagating through a calcite HWP. The optic axis in the QWP and HWP is along the y-direction as shown by lines parallel to the y-axis. Light polarized along the y-axis travels faster than light polarized along the x-axis; hence the symbols f (fast) and s (slow).

When an LCP is incident on such a HWP, the output polarization is given by

$$J_{\mathrm{HWP}} \left|\mathrm{LCP}\right\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} \tag{1.107}$$

which represents an RCP (Figure 1.10).

As the effect of each polarization component is represented by a (2×2) matrix, using Jones calculus becomes quite straightforward to determine the state of polarization of the light wave emerging from a series of components. Corresponding to each component, we need to multiply with the corresponding Jones matrix in the correct order. Since, in general, matrices do not commute, the output polarization state would depend on the order in which the light wave traverses them.

1.11 Lasers

Today lasers form one of the most important sources of light for various applications because of their special characteristics in terms of temporal and spatial coherence, brightness, and the possibility of extremely short pulse widths and power. Lasers spanning wavelengths from the ultraviolet to the infrared are available for various applications. Diode lasers, because of their size and efficiency, are one of the most important lasers. In this section, we discuss briefly the working of the laser and some of their properties. For more detailed discussions, readers are referred to books on lasers [9, 10].

1.11.1 Principle

Atoms and molecules are characterized by energy levels and they can interact with electromagnetic radiation through three primary processes:

1) Absorption: When radiation at an appropriate frequency falls on the atomic system atoms lying in the lower energy state can absorb the incident radiation and get excited to an upper energy level. If the energy of the two levels are E_1 and E_2 ($>E_1$), then for absorption the frequency of the incident radiation should be $\omega_0 = (E_2 - E_1)/\hbar$.
2) Spontaneous emission: An atom lying in the upper energy level can spontaneously emit radiation at the frequency ω_0 and get de-excited to the lower level. The average time spent by the atom in the excited level is called the *spontaneous life time* of the level.
3) Stimulated emission: An atom in the excited level can also be stimulated to emit light by an incident radiation at the frequency ω_0 and the emitted light is fully coherent with the incident radiation.

When we consider an atomic system at thermal equilibrium, the number of atoms in the lower state is always higher than that in the upper state. Thus there will be more absorptions than stimulated emissions when light interacts with the collection of atoms. If the population of the upper state can be made to be higher than that of the lower state (i.e., if we have population inversion between the two levels), then there would be greater number of stimulated emissions compared to absorptions and the incident radiation can get amplified. This is the basic principle of optical amplification and is the heart of the laser (short for light amplification by stimulated emission of radiation). An external source of energy is required to bring about the state of population inversion in the collection of atoms. This could be electrical energy such as in the case of diode lasers, it could be another laser such as in the case of Ti : sapphire lasers, it could be a flashlamp such as in the case of ruby laser, or it could be an electrical discharge such as in the case of an argon ion laser. The laser uses up the energy from the pump and converts it into coherent optical energy.

With the population inversion, we obtain an optical amplifier that can amplify an incident optical radiation at the appropriate frequency. However, we need a source of light which is similar to an oscillator rather than an amplifier. The optical amplifier can be converted to a laser by providing optical feedback. This is accomplished by the use of a pair of mirrors on either side of the optical amplifier. The pair of mirrors reflects a part of the energy back into the optical amplifier and is called an *optical resonator*.

As soon as the pump takes atoms from the lower to the higher energy state, the atoms in the higher energy state start to emit spontaneously. Part of the spontaneous emission that travels along the direction perpendicular to the mirrors of the cavity is reflected back into the cavity, which then gets amplified by the population inversion generated in the cavity by the pump. The amplified spontaneous emission is reflected back into the cavity by the other mirror and this process continues back and forth. When the loss of the optical radiation gets compensated completely by the gain provided by the population inversion in the medium, the laser starts to oscillate. Thus there is a threshold to start laser action and this depends on the losses in the cavity.

Owing to various mechanisms such as the finite lifetime of the excited level, or the motion of the atoms causing Doppler shift or collisions of the atoms, a range of frequencies, rather than a single frequency, can interact with the atoms. This is referred to as *line broadening* and depends on the atomic system, the temperature and pressure in the case of a gas, the environment of atoms in the case of solids, and the energy bands in the case of semiconductors. Typical bandwidths range from a few to hundreds of gigahertz. This implies that the atomic system is capable of amplifying a range of frequencies lying within this broadened line.

The optical resonator made up of the two mirrors supports modes of oscillation within the cavity. The transverse field distributions that repeat themselves after every roundtrip within the cavity are termed as *transverse modes* and the frequencies of oscillation supported by the cavity are termed as *longitudinal modes* of the cavity. Just as in the case of modes of oscillation of a string, the various frequencies of oscillation of the laser cavity are separated by an approximate value of $\Delta\nu$ where

$$\Delta\nu = \frac{c}{2nL} \tag{1.108}$$

where c is the velocity of light in free space, n is the refractive index of the medium filling the cavity, and L is the length of the cavity.

Since the atoms forming the optical amplifier can amplify a range of frequencies, there may be many frequencies supported by the cavity that can oscillate simultaneously leading to multi-longitudinal mode oscillation. This will lead to a drastic reduction of the coherence length of the laser. For interferometry applications where the path difference between the interfering beams may be large, it is important that the laser oscillate in a single frequency only.

There are many techniques to achieve single longitudinal mode operation of the laser. One of the standard techniques is to use a tilted Fabry–Perot etalon within the cavity of the laser. Since the Fabry–Perot etalon allows only certain frequencies to pass through, by choosing an appropriate etalon and adjusting its angle within the cavity, it is possible to allow only one of the longitudinal modes to oscillate. This leads to a single longitudinal mode of oscillation with a corresponding increase in the coherence length. In the case of semiconductor lasers, an external grating or a fiber Bragg grating can be used to achieve single longitudinal mode operation. In the distributed feedback diode laser a periodic variation of the structure along the direction of lasing ensures single frequency oscillation.

Box 1.5: Transversal Modes of a Laser Beam

Although the output beam of a laser is mostly a Gaussian, there exist higher order modes inside the laser cavity. The various transverse modes of the laser can be approximately described by Hermite–Gauss functions:

$$f_{mn}(x, y) = CH_m\left(\frac{\sqrt{2}x}{w_0}\right) H_n\left(\frac{\sqrt{2}y}{w_0}\right) \mathbf{e}^{-(x^2+y^2)/w_0^2}$$

where $H_n(\xi)$ represents a Hermite polynomial, and the lower order polynomials are

$$H_0(\xi) = 1, \qquad H_1(\xi) = 2\xi; \qquad H_2(\xi) = 4\xi^2 - 2\ldots$$

and n and m represent the mode numbers along the x- and y-directions, which are assumed to be transverse to the axis of the resonator.

The fundamental mode of the laser corresponds to $n = 0$ and $m = 0$ and is a Gaussian distribution:

$$f_{00}(x, y) = C\mathbf{e}^{-(x^2+y^2)/w_0^2}$$

This is the mode in which most lasers operate as it has the minimum diffraction divergence and also has a uniform phase front without any reversals of phase unlike the higher order transverse modes.

Since the fundamental mode has the smallest transverse size, any aperture within the cavity can be used to set the laser to oscillate only in the fundamental transverse mode by introducing additional losses to higher order modes as compared to the fundamental mode.

1.11.2
Coherence Properties of the Laser

The temporal coherence of the laser is determined by its spectral width. For a spectral width of $\Delta\nu$, the coherence length is given by

$$l_c = \frac{c}{\Delta\nu} = \frac{\lambda^2}{\Delta\lambda} \tag{1.109}$$

As discussed earlier, when the laser is used in an interference experiment, for proper contrast, the maximum path difference between the interfering beams must be much smaller than the coherence length. Hence the coherence length is a very important parameter in applications such as holography or interference measurements. Single longitudinal mode lasers with coherence lengths of more than a few meters to a few hundred meters are commercially available.

Tutorial Exercise 1.12

Consider a laser operating at 800 nm with a spectral width of 10 MHz. Calculate the coherence length. What is its spectral width in wavelength?

Solution:

The corresponding coherence length would be 30 m; the spectral width in wavelength is 21 fm.

When the laser oscillates in the fundamental Gaussian mode, it has very good spatial coherence. Sometimes, the laser is coupled into a single mode optical fiber and the fiber acts as a spatial filter to filter the spatial noise in the beam. The output of the single mode fiber is approximately Gaussian.

1.12 Optical Fibers

An optical fiber is a cylindrical structure consisting of a central core of refractive index n_1 surrounded by a cladding of slightly lower refractive index n_2 (Figure 1.11). A light wave can get trapped within the core of the fiber by the phenomenon of total internal reflection. Thus such a dielectric structure can guide light from one point to another and forms the heart of today's telecommunication system. More details on fiber optics and its application to communication and sensing can be found in Ref. [11].

An optical fiber can be characterized by its NA defined as

$$\mathrm{NA} = \sqrt{(n_1^2 - n_2^2)} \tag{1.110}$$

The maximum angle of acceptance of the fiber is given by

$$\theta_{\max} = \sin^{-1}\left(\frac{\mathrm{NA}}{n_0}\right) \tag{1.111}$$

where n_0 is the refractive index of the medium surrounding the fiber.

Tutorial Exercise 1.13

Consider a fiber with an NA of 0.2 and calculate the maximum acceptance angle of the fiber when the surrounding medium is (i) air and (ii) water.

Solution:

Using Eq. (1.111) and a refractive index of 1.0 for air and 1.33 for water, the acceptance angle is

1) $\sim 11.5^\circ$ and
2) $\sim 8.6^\circ$.

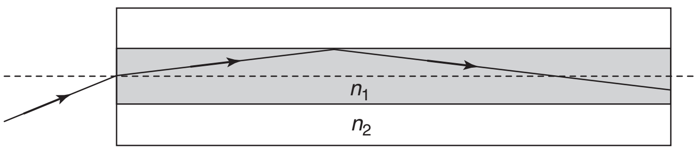

Figure 1.11 An optical fiber consists of a core of refractive index n_1 surrounded by a cladding of refractive index n_2.

Optical fibers can also be specified by a normalized frequency, also called the *V-number*, defined as

$$V = \frac{\omega}{c} a \text{ NA} = k_0 a \text{ NA} \tag{1.112}$$

where a is the radius of the core of the fiber.

Multimode fibers are characterized by core diameters of about 50 μm and an NA of about 0.2; they support many transverse modes of propagation. On the other hand, single mode fibers have much smaller core diameters and support only a single mode of propagation. For single mode operation, the V-number of the fiber should be less than 2.405. Thus a fiber can be single moded at a particular wavelength and if operated at lower wavelengths may have a V value of more than 2.405 and would then support more than one mode.

Since multimode fibers support many modes, the output from multimode fibers under laser illumination has a speckle pattern. Since a single mode fiber supports only a single mode, it has a well-defined transverse amplitude distribution, which can be very well approximated by a Gaussian distribution [12]:

$$\psi(r, z) = A\mathbf{e}^{-r^2/w_0^2}\mathbf{e}^{\mathrm{i}(\omega t - \beta z)} \tag{1.113}$$

where w_0 is referred to as the *spot size* of the mode and β represents the propagation constant of the mode. The ratio β/k_0 is referred to as the *effective index* of the mode that lies between n_1 and n_2. An empirical expression for the spot size is given by Marcuse [7]

$$w_0 = a\left(0.65 + \frac{1.619}{V^{1.5}} + \frac{2.879}{V^6}\right) \tag{1.114}$$

If we use this fiber at a wavelength of 500 nm, then its V-number would be 2.51 (assuming that NA remains the same) and the fiber would support more than one mode. The output from the fiber would not be a Gaussian but will also contain contribution from the higher order mode.

The spot size of the mode can be increased by taking fibers with a larger core radius. However, to keep the fiber as single moded the corresponding NA has to be reduced. This implies that the refractive index difference between the core and cladding needs to be reduced. This leads to poorer guidance by the fiber and hence such a fiber is more prone to bending-induced losses.

When a laser is used to couple light into a single mode fiber, the coupling efficiency would be determined by the transverse distribution of electric field of the laser beam and the modal field distribution. The maximum coupling efficiency (assuming perfect alignment between the laser beam and the fiber) from a laser oscillating in a Gaussian mode with a spot size w_L to a single mode fiber with a spot size w_0 is given by

$$T_\mathrm{max} = \left(\frac{2w_\mathrm{L}w_0}{w_\mathrm{L}^2 + w_0^2}\right)^2 \tag{1.115}$$

For unit coupling efficiency, the spot sizes of the laser beam and the fiber should be identical and the waist of the laser beam should be at the fiber end face. Any deviation from this condition would reduce the efficiency. The expression given in Eq. (1.115) is also valid for coupling efficiency between two single mode fibers with spot sizes w_L and w_0.

Standard single mode fibers do not maintain the state of polarization of light propagating through them. If linearly polarized light is launched into the fiber, then after propagating through the fiber, the state of polarization can become arbitrary because of the birefringence within the fiber induced by bending or residual stresses in the fiber. In interference experiments involving fibers, it may be necessary to maintain the state of polarization of the light. In such cases, one uses polarization maintaining fibers. Such fibers have a strong linear birefringence caused by introducing transverse stress within the fiber while fabricating them so that the fiber supports two orthogonal linearly polarized modes. Since the modes have significantly different propagation constants, only a spatially periodic perturbation with a small period (a few millimeters) can induce coupling between them. Since under normal circumstances such a perturbation does not exist, the fiber maintains the linear polarization of the propagating light beam. Such fibers are used in fiber optic sensors based on interference effects such as Mach Zehnder fiber interferometric sensor.

1.13 Summary

In this brief chapter, basic concepts in optics such as rays, image formation by optical systems, interference, and diffraction of light have been given. Laser is one of the most important sources of light and a brief outline of the working and its characteristics is also covered. Finally, a brief introduction to optical fibers has been given as optical fibers are very useful components that carry light between two different points and also provide for spatial filtering of the laser beam in experiments such as holography and interferometry.

Problems

1.1 What is the direction of propagation of a wave described by the following equation:

$$\boldsymbol{E} = \boldsymbol{y} E_0 \exp\left\{\mathrm{i}\left(\omega t - \frac{k}{\sqrt{2}}x - \frac{3k}{\sqrt{2}}z\right)\right\}$$

Give an expression for a wave propagating in the same direction but orthogonally polarized to this wave.

1.2 Consider a piece of parabolic index medium with flat faces and of thickness d. Consider a point source placed on the axis at the entrance face of such a medium. What is the minimum value of d so that the output rays form a parallel bundle? In which direction would the parallel beam be propagating?

1.3 Consider a symmetric double convex thick lens formed by surfaces of radii of curvatures 10 cm and separated by a distance of 1 cm. If the refractive index of the medium of the lens is 1.5, obtain the system matrix for the lens and calculate the focal length of the lens. Also obtain the positions of the principal planes.

1.4 A nonreflecting film to operate at a wavelength of 600 nm is to be coated on a material of refractive index 2.2. What is the thickness of the film that would be needed? What will be the optimum refractive index so that the reflection is minimum?

1.5 Consider a Fabry–Perot interferometer made of two mirrors of reflectivities 0.9 and separated by a distance of 1 mm. Obtain the resolving power of the interferometer if it is operated at a wavelength of 500 nm.

1.6 Estimate the lateral coherence width of the sun on the surface of the earth. Assume a wavelength of 550 nm.

1.7 A lens with a transverse diameter of 10 mm and a focal length of 20 mm is used to focus a laser beam at a wavelength of 600 nm and with a transverse diameter of 5 mm. What will be the approximate area of the focused spot? If the laser beam has a power of 5 mW, what will be the maximum intensity and maximum electric field at the center of the focused spot?

1.8 A Gaussian beam emerging from a laser with a wavelength of 800 nm has a transverse diameter of 1 mm. What will be the transverse size of the beam after traversing a distance of 10 m?

1.9 Consider a uniaxial medium with $n_o = 2.26$ and $n_e = 2.20$. For an incident wavelength of 600 nm, obtain the thickness of the HWP and QWP.

1.10 A circularly polarized beam is incident on a HWP. What will be the output state of polarization? If it is incident on a QWP what will be the state of polarization of the output wave?

1.11 An interference experiment is to be conducted using a He–Ne laser. We have two lasers, one oscillating in a single longitudinal mode with a linewidth of 10 MHz and the other with two modes with linewidths of 10 MHz and separated by a frequency of 600 MHz. Estimate the minimum path difference for which the interference pattern will disappear when either of the lasers is used.

1.12 Consider a laser resonator made of a plane mirror and a concave mirror of radius of curvature 1 m. For what range of separation between the mirrors would the resonator be stable? What will be the spot size of the Gaussian mode of the laser? Where will the waist of the laser beam lie?

1.13 A He–Ne laser has a gain bandwidth of 1 GHz. For what mirror separation would the laser operate in a single longitudinal mode?

1.14 Consider a single mode fiber with $a = 2.5\ \mu\text{m}$, NA $= 0.08$ operating at a wavelength of 600 nm. Obtain the V-number of the fiber and show that the fiber would be single moded. Obtain the corresponding spot size of the mode. What will be the angle of diffraction of the beam coming out of such a fiber? Above what wavelength will the fiber be single moded?